Emerging Products

# NMEC Pre-Qualification Pilot Feasibility Study

*ET19SCE7010 Final Report*

*Prepared by:*

*Emerging Products*
*Customer Service*
*Southern California Edison*

*[December 2019]*

SOUTHERN CALIFORNIA EDISON
An EDISON INTERNATIONAL® Company

## Acknowledgements

Southern California Edison's Emerging Products (EP) group is responsible for this project. It was developed as part of Southern California Edison's Emerging Technologies Program under internal project number ET19SCE7010. Sarah Monohon and Ted Helvoigt with Evergreen Economics conducted this technology evaluation with overall guidance and management from Dallen Coulter. For more information on this project, contact dallen.coulter@sce.com.

## Disclaimer

This report was prepared by Southern California Edison (SCE) and funded by California utility customers under the auspices of the California Public Utilities Commission. Reproduction or distribution of the whole or any part of the contents of this document without the express written permission of SCE is prohibited. This work was performed with reasonable care and in accordance with professional standards. However, neither SCE nor any entity performing the work pursuant to SCE's authority make any warranty or representation, expressed or implied, with regard to this report, the merchantability or fitness for a particular purpose of the results of the work, or any analyses, or conclusions contained in this report. The results reflected in the work are generally representative of operating conditions; however, the results in any other situation may vary depending upon operating conditions.

Formatted: Font: Verdana

# Executive Summary

## TECHNOLOGY DESCRIPTION

The normalized metered energy consumption (NMEC) program design enables program implementers to offer unique bundles of measures to individual customers, where the savings claimed (and incentives paid) depend on the savings observed at the meter. The California Public Utilities Commission requires that NMEC programs use the custom program and accompanying project review process, where each customer and project is required to submit a unique application, determine an appropriate baseline, and receive approval before proceeding with the project.

The proposed pilot program would target chain businesses that are interested in installing the same set of measures at a series of locations that have similar characteristics affecting energy consumption (e.g., building size, existing equipment, operating hours, and business practices). A single application and project plan would cover all branches, with some unique pre-screening criteria to confirm that the branches are sufficiently similar and a good fit for an NMEC program.

## PROJECT GOAL

In this initial proof-of-concept study, we will select a modeling approach and pre-screening algorithm to develop baseline models of energy consumption for each chain business (at all eligible branches) that will meet or exceed the NMEC requirements of model fit. We will also test whether a matched comparison group can be extracted from the remaining branches (i.e., non-participants) to estimate net savings in the post-period. Since the pilot has not begun implementation, this study does not cover performance payment calculations or savings claim estimates.

## PROJECT FINDINGS

We looked at three commercial chains, including two grocery chains and one retail chain, with 39 proposed participants, as well as a larger sample of non-participant branches from the same chain businesses.

Table ES 1 provides a summary of the results from the baseline modeling phase of the feasibility study. The model fit tests for each of these chains demonstrated that pooled and segmented baseline models are feasible but may not be a good fit for all types of chains. The individual baseline models met all of the NMEC model fit criteria for the vast majority of participant sites (n=38/39).[1] The one site with a failed individual model had a significant change in its energy consumption during the baseline period, which was identified during our pre-screening for non-routine events. These events will require a follow-up discussion with the customer to explain the event, and then adjust the baseline model prior to program intervention.

1 These NMEC model fit criteria are based on the current SCE site-level NMEC procedures manual and CPUC draft rulebook for population-level NMEC: CV(RMSE)<25%, NMBE<0.005%, FSU<25% at 90% confidence with bias correction, and preferably R-square>0.7.

