Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

REPORT

2025 SCE Real Time Pricing Demand Response Evaluation

Confidential information is redacted and denoted with black highlighting: [redacted]

April 1, 2026

Prepared for Southern California Edison

By:

Adriana Ciccone
Daniil Deych
Demand Side Analytics

CALMAC ID: SCE0508

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Real-Time Pricing Program offers commercial and industrial customers the opportunity to react daily to price signals and reduce loads when prices are high. Each day, the next days' hourly prices are tied directly to the daily maximum temperature in Downtown Los Angeles, grouped into one of seven day types: Hot Summer Weekday, Moderate Summer Weekday, Mild Summer Weekday, High Cost Winter Weekday, Low Cost Winter Weekday, High Cost Weekend and Low Cost Weekend. In PY2025, there were no days that qualified as a High Cost Winter Weekday.

Table 1 shows the summarized results of load impacts for the program under each RTP day type. The RTP program delivered [redacted] during the 4-9pm window on Hot Summer Weekdays, which was a [redacted] reduction in load. As RTP prices are the highest on these days relative to the otherwise applicable tariff (OAT), ex post impacts are higher on Hot Summer Weekdays, while impacts are lower on Moderate and Mild Summer Weekdays. High Cost Weekends also show a reduction in consumption during the peak period relative the OAT.

Table 1: Ex Post Peak Period Impacts by Day Type

RTP Day Type	# Dispatched	Average Customer (kW)					Agg. Impact (MW)
		Ref. Load	Obs. Load	Impact	95% CI	% Impact	
Hot Summer Weekday	86	[redacted]					
Moderate Summer Weekday	86	[redacted]					
Mild Summer Weekday	87	[redacted]					
Low Cost Winter Weekday	88	[redacted]					
High Cost Weekend	87	[redacted]					
Low Cost Weekend	88	[redacted]					

Table 2 shows impacts averaged across the RA window for an August monthly worst day in each year from 2026-2036 and for each set of ex ante weather conditions. RTP enrollments are expected to decline over time, from 83 in 2026 to 79 enrolled customers in 2036. Program load impacts of approximately [redacted] during the 4pm-9pm hours are projected in 2026. Load impacts by hour in the RA window are shown in Table 3. Due to the RTP treatment being determined by weather conditions, no weather variables are included in the ex ante model specification, so the only difference between these scenarios is the RTP day type associated with the CAISO and SCE 1-in-2 and 1-in-10 weather scenarios. Including weather variables in the modeling of RTP impacts would risk misattributing the effect of the price signals to the effect of weather. This would lead to incorrect estimates of program effects. All August Monthly Worst days are associated with the 'Hot Summer Weekday' RTP day type and have the same rate schedule applied. The decrease in impacts over time is attributable to a decline in program enrollment over the forecast horizon.

Table 2: RTP Aggregate Program Ex Ante Impacts (MW) - August Worst Day from 4pm-9pm

Forecast Year	SCE 1-in-2	SCE 1-in-10	CAISO 1-in-2	CAISO 1-in-10
2026	[redacted]			
2027				
2028				
2029				
2030				
2031				
2032				
2033				
2034				
2035				
2036				

Table 3: RTP Aggregate Program Ex Ante Impacts (MW) - 2026 August Worst Day

Hour Ending	SCE 1-in-2	SCE 1-in-10	CAISO 1-in-2	CAISO 1-in-10
Avg. 4pm-9pm	[redacted]			
17				
18				
19				
20				
21				

The RTP program can provide a small but measurable amount of demand response impacts during the 5pm-9pm period on Hot Summer Weekdays, when prices relative to the otherwise applicable tariff are high. The program has many customers who are dually enrolled in other demand response programs, making attribution of impacts challenging. Similarly, the program is dominated by several large industrial accounts that provide the majority of the load shed for the program. As a result, portfolio impacts averaged across the RA window tend to be small. Given the challenges of this evaluation – specifically the estimation of ex post and ex ante counterfactual loads – and the small portfolio load impacts, SCE should consider whether it is appropriate to evaluate this program on an annual basis going forward.

1 PROGRAM DESCRIPTION

The Real Time Pricing (RTP) program is a variable tariff-based demand response program for commercial and industrial customers in SCE's territory. The basis of the tariff is hour-specific generation energy prices that are set based on the prior day's daily maximum temperature in Downtown Los Angeles. Seven potential day types are available, including three summer weekday schedules, high and low-cost winter weekdays, and high and low-cost weekends. The rate is available to commercial, industrial, and agricultural customers on rates TOU-8, TOU-8 Standby, TOU-GS1, TOU-GS2, TOU-GS3, TOU-PA2 and TOU-PA3. Customers may be dually enrolled in other event-based demand response programs.

There were 86 customers enrolled on RTP rates at the end of PY2025, down from 93 in PY2024. As this program is rate-based, customer counts tend to fluctuate over time.

1.1 KEY RESEARCH QUESTIONS

The PY2025 evaluation of SCE's RTP program sought to answer these key research questions:

- What were the demand reductions for each RTP day type, monthly average weekday and monthly worst day? How do these results compare to the ex-post results from the prior year and why?
- How do load impacts differ for customers who have enabling technology and/or are dually enrolled in other programs?
- How do weather and event conditions influence the magnitude of demand response?
- How do load impacts vary for different customer sizes, locations, and customer segments?[1]
- What is the ex-ante load reduction capability for 1-in-2 and 1-in-10 weather conditions? And how well do these reductions align with ex-post results and prior ex-ante forecasts?
- What concrete steps can be undertaken to improve program performance?

1.2 PROGRAM DESCRIPTION

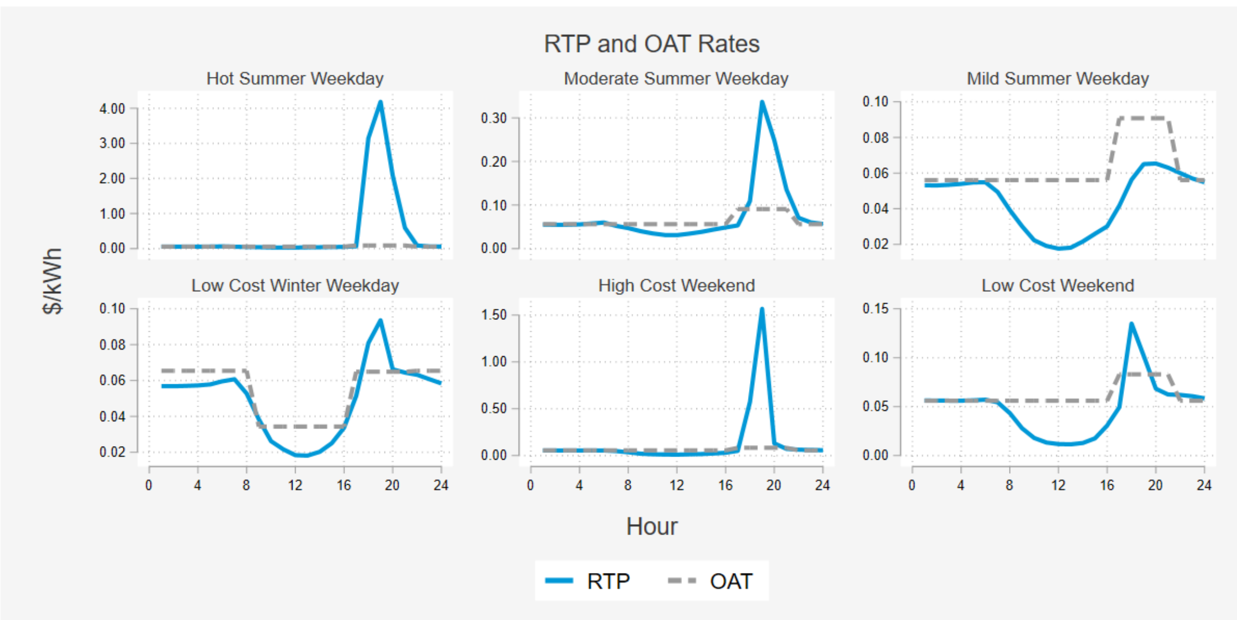
RTP offers commercial and industrial customers the opportunity to react daily to price signals and reduce loads when prices are high. Each day, the next days' hourly prices are tied directly to the daily maximum temperature in Downtown Los Angeles, grouped into one of seven day types: Hot Summer Weekday, Moderate Summer Weekday, Mild Summer Weekday, High Cost Winter Weekday, Low Cost Winter Weekday, High Cost Weekend and Low Cost Weekend. There were 86 customers enrolled in the RTP program at the end of the PY2025 season, the majority of which are on the TOU-8 rate, SCE's large industrial rate. While the analysis is performed for each customer using their specific RTP and OAT

[1] Pursuant to the Load Impact Protocol Process Guide (version 6.1, released by the Energy Division on March 5, 2026), large loads (e.g. data centers, EV fleet charging station load) should be reported as a distinct load type within ex-ante and ex-post table generators. Because the formal definition of "large loads" is still under development, this study does not incorporate large-load effects in either the ex post load impact estimates or the ex ante forecasts.

rates (i.e. GS-1 and GS-1-RTP), the graphs showing summary rate information in this report are constructed from TOU-8 and TOU-8-RTP rates, instead of showing the same graph for each combination of RTP and OAT rates for each of TOU-8, TOU-8-S, GS-1, GS2, GS-3, PA-2, and PA-3, for example. This is because the majority of RTP customers are on TOU-8-type rates and the differences in program rates are quite small.

