Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

FINAL REPORT

Early M&V Report for Program Year 2023
Energy Efficiency Summer Reliability Program

Prepared for: Pacific Gas and Electric Company
By: Demand Side Analytics

April 11, 2024
CALMAC ID: PGE0500.01

ACKNOWLEDGMENTS

DSA Research Team

- Stephanie Bieler
- Josh Bode
- Patrik Karlovic

Pacific Gas and Electric Company Team

- Megan Geraci
- Ezana Kemmer
- Larsen Plano
- Mark Salavitch
- Brian Arthur Smith

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY

Pacific Gas and Electric Company (PG&E) contracted with Sunrun to implement the Energy Efficiency Summer Reliability Program (EESRP) in 2023. Marketed to customers as the Peak Power Rewards Program, the objective of EESRP was to reduce Peak (5 p.m. to 9 p.m.) and Net Peak (7 P.M. to 9 p.m.) energy demands from August through October, seven days a week. This program involved recruiting single-family homeowners with rooftop solar and battery systems who were incentivized to allow Sunrun to manage their battery discharge efficiency, especially during Net Peak hours.

This report, an Early M&V[1] analysis commissioned by PG&E, aims to assess the accuracy of various Population-Based Normalized Metered Energy Consumption (Population NMEC) methods to estimate payable and claimable savings for battery storage within the EESRP. Additionally, it seeks to identify the most effective method for estimating future claimable savings.

An initial analysis was conducted to compare the results derived from hourly whole-building electric interval data (from PG&E) and 15-minute battery discharge data (from Sunrun). This approach was important to understand the impact of battery discharge on energy use at participating sites, especially considering how energy consumption patterns were influenced by factors beyond weather, season, and time of day. Key findings of this analysis include:

- EESRP saw engagement from 8,483 PG&E customers enrolled, which resulted in an estimated energy savings of 10,349 MWh and 10,563 MWh as measured by the battery discharge data (supplied by Sunrun) and hourly interval data (supplied by PG&E), respectively. These energy savings are the potential estimates of what claimed savings would be but were not claimed by PG&E toward its energy efficiency portfolio goals.
- The average estimated per customer peak reduction is 1.68 kW when estimated using the battery discharge data and 1.69 kW when estimated using hourly interval data.
- Customers with SolarEdge and Delta brand inverters showed markedly lower peak kW savings (0.70 kW) as compared to customers with the Tesla brand inverter (3.97 kW), when calculated using hourly meter data. This difference may be the direct result of customers with SolarEdge and Delta batteries experiencing an intervention in the pre-period, while customers with the Tesla batteries did not. This finding underscores the need to ensure that baselines are based on uniform conditions.

[1] An Early M&V evaluation commissioned by a program administrator "seeks to validate key savings assumptions and to better understand how savings are achieved for the purpose of improving programs." See Decision 10-04-029 (April 21, 2010), p. 25.

Public

The report also explores the accuracy of estimating energy savings using "end-use" data (that is, energy consumed from the discharge of batteries captured by Sunrun meters) and "whole home" data (that is, energy as measured using PG&E's net meters) through Population NMEC methods. The accuracy assessment of the Energy Efficiency Summer Reliability Program resulted in several key findings:

1. The Time-of-Week and Temperature (TOWT) model and the Difference-in-Difference (DiD) model with controls were the most effective, both for end-use and whole-home data sources.
2. Incorporating battery end-use data significantly enhanced model accuracy and precision, surpassing results of methods that incorporated only whole-home data.
3. In the context of model evaluation for battery programs, error metrics for peak times (7-9 pm) perform better than those calculated annually (over 8760 hours). This is because annual metrics are less reliable due to their reliance on small denominators. Focusing on peak demand periods provides more accurate insight into model performance.[2]
4. The large sample size (> 5,000 customers) allowed for robust estimations, meeting FSU targets with a range of savings from 3% to 15%.
5. Savings varied across inverter brands, with Tesla batteries showing more significant savings than SolarEdge and Delta. The TOWT model, while generally effective, was insufficient in capturing battery behavior during the atypical conditions in the baseline period that resulted in this finding.
6. The individual-matched controls DiD method demonstrated effectiveness in the whole-home evaluation, despite its limitations. Although individually matched controls provide value in a pre-post analysis, they do not capture the entirety of the impact due to a subset of participants exhibiting consumption patterns not observed within any of the control group members.
7. Recommendations resulting from the analyses include:
    a. Revising baseline construction methods to consider undisturbed load patterns;
    b. Incorporating battery end-use data for more accurate baseline establishment and model evaluation, and
    c. Adding additional right-hand weather variables, such as Solar Irradiance and Cloud Cover, to bolster Population NMEC analysis models for battery programs.

[2] Many readings of battery charge/discharge throughout a year centralize around zero and so measurement of annual effects are low but traditional error metrics (e.g., Mean Percent Error) produce large measurement errors due to having denominators very close to zero even when in reality measurement error could be +/-.01 kW.

2 INTRODUCTION

On July 30, 2021, Governor Newsom signed an emergency proclamation to "free up energy supply to meet demand during extreme heat events and wildfires that are becoming more intense and to expedite deployment of clean energy resources this year and next year."[3] The proclamation directs all energy agencies, including the California Public Utilities Commission (CPUC or Commission) to act immediately to achieve energy stability during this emergency.

In response, the CPUC received comments from parties to the assigned Administrative Law Judge's e-mail ruling seeking input on actions that the Commission could take specific to energy efficiency (EE) and reliability to help support the intent of the proclamation and the Commission's overall goals. Then, the Commission issued Decision (D.) 21-12-011 (12-8-2021)[4] which orders the IOUs to take action to prepare for potential extreme weather in the summers of 2022 and 2023.

The PG&E Energy Efficiency Summer Reliability (EESRP), marketed as Peak Power Rewards Program, is a direct result of that decision. EESRP was designed to support PG&E's summer reliability efforts by providing sustained and scheduled customer load shifting/modification services to PG&E during the 2023 summer period.

PROGRAM DESCRIPTION

PG&E collaborated with Sunrun, a provider in the residential solar and battery space, to launch the EESRP with the objective of mitigating Peak (4 p.m. to 9 p.m.) and Net Peak (7 p.m. to 9 p.m.) energy demands during the months of August – October, seven days a week in 2023.This initiative was part of a broader strategic approach to optimize energy usage across various timeframes, classified distinctly by PG&E as follows:

- **Peak Hours:** The interval between 4 p.m. and 9 p.m. on business days, extending from June 1st to September 30th.
- **Net Peak Hours:** This category encompasses the period from 7 p.m. to 9 p.m. on weekdays, within the same date range as mentioned above.
- **Non-Peak Hours:** This term refers to all remaining hours that do not fall within the peak periods.

Sunrun recruited single-family home participants owning battery systems. The primary objective was to leverage advanced charge/discharge algorithms, aiming to maximize the discharge efficiency of on-site batteries particularly during PG&E's Net Peak Hours. This operational focus of EESRP spanned from August through October of 2023.

Participants were notified of their enrollment for this initiative by Sunrun and were compensated with an incentive of $750 along with a Smart Thermostat (if one hadn't been provided one as part of a PG&E

[3] Available at https://www.gov.ca.gov/wp-content/uploads/2021/07/Energy-Emergency-Proc-7-30-21.pdf?emrc=fe927f

[4] Available at https://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M429/K805/429805997.PDF

EE rebate previously). This compensation was in exchange for granting Sunrun the authorization to manage their battery systems. It is noteworthy that customer inclusion in the program was defaulted to qualified customers, yet it provided an opt-out clause for those preferring not to participate.[5]

REPORT OBJECTIVE

The purpose of this Early M&V report is twofold. Its primary purpose is to provide an accuracy assessment to determine the best method for estimating payable and claimable savings for battery storage for the EESRP. Its secondary purpose is to provide an estimate of potential claimable savings utilizing the best method identified in the accuracy assessment.

It is important to note that this program did not claim any EE savings, and that the objective of this specific early M&V effort was to evaluate method accuracy and suitability of the program design for NMEC. The non-claimable program savings in this report span from August through October 2023, as well as for an entire weather-normalized year.

The Demand Side Analytics (DSA) team produced estimates of potential gross electric savings that would have been attributed to the EESR program using Population NMEC methods.[6] These methods relied on whole-building granular (hourly) electric AMI data from PG&E and battery end-use discharge data from Sunrun to estimate the savings associated with the battery discharge savings at participating sites.

This analysis also included an accuracy assessment to determine which population NMEC methods were the most accurate for estimating program savings. Prior to the year 2020, most Population NMEC savings methods were simple pre/post models that included weather, seasonal, and hourly components as explanatory variables in the regression model. Inherent to these models is the assumption that weather, season, and time of day factors were sufficient to build robust counterfactual models of participant energy consumption for large, relatively homogenous residential customer groups. With the effects of shelter-in-place orders across California during the height of the COVID-19 pandemic that resulted in profound disruptions in population energy use, it became clear that this assumption could not hold.

As post-COVID research analyses confirms, Difference-in-Differences(DiD) Population NMEC models that incorporate information about non-participant consumption significantly improves the quality of Population NMEC methods.[7] To quantify the improvements of adding a comparison group, a variety of

---

[5] The program enrollment strategy used by EESRP (whereby qualified customers were enrolled without explicit consent and provided the opportunity to disenroll) is referred to as "opt-out" or "default" recruitment. Offering customers an opportunity to participate in a program, and requiring their explicit consent prior to enrollment, is referred to as "opt-in" recruitment.

[6] For further reading on population NMEC, refer to the California Public Utility Commission's NMEC Rulebook. The latest version of the rulebook can be found at https://www.cpuc.ca.gov/-/media/cpuc-website/files/legacyfiles/n/6442463694-nmec-rulebook2-0.pdf. Note that an updated rulebook is forthcoming in 2024.

[7] For example, see Pacific Gas and Electric Co.: NMEC Control Group Accuracy Assessment (Demand Side Analytics, 2022), available at https://www.calmac.org/publications/PGE0476.01.pdf.

regression models with controls were tested, and the results of that assessment are also included in this report.

DATA SOURCES USED IN THIS EVALUATION

The program's primary goal is to reduce energy consumption during the Net Peak hours of 7 p.m. to 9 p.m. There are two primary sources of data used to estimate savings resulting from EESRP:

- "Whole-Home" electricity usage as tracked by PG&E Smart Meters. This refers to the hourly AMI interval data collected by PG&E's net meters that track all of the electricity flowing into, and out of, participant households.
- "Battery End-Use" data as tracked by Sunrun inverters. This refers to the 15-minute battery charge/discharge data collected by Sunrun.

In addition to the data referenced above, Sunrun provided participant household level electricity consumption data pulled from customer inverters. It was compared to PG&E AMI data and found to be sufficiently similar (within ±2%). As both the Sunrun household data and the Sunrun battery charge/discharge data are pulled from the same inverters, we assume that both data sets will have the same level of accuracy.

As requested, Sunrun provided additional data on the batteries as available, including state of charge; share (%) of battery reserved for backup; whether the premise was being dispatched for a DR event at any point in time, and customer battery operation settings (e.g., backup only, self-powered, time-based control balanced, time-based control cost-savings). If the customer has enabled time-based control, the peak hours set by the customer were requested as well.

POTENTIAL ANNUAL WEATHER-NORMALIZED NET PEAK SAVINGS FROM EESR PROGRAM COHORT

Using the data from the program period, DSA projected the potential savings for the EESRP if it were to run for an entire calendar year, assuming a similar rate of participation but under typical meteorological conditions. Table 1 presents the potential weather-normalized Net Peak (kWh/MWh) savings for all customers participating in the EESRP. Savings for each site are estimated using actual program performance (August 2023 – October 2023) and modeled for the entire year. This projection suggests that the participant cohort would be net contributors of energy (kWh) to the grid, which is reflected in net negative annual energy consumption. The variance in predicted impacts by data source are discussed in other sections of the report.