**TABLE ES 1. SUMMARY OF BASELINE MODEL FIT**

| BUSINESS CHAIN | SUBTYPE | N PART SITES | N COMP SITES | POOLED (ALL) | Pooled (OF MATCHED) | SEGMENTED[1] | INDIVIDUAL[2] |
|---|---|---|---|---|---|---|---|
| Grocery 1 | A | 8 | 2 | Pass | Pass | Pass (n=4) | Pass (all) |
| | B | 5 | 4 | | Pass | Pass (n=2) | Pass (all) |
| Grocery 2 | A | 12 | 4 | Pass | Pass | Pass (n=10) | Pass (all) |
| Retail 1 | A | 9 | 15 | **Fail** | Pass | Pass (n=7) | Pass (all) |
| | B | 5 | 2 | | **Fail** | N/a | Pass (n=4) + **Fail (n=1)** |

1. N refers to number of participant sites that were successfully segmented. The remaining sites were too unique for segmentation. These sites would need a larger non-participant sample to draw from or rely on an individual model.
2. Only a single participant site failed to meet the NMEC criteria for a successful individual baseline model.

## CONCLUSIONS & RECOMMENDATIONS

- **Pre-Screening** – Identify concurrent program participation and any non-routine events in the baseline year of energy consumption. Additional data collection will be required to produce accurate savings estimates.
- **Comparison Group** – While a matched comparison group of non-participant branches is feasible, this will require a much larger sample or synthetic comparison customers to ensure a match for every participant branch.
- **Baseline Models** – Individual baseline models consistently provide the most accurate predictions. Pooled and segmented models may be considered for populations that are relatively homogenous, such as grocery chains.

# Abbreviations and Acronyms

| | |
|---|---|
| 3P DRC | Third-party implemented demand response contracts |
| AMI | Advanced metering infrastructure |
| AMICS | AMI customer segmentation model |
| CDD | Cooling degree-day |
| CEC | California Energy Commission |
| CPUC | California Public Utilities Commission |
| CV(RMSE) | Coefficient of variation of the root mean square error |
| FSU | Fractional savings uncertainty |
| IQR | Interquartile range |
| NEM | Net energy metering |
| NMEC | Normalized metered energy consumption |
| NMBE | Normalized mean bias error |
| NRE | Non-routine event |
| NTG | Net-to-gross |
| OLS | Ordinary least-squares |
| SCE | Southern California Edison |

# CONTENTS

# FIGURES



# TABLES

# EQUATIONS

# Background

Southern California Edison (SCE) contracted with Evergreen Economics to conduct interval energy usage data analysis and determine if a modified normalized metered energy consumption (NMEC) approach might be feasible to estimate energy savings for multiple locations across a single business entity. This analysis will use a streamlined NMEC method to estimate savings for retail chains that may wish to retrofit multiple buildings with similar consumption levels and load profiles.

## CURRENT NMEC PROCEDURES AND GUIDELINES

The NMEC option enables program implementers to offer unique bundles of measures to individual customers, where the savings claimed (and incentives paid) depend on the savings observed at the meter. The California Public Utilities Commission (CPUC) requires that site-level NMEC programs use the custom program and accompanying project review process, where each customer and project is required to submit a unique application, determine an appropriate baseline, and receive approval before proceeding with the project.[2] SCE has received feedback from its commercial customers that they are willing and able to pursue efficiency projects, but cannot wait for the complicated custom M&V process to be completed before they find out which projects are eligible.

If the proposed pilot utilizes individual site-level baseline models and savings claims, then the current CPUC and SCE site-level NMEC guidelines will apply.[3] If the pilot utilizes segmented or pooled baseline models with aggregate savings claims for each chain, then population-level NMEC rules and guidelines will apply.

The criteria for population-level baseline models are not yet well defined. The CPUC hosted an NMEC working group in June 2019 to suggest rules and recommendations for population-level analysis with aggregate NMEC savings claims.[4] A draft ruling was issued in August 2019, which includes a draft rulebook for population-level NMEC prepared by CPUC Energy Division staff.[5] SCE and other parties issued comments on the draft ruling in September 2019.[6] While the population-based NMEC rules are still under review, a program-level measurement and verification (M&V) plan must be

---

[2] *Rulebook for Custom Program and Projects Based on Normalized Metered Energy Consumption (NMEC)*, March 2019. ftp://ftp.cpuc.ca.gov/gopher-data/energy_division/EnergyEfficiency/RollingPortfolioPgmGuidance/Draft_Rulebook_OUT.pdf

[3] *Normalized Metered Energy Consumption Savings Procedures Manual (ET15SCE1130)*, December 2017. https://www.etcc-ca.com/reports/normalized-metered-energy-consumption-savings-procedures-manual