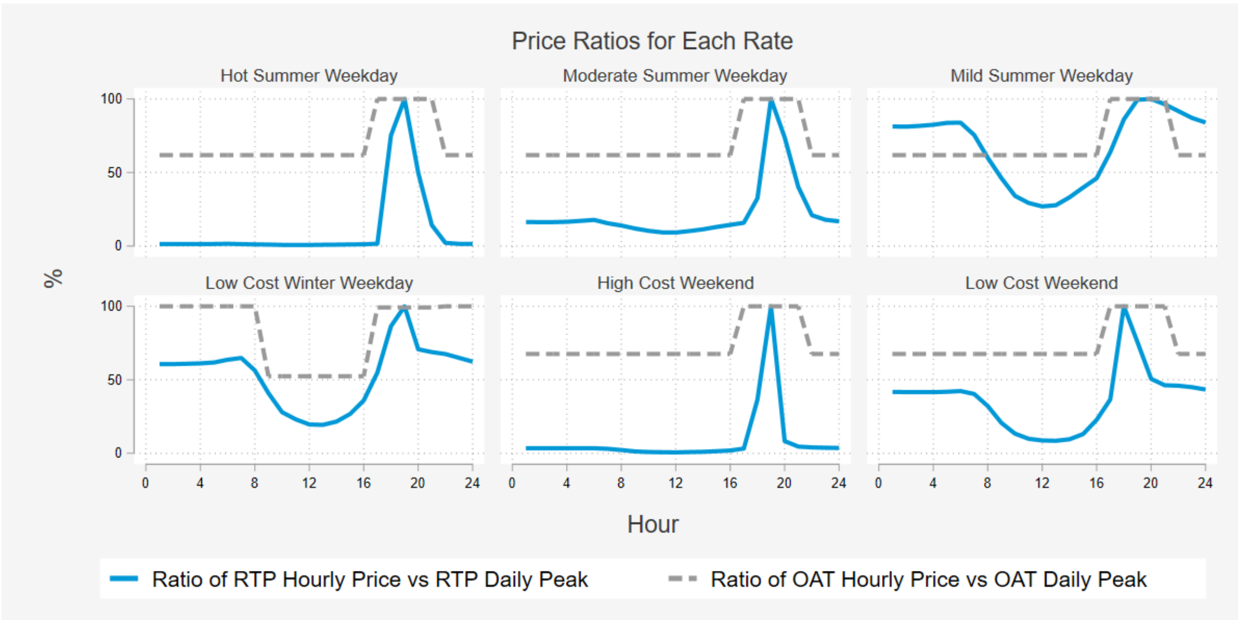
Figure 1 shows the rates experienced by day type for both the TOU-8 RTP and OAT, including normalized demand charges. In general, there is minimal difference between the RTP and OAT rates except for Hot Summer Weekdays and High Cost Weekends, where the difference between the two rates can exceed several dollars per kilowatt-hour.

Figure 1: Comparison of RTP and OAT Rates (TOU-8)



While the main goal of this evaluation is to assess the impact of being on the RTP rate compared to the OAT rate, it may also be helpful to assess the impact of the various RTP day types on customer consumption. Figure 2 shows the price ratios associated with each of the two rates, normalized to each rate's maximum value. In both cases, the highest rates that a customer experiences occur during the 4pm-9pm peak window on hot summer weekdays. However, the RTP peak rate is at its peak between 6-9pm only, a narrower peak than the OAT rate. In addition, the concentration of the price signal in those peak hours stands in contrast with the OAT rate, where in the summer period, peak prices are in place every weekday during the full 4pm-9pm window. The structure of the RTP rate concentrates prices exclusively during hours where the grid experiences peak capacity constraints, offset by very low prices in all other hours. Non-RTP rates, in contrast, do not have as strong of a price signal during peak hours, and therefore have less variability between peak and off-peak prices.

Figure 2: Ratio of Peak to Off-Peak Rates for RTP and OAT (TOU-8)



1.3 PARTICIPANT CHARACTERISTICS

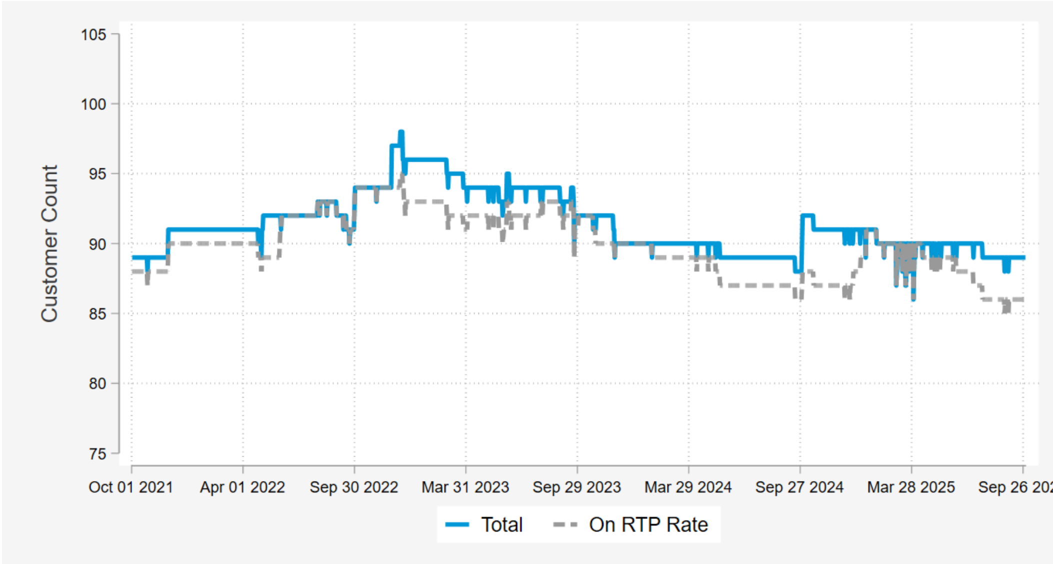
There were 86 commercial, industrial, and agricultural customers active on RTP as of the 2025 SCE peak day, August 21st, 2025. Table 4 summarizes their key characteristics. "Manufacturing" was the most common customer industry, followed by "Agriculture, Mining, and Construction" and "Offices, Hotels, Finance, and Services". Most customers are on the industrial TOU-8 rate.

Table 4: Participant Characteristics on 8/21/2025 (SCE Peak Day)

Category	Subcategory	Customer Mix
Industry	Manufacturing	34.9%
	Agriculture, Mining, Construction	20.9%
	Offices, Hotels, Finance, Services	19.8%
	Wholesale, Transport, Other Utilities	12.8%
	Schools	5.8%
	Institutional/Government	2.3%
	Unknown/Other	2.3%
	Retail Stores	1.2%
Local Capacity Area (LCA)	La Basin	81.4%
	Big Creek/Ventura	12.8%
	Non-Lca	5.8%
Rate Family	TOU-8	60.5%
	TOU-GS1	14.0%
	TOU-GS3	11.6%
	TOU-GS2	5.8%
	TOU-PA-2	4.7%
	TOU-8-S	3.5%
Size	Greater Than 200kW	74.4%
	20kW Or Lower	17.4%
	20-200kW	8.1%
Zone	Remainder Of System	64.0%
	South Of Lugo	19.8%
	South Orange County	16.3%

Figure 3 plots the RTP enrollment over time from the end of 2021 through the end of the PY2025 evaluation period. Enrollment in RTP has been steady since 2021, fluctuating between upper 80's and lower 90's counts. The peak enrollment was in early 2023 with 95 customers enrolled but has been steadily declining since that time. By the end of the 2025 evaluation period, 86 customers were enrolled in the RTP program.