Table 1: Peak Weather Normal kWh/MWh Savings by Data Source and Vendor[8]

Data Source	Cohort	Premises in Enrollment Data	Premises in Analysis	% Coverage	Average Annual Predicted Consumption (kWh)	Average Annual Actual Consumption (kWh)	Per Customer Impact (kWh)	Percent Impact	Energy Savings (MWh)	Analysis Days
Battery End-Use	SolarEdge + Delta	5,917	2,988	50%	-1357.5	-1734.2	376.7	-28%	2,229	365
	Tesla	2,566	1,457	57%	-400.5	-3565.1	3,164.6	-790%	8,120	365
	All EESRP	**8,483**	**4,445**	**52%**	**-1068.0**	**-2288.0**	**1,220.0**	**-201%**	**10,349**	**365**
Whole-Home AMI	SolarEdge + Delta	5,917	3,683	62%	218.3	-291.6	509.8	-234%	3,014	365
	Tesla	2,566	1,436	56%	654.4	-2226.7	2,881.0	-440%	7,548	365
	All EESRP	**8,483**	**5,119**	**60%**	**350.2**	**-876.9**	**1,227.1**	**-389%**	**10,563**	**365**

[8] Premises are included in the analysis only if a complete year of interval data was available before any intervention was implemented.

3 ACCURACY ASSESSMENT METHODOLOGY AND RESULTS

To date, studies using Population NMEC methods have relied on the use of utility AMI (whole building) data and excluded battery end-use data from their models. Sites with solar and battery installations have been excluded from participant cohorts due to their relatively low incidence in customer populations and as a result Population NMEC-based models that incorporate battery end-use data have yet to be well explored. One of the main methods used for Population NMEC, CalTRACK,[9] explicitly excludes solar and battery storage and, based on initial tests, does not perform well even when comparison groups are introduced. In addition, sites with solar and battery are fundamentally different than most residential sites. With AMI data, the solar and battery patterns dominate, and the share of weather sensitive loads is smaller. Moreover, homes with solar and battery are more prone to non-routine events (such as the adoption of electric vehicles during the baseline or performance years).

As the adoption of solar and battery installations continues to grow, so does the need for a test of the accuracy of Population NMEC methods to assess programs that include customers as participants. Thus, the test includes an accuracy assessment of six different methods that vary by regression model, the use of comparison groups, an examination of how controls are incorporated, and the inclusion/exclusion of battery end-use data along with AMI data. Once the most accurate method is identified, it will be applied to estimate energy savings for EESRP and serve as a model for future evaluations moving forward.

To determine the best-practice model to assess battery storage savings estimates for the program, the DSA team conducted an accuracy assessment to test the performance of various regression models. Accuracy assessments provide critical information about how well a particular model and its specifications can account for relative drivers of energy consumption in the context of battery technology.

Accuracy assessments involve judging how well a statistical model represents participant consumption patterns absent an intervention. These approaches will typically require either:

1. Contemporary data from pseudo-participants – customers who look like actual participants based on eligibility screening or observable characteristics. This is typically referred to as a quasi-experimental design.
2. Pre-intervention data from actual participants – consumption data for periods prior to the intervention. This is typically referred to as a pre/post (within groups) approach.

Both approaches work well; however, whenever possible, option 2 should be selected as it avoids any assumptions about how similar pseudo-participants are to the actual participants. However, the within

[9] This model is a variant of the Time-of-Week and Temperature (TOWT) model developed by Lawrence Berkeley National Laboratory (Matthieu, J.L., P.N. Price, S. Kiliccote, and M.A. Piette, "Quantifying Changes in Building Electricity Use, With Application to Demand Response" (2011) available at: https://eta-publications.lbl.gov/sites/default/files/LBNL-4944E.pdf. More information about CALTRACK is available at: https://docs.caltrack.org/en/latest/methods.html

Public

groups approach relies on the assumption that exogenous impacts on energy use (e.g., macroeconomic factors) have similar impacts on the entire cohort of participants. PG&E has provided DSA with two years of pre-intervention data for EESRP participants, which allows for estimating the premise-level counterfactual using unperturbed loads. This historical usage data is the critical component of an accuracy assessment: the model is developed based on a period prior to an intervention was in place, so the "true answer" is known.

The procedure for such an estimation mimics the savings estimation: the regression model of interest is estimated on one year of pre-intervention data and the resulting coefficients are used to predict a back-casted counterfactual in the year prior to the pre-intervention year. An example of this approach for a batch of residential sites is shown in Figure 1. The gray highlighted period is the baseline year before any energy efficiency measures are installed. The model is fit on the observed data in this period and then predicted both for the post-intervention year (highlighted in orange) and the out-of-sample year (in blue) before the baseline period. Any difference between the observed and counterfactual loads in the pre-intervention periods is a model error, while post-intervention, the difference also incorporates actual program savings. While the error is lowest in-sample – during the baseline period, the error in the out-of-sample period reflects the ability of the model to explain participant consumption trends.

Figure 1: Out-of-Sample Testing Example

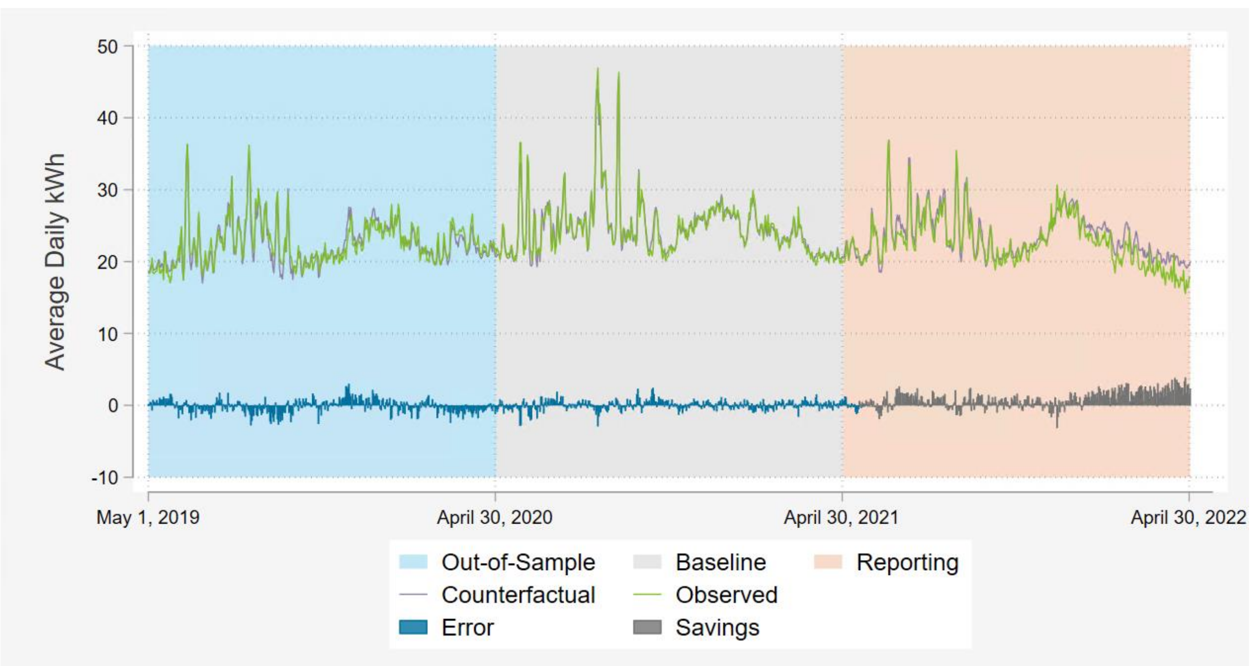


Table 2 provides an overview of the tested methodologies. In the remainder of this section, the model frameworks will be discussed in detail, along with a summary of how each model is assessed. Results are discussed in section 3, Accuracy Assessment Methodology and Results.

Table 2: Summary of Tested Methodologies

Parameter	Options Tested
Sectors	Residential Battery Customers participating in EESRP
Fuels	Electric
Control Customers	The control customers come from the pool of sites selected to be part of the granular profiles. Only includes customers with a battery installed a year prior to Program Start. ▪ Control pool of 32,000 Battery Customers
Regression Frameworks	Four sets of models are tested: • CalTRACK Daily and Hourly Methods using Difference-in-Differences • Manual Difference-in-Differences with a control group • Seasonal Time-of-Week and Temperature Model with control profiles as right-hand-side (explanatory) variables, including a second difference, with and without temperature lags[10]
Comparison Groups	Comparison groups are formed either through: • Individual customer matching • Granular profiles
Period	DSA was provided data for EESSRP participants in the years prior to their program intervention. The models were run on this period to assess accuracy when no interventions were in place. Sites began participating in the program from July 1, 2023 to October 31,2023, their baseline period would be July 1, 2022 to June 30, 2023.
Accuracy assessed on	• Fractional Savings Uncertainty (<25% at 90% confidence) • Mean Percent Error • Normalized Root Mean Squared Error • Sum of Squared Errors • Percent Bias

[10] Temperature lags and therm interpolation tests were requested specifically to be analyzed for the residential gas sector only. All other sectors and fuels relied on same-day only temperature splines.

POPULATION NMEC ACCURACY AND PRECISION REQUIREMENTS

PG&E relies on Population NMEC methods to determine net and gross savings associated with several EE programs. While the details of the regression methods may vary from program to program, there are fundamental principles that underlie any assessment of program savings. The ability to measure energy savings accurately using these methods depends on four key components:

1. **The effect or signal size** – The effect size is most easily understood as the percent change in energy use following the intervention. It is easier to detect significant changes than it is to identify small ones.
2. **Inherent data volatility or background noise** – The more volatile the load, the more difficult it is to detect small changes attributable to the intervention. Non-routine events effectively add noise to the data (or signal).
3. **The ability to filter out noise or control for volatility** – Statistical models – no matter how simple or complex – are tools to reduce noise (or unexplained variation) and allow the effect or impact to be detected more easily.
4. **Sample/population size** – The full participant population is analyzed as a group. Regardless, it is easier to precisely estimate average impacts for a larger population than for a smaller population because individual customer behavior patterns "smooth out" and offset individual customer volatility across larger populations.

Largely, the signal size (1) and data volatility (2) are fixed consequences of program design: specifically, the targeted participant sector and the types of installed measures. Similarly, program population size (4) is a function of the available participants meeting eligibility criteria, incentive levels, and the amount and type of outreach done. At the time of evaluation, the only controllable component is the choice of statistical model (3).

The statistical model should be selected before the savings estimation to avoid "cherry picking" among savings estimates to select the method with the most favorable outcome (least amount of noise). In particular, the NMEC Rulebook sets precision requirements for any Population NMEC method to ensure that savings are accurately estimated. The specific statistic for assessing precision is the Fractional Savings Uncertainty (FSU), which measures the relative margin of error of the estimate at a given level of confidence. Population NMEC methods must not have more than a 25% FSU at the 90% confidence level. For example, a group expected to save 5,000 MWh must not have a margin of error of more than ±1,250 MWh at 90% confidence to meet this requirement.

MODEL FRAMEWORKS ASSESSED

Variations of four main model frameworks were tested in this assessment for whole-home data individual customer regressions with the inclusion of synthetic controls, manual difference-in-differences models, and CalTRACK difference-in-differences models with controls and individually matched controls and synthetic controls. The details on the construction of these models are described in more detail below. These models were tested with one of two types of comparison groups: individual customer-matched controls and aggregated 8760 (hourly) profiles of non-participant data. More detail about how these comparison groups were constructed can be found in subsequent sections of this report. Table 3 provides a summary of the model frameworks for each method tested.