[4] *Normalized Metered Energy Consumption Working Group Recommendations for Population-Level Approaches*, Common Spark Consulting, June 20, 2019. https://www.cpuc.ca.gov/uploadedFiles/CPUC_Public_Website/Content/Utilities_and_Industries/Energy/Energy_Programs/Demand_Side_Management/EE_and_Energy_Savings_Assist/PopNMEC_Working_Group_Report_June2019_FINAL.pdf

[5] *Administrative Law Judge's Ruling for Issuing Draft Revised Rulebook for Normalized Metered Energy Consumption and Inviting Comments on Population-Level Rules, Measurement Methods and Calculation Software (Rulemaking 13-11-005)*, Valarie Kao, August 29, 2019. http://docs.cpuc.ca.gov/SearchRes.aspx?DocFormat=ALL&DocID=311581553

[6] *Comments filed on Proceeding R1311005*, September 2019. https://apps.cpuc.ca.gov/apex/f?p=401:56:0::NO:RP,57,RIR:P5_PROCEEDING_SELECT:R1311005

included with any advice letter and implementation plan (IP) submissions for approval by the CPUC.

## STUDY OBJECTIVE

Ultimately, the goal of the proposed analysis is to determine if there is a way to use a streamlined and standardized advanced metering infrastructure (AMI) interval modeling approach to group together candidate locations within a business chain that will have similar savings potential from an NMEC analysis. All projects will be pre-identified at the pilot or program outset. This pre-qualification is analogous to TSA Pre Check at the airport, where once an individual goes through an initial background review, they are able to enjoy a streamlined experience going through airport security. Similarly, if we are able to use the AMI model to identify those individual branches within a chain that have similar and consistent energy consumption patterns, a streamlined NMEC approach might be appropriate and lead to increased participation by customers.

The proposed pilot program would target chain businesses that are interested in installing the same set of measures at a series of locations that have similar characteristics affecting energy consumption (e.g., building size, existing equipment, operating hours, and business practices). A single application that includes a shared measurement and verification plan would cover all branches, with some unique pre-screening criteria to validate that the branches are sufficiently similar.

## SUMMARY OF APPROACH

In this initial proof-of-concept study, we will select a modeling approach and pre-screening algorithm to develop baseline models of energy consumption for each customer (at all eligible branches) that will meet or exceed the NMEC requirements of model fit for individual branches. We will also test whether a matched comparison group can be extracted from the remaining branches (i.e., non-participants) to estimate net savings in the post-period. Since the pilot has not begun implementation, this study does not cover performance payment calculations or savings claim estimates.

# Methods

## Database Creation

### Customer and Billing Data

SCE recruited three chain businesses for the proposed pilot. Each business has around 100 to 150 branches within SCE's service territory. The business managers identified a subset of 10 to 15 branches that they believe will be the best fit for the program pilot, which we will refer to as the participating branches.

Evergreen requested a full year of hourly interval AMI energy usage data for each of these participating branches, as well as a random sample of the remaining branches (i.e., non-participants) from each chain business. We received the following data from SCE:

- Customer Account Details
  - Service account and premise ID
  - Service start date
  - Full service address
  - CEC building climate zone
  - Net Energy Metering (NEM) indicator with connection date
  - Demand response participation indicator(s) with enrollment date(s)
- Energy Usage Data
  - 15-minute interval electricity usage data from May 1, 2018 to April 30, 2019
- Billing Data
  - Billing rate codes
  - Monthly billing date, total usage (kWh), and bill amount ($)
- Pilot Recruitment Records (for participants only)
  - Program eligibility criteria
  - Identifiers for each branch selected by the business manager for participation in the proposed pilot program

Evergreen requested building characteristics (e.g., square footage), major electric end uses, and business/operational details from SCE, but these were not available at the time of the data request. The pilot advice letter filing was put on indefinite hold in July 2019. This prevented any direct follow-up communication with customers; we did not receive detailed site characteristics from the program application form (self-reported), nor the detailed on-site survey results that were to be collected by SCE engineers at one or two sites per chain. These items may be available for future phases of analysis.