Figure 3: RTP Enrollment over Time



1.4 2025 SUMMER CONDITIONS

RTP rate schedules are determined based on temperature conditions on the prior day in Downtown Los Angeles; essentially every day experiences a treatment, though the treatments themselves vary. The program year is divided into 7 day types: summer weekdays are classified as Hot, Moderate, or Mild; winter weekdays are designated as High-Cost or Low-Cost; and weekends are designated as High-Cost or Low-Cost. The temperature ranges for these dispatch types are shown in Table 5.

Table 5: Event Dispatch Criteria

Day Type	Dispatch Criteria (°F)
Hot Summer Weekday	≥ 91
Moderate Summer Weekday	81-90
Mild Summer Weekday	≤80
High Cost Winter Weekday	>90
Low Cost Winter Weekday	≤90
High Cost Weekend	≥78
Low Cost Weekend	<78

Table 6 gives a count comparison from year to year. PY2025 was comparable to PY 2024 but there were more Hot Summer Weekday day types in 2025.

Table 6: Count of Summer Days by Program Year

RTP Daytype	PY 2022	PY 2023	PY 2024	PY 2025
HOT SUMMER WEEKDAY	12	8	11	14
MODERATE SUMMER WEEKDAY	45	34	48	45
MILD SUMMER WEEKDAY	31	45	26	28
HIGH COST WINTER WEEKDAY	2	1	1	0
LOW COST WINTER WEEKDAY	168	171	172	171
HIGH COST WEEKEND	42	29	43	46
LOW COST WEEKEND	62	76	63	58

Figure 4 provides an initial indication of participants' price responsiveness relative to prior years. Average peak demand declines as temperature increases, mirroring the established relationship in prior years between RTP prices and temperature. This pattern suggests that participant demand responds to rising prices, with higher prices coinciding with lower peak-period load.

Figure 4: Relationship between Temperature and Average Peak Demand in PY2025 Summer

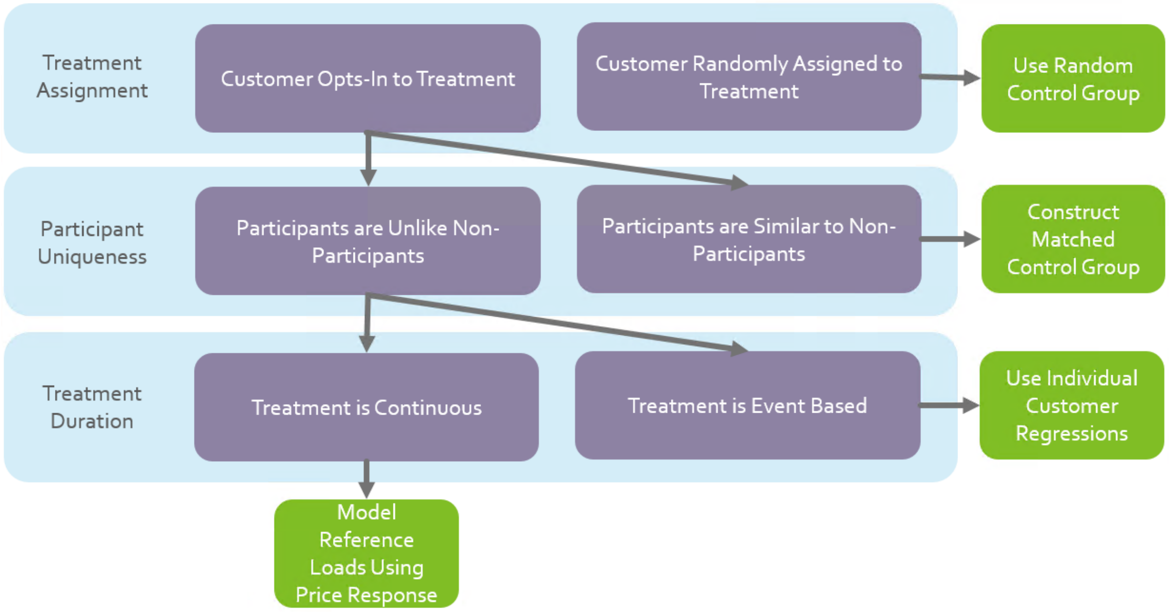
1.5 PROGRAM CHARACTERISTICS THAT INFLUENCE EVALUATION

A substantial challenge for the evaluation of rate-based demand response, especially when the program is one that a customer can opt in to, is the difficulty of finding a valid counterfactual. The counterfactual load for a customer enrolled in RTP is what the customer would consume if they were billed on their OAT. Because we cannot observe customers on the OAT, we must estimate it. The characteristics of the RTP participants and program design make this challenging and should be carefully considered as part of the evaluation planning process. The three characteristics that most affect the evaluation choice are:

- **Treatment assignment:** RTP customers opt into the program, which creates potential selection effects when comparing to customers who do not opt in to the program. Said another way, customers who opt in to RTP may be those that are more able to benefit from the program intervention, such as having flexible scheduling during peak hours or a dedicated on-site energy manager.
- **Uniqueness:** Participants are large and have unique loads and processes that make finding comparable customers difficult.
- **Treatment duration:** Unlike an event-based program (such as BIP or AP-I) where demand response is called on a handful of days every year, rate-based demand response is continuous. That is, once on the rate, customers generally remain on the rate. This presents a challenge for estimating load reductions, because pre-treatment data should not be used to construct a counterfactual. This is because doing so would make the strong assumption that no other conditions that affect energy use would have changed for each customer since the customer came on the RTP rate. As an example, using the pre-post approach for a customer who hypothetically enrolled in RTP at the beginning of March 2020 would misattribute the effects of the COVID-19 pandemic to the effect of being on the RTP rate.

A summary of the implications of these characteristics is shown in Figure 5. When customers can be randomly assigned a rate, such as when a default Time-of-Use rate is rolled out in staggered waves, there are customers who experience the OAT and who can function as a control. For the RTP program, however, customers opt into the program. Customers who opt in tend to be different than customers who do not. They likely have more flexibility in their loads, may be larger or smaller, may be more likely to be a standby customer, or in a particular industry or location. In some cases, a matched control group could be constructed to find a statistically similar population of customers to participants, however that approach requires that a similar group of non-participants exist in the population. For programs like RTP, which is mostly made up of large, unique customers, this is unlikely to be the case. What remains, then, is to use participant consumption data to model the counterfactual. This approach requires a sufficient amount of data to fit the model. This can be easy, as in the evaluation of the AP-I program, where events occur one or two days out of the year and the remaining days are unperturbed. When a demand response program operates continuously, as with RTP, pre-treatment data is likely to reflect an outdated model of how a customer operates. For a longstanding program such as RTP, there is very little validity to using this approach.

Figure 5: Evaluation Options for Non-Weather Sensitive Demand Response Programs



What remains, then, is a modeling exercise that will be described in the following section. Because RTP participants are exposed to a wide variety of prices while on the rate, the relationship between price signal and consumption can be estimated. By substituting the RTP price signal with the OAT price signal, a counterfactual reference load can be constructed.

One further complicating factor for the RTP evaluation concerns the inclusion of weather variables in both the ex post and ex ante regression modelling. For many individual customer regression methods, it is standard to use weather variables to explain variation in customer loads. However, because RTP day types are inherently dependent on weather, as they are defined by it, including weather as an explanatory variable in the regression can introduce confounding bias. Including weather variables in the model will misattribute the effect of the price signal to the change in weather, making the incorrect assumption that prices and weather are independent.

2 EVALUATION METHODOLOGY

Because of the long-standing RTP program option for commercial customers, and because the program is not dispatched on only a subset of days, the evaluation options to estimate load impacts are quite different than many other demand response programs. What is similar, however, is that to assess program impacts, we must construct load profiles for what the customer would have done had they not been on the RTP tariff. The appropriate counterfactual is the customer's consumption patterns on the OAT. For example, a customer on the GS-2 RTP tariff would otherwise be metered on the standard GS2 tariff.