Table 3: Summary of Model Frameworks Tested

No.	Method	Regression Equation Used	Who are participants being matched to?	How is the control group incorporated?	How is DiD calculated?
1	CalTrack DiD w/GP	CalTRACK TOWT	GP Segment	Difference in differences	Recurve Difference in differences
2	Synthetic Control with Matched GP	Alternative TOWT	GP Segment	Right-hand side variable	N/A
3	Individually Matched Control DiD	N/A	Stratified Euclidean distance matching with Sunrun non-participants	Difference in differences	Standard Difference in differences
4	CalTRACK DiD with Individual Matched Control	CalTRACK TOWT	Stratified Euclidean distance matching with Sunrun non-participants	Difference in differences	Recurve Difference in differences
5	TOWT model without controls	Alternative TOWT	N/A	N/A	N/A
6	CalTRACK model without controls	CalTRACK TOWT	N/A	N/A	N/A

Public

SYNTHETIC CONTROLS

Synthetic control models are premise-level regressions that incorporate standard temperature, season, day-of-week, and hour features, but also add in one or more aggregated profiles[11] of hourly consumption of non-participants as additional explanatory variables. When included in the regression, these profiles allow exogenous factors that influence energy use to be accounted for in the model. The granular profiles (GPs) are constructed from the energy data of multiple customers who have not installed energy efficiency measures over the evaluation period and are in the same region, have the same solar status (with or without rooftop solar installed), and are of similar size to the participant.[12] Their use is aggregated together and uses the same time basis as the participant's consumption. That is, if the participant's baseline period spans July 1, 2022 to June 30, 2023 and their post-treatment period goes through program end, the granular profile would contain hourly data of the constituent sites during that same time range.

A typical regression specification for this approach is shown in Equation 1, with terms explained in Table 4. This regression would be used in the baseline period to estimate the coefficients listed, which would then be used to predict in the out-of-sample period to assess accuracy. It is referred to as a Seasonal Time-of-Week and Temperature model because it can be run independently for each season. Currently, seasons are defined as Summer (June-September), Winter (November-February) and Shoulder (all other months). Note that this equation can be easily adjusted to be estimated on a daily, rather than hourly, basis for gas data (daily gas data is available for residential PG&E customers).

Equation 1: Seasonal Time of Week and Temperature Model

$$kWh_{p,t} = \sum_{i=1}^{168} (\beta_i * I_{i,t}) + \sum_{b=1}^{b=[2,7]} (\gamma_b * B_{b,t}) + \sum_{g=0}^{n} (\delta_g * GP_{g,t}) + \varepsilon_{p,t}$$

[11] Aggregated profiles, also referred to as granular profiles (GPs), are hourly (8760 for electric) load profiles that are constructed from actual PG&E customer interval data so as to be representative of the loads of customers within specific customer segments.

[12] Residential GP candidates are segmented into groups based on electric heat status (as defined by the electric end-use rate code within PG&E's customer database). For sites without electric heat, they are further segmented based on size bins within each climate zone group. The size bins are constructed on the basis of relevant premise characteristics: installed solar capacity for solar customers, annual kWh for non-solar customers, and annual therms for gas customers. More information is available at: https://www.calmac.org/granular/Granular_Profile_Overview_and_Background.docx

Table 4: Definition of Equation 1 Terms

Symbol	Interpretation
$kWh_{p,t}$	The observed kWh consumption for participant *p* in date-hour *t* (or date *t*, for daily models). Note that for gas models, the analysis is conducted in therms, not kWh.
β_i	The coefficient representing the base energy consumption for hour-of-week *i*, above or below the participant average. Note that for daily models, this represents day-of-week-specific base consumption
$I_{i,t}$	A dummy variable for each hour-of-week *i*. Equal to 1 when date-hour *t* is in that hour-of-week, and 0 otherwise. Note that for daily models, this represents a day-of-week dummy variable
γ_b	The coefficient representing the marginal consumption associated with a one-degree change in outdoor temperature for temperature bin *b*
$B_{b,t}$	The value of temperature bin *b*.
δ_g	The coefficient representing the marginal effect of one kWh (therm) change in the comparison group granular profile *g*.
$GP_{g,t}$	The average consumption of the granular comparison group profile *g* in date-hour *t* (or date *t*, for daily models).
$\varepsilon_{p,t}$	The error term for participant *p* in date-hour *t* (or date *t*, for daily models)

Difference-in-Differences

The simple difference-in-differences (DiD) model relies on a straightforward assumption: any exogenous factors affecting energy use in the comparison group apply equally to the participants. Said another way, savings for a given participant can be estimated by differencing out the model's estimate of comparison group "savings" from the model's estimate of participant energy change.

The mechanics of a DiD model involves comparing usage among a comparison group, selected independently for each participant, and treatment group in both the pre-intervention (baseline) and post-intervention (reporting) periods. The pre-intervention difference in usage between the treatment and comparison group is subtracted from the post-treatment difference in usage between those same groups to estimate the effect of treatment, and to control for exogenous impacts of energy use that are unrelated to the intervention (the counterfactual). In this way, the DiD model provides both an estimate of savings (or error, in the out-of-sample period) and a counterfactual.

CalTRACK Difference-in-Differences

Broadly, CalTRACK methods are a set of procedures that rely on temperature and time variables to model premise-level loads. There are two main modeling strategies defined in this framework: 1) analyze consumption by hour and 2) model energy usage at the daily or billing-period level. For simplicity, this analysis focuses on hourly methods as prior testing has indicated that results between daily and hourly methods are quite similar. The recommended framework for this strategy is to run a premise-level regression using time-of-week dummy variables, temperature bins, and an occupancy flag as right-hand-side variables. This model is then fit on a pre-intervention period and used to predict for the post-intervention period. The CalTRACK Hourly regression specification is displayed in Equation 2, with terms explained in Table 5.

Equation 2: CalTRACK Hourly Model

$$kWh_{p,t} = \sum_{i=1}^{168}(\beta_i * I_{i,t}) + \sum_{b=1}^{b=[2,7]}(\gamma_b * B_{b,t}) + \sum_{i=1}^{168}(occupied_i * \delta_i * I_{i,t}) + \sum_{i=1}^{168}\sum_{b=1}^{b=[2,7]}(occupied_i * \theta_{b,i} * B_{b,t}) + \varepsilon_{p,t}$$

Table 5: Definition of Equation 2 Terms

Symbol	Interpretation
$kWh_{p,t}$	The observed kWh consumption for participant *p* in date-hour *t*
β_i	The coefficient representing the base energy consumption for hour-of-week *i*, above or below the participant average when the premise is unoccupied
$I_{i,t}$	A dummy variable for each hour-of-week *i*. Equal to 1 when date-hour *t* is in that hour-of-week, and 0 otherwise
γ_b	The coefficient representing the marginal consumption associated with a one-degree change in outdoor temperature for temperature bin *b* when the premise is unoccupied
$B_{b,t}$	The value of temperature bin *b*.
$occupied_i$	A dummy variable indicating high energy usage in hour-of-week *i*.
δ_i	The coefficient representing the additional base energy consumption for hour-of-week *i* when a premise is occupied
$\theta_{b,i}$	The coefficient representing the additional marginal consumption associated with a one-degree change in outdoor temperature for temperature bin b when the premise is occupied during hour *i*.
$\varepsilon_{p,t}$	The error term for participant p in date-hour *t*

To incorporate a comparison group via a difference-in-differences approach, this regression specification is run for both the participant and either a matched control (pseudo-participant) or an aggregated profile as if that control profile had the same treatment start and end dates as their matched participant. The results for the participants and their controls are then differenced from each other, on a percent basis, in the post-treatment period.

GRANULAR PROFILE SEGMENTATION AND VALIDATION

Each tested model required the construction of a comparison group of customers like those participating in the interventions. Traditionally, these comparison groups have been constructed based on individual customer-matched controls selected via propensity score matching. A challenge to the application of the individual customer-matched controls method is the amount of non-participant customer energy usage data required, first to identify the customers for the comparison group and ongoing through the reporting period to complete the analyses. While the DSA team tested the performance of individual matched controls, granular profiles offer an alternative approach in place of individual non-participant data.

Using granular nonparticipant profiles ("granular profiles" or "profiles") is one way to overcome the challenges of using individual customer matched controls. Granular profiles are interval data streams that are representative of the energy use of discrete segments of customers that can be used as the building blocks for comparison groups. Granular profiles, in contrast to non-participant interval data, are anonymized (that is, they contain no personally-identifiably information) so they can be made available to the public.

PG&E's granular profiles are aggregations of loads from customers that have not recently installed any energy efficiency measures and that have been grouped into relevant segments based on an extensive segmentation assessment conducted by DSA in the summer of 2023. The granular profiles available for the PG&E service territory are comprehensive and representative of unique residential and non-residential segments of PG&E's service population. Although the granular profiles developed do not include battery customers, solar profiles were used for model testing to monitor exogenous changes between battery and solar participants.

SEGMENTATION OF GRANULAR PROFILES

The number of granular profiles for the PG&E service territory has increased since their original development. A summary of the currently available non-participant profile segments is provided in Table 6. Climate Zone Groups are based on CEC Building Climate Zones[13] and are grouped in to four categories based on their similarity: Coastal (Zones 1, 3, and 5); Inland (Zones 2 and 4); North Central Valley (Zones 11 and 12), and South-Central Valley (Zone 13). The result is 160 residential electric profiles. Because granular profiles for battery customers are limited, the testing only included solar profiles which resulted in 80 residential electric profiles.

[13] See https://www.energy.ca.gov/programs-and-topics/programs/building-energy-efficiency-standards/climate-zone-tool-maps-and

Table 6: Granular Profile Segments

Electric	Gas
▪ Climate Zone Groups (4 Groups) ▪ Solar Status (2 Groups) ▪ Size (Annual kWh for Non-Solar, Solar Install Size for Solar Customers), or Electric Heat Designation (5 Groups) ▪ % of Consumption in Mid-Day (4 Groups)	▪ Climate Zone Groups (4 Groups) ▪ Size based on annual consumption (4 Groups)

CONSTRUCTION AND VALIDATION OF THE GRANULAR PROFILES

Production of granular profiles relies on identifying valid premises that can be sampled into the profiles. At each monthly refresh of profile data, the existing sites selected to be part of the profiles at the prior generation are re-screened to ensure they still meet the criteria defined in Table 7. If they do not, they are removed from the profiles going forward, and a pre-selected alternate site is introduced in its place.

Table 7: Granular Profile Candidate Requirements

Electric	Gas
▪ Full panel of data for the prior year ▪ No change in solar status (no addition of onsite solar, no adding of incremental capacity) ▪ No EV rates ▪ No other DERs (e.g., batteries) ▪ No EE program participation in the last 12 months	▪ Full panel of data for the prior year ▪ No EE program participation in the last 12 months

Validation of each set of granular profiles is conducted upon each data refresh to ensure outlier usage readings, unusually large sites, interval data gaps and/or other data issues do not compromise the integrity of the profiles. Profiles are additionally screened to ensure that they conform to historic weather sensitivity, do not exhibit uncharacteristic volatility, and/or exhibit abrupt changes in weekly consumption patterns.

MAPPING PARTICIPANTS TO COMPARISON GROUPS

Each regression model framework described above requires comparison groups to be constructed. In all cases, the choice of comparison groups to incorporate in modeling is another parameter to vary in an accuracy assessment. For each participant in a dataset, various comparison groups can be mapped to any customer. With synthetic controls, because profiles function as right-hand-side (predictive) variables, more than one non-participant profile can be included as part of the comparison group. By contrast, for a DiD method, only one non-participant profile can be assigned to each participant.

Table 8 provides a summary of the mapping strategies that were assessed in this research. In all frameworks, the matched granular profile was tested – that is, the profile which corresponds to the segment the participant is in. More concretely, if a EESRP participant resides in the Coastal Climate Region, is in the smallest consumption size group and in the group representing the largest % of midday consumption, then the matched profile assigned to that participant is the aggregated group of non-participants with all of those characteristics. Similarly, in all cases, accuracy was also assessed using the comparison group for each customer comprised of their individually matched non-participant control. The individual customer match was constructed using Euclidean distance matching of AMI data characteristics within fixed segments of Climate Region, Solar Status, Size, Industry and Load Shape.

Additional options are explained in the table below. For synthetic controls, DSA tested expanding the set of profiles included, but restricted them to be within the same climate zone and solar status for residential participants. The DiD approaches can only accommodate one non-participant profile for each participant; as a result, DSA tested only either individual matched controls or the participant's matched granular profile.