#### Business Subtypes

Evergreen performed manual lookups by service address to identify business chain name (i.e., parent company), branch name (there are multiple business subtypes under the same management), retail hours, and departments (e.g., food, apparel, home goods). We received data for 39 participants as shown in Table 2~~Table 2~~, and all but one site were assigned to a branch subtype with these manual lookups. This unknown site is supposedly a branch of the retail chain, but it was not listed on the parent company website; we were unable to verify that this site is currently operating as a retail store.

We requested data for 50 non-participants from each of the three business chains, but only received data for 43 non-participant sites. The manual lookups revealed that 34 (80%) had a branch name corresponding to our participant sites. The remaining nine sites (20%) were different branch types that were excluded from the sample for comparison group selection; these are indicated in red. This sample was especially limited for the retail business, which was concerning, as the retail participant branches had highly variable energy usage.

We submitted a revised data request with addresses of the known Retail 1 A and B branches in SCE's service territory (based on a manual search on the parent company website). SCE provided data for 54 of these non-participant retail sites. Our final sample had 39 participant and 97 non-participant branches, as shown in Table 2~~Table 2~~.

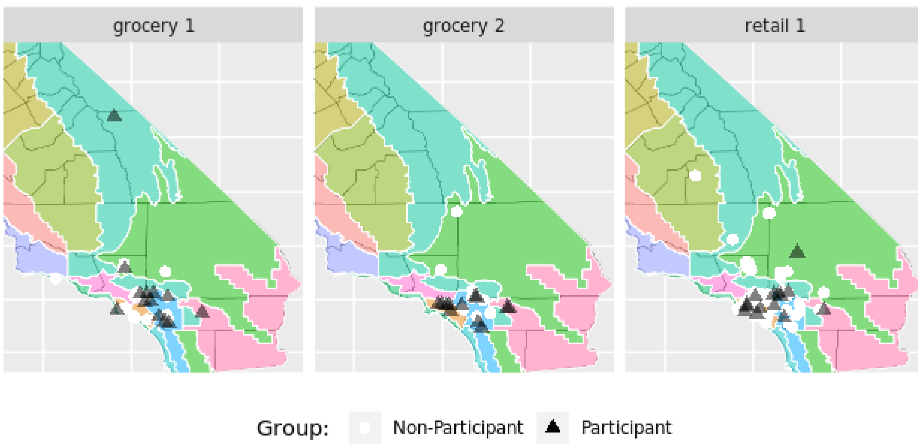
**TABLE 2: CUSTOMER ACCOUNT SAMPLE BY BUSINESS CHAIN AND SUBTYPE**

| BUSINESS CHAIN (PARENT COMPANY) | CHAIN SUBTYPE (BRANCH NAME) | N PARTICIPANT BRANCHES | N NON-PARTICIPANT BRANCHES |
|---|---|---|---|
| **Grocery 1** | A | 8 | 7 |
| | B | 5 | 8 |
| | C | 0 | **1** |
| | **Subtotal** | **13** | **16** |
| **Grocery 2** | A | 12 | 11 |
| | B | 0 | **1** |
| | **Subtotal** | **12** | **12** |
| **Retail 1** | A | 9 | 47 |
| | B | 4 | 15 |
| | C | 0 | **3** |
| | D | 0 | **1** |
| | Unknown | 1 | **3** |
| | **Subtotal** | **14** | **69** |
| **Total** | | **39** | **97** |

### CLIMATE ZONE

Figure 1 provides a map for each of the three businesses, showing the location of each participant (black triangle) and non-participant (white circle) branch in our sample by climate zone. The black lines represent county boundaries, and the colorful shapes show the different building climate zones. These zones were defined by the California Energy Commission to categorize regions by the typical climate, such as temperature, humidity, and wind speed.[7] We do not have non-participants (white circles) in every climate zone containing a participant site (black triangles). However, there may be sufficient sample across regions that the models will be able to adjust predictions to account for the remaining differences in climate.

FIGURE 1: SAMPLE SITES BY CLIMATE ZONE




### ONSITE GENERATION AND NET ENERGY METERING

We created three distinct approaches to look for evidence of onsite generation at each branch in the sample.