The counterfactual was modeled using a price model that estimates the relationship between the price each customer segment is exposed to and their load. From that model reference loads can be constructed by predicting what customers would have done on the OAT using individual sector regressions. Table 7 and Table 8 summarize the evaluation approaches for the ex post and ex ante evaluations, respectively.

Table 7: Real-Time Pricing Ex-Post Approach

Methodology Component	Demand Side Analytics Approach
1. **Population or sample analyzed**	Analyze the full population of participants. Because most participants have been on the program for a long time, there is little available data from which to construct any comparison group. For that reason, we relied on individual segment regressions using a price model. These segment results were applied to aggregated datasets of all customer loads in that group. That is, to get customer response for a specific category, all consumption and price data for customers in that category were averaged together and the regression was run on that group.
2. **Data included in the analysis**	All 2024-2025 data for participants. Data from October 2024 through September 2025 was included.
3. **Use of control groups**	Because of the uniqueness of the target population, we relied on a quasi-within-subjects method for developing ex post impacts. Synthetic controls were added to the ex post model for each segment to explain other variation in loads.
4. **Model selection**	The final matching model is identified based on out-of-sample metrics for bias and fit. The process relies on splitting the dataset into training and testing data. The models are developed using the training data and applied, out-of-sample, to the testing data. For each of the models specified, we produce standard metrics for bias and goodness of fit. The best model is identified by first narrowing the candidate models to the three with the least bias and then selecting the model with the highest precision.
5. **Segmentation of impact results**	The results are segmented by: ▪ Rate/Otherwise Applicable Tariff ▪ LCA ▪ Enabling technology (Y/N) ▪ Dual enrollment (by program)

Methodology Component	Demand Side Analytics Approach
	▪ SubLAP The main segment categories are building blocks. They are designed to ensure segment level results add up to the total and to enable production of ex ante impacts, including busbar level results. We also produced results for additional categories, such as industry type.

Ex ante impacts for the RTP program are straightforward. Leveraging the model estimated for each customer in the ex-post analysis, both the predicted observed load and counterfactual reference load can be predicted using projected prices and weather scenarios.

Table 8: Real Time Pricing Ex Ante Approach

Methodology Component	Demand Side Analytics Approach
1. **Years of historical performance used**	PY2024 and PY2025 data were used to model ex-ante.
2. **Process for producing ex ante impacts**	The key steps were: ▪ Collect data on the current or future RTP and OAT tariffs for each rate class ▪ Estimate price sensitivity of participants during PY2024 and PY2025 ▪ Construct the price ratios associated with the ex-ante rates for the RTP day type associated with each ex ante day-type ▪ Use the price sensitivities to predict loads for RTP and OAT scenarios ▪ Combine the ex-ante reference loads, predicted RTP loads, and enrollment forecasts for each segment ▪ Aggregate to produce overall ex ante load impacts
3. **Accounting for changes in the participant mix**	Because the customer mix may evolve, changes in the participant mix need to be accounted for developing forecasts of reduction capability under planning conditions. From the outset, we produced a detailed segmentation – building blocks – so we can account for changes in the customer mix over the historical and forecast periods.
4. **Producing busbar level impacts**	The requirement to produce granular results for distribution planning is relatively recent. Because impacts are modeled, using individual customer regressions, impacts can easily be aggregated to whatever level of granularity is required, including at the busbar level. Unless other information is provided, we will scale impacts proportionately for even participation changes across busbars according to the ex-ante participation forecast.

2.1 OVERVIEW OF EVALUATION METHOD SELECTED

As discussed above, RTP impacts were modeled using individual cost category regressions that related price variations on a tariff to changes in hourly consumption. The first step in performing this estimation is to determine the prices that customers face during RTP participation and the otherwise-applicable tariff. Rates have several components that add up to what a customer must respond to in each hour. The approach taken for each category is summarized in Table 9.

Table 9: Rate Component and Approach

Cost Component	Category	Applies to	In Which Rate?	Approach
Delivery	Customer Charge	One-Time Monthly	Both	This charge does not vary with consumption and is identical in both RTP and OAT. This category can be ignored for this analysis.
Delivery	Energy Charge	TOU Rate Blocks	Both	Multiply kWh consumed in each rate block by TOU price
Delivery	Demand Charge	Overall	Both	Convert to kWh equivalent by dividing by total hours in month and spreading out
Delivery	Demand Charge	TOU Rate Blocks	Both	Convert to kWh equivalent by dividing by total hours in each rate block by month and spreading out
Generation	RTP Energy Charge	Hourly (Variable)	RTP	Apply to hourly consumption in appropriate day type/hour
Generation	OAT Energy Charge	TOU Rate Blocks	OAT	Multiply kWh consumed in each rate block by TOU price
Generation	Demand Charge	Overall	OAT	Convert to kWh equivalent by dividing by total hours in month and spreading out
Generation	Demand Charge	TOU Rate Blocks	OAT	Convert to kWh equivalent by dividing by total hours in each rate block by month and spreading out

Once each component has been normalized to an hourly per-kWh value, the components for either the RTP or OAT rates are summed.

SYNTHETIC CONTROLS

The RTP evaluation in PY2025 once again made use of synthetic control profiles to improve the accuracy of the ex-post impact estimation. Synthetic controls are included in the regression specification as right-hand-side variables and serve as a proxy for other unobserved characteristics that can affect customer loads. The synthetic control for each RTP customer was the hourly mean demand of all control customers who were in the same industry and rate type. For the eight industry / rate type combinations that had more than 300 customers, only the 300 customers that had demand closest to the mean treatment consumption were used. Figure 6 shows how control loads are highly correlated with participant loads. In effect, the control customer profiles can explain much of the variation in customer usage on a day-to-day basis, improving the accuracy of the predictions.

Figure 6: Synthetic Control Variation

[redacted]

Out of Sample Testing

To ensure that the model selected is accurately capturing the relationship between prices and consumption, each model was fitted on a training data set that included all of PY2025, withholding 3 of each RTP day type to use for testing the goodness of fit of the models fit using the training data. A comparison of the training days to the average day for RTP participants is shown in Figure 7.

Figure 7: Comparison of Training and Testing Days Selected for Out of Sample Testing

[redacted]

Ex Post Model

The framework for tested models is shown in Table 10 and Table 11. Each model had a price detail component, a day-type component, and a period fixed effect, and additionally were tested both with and without a synthetic control. Period definitions were evaluated under five alternative structures: (1) all hours treated as a single period, (2) two-period split distinguishing Resource Adequacy (RA) hours from non-RA hours, (3) a three-period split consistent with typical commercial occupancy patterns, (4) individual hours treated as their own periods, and (5) two-period split between Peak RTP hours (5pm – 8pm) and the rest. There were 14 x 5 x 2 x 5 unique combinations of these components, meaning that 700 models were assessed. The regressions were run separately for each site, and a winning model was chosen for each site.

Table 10: Regression Models Tested and Best Model by Customer

Model	Price Detail
1	price
2	log(price)
3	price + price ratio
4	price + price ratio + price squared
5	ln(price) + ln(price ratio)
6	ln(daily avg price) + ln(price ratio)
7	hour + price + price ratio
8	hour + price + price ratio + price squared
9	hour + ln(price) + ln(price ratio)
10	hour + ln(daily avg price) + ln(price ratio)
11	hour x price + hour x price ratio
12	hour x price + hour x price squared + hour x price ratio
13	hour x ln(price) + hour x ln(price ratio)
14	hour x ln(daily avg price) + hour x ln(price ratio)

Day Type	Day Type Detail
1	month
2	month + weekday
3	month x weekday
4	hour x month + weekday
5	hour x month x weekday
Period Option	**Detail**
1	All hours in one period
2	RA (17-21) vs non-RA
3	5am-4pm, 5pm-9pm, all other hours
4	Each hour is a period
5	RTP Peak Hours (18-20) vs others

Table 11: Definition of Regression Terms

Category	Model Term	Description
Base	kW_{ih}	Electricity delivered in kW for customer i, in hour h
	α_{0h}	Intercept
	ε_{ih}	Error term
Price	price	Hourly energy price inclusive of demand charges
	price squared	Square of hourly energy price
	price ratio	Ratio of hourly price to the daily max price
	ln(price)	Natural log of hourly price
	ln(price ratio)	Natural log of the price ratio
	ln(daily avg price)	Natural log of the daily average price
Month/Week	weekday	Day of week indicators grouping Monday, Tuesday-Thursday, Friday, and Weekends/Holidays
	month	Month indicator variable
Synthetic Control	ctrl_kwh_all	Profile of average RTP-like control customer
Hour	hour	Hour fixed effect

As discussed at the end of Section 1.5, including weather variables in the regression models can introduce bias in the estimates – even for weather sensitive customers – and should be avoided. The best[2] model for each site was then used to predict ex-post loads on the withheld days. Figure 8 shows the predicted loads for the training data days. More detail, including a summary of model fit statistics, can be found in the appendix.