Table 8: Options for Assigning Non-Participants to Participants

Synthetic Control	Difference-in-Differences	CalTRACK DiD
▪ Matched GP (1) ▪ All GPs in Climate Region, Solar Status, and Size Groups (4) ▪ All GPs in Climate Region and Solar Status (20) ▪ Individual matched control	▪ Individual matched control via Euclidian distance matching within Climate Region, Solar Status, Size Group and Load Shape by: • Annual usage • Monthly usage profiles • Load factor • Peak kWh	▪ Matched GP ▪ Individual matched control via Euclidian distance matching within Climate Region, Solar Status, Size Group and Load Shape by: • Annual usage • Monthly usage profiles • Load factor • Peak kWh

ACCURACY METRICS

It is often helpful to conceptualize the process of conducting an accuracy assessment as a tournament: the candidates are defined in advance and the rules for how the contest will be conducted and judged are not changed after the fact. This section defines the statistics that were used to judge the performance of each model described above.

FRACTIONAL SAVINGS UNCERTAINTY

Savings estimated using Population NMEC methods will have uncertainty in their estimates. While uncertainty is a statistical feature of regression-based savings calculations, aggregating site-level estimates of program performance can mitigate the uncertainty by having noise cancel out noise. Nevertheless, it is important that the chosen analysis method can accurately detect an effect given the expected participant population size. Table 9 below provides an example of the relationship between population and effect sizes, quantifying the settlement risks associated with Population NMEC methods for battery end-use data FSU calculations. These values were constructed on EESRP participants using the synthetic control method described above by bootstrapping 200 iterations of each number of sites, aggregating the loads, and computing the distribution of errors in the year prior to their baseline year. The table values represent the FSU, or the expected margin of error divided by the effect size. Values are color-coded to ensure correct interpretation, where green indicates that the FSU target has been met.

Table 9: Settlement Risk as a Function of Effect Size and Population Size for Battery End-Use Data – Annual kWh

Sample Size	Fractional Savings Uncertainty Using Battery End-Use Data (Time of Week and Temperature Method)			
	3% Savings	5% Savings	10% Savings	15% Savings
5	24.0%	14.0%	7.0%	5.0%
10	19.0%	12.0%	6.0%	4.0%
25	10.0%	6.0%	3.0%	2.0%
50	7.0%	4.0%	2.0%	1.0%
100	5.0%	3.0%	2.0%	1.0%
150	4.0%	3.0%	1.0%	1.0%
200	4.0%	2.0%	1.0%	1.0%
500	2.0%	1.0%	1.0%	0.0%
1,000	2.0%	1.0%	0.0%	0.0%
5,000	0.0%	0.0%	0.0%	0.0%

Table 10 shows the same methods for calculating FSU using the whole-home data source. When evaluating Fractional Savings Uncertainty (FSU) using whole-home data, incorporating end-use measurements for battery consumption enhances the confidence in the results. The reduced noise in end-use (battery) data leads to a clearer understanding of the battery's energy savings, thereby improving the precision of FSU calculations.

Table 10: Settlement Risk as a Function of Effect Size and Population Size for AMI Data – Annual kWh

Sample Size	Fractional Savings Uncertainty Using AMI Data (Difference in Difference with Matched Control Method)			
	3% Savings	5% Savings	10% Savings	15% Savings
5	190.2%	114.1%	57.1%	38.0%
10	121.0%	72.6%	36.3%	24.2%
25	61.7%	37.0%	18.5%	12.3%
50	51.6%	31.0%	15.5%	10.3%
100	39.4%	23.6%	11.8%	7.9%
150	30.6%	18.3%	9.2%	6.1%
200	24.4%	14.7%	7.3%	4.9%
500	15.5%	9.3%	4.6%	3.1%
1,000	9.5%	5.7%	2.8%	1.9%
5,000	2.6%	1.6%	0.8%	0.5%

Public

ACCURACY AND PRECISION

In the quantitative assessment, bias and CVRMSE are measured as key metrics to evaluate the program's effectiveness. Bias indicates the tendency of a method to over or underpredict savings, while CVRMSE gauges how closely these predictions align with actual results, regardless of the direction of error. The FSU statistic, as described earlier, serves as an important measure of expected results and is used as a screening tool to ensure that methods meet the established sufficiency criteria. Table 11 presents a detailed summary of metrics for accuracy (bias) and CVRMSE, which are essential for evaluating the performance of each method.

However, it is important to acknowledge that in instances where the denominators in the calculations are very close to zero, the accuracy metrics may become skewed. This issue can lead to metrics that do not convey the annual trends. This is less of an issue for peak savings metrics, as the metric considers the change in energy consumption during peak hours.

This issue is particularly evident in the context of annual savings, where the net discharge of batteries over a day often sums to approximately zero. This near-zero denominator in the calculations can significantly distort the accuracy metrics. Therefore, a revision of these metrics might be necessary in future assessments to ensure a more accurate representation of annual savings.

Furthermore, considering the offset of household consumption by battery charge and discharge, as indicated by AMI data, the interaction between these factors becomes evident. The balancing of charging and discharging within a household can complicate the accuracy of measurements, reinforcing the need for a careful review of these metrics and potentially revising them to better suit the data and the program's objectives.

Table 11 summarizes metrics for accuracy (bias) and precision that will be used to assess performance. Assessing both accuracy and precision is clearly useful for quantifying errors in each method. It is important to distinguish the level at which these values can be computed, however. For example, bias and precision can be calculated for an individual site, where the % Bias indicates the percent by which the method tends to overstate or understate the savings for that site and the relative RMSE (CV(RMSE)) represents the relative "noisiness" of errors for an individual hour. Nevertheless, results in this report are produced at the portfolio level as that is most appropriate for a population-based program.

Table 11: Accuracy and Precision Metrics

Type of Metric	Metric	Description	Mathematical Expression
Bias	% Bias	Indicates the percentage by which the measurement, on average, over or underestimates the true energy savings.	$\% \, Bias = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)}{\bar{y}}$
Precision	Relative RMSE or CVRMSE	Measures the relative magnitude of errors, weighting more extreme errors more heavily.	$RRMSE = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\bar{y}}$

ACCURACY ASSESSMENT RESULTS

Overall, the methods tested under the accuracy assessment performed well. When using a sample population of 5,000 individuals and holding the effect size at 5%, each met FSU targets. Additionally, most models had low estimates of percent bias. Table 12 shows the overall results of the accuracy assessment with each model broken out by bias and FSU. At the population level, all models meet FSU criteria, and many perform well, with unbiased and precise results. The method proposed in the EESRP M&V plan (TOWT models without controls for end-use) and Individually Matched Control difference-in-differences for whole home are bolded on the table.

Table 12: Accuracy Assessment Results[14]

Data Source	Method	Annual kWh			Peak kWh		
		% Bias	CVRMSE	FSU	% Bias	CVRMSE	FSU
Battery end-use data	Battery CalTRACK Hourly no controls	3%	0.11	11%	0%	0.01	-1%
	Battery TOWT no controls	1%	0.03	-2%	0%	0.00	0%
Whole-home AMI data	CalTRACK DiD w/ GP	13613%	4.14	1%	591%	0.18	-13%
	CalTRACK DiD w/ Individual Matched Control	13605%	4.16	-5%	13721%	4.19	-4%
	Matched Control DiD	2%	0.00	0%	27%	0.08	8%
	Synthetic Control with Matched GP	21%	0.11	11%	21%	0.09	9%

While overall program performance is important, the final model selected should also produce results that are accurate for subsets of customers with different attributes.

[14] Green shading shows that the model was +/- 1.5% Bias and < 25% FSU. The sample population was held fixed at 5,000, and the assumed savings was 5%.

Discussion of Findings: Methods Accuracy Assessment

These results of the accuracy assessment of Population NMEC models support several main conclusions:

1. Temperature and time of week and a Difference-in-Difference with controls models performed best for models incorporating both battery end-use and whole-home data sources.
2. Incorporating battery end-use data markedly enhances the accuracy and precision of the models, outperforming whole-home metrics even in the absence of a control group. This improvement is likely due to the reduced data noise and clearer insights into energy usage patterns provided by battery-specific measurements, thereby reinforcing the robustness of the M&V plan's methodologies.
3. The Peak FSU metrics demonstrate better performance compared to the annual FSU metrics, largely due to the focused time-window they encompass. A switch to an alternative metric for average error in kWh could result in improved insights into model performance
4. Given the substantial sample size of > 5,000 customers in the program, FSU targets can be achieved for both peak and annual measurements, accommodating a range of realized savings from 3% to 15%. This extensive participant base provides a robust dataset, enabling precise and accurate estimations that meet FSU criteria across different time frames and savings scenarios.
5. The approach of using an individual matched control appears to perform better than using an aggregated profile of non-participants for AMI measures of accuracy and precision. Because individual customer consumption is more representative of battery behavior compared to solar profiles, it's more likely to produce a good fit on a participant-by-participant basis.
6. The customer mix of EESRP is fairly unique as compared to other programs, and while these methods performed best for this customer cohort, results of these models may vary as they are applied to other customer cohorts.

4 PEAK ENERGY SAVINGS

After selecting the final models to be used for the electric analyses, the DSA team estimated program and weather-normalized year savings for each of the data sources. Participants included all customers who participated in the EESR program. Normalized savings represent all estimated savings for a complete weather-normalized year, and program savings represent all estimated savings for the program duration of August 1st, 2023, through October 31, 2023. To reach these estimates, the models were trained on a year of pre-installation data, then were used to predict consumption in the post-period. The difference between the predicted values and what was observed is the avoided energy use for each participant.

Once the avoided energy use for participants was estimated, the DSA team estimated weather-normalized savings using CZ2022 weather files.[15] The weather-normalized impacts were developed by estimating the relationship between weather and energy savings during the post period for all participants.

Participants in the EESRP were consistent over time, with 8,552 customers. A breakdown of sites by key segments is shown in Table 13.

Table 13: Participation Site Counts

Category	Subcategory	Vendor	
		SolarEdge/Delta	Tesla
All Participants	All Participants	5,911	2,621
Climate Region	Coastal	808	507
	Inland	1,648	1,000
	North Central Valley	2,385	896
	South Central Valley	1,070	218

The remainder of this section will explore the methods and outcomes of the EESR program with a breakdown by battery vendor and data source. The intention is to evaluate the potential savings, with particular emphasis on data source. Furthermore, the results are categorized based on battery vendor.

[15] CZ2022 weather files are available for public use at http://calmac.org/weather.asp as recommended by Version 2.0 of the CPUC's Rulebook for Programs and Projects Based on Normalized Metered Energy Consumption NMEC Rulebook available at: https://www.cpuc.ca.gov/-/media/cpuc-website/files/legacyfiles/n/6442463694-nmec-rulebook2-0.pdf

Public

Savings Estimation Methodology

The final methodology used for this analysis was documented in PG&E's filed M&V plan for the EESR program. A summary of the approach is reviewed in this document for clarity.

Methods for Estimating Gross Savings

The regression specification used for estimating participant impact for the end-use data source is based on the time of week and temperature (TOWT) model developed by LBNL.[16] There are five components to the regression, which is run on the hourly participant consumption data:

1. The regression constant term, representing the average base consumption for the participant.
2. Hour-of-week fixed effects. There are 7 x 24 = 168 dummy variables that capture deviations from the base consumption in each hour of the week.
3. Temperature spline. Between one and seven bins of temperature, with cut points for each temperature bin set algorithmically to ensure sufficient coverage.
4. Granular profiles. These are average hourly consumption profiles for a sample of non-participants in similar segments to the participant. The role of the granular profile is to capture information about non-weather characteristics of each date-hour that may influence participant energy consumption. Excluding these granular profiles from the model result in a simple pre-post model.
5. The error term.