1. **Known Net Energy Metering (NEM)** – SCE NEM connection date listed in customer account characteristics;
2. **Net Generation** – History of electricity received by SCE (not just delivered by SCE to the customer) in AMI interval data during the study period; and
3. **Load Shape** – Electricity usage approaches zero during midday hours, while sun is shining, and the business is operating.

Table 3~~Table 3~~ provides a summary of these three indicators of onsite generation among the participant branches within each business chain.

Not all sites with known NEM (#1) tested positive for onsite generation using the other two detection approaches (#2 and #3). It is possible that these branches have onsite generation, but the generation is not sufficient to offset a significant proportion of the building's energy usage. There were no participant branches that tested positive for generation with indicators #2 and #3 without being on an NEM rate (#1).

[7] A description of the CEC climate zones can be found at https://ww2.energy.ca.gov/maps/renewable/building_climate_zones.html

**TABLE 3: ONSITE GENERATION INDICATORS BY BUSINESS CHAIN**

| BUSINESS CHAIN (PARENT COMPANY) | CHAIN SUBTYPE (BRANCH NAME) | N PARTICIPANTS | 1. KNOWN NEM | 2. NET GENERATION | 3. LOAD SHAPE |
|---|---|---|---|---|---|
| **Grocery 1** | A | 8 | 0 | 0 | 0 |
| | B | 5 | 2 | 0 | 2 |
| | **Subtotal** | **13** | **2** | **0** | **2** |
| **Grocery 2** | A | 12 | 0 | 0 | 0 |
| | **Subtotal** | **12** | **0** | **0** | **0** |
| **Retail 1** | A | 9 | 5 | 5 | 2 |
| | B | 4 | 3 | 3 | 3 |
| | E | 1 | 0 | 0 | 0 |
| | **Subtotal** | **14** | **8** | **8** | **5** |
| **Total** | | **39** | **10** | **8** | **7** |

All the participant branches with onsite generation had NEM connection dates that occurred between 2008 and 2011, prior to the beginning of the baseline year. We retained all these branches in the sample for modeling under the assumption that any generation that occurs during the baseline year will be representative of the generation that will continue throughout the program intervention and reporting period. Any major changes in onsite generation (e.g., decommissioning of rooftop solar) should be detected during the testing for non-routine events in the baseline period (Non-Routine Event Detection~~Non-Routine Event Detection~~).

Formatted: Font: 10 pt

### DEMAND RESPONSE PARTICIPATION

Table 4~~Table 4~~ provides an overview of the demand response participation among participant sites by program. The capacity bidding program (CBP) will impact the site energy usage on event days, if and only if their bid for energy reduction was accepted by SCE. To be conservative, we can exclude all event days from the baseline analysis at these three participant sites. Third party implementers (NRG and Stem) are responsible for the remaining programs. These companies provide demand response services directly to SCE customers and receive payment from SCE through demand response contracts (DRCs).[8] Unlike the CBP, we do not know whether these third party programs will impact energy usage on all demand response event days or for a subset of events, or will lead to long term energy savings with efficiency measures and/or behavioral changes. Without this detail, we have no way of knowing whether this program participation will have the same impact on the baseline period as on the reporting period, impacting our ability to produce accurate energy savings estimates.

[8] https://www.sce.com/business/demand-response

**TABLE 4: DEMAND RESPONSE PARTICIPATION BY BUSINESS CHAIN**

| PROGRAM NAME | DESCRIPTION | GROCERY 1 (N=13) | GROCERY 2 (N=12) | RETAIL 1 (N=14) | SUBTOTAL (N=39) |
|---|---|---|---|---|---|
| Capacity Bidding Program (CBP) | Participant submits "bids" each month with proposed energy usage reduction and compensation amount for events. | 0 | 3 | 0 | **3** |
| NRG Demand Response Contract (NRG-DRC) | Third party curtailment with customized recommendations for reduction strategies, real-time training and support. | 0 | 5 | 0 | **5** |
| Stem Demand Response Contract (Stem-DRC) | Third party real-time energy optimization, automated response to events with discounted service rates. | 0 | 0 | 2 | **2** |
| **Total Demand Response Participants*** | | **0** | **5** | **2** | **7** |

* N refers to number of distinct participant sites. Three branches of the Grocery 2 chain participated in both CBP and Stem-DRC.