[2] Method for selecting best model is described in the appendix.

Figure 8: Out of Sample Predictions on Proxy Days, Ex Post

[redacted]

Ex Ante Reference Load Model

The reference load modeling approach for ex ante is very similar to ex post, with the notable exclusion of synthetic control profiles, as these do not have an ex ante equivalent data stream. The coefficients for the price components were captured from the ex post model and applied to the ex ante rates in combination with a model that fit category usage by month and hour. This approach has the benefit of leveraging the mode with the best explanatory power – from the ex post analysis – and simply incorporating an additional year of data to develop ex ante RTP and OAT predictions. As such, no ex ante model fit metrics were developed as the results would be essentially identical to the ex post out of sample testing as described in the previous section.

To produce ex ante load shapes, updated rates[3] were used to predict both the reference load, under the otherwise applicable tariff, and the expected observed load, under the RTP rate. Because no weather variables were included, the models only depend upon day type (weekday or weekend) and price signals to estimate variation in loads. As ex ante weather scenarios all have different weather conditions, small changes in temperature may categorize the average weekday or monthly worst day into different RTP day types, however the loads themselves do not depend upon daily weather conditions.

---

[3] The rates used for ex ante modeling were taken from SCE's website as effective from January 1, 2026.

The California load impact protocols strongly suggest using multiple years of data to provide the model with a wider range of weather and economic conditions to estimate the relationship of various factors to load changes. For the RTP program, however, no weather variables were included in the ex post model for the reasons outlined above. As such, variability in weather conditions are not applicable to producing ex ante reference loads.

3 EX POST RESULTS

This section details the results of the ex post analysis, with particular attention paid to the program's performance during the summer months, and the general impact of RTP prices on customer loads.

The RTP rate is designed to produce load reductions during key hours on hot days. This targeted approach is shown in the RTP rates overall, where customers experience high rates between 5-8pm on hot summer weekdays, and relatively discounted rates in all other hours. However, the OATs for these customers would expose them to relatively higher rates in the 4pm-5pm window, therefore resulting in relatively higher loads for RTP customers in this period, as shown in Figure 9. Because program impacts are reported over the full peak window (4pm–9pm), observed load reductions during the highest-priced RTP hours are substantially smaller in the shoulder hours, resulting in a dilution of average peak-period impacts on Hot Summer Weekdays.

Figure 9: OAT Peak Hours vs RTP Peak Hours on the Average Hot Summer Weekday

The same information for Moderate Summer Weekdays is shown below in Figure 10. In the peak hours, the overall OAT rate is higher than the RTP rate, leading to relatively higher loads for RTP customers and load impacts that are close to zero for the RTP program.

Figure 10: OAT Peak Hours vs RTP Peak Hours on the Average Moderate Summer Weekday

[redacted]

3.1 OVERALL RESULTS

At the program level, RTP participants curtailed [redacted] of load reduction during the 4pm-9pm window during Hot Summer Weekdays. However, it is important to keep in mind that most of the large RTP customers are dually enrolled in BIP and therefore operate with lower and less flexible peak-period load when BIP events are called within the period where RTP rates are highest. Nevertheless, these RTP participants still respond on Hot Summer Weekdays when BIP events are not called and can deliver value to the grid across the 10-12 non-BIP event Hot Summer Weekdays that occur each year.

The average ex post impacts by RTP day type are shown in Figure 11. Most RTP day types experience essentially no impacts while Hot Summer Weekdays show a load reduction during peak hours. When OAT prices are higher than RTP prices, load increases relative to the otherwise applicable tariff can occur.

Figure 11: Average Ex Post Impacts by RTP Day Type

[redacted]

Table 12 summarizes ex-post impacts by month and RTP day type. Across most of the year, both average weekdays and monthly peak days fall under Low Cost Winter, Mild Summer, or Moderate Summer Weekdays, during which estimated impacts are small and not statistically distinguishable from zero. In contrast, substantial load reductions are observed during Hot Summer Weekdays in August and September, when RTP prices diverge most sharply from the otherwise applicable tariff.

Table 12: Ex Post Impacts by Day Type for All Customers*

RTP Day Type	# Cust	Average Customer (kW)				% Impact	Agg. Impact (MW)
		Ref. Load	Obs. Load	Impact	95% CI		
January - Average Weekday: Low Cost Winter Weekday	88	[redacted]					
January - Monthly Worst Day: Low Cost Winter Weekday	88						
February - Average Weekday: Low Cost Winter Weekday	87						
February - Monthly Worst Day: Low Cost Winter Weekday	87						
March - Average Weekday: Low Cost Winter Weekday	90						

RTP Day Type	# Cust	Average Customer (kW)				% Impact	Agg. Impact (MW)
		Ref. Load	Obs. Load	Impact	95% CI		
March - Monthly Worst Day: Low Cost Winter Weekday	90	[redacted]					
April - Average Weekday: Low Cost Winter Weekday	90	[redacted]					
April - Monthly Worst Day: Low Cost Winter Weekday	90	[redacted]					
May - Average Weekday: Low Cost Winter Weekday	89	[redacted]					
May - Monthly Worst Day: Low Cost Winter Weekday	89	[redacted]					
June - Average Weekday: Mild Summer Weekday	87	[redacted]					
June - Monthly Worst Day: Moderate Summer Weekday	87	[redacted]					
July - Average Weekday: Mild Summer Weekday	87	[redacted]					
July - Monthly Worst Day: Moderate Summer Weekday	87	[redacted]					
August - Average Weekday: Moderate Summer Weekday	86	[redacted]					
August - Monthly Worst Day: Hot Summer Weekday	86	[redacted]					
September - Average Weekday: Hot Summer Weekday	86	[redacted]					
September - Monthly Worst Day: Hot Summer Weekday	86	[redacted]					
October - Average Weekday: Low Cost Winter Weekday	88	[redacted]					
October - Monthly Worst Day: Low Cost Winter Weekday	88	[redacted]					
November - Average Weekday: Low Cost Winter Weekday	87	[redacted]					
November - Monthly Worst Day: Low Cost Winter Weekday	87	[redacted]					

RTP Day Type	# Cust	Average Customer (kW)				% Impact	Agg. Impact (MW)
		Ref. Load	Obs. Load	Impact	95% CI		
December - Average Weekday: Low Cost Winter Weekday	87	[redacted]					
December - Monthly Worst Day: Low Cost Winter Weekday	87						

** Results here are shown for the RA window (4pm-9pm from June-October and 5pm-10pm in all other months)*

On the following pages load profiles for August 21st, 2025, the SCE System Peak Day, are shown in Figure 12 and Figure 13.

Figure 12: Average Customer Ex Post Impacts on August 21st, 2025

[redacted]

Figure 13: Aggregate Ex Post Impacts on August 21st, 2025

[redacted]

To get a better sense of the average program impacts across day types, the average PY2025 ex post peak period impacts are summarized in Table 13. Hot Summer Weekdays deliver the largest savings, with impacts declining on Moderate and Mild Summer Weekdays. High-Cost Weekends also show reduced peak-period consumption relative to Low-Cost Weekends. While the program can deliver up to [redacted] during peak periods on average, performance on individual days will vary.

Table 13: Ex Post Peak Period Impacts by Average Day Type

RTP Day Type	# Dispatched	Average Customer (kW)					Agg. Impact (MW)
		Ref. Load	Obs. Load	Impact	95% CI	% Impact	
Hot Summer Weekday	86	[redacted]					
Moderate Summer Weekday	86						
Mild Summer Weekday	87						
Low Cost Winter Weekday	88						
High Cost Weekend	87						
Low Cost Weekend	88						

3.2 RESULTS BY CATEGORY

In the following tables, values are reported for key RTP customer segments on the average Hot Summer Weekday. It's important to note that these results will not match the load impact tables, as the load impact tables show only an example 'monthly worst day' and 'average weekday' on a given day per month. This change was done for several reasons:

- It is a more representative summary of the ex post performance over the prior year.
- The individual ex post days are now noisier on a day-to-day basis with the inclusion of synthetic controls. The synthetic controls provide more estimation precision at the average event day level but can obscure the day-to-day effects of the program.
- It helps facilitate the comparison to ex ante impacts, since ex ante relies on all of the ex post data rather than just snapshots of individual days.