The exact specification is shown in Equation 3:

Equation 3: Seasonal Time of Week and Temperature Model

$$kWh_{p,t} = \sum_{i=1}^{168}(\beta_i * I_{i,t}) + \sum_{b=1}^{b=[2,7]}(\gamma_b * B_{b,t}) + \sum_{g=0}^{n}(\delta_g * GP_{g,t}) + \varepsilon_{p,t}$$

[16] Matthieu, J.L., P.N. Price, S. Kiliccote, and M.A. Piette, "Quantifying Changes in Building Electricity Use, With Application to Demand Response" (2011) available at: https://eta-publications.lbl.gov/sites/default/files/LBNL-4944E.pdf

Table 14: Definition of Equation Terms

Symbol	Interpretation
$kWh_{p,t}$	The observed kWh consumption for participant p in date-hour t
β_i	The coefficient representing the base energy consumption for hour-of-week i, above or below the participant average
$I_{i,t}$	A dummy variable for each hour-of-week i. Equal to 1 when date-hour t is in that hour-of-week, and 0 otherwise
γ_b	The coefficient representing the marginal consumption associated with a one-degree change in outdoor temperature for temperature bin b
$B_{b,t}$	The value of temperature bin b. The construction of temperature bins is described in more detail below.
δ_g	The coefficient representing the marginal effect of one kWh change in the control group granular profile g.
$GP_{g,t}$	The average consumption of the granular control group profile g in date-hour t.
$\varepsilon_{p,t}$	The error term for participant p in date-hour t

The temperature spline is comprised of between one and seven temperature bins that relate outside air temperature to participant consumption. A spline model splits temperature from a single value into ordered bins that correspond to the degrees Fahrenheit that fall in that bin. As examples, Table 15 shows how a range of temperatures can be represented as temperature bins.

Table 15: Relationship Between Temperature and Spline Temperature Bins

Temperature Condition (F)	B_1 < 30	B_2 30-45	B_3 45-55	B_4 55-65	B_5 65-75	B_6 75-90	B_7 > 90
25F	25						
47F	30	15	2				
65F	30	15	10	10			
83F	30	15	10	10	10	8	
101F	30	15	10	10	10	15	11

To ensure that the relationship between temperature and consumption can be robustly estimated, there must be sufficient data in each temperature bin. To that effect, the number of bins used in the regression are modified dynamically by algorithmically removing cut points between the bins. The procedure for this pruning is described in further detail in Section 3.9 of the CalTRACK methods.[17] In brief, the procedure involves:

1. Count the number of hours in each temperature bin $\boldsymbol{B_1}$ through $\boldsymbol{B_7}$
2. If any of bins $\boldsymbol{B_1}$ through $\boldsymbol{B_6}$ have fewer than 20 observations in that range, combine the observations in that bin with the next highest bin:

[17] http://docs.caltrack.org/en/latest/methods.html

Public

3. For example, if bin B_2 (30-45F) had 17 observations and bin B_3 (45-55F) has 30 observations, combine B_2 and B_3 to create one bin from 30-55F with 47 observations
4. If B_7 has fewer than 20 observations, combine it with the next lowest bin until the 20-observation criteria is met
5. Continue pruning the bins until each bin contains at least 20 observations.

An example of this pruning procedure is shown in Figure 2, below.

Figure 2: Pruning of Temperature Bins

<table>
<tr><th colspan="2">Starting</th><th colspan="2">Iteration 1</th><th colspan="2">Iteration 2</th><th colspan="2">Termination</th></tr>
<tr><th>Bin Conditions</th><th>Count</th><th>Bin Conditions</th><th>Count</th><th>Bin Conditions</th><th>Count</th><th>Bin Conditions</th><th>Count</th></tr>
<tr><td>< 30</td><td>5</td><td rowspan="2"><45</td><td rowspan="2">35</td><td rowspan="2"><45</td><td rowspan="2">35</td><td rowspan="2"><45</td><td rowspan="2">35</td></tr>
<tr><td>30-45</td><td>30</td></tr>
<tr><td>45-55</td><td>50</td><td>45-55</td><td>50</td><td>45-55</td><td>50</td><td rowspan="5">>45</td><td rowspan="5">59</td></tr>
<tr><td>55-65</td><td>5</td><td rowspan="2">55-75</td><td rowspan="2">6</td><td rowspan="4">>55</td><td rowspan="4">9</td></tr>
<tr><td>65-75</td><td>1</td></tr>
<tr><td>75-90</td><td>2</td><td rowspan="2">> 75</td><td rowspan="2">3</td></tr>
<tr><td>> 90</td><td>1</td></tr>
<tr><td>Total Bins</td><td>7</td><td colspan="2">4</td><td colspan="2">3</td><td colspan="2">2</td></tr>
<tr><td>Count</td><td>94</td><td colspan="2">94</td><td colspan="2">94</td><td colspan="2">94</td></tr>
</table>

If applicable, the final element in this Seasonal TOWT model are the granular profiles. These represent the average granular (8760) consumption of a group of non-participants. Participants are matched to the correct granular profile(s) based on having similar segmentation. The regression will have one or multiple granular profiles added as explanatory (right-hand-side) variables. This approach is called a synthetic control and relies on exploiting the correlations that exist between participant loads and nearby similar customers. These customers experience similar economic conditions and other unobserved conditions that may influence energy use.

The regression model is estimated independently for each season[18] in the baseline period, and then predicted for that season in the reporting period. The predicted hourly consumption in the reporting period is called the counterfactual consumption. These values represent what the consumption would have been had the premises not participated in EESRP. Gross savings in the reporting period are simple summations of the hourly impacts by period of interest. As this modeling is done at the hourly level, peak period kW and kWh values can be easily estimated by summing or averaging the appropriate hourly impacts. For the purpose of the EESRP end-use evaluation, granular profiles and controls were not used as explanatory variables, and savings were estimated strictly as a function of temperature, past usage, and program effect.

[18] Seasons are defined as: Summer: June through September. Winter: December through March. Shoulder: April, May, October, November.

For whole-home AMI data, a difference-in-differences with individually matched control was used. The standard DiD calculation requires that each participant and the corresponding matched control have before and after data for the same periods. It can be implemented at different levels of temporal granularity – e.g., hourly 8760, daily, by peak period, or annually. The first step is to aggregate the usage for each site and the corresponding matched control to before and after data at the level of temporal granularity desired. At that point, the following equation is applied:

Equation 4: Standard Difference-in-Differences Calculation

$$Savings_{p,t} = (kWh\_treat\_post_t - kWh\_treat\_pre_t) - (kWh\_control\_post_t - kWh\_control\_pre_t)$$

Table 16: Definition of Equation 4 Terms

Symbol	Interpretation
$Savings_{p,t}$	The kWh savings for participant *p* at time period *t*
$kWh\_treat\_post_i$	Observed participant kWh during the reporting period for time period *t*
$kWh\_treat\_pre_i$	Observed participant kWh during the training (pre-treatment) period *t*
$kWh\_control\_post_i$	Observed control kWh during the reporting period for time period *t*
$kWh\_control\_pre_i$	Observed control kWh during the training period for time period *t*
t	Is the level of temporal granularity used for the analysis. For the analysis, the time dimension will be hour-of-year (same week of year and same hour of week), so it is comparable for the training (pre-treatment) and reporting period

The standard Difference-in-Differences (DiD) approach is notable for its consistency, easy to understand, and straightforward replication, making it particularly suitable for various applications, including solar customers. Its consistency ensures that results across different segments match up, providing a clear overall picture of the intervention's impact. This method is easy to comprehend and replicate, crucial for widespread application and validation of findings. Unlike methods reliant on regression analyses, DiD is more straightforward to implement and quicker to execute. Conceptually, the Difference-in-Differences approach is based on the idea that before any intervention occurs, the behavior or usage patterns of the group receiving the intervention (the participant group) and a similar group not receiving it (the control group) are almost the same. This similarity is crucial for comparing the two groups accurately after the intervention. This similarity establishes a baseline against which post-intervention divergences are measured. Any pre-existing differences between the groups observed during the pre-intervention or training period are accounted for, ensuring that the observed changes are attributable to the intervention.

Weather Normalization

The claimed savings in this report are normalized to a typical weather year, per the M&V plan. The normal weather year data source is CALMAC's historic and normalized weather data (specifically, the

CZ2022 Weather Data).[19] This data is publicly available on CALMAC's website and contains data for 97 stations that map to climate zones across the state.

The weather-normalized estimates of savings are produced using a second-stage model of 8,760 energy savings at the site level. The procedure involves:

1. Construct the site-level avoided energy use according to the specifications described above.
2. Run the second-stage model in the performance period for each site. This model uses the 8,760 savings values from Step 1 as the dependent variable and estimates coefficients using historical weather.
3. Predict weather-normalized savings using the CZ2022 weather data and the estimated coefficients.
4. Summarize the 8,760 weather-normalized savings for each site, enrollment group, and EESRP in total by summarizing kWh savings during specific periods of interest (peak, net peak, and off-peak) as appropriate.

Weather-normalization procedures necessarily involve predicting savings for periods that represent a hypothetical weather year. They represent savings under a standard scenario, not specific historical conditions. For this reason, the control granular profiles that support the development of avoided energy use cannot be used as part of the weather-normalized estimates. The value of their inclusion is in helping ensure that the savings based on historical weather data are accurate. These savings estimates are then used in the second-stage model, where variation in savings can be explained according to the time of week and weather conditions. The regression model for the weather-normalized estimates is shown in Equation 5. Note that this model mirrors that used to construct the avoided energy use but removes the granular profiles from the list of covariates.

Equation 5: Second Stage Model for Weather Normalization

$$AEU_{p,t} = \alpha_p + \sum_{i=1}^{168}(\beta_i * I_{i,t}) + \sum_{b=1}^{b=[2,7]}(\gamma_b * B_{b,t}) + \varepsilon_{p,t}$$

[19] Per California State Law, Title 24 2022 updates must be in effect by January 1, 2023. Since SRP installations span 2022 and 2023, but claimable savings for the program are not expected until 2023 or 2024 at the earliest, the CZ2022 data is the most appropriate normalized weather file.

Table 17: Definition of Equation 4 Terms

Symbol	Interpretation
$AEU_{p,t}$	The avoided energy use from for participant p in date-hour *t*
α_p	The constant for participant *p*
β_i	The coefficient representing the base savings for hour-of-week *i*, above or below the participant average
$I_{i,t}$	A dummy variable for each hour-of-week *i*. Equal to 1 when date-hour *t* is in that hour-of-week, and 0 otherwise
γ_b	The coefficient representing the marginal savings associated with a one-degree change in outdoor temperature for temperature bin *b*
$B_{b,t}$	The value of temperature bin *b*. The construction of temperature bins is described in more detail below.
$\varepsilon_{p,t}$	The error term for participant *p* in date-hour *t*

These models were applied to each participant from August 1, 2023, through October 31, 2023. The model's effectiveness in predicting avoided energy use is showcased in Figure 3 for battery end-use and in Figure 4 for whole-home AMI data. It's important to note that the analysis excludes the period from July 1, 2023, to July 31, 2023, due to the program's roll-out schedule. The left side of these figures illustrates the baseline fit against historical data, providing a clear view of the model's ability to replicate past energy usage trends.

While both models yield valuable insights into the counterfactual, they come with certain limitations. For instance, the Time of Weather Temperature (ToWT) model simplifies its approach by only accounting for temperature's influence on battery discharge, neglecting other significant factors like cloud cover and solar irradiance, which can greatly affect not only solar performance, but battery charging behaviors as well. For the Difference in Differences (DiD) approach, the model assumes that, without the intervention, the average outcomes for the treated and comparison groups would have moved along similar paths over time. This assumption generally holds up in the data presented, apart from specific segments of the pre-treatment phase. The discrepancy arises because of a particular charging dispatch setting applied to the participant group, which was not present in the control group's load.

Adding to this context, it's noteworthy that during the pre-year baseline period, the participants batteries were specifically dispatched during the hour ending at 20:00 for SolarEdge and Delta participants. This operational detail highlights a pre-existing difference in how participant batteries were managed compared to the control group, further explaining some of the observed variations in the pre-treatment period data.