We recommend that the pilot program staff collect additional detail on these DRC programs from the third-party implementers and the participating sites. If detailed information about the program's impact on both short- and long-term energy usage cannot be adequately determined, these sites will need to be disqualified from participation in this NMEC program. SCE program staff have indicated that measure descriptions and estimated useful life (EUL) should be available for all participants in third party programs, as this is regularly provided to SCE for internal review.

### WEATHER DATA

We geocoded each of the customer service addresses (to latitude and longitude coordinates) and then identified the closest National Oceanic and Atmospheric Administration (NOAA) weather station for each site.

Evergreen appended weather data obtained from NOAA to develop a database that links hourly weather data to the AMI energy consumption data for each site. We identified weather stations for each customer based on the stations' proximity to the zip code of the customer's building, within the same CEC climate zone. Next, we identified unreasonably high or low outdoor temperature readings, based on the record high and low temperatures in each climate zone.[9] Missing temperature readings and those identified as unreasonable were imputed with the average of the preceding and following temperature reads. In the rare instances where this imputation was not sufficient, we relied on temperature readings from the next closest weather station.

## OUTLIERS AND NON-ROUTINE EVENTS

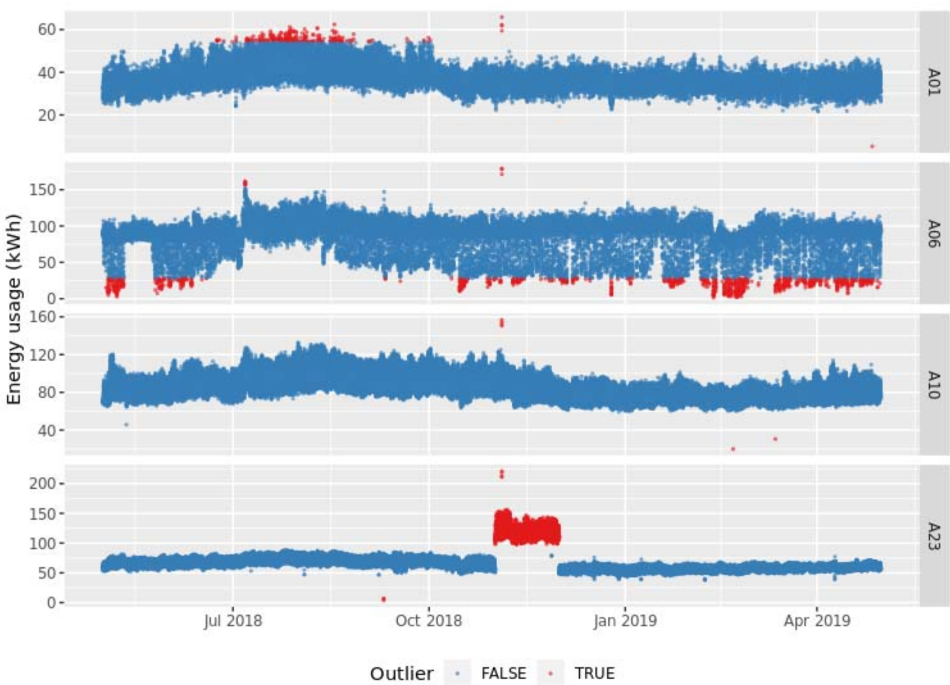
This section describes our methods for outlier and non-routine event (NRE) detection and correction, where applicable.

[9] *The Pacific Energy Center's Guide to: California Climate Zones and Bioclimatic Design*, October 2006. https://www.pge.com/includes/docs/pdfs/about/edusafety/training/pec/toolbox/arch/climate/california_climate_zones_01-16.pdf

## OUTLIER DETECTION

To start, we defined an outlier as any kWh reading that was more than three times the distance of the interquartile range (IQR) from the median interval measurement, based on a full year of hourly baseline interval AMI data for each site.[10] Figure 2~~Figure 2~~ shows the baseline energy usage of four participant sites with all of the outliers indicated in red.