Table 14 shows the impacts from the program on the average Hot Summer Weekday by Local Capacity Area (LCA). The highest impacts came from the [redacted], which delivered [redacted] from 4pm-9pm on the average Hot Summer Weekday. This was primarily due to the large customer size and price responsiveness of these customers. Average reference loads for the LA Basin customers were nearly [redacted]

Table 14: Ex Post Impacts by LCA on Average Hot Summer Weekday

LCA	# Enrolled	Average Customer (kW)					Agg. Impact (MW)
		Ref. Load	Obs. Load	Impact	95% CI	% Impact	
Outside LA Basin	5	[redacted]					
Big Creek/Ventura	11	[redacted]					
LA Basin	70	[redacted]					
All Customers	86	[redacted]					

Table 15 shows the results from PY2025 for the average Hot Summer Weekday by Zone. In the zones affected by the San Onofre Nuclear Generating Station (SONGS) closing, customers delivered [redacted] of load reduction during the full event hours. The highest average customer impacts were delivered by customers in the [redacted], who delivered on average [redacted] of load relief per participant.

Table 15: Ex Post Impacts by Zone on Average Hot Summer Weekday

Size	# Enrolled	Average Customer (kW)					Agg. Impact (MW)
		Ref. Load	Obs. Load	Impact	95% CI	% Impact	
South Orange County	14	[redacted]					
South of Lugo	17	[redacted]					
Remainder of System	55	[redacted]					
All Customers	86	[redacted]					

3.3 COMPARISON TO PRIOR YEAR

Participant reference loads have increased in PY2025 compared to PY2024. This increase in load is associated [redacted]. This change can also be seen in Figure 14 which compares PY2024 ex post, PY2025 ex post and PY2024 ex ante usage loads.

Figure 14: Comparison of PY2024 Ex Post and Ex Ante to PY2025 Ex Post

[redacted]

Table 16 compares PY2024 Ex Post and Ex Ante with PY2025 Ex Post on Hot Summer Weekdays. This table summarizes the average across all days of each month for Hot Summer Weekdays to capture the distributions of peak period impacts. There were no Hot Summer Weekdays in June or July of 2025 or July of 2024.

Table 16: Comparison of PY2025 to PY2024 Ex Post and Ex Ante Average Customer Reference Loads and Impacts (kW)

Day Type	Year	Type	Portfolio	Average # Customers	June		July		August		September	
					Ref	Imp	Ref	Imp	Ref	Imp	Ref	Imp
Hot Summer Weekday	PY2025	Ex Post	Portfolio	86	[redacted]							
Hot Summer Weekday	PY2025	Ex Post	Program	86								
Hot Summer Weekday	PY2024	Ex Post	Portfolio	88								
Hot Summer Weekday	PY2024	Ex Post	Program	88								
Hot Summer Weekday	PY2024	Ex Ante	Portfolio	65								
Hot Summer Weekday	PY2024	Ex Ante	Program	85								

4 EX ANTE RESULTS

This section summarizes the results of the ex ante impact estimation process for RTP from 2026 to 2036.

4.1 ENROLLMENT FORECAST

Table 17 shows the expected future enrollment for the RTP program. Enrollment is expected to decrease over the next few years with 86 participants enrolled at the end of PY2025, 83 in August of 2026, and eventually stabilizing at 79 participants. Enrollment trends in this forecast are extrapolated from historic net de-enrollment rates.

Table 17: RTP Ex Ante Enrollment Forecast in August from 2025-2035

Program/Portfolio	2026	2027	2028	2029	2030	2031	2032-2036
Portfolio	65	62	62	62	62	62	62
Program	83	79	79	79	79	79	79

4.2 OVERALL RESULTS

As RTP is a rate-based program, any changes in the RTP rate significantly affect the forecasted load impacts. As of January 1st, 2026 the prices are the same as they were the previous year, shown in Figure 15.

Figure 15: Comparison of Ex Post to Ex Ante RTP Rates

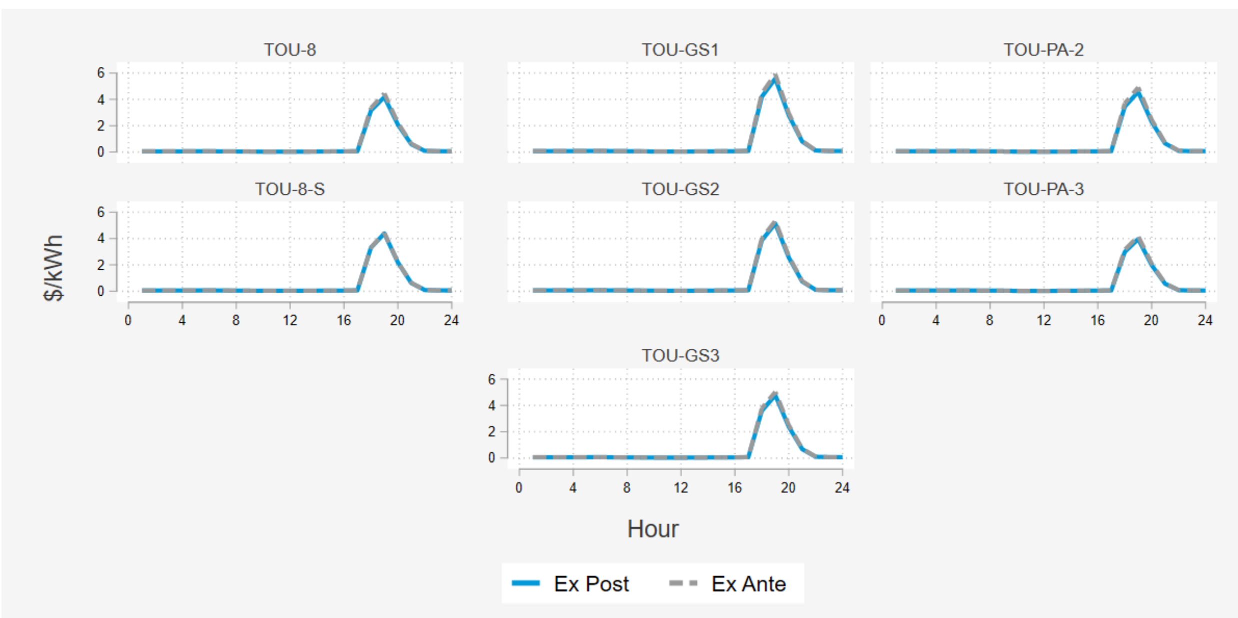


Figure 16 shows the average Program Ex Ante Profiles for RTP and OAT on hot summer days by month. Reference loads and expected program impacts are generally similar across each of these months.

Figure 16: Average Customer Program Ex Ante Profiles by Month on Hot Summer Weekdays

[redacted]

Table 18 and Table 19 contain a summary of the impacts by forecast year for both the program and portfolio values. Similar to the ex-post modeling, no weather variables are included in the ex-ante specification, so the only difference between these scenarios is the RTP day type associated with the CAISO and SCE 1-in-2 and 1-in-10 weather scenarios. All August Monthly Worst days are associated with the 'Hot Summer Weekday' RTP day type and have the same rate schedule applied, which is why impacts are the same across each weather scenario. Finally, the decrease in impacts over time is attributable to a decline in program enrollment over the forecast horizon. [redacted]

- [redacted]
- [redacted]
- [redacted]

Table 18: RTP Aggregate Program Ex Ante Impacts – Average over RA Hours on August Worst Day (MW)

Forecast Year	SCE 1-in-2	SCE 1-in-10	CAISO 1-in-2	CAISO 1-in-10
2026	[redacted]	[redacted]	[redacted]	[redacted]
2027	[redacted]	[redacted]	[redacted]	[redacted]
2028	[redacted]	[redacted]	[redacted]	[redacted]
2029	[redacted]	[redacted]	[redacted]	[redacted]
2030	[redacted]	[redacted]	[redacted]	[redacted]
2031	[redacted]	[redacted]	[redacted]	[redacted]
2032	[redacted]	[redacted]	[redacted]	[redacted]
2033	[redacted]	[redacted]	[redacted]	[redacted]
2034	[redacted]	[redacted]	[redacted]	[redacted]
2035	[redacted]	[redacted]	[redacted]	[redacted]
2036	[redacted]	[redacted]	[redacted]	[redacted]

Table 19: RTP Aggregate Portfolio Ex Ante Impacts – Average over RA Hours on August Worst Day (MW)

Forecast Year	SCE 1-in-2	SCE 1-in-10	CAISO 1-in-2	CAISO 1-in-10
2026	[redacted]			
2027	[redacted]			
2028	[redacted]			
2029	[redacted]			
2030	[redacted]			
2031	[redacted]			
2032	[redacted]			
2033	[redacted]			
2034	[redacted]			
2035	[redacted]			
2036	[redacted]			

Load impacts also vary by month, as weather patterns change the mix of RTP day types that are dispatched in the ex ante scenario. Table 20 shows the average customer impacts for a monthly worst day. In some cases the difference between an average (1-in-2) year compared to an extreme (1-in-10) year are enough to shift the RTP day type customers are subjected to. In those cases, impacts are expected to change as well.