Public

Figure 3: Baseline Model Fit for Time-of-Week and Temperature Model with Battery End-Use Data

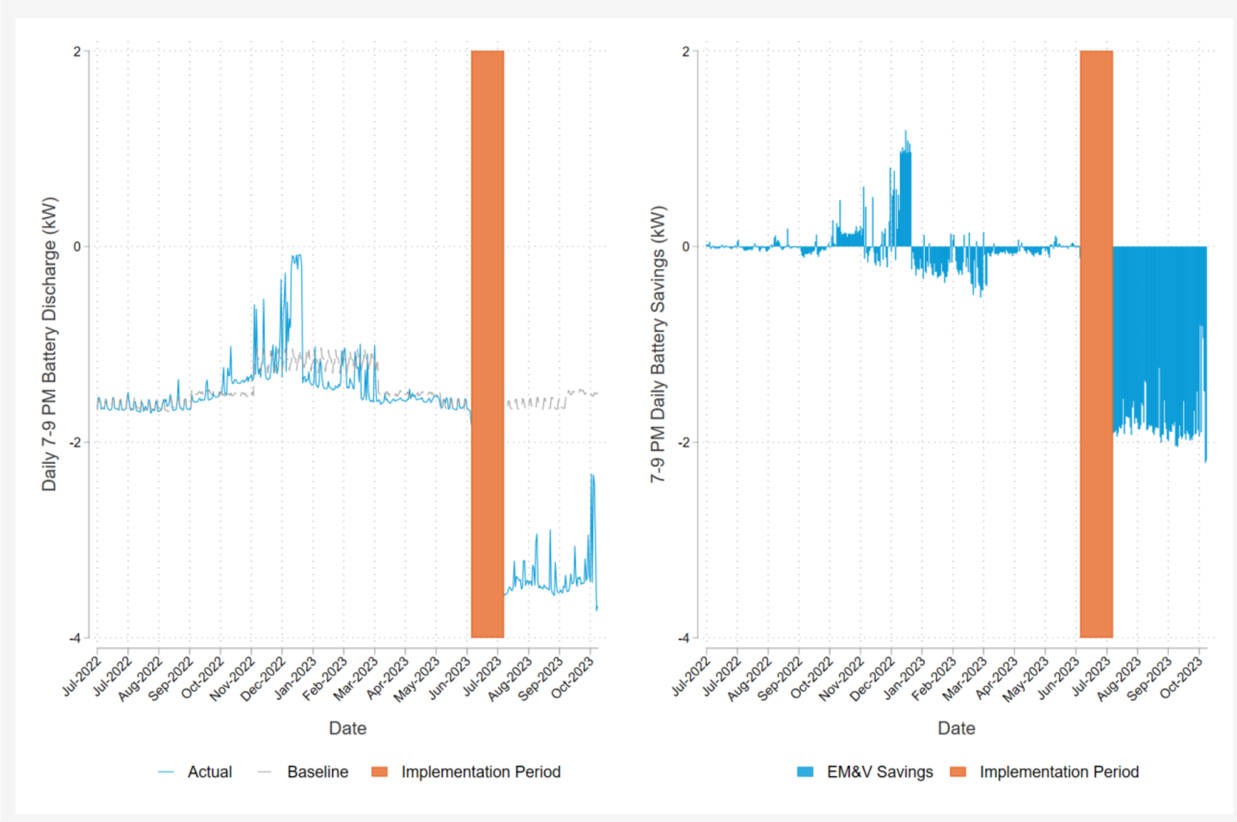
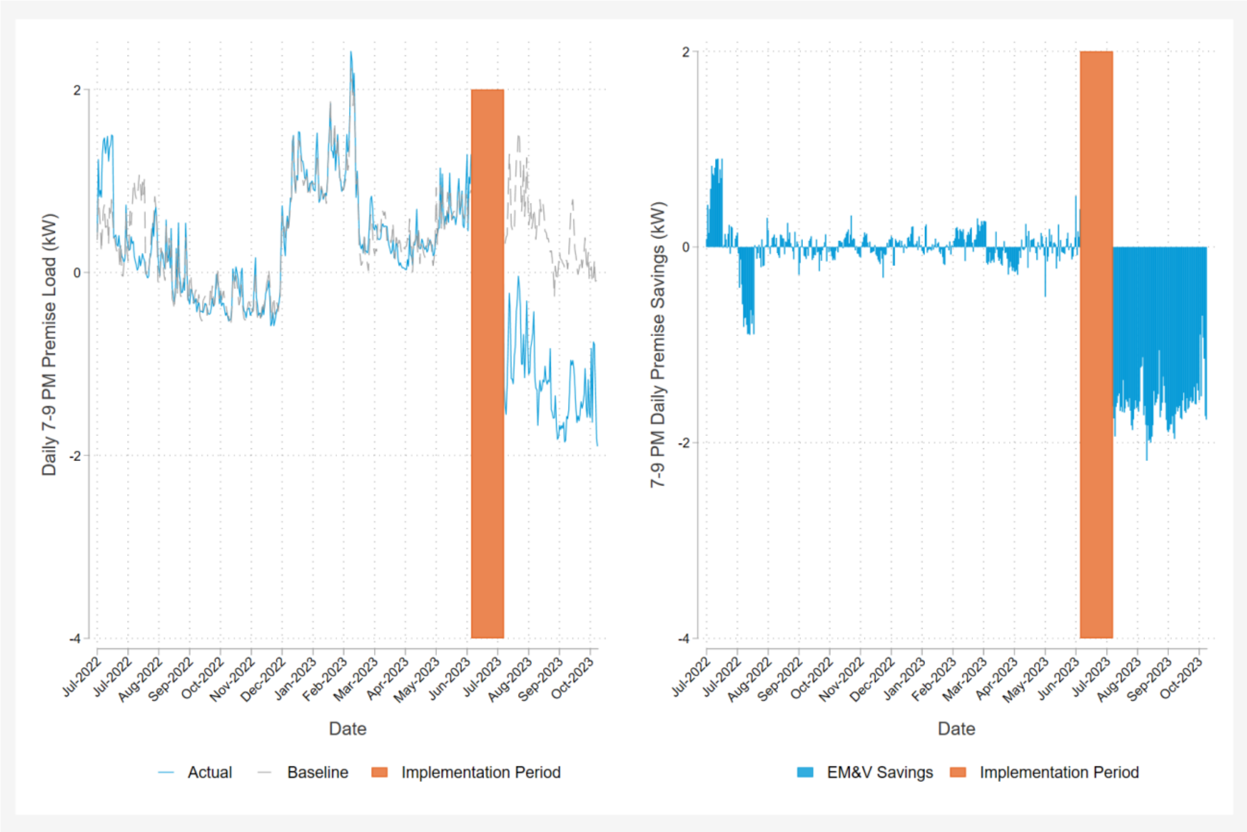


Figure 4: Baseline Model Fit for Difference-in-Difference Model with a Control Group for Whole-Home AMI Data



MODELLING TRENDS

Modelling battery discharge as a function of temperature does not capture any significant temperature trends. Figure 5 depicts the daily average battery discharge under both the observed conditions and the counterfactual throughout the program's duration. The data shows a discrepancy in discharge levels between the counterfactual and the observed outcomes suggesting a clear program effect. However, these levels do not exhibit a strong correlation with participant temperatures.

Figure 5: EESR per Participant 7-9 PM Discharge for Program and Counterfactual

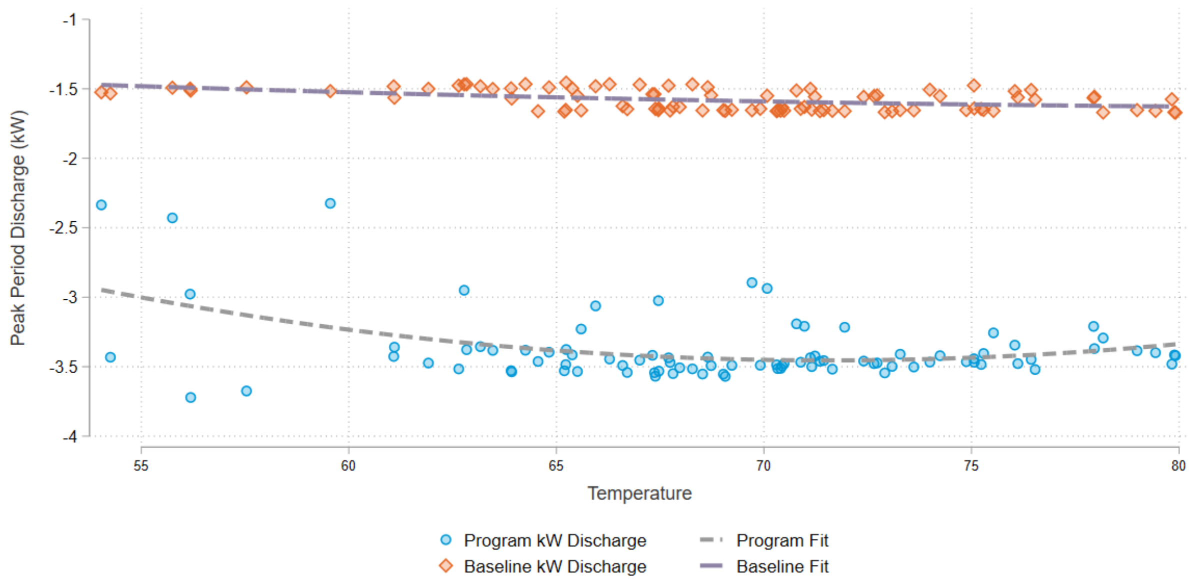
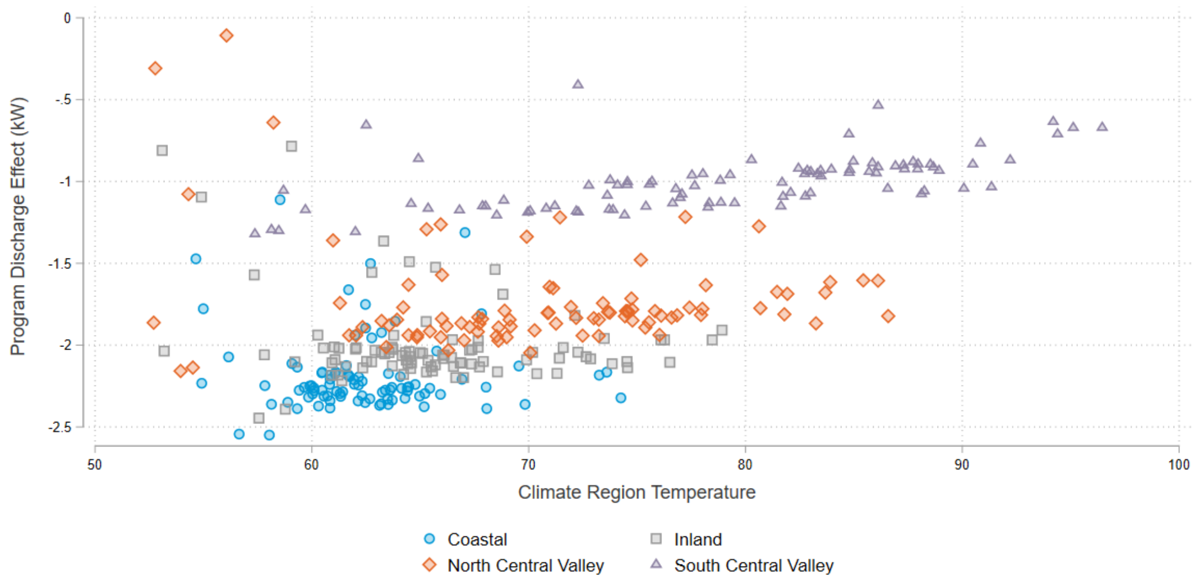


Figure 6 illustrates the relationship between EM&V savings, i.e., the incremental 7-9 p.m. discharge from the modelled baseline, and temperature categorized into climate regions. It shows four climate regions: Coastal (circles), Inland (squares), North Central Valley (diamonds), and South-Central Valley (triangles), with temperatures ranging approximately from 50 to 100 degrees. There is a spread of discharge values across all regions without a clear temperature trend, but certain climate regions perform better than others in terms of program discharge. These figures indicate the potential for alternative right-hand variables to model discharge effect.