**FIGURE 2: CUSTOMER ENERGY USAGE WITH OUTLIERS IDENTIFIED BY IQR LIMITS**



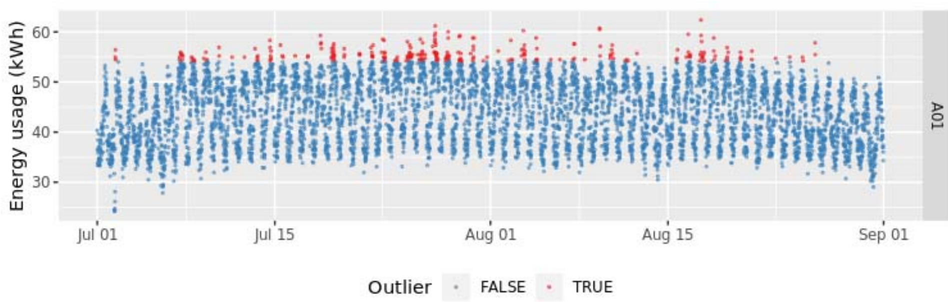

Next, we manually reviewed all the flagged outliers with time-series plots and adjusted any flags that appeared to be too sensitive (false positives) or not sensitive enough (false negatives) by site.

Many of the high outliers occurred on summer days during peak hours. When we consistently observed the hourly interval kWh gradually ramping up to each outlier kWh, we dismissed all high kWh outliers during the summer months. Figure 3~~Figure 3~~ provides the summer energy usage of customer A01 during the summer months, with outliers indicated in red. As seen in this example, the outliers are preceded with a gradual ramp up in energy usage and followed by a gradual decline in energy

[10] This definition of an outlier is based on Caltrack rule 2.3.6. The interquartile range (IQR) is a measurement of variability. The rank-ordered data is divided into four equal parts, called quartiles. The IQR measures the distance between the first and third quartiles, corresponding to the 25th and 75th percentiles, containing the middle 50 percent of observations.

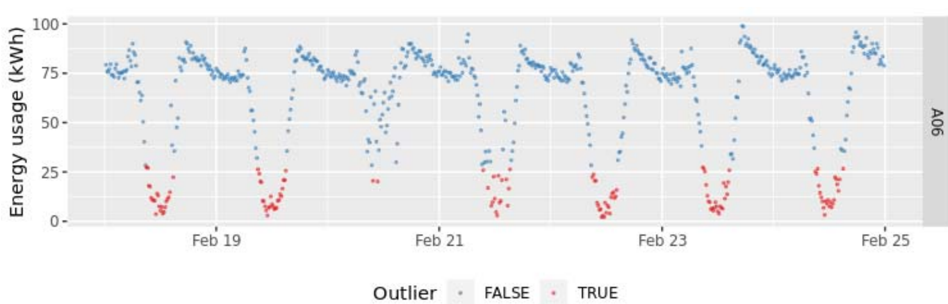
usage. These do not appear to be data errors or extreme values that would introduce any bias into the models, but rather a consistent pattern in summer peak usage.

**FIGURE 3: CUSTOMER A01 WITH SUMMER PEAK OUTLIERS**




Many of the low outliers were observed at sites with onsite generation (e.g., rooftop solar), which makes a near-zero kWh usage measurement feasible for that site. If we could confirm the site has onsite generation via NEM date, net generation in AMI, and load shape, then we dismissed all low outliers at the site. Figure 4~~Figure 4~~ provides the summer energy usage of customer A06 during one week of February 2019, with outliers indicated in red. As seen in this example, the energy usage in this building approaches zero during the middle of the day, during the hours where we would expect to see the highest levels of solar generation. The service account information provided by SCE confirms that this site has an NEM that was connected on January 6, 2010. We dismissed all outliers flagged for low kWh at this site.