Table 20: RTP Average Customer Portfolio Ex Ante Impacts – Average over RA Hours By Monthly Worst Day in 2036 (kW)

Day Type	SCE 1-in-2	SCE 1-in-10	CAISO 1-in-2	CAISO 1-in-10
January Worst Day	[redacted]			
February Worst Day	[redacted]			
March Worst Day	[redacted]			
April Worst Day	[redacted]			
May Worst Day	[redacted]			
June Worst Day	[redacted]			
July Worst Day	[redacted]			
August Worst Day	[redacted]			
September Worst Day	[redacted]			
October Worst Day	[redacted]			
November Worst Day	[redacted]			
December Worst Day	[redacted]			

Figure 17 and Figure 18 show the results on an August monthly worst day under SCE 1-in-2 conditions at the program and portfolio level.

Figure 17: Portfolio Aggregate Ex Ante Impacts for SCE 1-in-2 August Worst Day

[redacted]

Figure 18: Program Aggregate Ex Ante Impacts for SCE 1-in-2 August Worst Day

[redacted]

4.3 RESULTS BY CATEGORY

Table 21 shows the average program forecasted ex ante impacts for the August monthly worst day by ex ante LCA and weather conditions. The majority of ex ante impacts are expected to come from the [redacted]. This group of customers is both large and price-sensitive, which means that they can contribute significant load reductions.

Table 21: RTP Aggregate Program Ex Ante Impacts (MW) – Average over RA Hours on August Worst Day by LCA

LCA	Weather Year	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036
Big Creek/ Ventura	CAISO 1-in-10	[redacted]										
	CAISO 1-in-2	[redacted]										
	SCE 1-in-10	[redacted]										
	SCE 1-in-2	[redacted]										
LA Basin	CAISO 1-in-10	[redacted]										
	CAISO 1-in-2	[redacted]										
	SCE 1-in-10	[redacted]										
	SCE 1-in-2	[redacted]										
Outside LA Basin	CAISO 1-in-10	[redacted]										
	CAISO 1-in-2	[redacted]										
	SCE 1-in-10	[redacted]										
	SCE 1-in-2	[redacted]										

4.4 COMPARISON TO PRIOR YEAR

Table 22 compares prior-year and current year ex-ante forecasts with realized ex-post results for Hot Summer Weekdays. While somewhat informative, interpreting results still presents a challenge due to the extent that the patterns of large customers on any given year can dominate the results. Average customer loads and impacts are highly sensitive to changes in the largest participating customers, resulting in a small number of high-load accounts materially shifting portfolio-level averages from year to year, even when underlying program operations remain consistent.

[redacted]

Table 22: Comparison of PY2024 and PY2025 Ex Post and Ex Ante Average Customer Reference Loads and Impacts (kW) on Hot Summer Weekdays

Year	Type	Portfolio	Average # Customers	June		July		August		September	
				Ref.	Impact	Ref.	Impact	Ref.	Impact	Ref.	Impact
PY2025	Ex Post	Portfolio	[redacted]								
		Program	[redacted]								
PY2025	Ex Ante	Portfolio	[redacted]								
		Program	[redacted]								
PY2024	Ex Post	Portfolio	[redacted]								
		Program	[redacted]								
PY2024	Ex Ante	Portfolio	[redacted]								
		Program	[redacted]								

Figure 19 compares ex ante predictions between PY2024 and PY 2025. Because in both cases, ex ante estimates are based on two years of data, the load shapes between the years are broadly similar across months and hours.

Figure 19: Comparison of PY2024 and PY2025 Program and Portfolio Average Ex Ante Loads on Hot Summer Weekdays

[redacted]

[redacted] In addition, the ex post results are estimated using only the most recent year of data, whereas ex ante estimates are result of the combined PY2024 and PY2025 load data. These differences in participant composition and training data drive differences between ex post and ex ante, but in general, ex ante results between PY2024 and PY2025 are comparable.

Figure 20: PY2024 to PY2025 Ex Post and Ex Ante Comparison

[redacted]

5 DISCUSSION

RTP delivered approximately [redacted] of load reduction during 4-9pm on the average Hot Summer Weekday in PY2025. Ex ante prediction of capability during August Worst Monthly Days is approximately [redacted] at the program level. Differences between ex post and ex ante results are due to [redacted] and planned enrollment declines. In general, RTP loads are dependent on:

- Large, unique customers with operating schedules that vary from year to year and season to season
- Lower RTP prices relative to the participant's otherwise applicable tariffs
- A mix of dual enrollments in other programs
- Weather conditions in a given summer, including if the hottest days are consecutive

The largest concentrations of impacts were among [redacted]. Outside of hot summer weekdays, the modelling did not detect statistically and economically significant impacts of the program.

The RTP program can provide a measurable amount of demand response impacts during the 5pm-8pm period on Hot Summer Weekdays, when prices relative to the otherwise applicable tariff are high. The program's biggest customers are dually enrolled in other demand response programs, making attribution of impacts challenging. Similarly, the program is dominated by several large industrial accounts that provide the majority of the load shed for the program. [redacted]. As a result, portfolio impacts averaged across the RA window tend to be small. Given the challenges of this evaluation – specifically the estimation of ex post and ex ante counterfactual loads – and the small portfolio load impacts, SCE should consider whether it is appropriate to evaluate this program on an annual basis going forward.

Furthermore, consideration should be given to simplifying the RTP rates. Moderate price days do not clearly impact consumption, and are so close to the OAT rates, that it is unclear that customers are making any changes when these prices are in effect. Previous years have also showed how the weather trigger, not load forecasts, for next day pricing meant that strong price signals were not sent on all peak days. A day-ahead CAISO forecast would better align prices with the goals of the program.

6 APPENDIX: EVALUATION METHODOLOGY

6.1 DEMAND RESPONSE EVALUATION METHODS

The primary challenge of impact evaluation is the need to accurately detect changes in energy consumption while systematically eliminating plausible alternative explanations for those changes, including random chance. Did the dispatch of demand response resources cause a decrease in hourly demand? Alternatively, can the differences be explained by other factors? To estimate demand reductions, it is necessary to estimate what demand patterns would have been in the absence of dispatch – this is called the counterfactual or reference load. At a fundamental level, the ability to measure demand reductions accurately depends on four key components:

- The effect or signal size – The effect size is most easily understood as the percent change. It is easier to detect large changes than it is to detect small ones. For most DR programs, the percentage change in demand is relatively large.
- Inherent data volatility or background noise – The more volatile the load, the more difficult it is to detect small changes. Energy use patterns of homes with air conditioners tend to be more predictable than industrial or agricultural load patterns.
- The ability to filter out noise or control for volatility – At a fundamental level, statistical models, baseline techniques, and control groups – no matter how simple or complex – are tools to filter out noise (or explain variation) and allow the effect or impact to be more easily detected.
- Sample/population size – For most of the programs in question, sample sizes are irrelevant because we analyzed data for the full population of participants using AMI data. Sample size considerations aside, it is easier to precisely estimate average impacts for a large population than for a small population because individual customer behavior patterns smooth out and offset across large populations.