Figure 6: EESR Peak 7-9 PM Effect by Climate Region



COUNTERFACTUALS COMPARISON BETWEEN CUSTOMER SEGMENTS

Throughout the analysis of the program, it became evident that a structural difference existed between specific customer segments, specifically battery inverter brands. SolarEdge and Delta batteries were impacted by previous years of 7 p.m. to 8 p.m. dispatch, resulting in lower predicted EM&V savings. Figure 7 and Figure 8 illustrate the counterfactual differences between these segments by the different data sources. The observation that participants under Tesla inverters discharge more kW from the 7-9 p.m. window is a false notion, as Tesla customers did not experience an intervention in the pre-period. This is clear in the left pane as there is a definite intervention taking place in the counterfactual period.

Figure 7: EESRP Whole-Home AMI Baselines by Battery Inverter

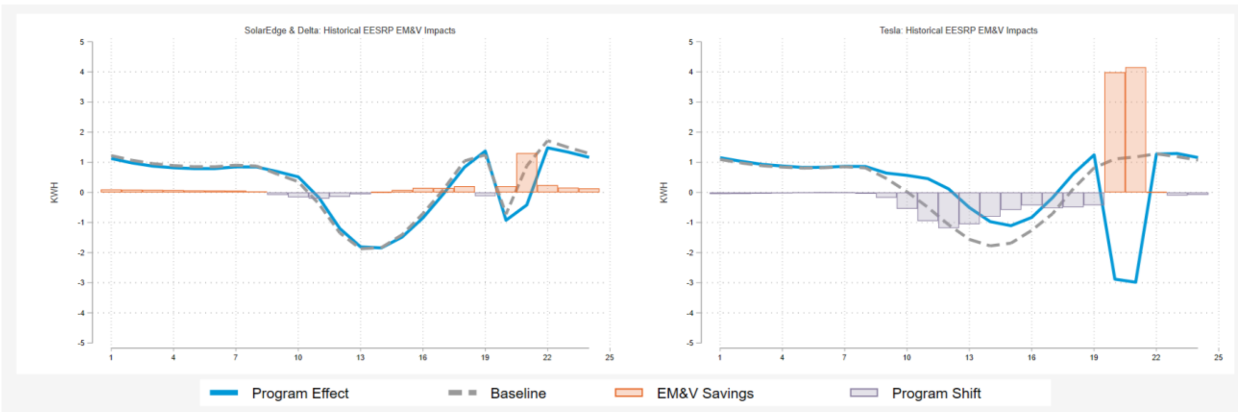
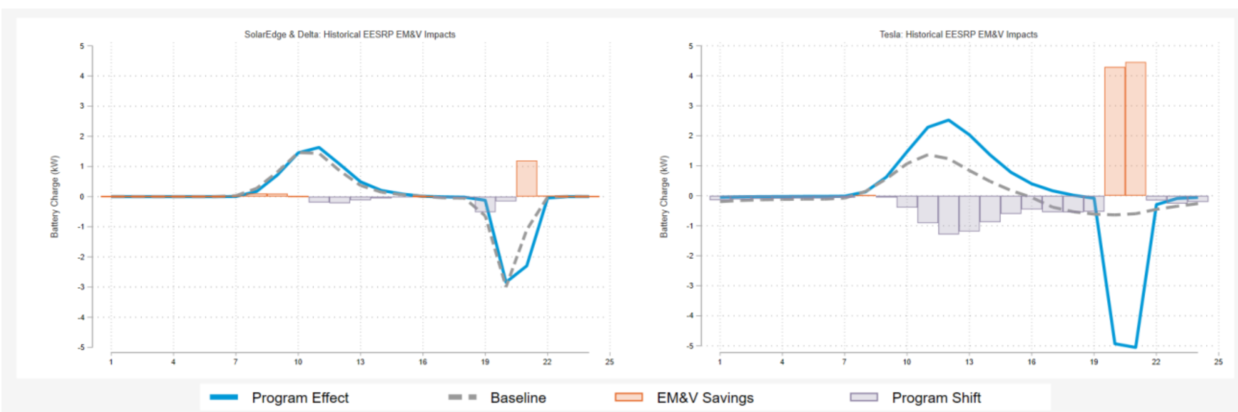


Figure 8: EESRP End-Use Battery Baselines by Battery Inverter



HISTORICAL EESR PROGRAM RESULTS

The tables below present a comparative analysis of energy consumption and impact across the different data sources of the Energy Efficiency Summer Reliability Program based on battery inverter brands, SolarEdge and Delta, and Tesla, as well as an aggregated view of all EESRP participants. For the battery end-use data source, the Tesla cohort exhibits a more substantial reduction in both predicted and actual annual consumption compared to SolarEdge and Delta, translating into a higher per customer impact and a more significant percentage reduction. The whole-home AMI data mirrors this trend, with Tesla inverters again showing greater average reductions in consumption and higher percentage impacts. The table further indicates the breadth of the program's coverage and the peak reduction days analyzed, underscoring the differential response in counterfactual development by inverter brand within the EESRP. It is important to note that none of these savings were claimed by PG&E toward its EE goals but are the potential estimates of what claimed savings would look like.

Table 18 illustrates the peak kW savings across different segments and data sources. With 8,483 participants, the per customer aggregated peak reduction impacts falls to 1.68 kW for end-use data and 1.69 kW for whole-home AMI. As with prior sections, the SolarEdge and Delta impacts are based on the deviance from the counterfactual, and don't represent the actual change from an untreated baseline. However, they do represent an incremental difference from the prior year compared to the program.

Table 18: Historical EESRP Peak 7-9 p.m. kW Savings by Data Source

Data Source	Cohort	Premises in Enrollment Data	Premises in Analysis	% Coverage	Average Annual Predicted Consumption (KW)	Average Annual Actual Consumption (KW)	Per Customer Impact (KW)	Percent Impact	Peak Reduction (MW)	Analysis Days
Battery End-Use	SolarEdge + Delta	5,917	2,988	50%	-2.1	-2.6	0.52	-25%	3.06	92
	Tesla	2,566	1,457	57%	-0.6	-5.0	4.37	-707%	11.21	92
	All EESRP	**8,483**	**4,445**	**52%**	**-1.6**	**-3.3**	**1.68**	**-183%**	**14.26**	**92**
Whole-Home AMI	SolarEdge + Delta	5,917	3,683	62%	0.1	-0.6	0.70	506%	4.15	92
	Tesla	2,566	1,436	56%	1.2	-2.8	3.97	339%	10.39	92
	All EESRP	**8,483**	**5,119**	**60%**	**0.5**	**-1.2**	**1.69**	**357%**	**14.54**	**92**

Table 19 presents the historical kWh savings for the EESRP program duration by data source and vendor. In this case, the end-use data presented higher MWh savings due to nature of being unperturbed by the noise of whole-home data. The EESRP total for the end-use falls at 2,619 and 2,598 MWh during the peak period for end-use and whole-home methods.

Table 19: Historical EESRP Peak 7-9 p.m. kWh Savings by Data Source

Data Source	Cohort	Premises in Enrollment Data	Premises in Analysis	% Coverage	Average Annual Predicted Consumption (kWh)	Average Annual Actual Consumption (kWh)	Per Customer Impact (kWh)	Percent Impact	Energy Savings (MWh)	Analysis Days
Battery End-Use	SolarEdge + Delta	5,917	2,988	50%	-376.3	-471.5	95.1	-25%	563	92
	Tesla	2,566	1,457	57%	-113.4	-914.8	801.4	-707%	2,056	92
	All EESRP	**8,483**	**4,445**	**52%**	**-296.8**	**-605.6**	**308.8**	**-183%**	**2,619**	**92**
Whole-Home AMI	SolarEdge + Delta	5,917	3,683	62%	24.7	-104.3	129.0	523%	763	92
	Tesla	2,566	1,436	56%	205.4	-495.1	700.5	341%	1,835	92
	All EESRP	**8,483**	**5,119**	**60%**	**79.3**	**-222.5**	**301.9**	**361%**	**2,598**	**92**

Table 20 introduces the total energy savings for every hour of the program for the 8,483 participants. Whole-home AMI performs significantly better compared to battery end-use as the whole-home data captures the full scope of kWh reductions. Analyzing batteries and the end-use data associated, it is expected for very minimal shifts as batteries are meant to equalize loads throughout the day. Batteries operate by essentially discharging the solar generation that is received. In this case, 76 MWh and 3,406 MWh were saved by battery end-use and whole-home AMI data sources.

Table 20: Historical EESRP Program Length kWh Savings by Data Source

Data Source	Cohort	Premises in Enrollment Data	Premises in Analysis	% Coverage	Average Annual Predicted Consumption (kWh)	Average Annual Actual Consumption (kWh)	Per Customer Impact (kWh)	Percent Impact	Energy Savings (MWh)	Analysis Days
Battery End-Use	SolarEdge + Delta	5,917	2,988	50%	59.3	47.1	12.18	21%	72	92
	Tesla	2,566	1,457	57%	100.8	99.4	1.40	1%	4	92
	All EESRP	**8,483**	**4,445**	**52%**	**71.8**	**62.9**	**8.92**	**8%**	**76**	**92**
Whole-Home AMI	SolarEdge + Delta	5,917	3,683	62%	810.8	352.5	458.30	57%	2,709	92
	Tesla	2,566	1,436	56%	497.7	232.0	265.73	53%	696	92
	All EESRP	**8,483**	**5,119**	**60%**	**716.1**	**316.1**	**400.05**	**55%**	**3,406**	**92**

WEATHER NORMALIZED RESULTS FOR EESR PROGRAM YEAR 2023

The figures below demonstrate the weather-normalized 24-hour shift in consumption patterns resulting from the intervention, as observed across various data sources. The color coding is blue for observed values, grey for the counterfactual estimates among participants, and orange for the calculated savings. In the year before the intervention, the trend lines superimposed on the individual data points are expected to match the observed data and the counterfactual predictions.

Figure 9 showcases the adjustment in battery behavior, notably the increase in discharge during hour ending 20 and the extension of this discharge period to hour ending 21. Simultaneously, Figure 10 depicts the shift in the overall premise load, echoing the changes observed in the battery behavior. It is important to note that savings are not claimed but are an estimation of savings in NMEC context.

Figure 9: Battery End-Use Weather Normalized 24-Hour Load for EESRP

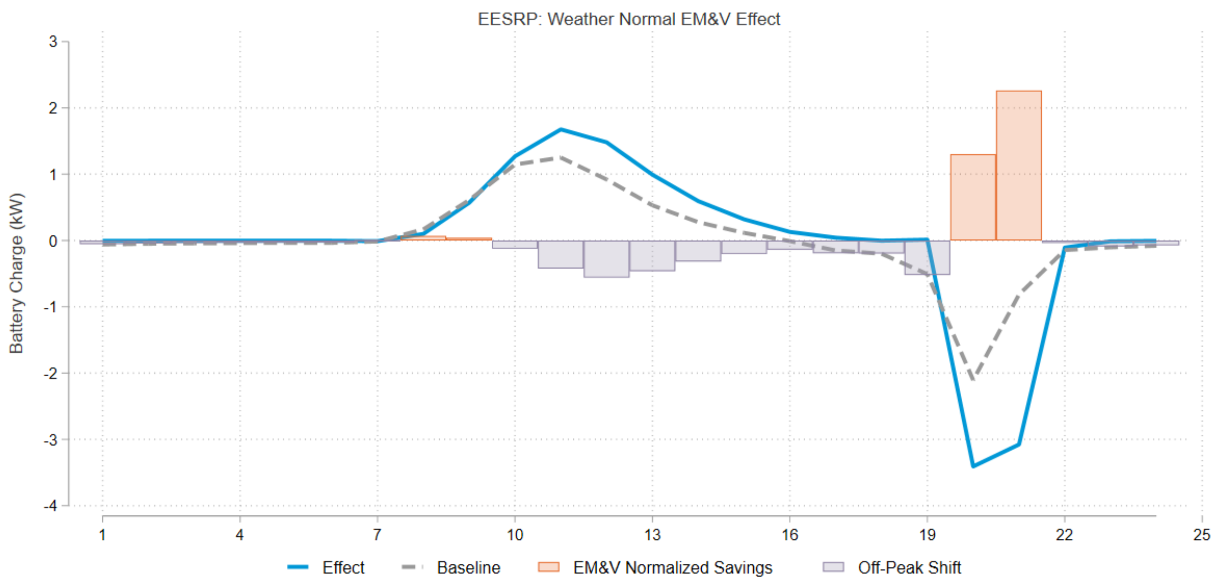
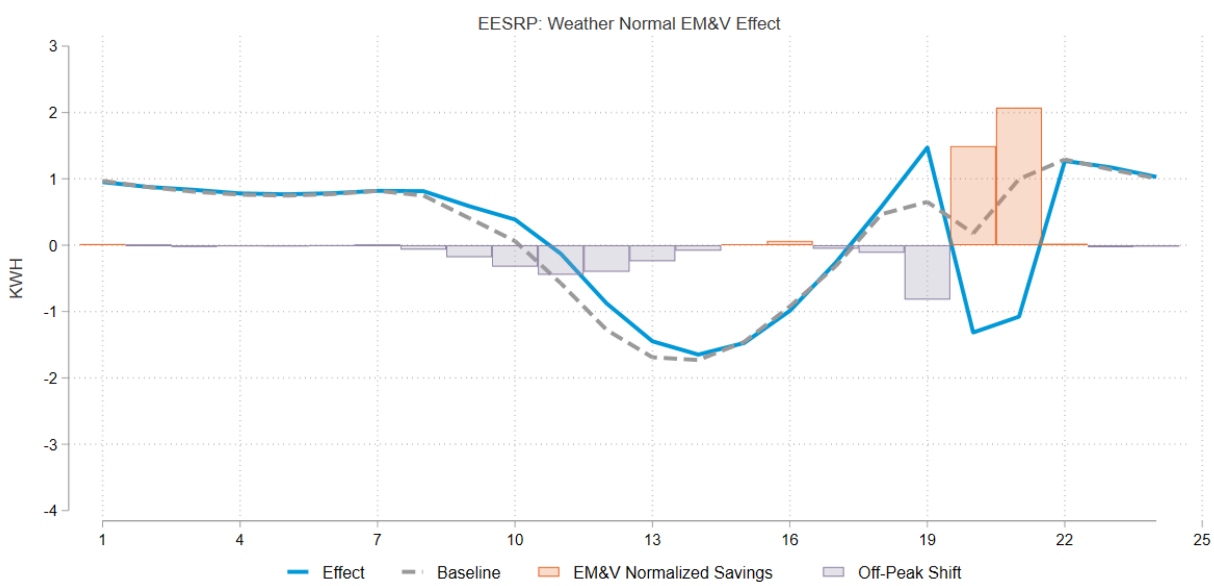


Figure 10: Whole-Home AMI Weather Normal 24-Hour Load for EESRP



Public

EESR Weather Normalized Program Savings

Table 21 illustrates the annual weather-normalized kWh savings during the Net Peak dispatch window of 7 p.m. to 9 p.m., utilizing both end-use and whole-home approaches. The EESRP resulted in engagement from 8,483 customers, leading to significant energy savings of 10,349 MWh and 10,563 MWh for the battery end-use and whole-home AMI methodologies, respectively.

Table 21: Peak kWh/MWh Savings by Data Source and Vendor

Data Source	Cohort	Premises in Enrollment Data	Premises in Analysis	% Coverage	Average Annual Predicted Consumption (kWh)	Average Annual Actual Consumption (kWh)	Per Customer Impact (kWh)	Percent Impact	Energy Savings (MWh)	Analysis Days
Battery End-Use	SolarEdge + Delta	5,917	2,988	50%	-1357.5	-1734.2	376.7	-28%	2,229	365
	Tesla	2,566	1,457	57%	-400.5	-3565.1	3,164.6	-790%	8,120	365
	All EESRP	**8,483**	**4,445**	**52%**	**-1068.0**	**-2288.0**	**1,220.0**	**-201%**	**10,349**	**365**
Whole-Home AMI	SolarEdge + Delta	5,917	3,683	62%	218.3	-291.6	509.8	-234%	3,014	365
	Tesla	2,566	1,436	56%	654.4	-2226.7	2,881.0	-440%	7,548	365
	All EESRP	**8,483**	**5,119**	**60%**	**350.2**	**-876.9**	**1,227.1**	**-389%**	**10,563**	**365**

Table 22 presents the annual weather-normalized kWh savings during the 7 p.m. to 9 p.m. dispatch period, analyzed through both battery end-use and whole-home methodologies. In this period, the EESR program engaged 8,483 customers, achieving a potential Net Peak demand reduction of 14.21 MW and 14.52 MW for the battery end-use and whole-home AMI data sources, respectively, over the course of the weather-normalized year.

Table 22: Peak kW/MW Savings by Data Source and Vendor

Data Source	Cohort	Premises in Enrollment Data	Premises in Analysis	% Coverage	Average Annual Predicted Consumption (KW)	Average Annual Actual Consumption (KW)	Per Customer Impact (KW)	Percent Impact	Peak Reduction (MW)	Analysis Days
Battery End-Use	SolarEdge + Delta	5,917	2,988	50%	-1.9	-2.4	0.52	-28%	3.07	365
	Tesla	2,566	1,457	57%	-0.5	-4.9	4.34	-790%	11.14	365
	All EESRP	**8,483**	**4,445**	**52%**	**-1.5**	**-3.1**	**1.68**	**-201%**	**14.21**	**365**
Whole-Home AMI	SolarEdge + Delta	5,917	3,683	62%	0.3	-0.4	0.70	232%	4.14	365
	Tesla	2,566	1,436	56%	0.9	-3.0	3.96	419%	10.38	365
	All EESRP	**8,483**	**5,119**	**60%**	**0.5**	**-1.2**	**1.69**	**374%**	**14.52**	**365**

Table 23 details the full-year weather-normalized kWh savings for every hour of the year, encompassing both end-use and whole-home methods. With the EESR program's current enrollment of customers, there is an expected weather normal annual energy savings of 336 MWh and 15,310 MWh for battery end-use and whole-home AMI data sources throughout the weather-normalized year.

Table 23: Annual Peak kWh/MWh Savings by Data Source and Vendor

Data Source	Cohort	Premises in Enrollment Data	Premises in Analysis	% Coverage	Average Annual Predicted Consumption (kWh)	Average Annual Actual Consumption (kWh)	Per Customer Impact (kWh)	Percent Impact	Energy Savings (MWh)	Analysis Days
Battery End-Use	SolarEdge + Delta	5,917	2,988	50%	184.1	115.4	68.71	37%	407	365
	Tesla	2,566	1,457	57%	288.0	315.6	-27.55	-10%	-71	365
	All EESRP	**8,483**	**4,445**	**52%**	**215.5**	**175.9**	**39.59**	**9%**	**336**	**365**
Whole-Home AMI	SolarEdge + Delta	5,917	3,683	62%	1629.1	-366.0	1995.05	122%	11,795	365
	Tesla	2,566	1,436	56%	1594.4	252.7	1341.76	84%	3,515	365
	All EESRP	**8,483**	**5,119**	**60%**	**1618.6**	**-178.8**	**1797.44**	**104%**	**15,310**	**365**

5 DISCUSSION AND RECOMMENDATIONS

This analysis demonstrates potential savings across both data sources for customers engaged in the EESR program. It is important to note that these savings are non-claimable and are only estimates for the year based on NMEC methods. Several critical insights emerge from these findings, as detailed in this section. Based on these insights, a set of key recommendations has been formulated, presented in the recommendations section that follows.

Key Findings Resulting from the Analysis

1. **The EESR program successfully increased battery discharge during the Net Peak hours of 7 p.m. to 9 p.m.** The battery end-use data source resulted in a weather normal 1.68 kW savings per customer for the discharge period, and the whole-home AMI data source resulted in a 1.69 kW savings per customer during that period. Potential estimated savings for these programs are seen during peak periods and are not necessarily realized during the 8760 period.

2. **There is a variation of savings across EESR participants based on battery inverter brand.** Tesla battery loads were unperturbed by discharge programs implemented during the prior year, resulting in more significant EM&V savings. SolarEdge and Delta batteries were impacted by previous years of 7 p.m. to 8 p.m. dispatch, resulting in lower predicted EM&V savings.

3. **The current method proposed in the M&V plan for whole-home and end-use savings estimation meets FSU targets at the expected sample size (fixed at 5,000 participants) and savings threshold (about 5%).** In essentially all cases, the FSU thresholds are also met from FSU savings of 3% to 15%.

4. **The Time-of-Week and Temperature Model proposed in the M&V plan and used for the NMEC analysis for the battery end-use data source failed to capture exogenous battery patterns during extraneous events.** Batteries generally show a relatively consistent performance throughout the year, as they are not significantly sensitive to temperature variations. In the context of the EESR program, the effectiveness of battery performance is predominantly influenced by the efficiency of solar energy generation rather than factors such as temperature.

5. **The individual-matched controls and granular profiles, as currently constructed, do not follow a similar consumption pattern as participants in the program.** Because individual customer consumption is volatile, it’s less likely to produce a good fit on a participant-by-participant basis. This is exaggerated for SolarEdge and Delta participants, since batteries independently controlled had a vastly different behavior than those operating under Sunrun’s prior dispatch patterns.

KEY STUDY RECOMMENDATIONS

1. **The current Population NMEC analysis and baseline penalize prior good faith actors.** In their baseline periods, SolarEdge and Delta batteries exhibited disrupted load patterns due to being dispatched as part of prior programs. Consequently, the observed EM&V impacts in these cases are less pronounced than other programs, due to pre-existing interventions. When analyzing whole-home AMI data, it was found that participants using SolarEdge and Delta inverters achieved an average reduction in consumption of 0.70 kW. In contrast, participants with Tesla inverters, which had undisturbed load profiles, showed a more significant average decrease of 3.96 kW, indicating a notable difference in outcomes between the two groups. Adjusting the methods to create baselines based on actual, non-disturbed participant loads is recommended.

2. **The analysis of battery end-use data for the creation of baselines, as well as the evaluation of model accuracy, led to results with better valuation metrics.** The analysis of battery end-use data for establishing baselines and verifying model accuracy has led to more confidence in valuation metrics, partly because it minimizes noise in the data. By determining baselines, the program's effect is more isolated and results in a measure of genuine battery performance, filtering out extraneous fluctuations or anomalies that might otherwise skew results. This method of filtering the data to its most relevant parts ensures that the insights and conclusions are based on clear, noise-reduced information.

3. **Future analysis in battery-centered programs for Population NMEC purposes should feature adjustments that include alternative right-hand variables.** Incorporating variables such as solar irradiation and cloud cover into models for future study of battery-centered programs, especially for Normalized Metered Energy Consumption (NMEC) purposes, is crucial due to their direct impact on battery performance. Solar irradiation, which represents the amount of sunlight reaching the solar panels, is a key factor in determining the energy production capacity of solar-powered battery systems. Including this variable allows for a more accurate assessment of the energy that the batteries can store and utilize. Similarly, cloud cover significantly influences the solar radiation reaching the panels. On cloudy days, reduced solar irradiation can lead to lower energy generation, directly impacting how batteries charge and discharge. By accounting for these environmental factors, the models potentially become more reflective of real-world conditions, leading to more precise evaluations and predictions of battery performance in various weather scenarios. This enhancement of current NMEC methods could allow for better assessment of both solar and battery programs.

Public