In general, there are seven main methods for estimating demand reductions, as summarized in Table 23. The first four only make use of use patterns during days when DR is not dispatched to calculate the baseline. The latter three methods incorporate non-event data but also use an external control group to establish the baseline. The control group consists of customers who are similar to participants, experienced the same event day conditions, but are not dispatched during events (or were not transitioned to time-varying pricing). Control and participant groups should have similar energy usage patterns when the intervention is not in place and diverge when the intervention is in effect. The only systematic difference between the two groups should be that one is dispatched for events (or transitioned to time-varying prices) while the other group is not.

Table 23: Methods for Demand Response Evaluation

General Approach		Method	Method Description
Use non-event days only to establish the baseline	1	Day matching baseline	This approach relies on electricity use in the days leading up to the event to establish the baseline. A subset of non-event days in close proximity to the event day are identified (e.g., Top 3 of 10 prior days). The electricity use in each hour of the identified days is averaged to produce a baseline. Day matching baselines are often supplemented with corrections to calibrate the baseline to usage patterns in the hours preceding an event – usually referred to as in-day or same-day adjustments.
	2	Weather matching baseline	The process for weather matching baselines is similar to day-matching except that the baseline load profile is selected from non-event days with similar temperature conditions and then calibrated with an in-day adjustment.
	3	Regression models (interrupted time series)	Regression models quantify how different observable factors such as weather, hour of day, day of week, and location influence energy use patterns. Regression models can be informed by electricity use patterns in the day prior (day lags) and in the hours before or after an event (lags or leads) and can replicate many of the elements of day and weather matching baselines.
	4	Machine learning (w/o external controls)	Most machine learning approaches (e.g., random forest, neural networks, etc.) rely exclusively on non-event day data to establish the baselines. The algorithms test different model specifications and rely on a training and testing datasets (out-of-sample testing) to identify the best model and avoid overfitting.
Use non-event days plus a control group to establish the baseline	5	Matched control groups	Matching is a method used to create a control group out of a pool of nonparticipant customers. This approach relies on choosing customers who have very similar energy use patterns on non-event days and a similar demographic and geographic footprint. The non-event day data is incorporated by either analyzing the data using a regression model, a difference-in-differences model, or both.
	6	Synthetic control groups	This approach is similar to matching except that multiple controls are used and weighted according to their predictive power during a training period. A key advantage of this approach is that it can be used to produce results for individual customers.
	7	Randomized control trials	Participants are randomly assigned to different groups, and one group (the "control" group) is withheld from dispatch to establish the baseline. The control group provides information about what electricity use would have been in the absence of DR dispatch – the baseline. The estimate is refined by netting out any differences between the two groups on hot non-event days (difference-in-differences).

Approaches that use an external control group typically provide more accurate and precise results on an aggregate level when there are many customers (i.e., several hundred). They also make use of non-event days to establish the baseline but have the advantage of also being informed by the behavior of the external control group during both event and non-event days. Except for synthetic controls, the two fundamental limitations to control groups have been: the limited ability to disaggregate results, and the inability to use control groups for large, unique customers. The precision of results for control group methods rapidly decreases when results are disaggregated, and a control group cannot be used to estimate outcomes for individual customers (except for synthetic controls).

Methods that rely only on non-event days to establish the baseline – such as individual customer regressions – are typically more useful for more granular segmentation. Individual customer regressions have the benefit of easily producing impact estimates for any number of customer segments. Because they are aggregated from the bottom up, the results from segments add up to the totals. However, the success of individual customer regression hinges on having non-event days comparable to event days. When most of the hottest days are event days, as has been the case historically, estimating the counterfactual requires extrapolating trends to temperature ranges that were not experienced during non-event days. This produces less accurate and less reliable demand reduction estimates for the hottest days when resources are needed most.

6.2 MODEL SELECTION

A key question every evaluator must address is how to decide which model produces the most accurate and precise counterfactual. In many instances, multiple counterfactuals are plausible but provide different estimated demand reductions. Model selection plays a role both in developing matching models and for individual customer regressions.

Our process for model selection relies on splitting the data into testing and training days and implementing an out-of-sample testing process. First, we define testing and training days. Days with actual events are not included in either the training or testing days. Next, ten or more model specifications are defined. Because the treatment is not activated during either the training or testing days, the impacts are by definition zero. Any estimated impact by models is in fact due to model error. Third, we run each of the models using the training data and predict out-of-sample loads for the testing days. Fourth, the testing data out-of-sample predictions are compared to actual electricity use and used to calculate metrics for bias and fit. Next, the best model is identified by first narrowing the candidate models to the three with least bias (or with % bias less than 1%) and then selecting the model with the best fit. Finally, the best performing model is applied to all days and used to estimate the counterfactual for actual event days. The final model is designed to produce load impacts (treatment effects) for each event day and hour. Figure 21 illustrates the process.

Figure 21: Model Selection and Validation

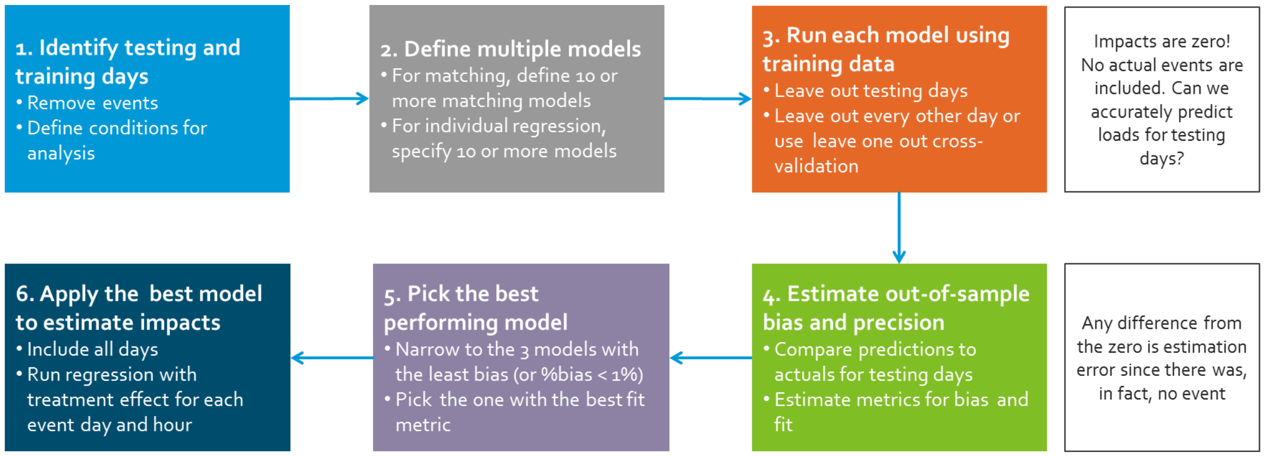


Table 24 summarizes the metrics for bias and precision we employ. Bias metrics measure the tendency of different approaches to over or under predict and are measured over multiple days. The mean percent error describes the relative magnitude and direction of the bias. A negative value indicates a tendency to under predict, and a positive value indicates a tendency to over predict. This tendency is best measured using multiple days and hours. The precision metrics describe the magnitude of errors for individual events days and are always positive. The closer they are to zero, the more precise the results. The mean percentage error is used to narrow down to the three models with the least bias. The Relative RMSE metric is used to identify the most precise and final model among the remaining candidates.

Table 24: Definition of Bias and Precision Metrics

Type of Metric	Metric	Description	Mathematical Expression
Bias	Average Error	Absolute error, on average	$AE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)$
	% Bias	Indicates the percentage by which the measurement, on average, over or underestimates the true demand reduction.	$\%\ Bias = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)}{\bar{y}}$
Precision	Root mean squared error (RMSE)	Measures how close the results are to the actual answer in absolute terms, penalizes large errors more heavily	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$
	Relative RMSE	Measures the relative magnitude of errors across event days, regardless of positive or negative direction. It can be thought of as the typical percent error, but with heavy penalties for large errors.	$CV(RMSE) = \frac{RMSE}{\bar{y}}$

Table 25 shows the out of sample testing results for the best model for each participant, summarized by rate family. The process to pick the best model overall relied on a combination of visual and statistical tests to identify the best model. The results of the out of sample fit metrics are listed below.

Table 25: Best Model Out of Sample Fit by Rate Family

Rate	Observed Usage	Avg. Error	% Bias	cvRMSE
TOU-8	[redacted]			
TOU-8-S				
TOU-GS1				
TOU-GS2				
TOU-GS3				
TOU-PA-2				