**FIGURE 4: CUSTOMER A06 WITH NEAR-ZERO USAGE**




This outlier detection method revealed a surprising pattern at a subset of sites, where energy usage would suddenly double (i.e., no ramp up), cycle at this high level for around 30 days, and then suddenly drop back down to the prior range. This was usually observed in November or April, but the start and end dates varied across sites. Figure 5~~Figure 5~~ provides the energy usage of customer A23 spanning October 15 to December 15, 2019, with outliers indicated in red. As seen in this example, energy usage suddenly increases on November 1, continues at a high level for 30 days, and then drops back down to prior values on December 1. Unlike the prior examples, these outliers were not dismissed, but rather led us to investigate possible errors in data collection, transmission, analysis, and reporting.

FIGURE 5: CUSTOMER A23 WITH OUTLIERS THAT SPAN A FULL BILLING CYCLE

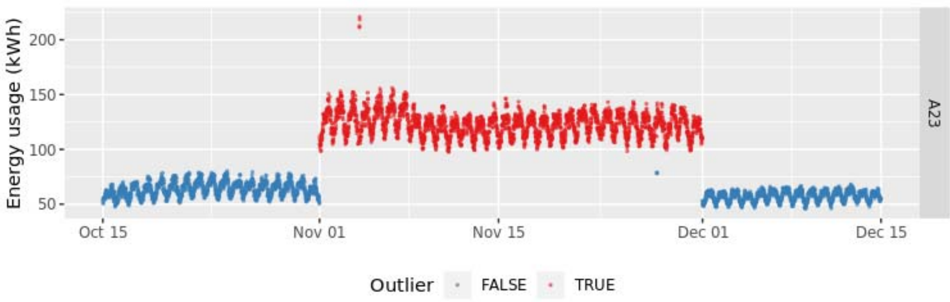



Table 5 provides the output from our quality control test for customer A23, comparing the kWh from SCE's billing records to the total kWh listed in the AMI interval data for each billing cycle. The billing records for the cycle starting on 2018-11-01 listed 177,348 kWh as the billed energy usage quantity for the period ending 2018-12-01. The AMI 15-minute interval data for this same time period has a total of 354,696 kWh, almost exactly double the billed usage. In other words, despite a complete coverage of the billing period in the AMI data (coverage factor = 1.00), the bills indicate that the kWh listed in the AMI are not accurate and should be reduced by half (adjustment factor = 0.50).

TABLE 5: CUSTOMER A23 DATA QUALITY CONTROL TEST BY BILLING CYCLE

| Billing Cycle Start Date | N Days Billed | N Days in AMI | Billed kWh | AMI Interval kWh | Coverage Factor (N Days) | Adjustment Factor (kWh) |
|---|---|---|---|---|---|---|
| 2018-05-01 | 31 | 31 | 200,880 | 200,880 | 1.00 | 1.00 |
| 2018-06-01 | 30 | 30 | 199,564 | 199,564 | 1.00 | 1.00 |
| 2018-07-01 | 31 | 31 | 219,242 | 219,242 | 1.00 | 1.00 |
| 2018-08-01 | 31 | 31 | 216,902 | 216,872 | 1.00 | 1.00 |
| 2018-09-01 | 30 | 30 | 200,634 | 200,635 | 1.00 | 1.00 |
| 2018-10-01 | 31 | 31 | 197,000 | 196,909 | 1.00 | 1.00 |
| **2018-11-01** | **30** | **30** | **177,348** | **354,696** | **1.00** | **0.50** |
| 2018-12-01 | 31 | 31 | 170,010 | 170,010 | 1.00 | 1.00 |
| 2019-01-01 | 31 | 31 | 167,381 | 167,381 | 1.00 | 1.00 |
| 2019-02-01 | 28 | 28 | 153,170 | 153,170 | 1.00 | 1.00 |
| 2019-03-01 | 30 | 27 | 175,210 | 158,282 | 1.11 | 1.11 |

We repeated this quality control test for every customer and billing cycle. SCE investigated these discrepancies and confirmed that this is a valid data error but was unable to find and correct the error at its source. In all instances where the coverage factor and adjustment factor disagreed (note: the coverage was equal to 1.00 in all these cases), we applied the adjustment factor to all the AMI interval data for that billing cycle.

It was not possible to validate and correct suspected data errors in a few non-participant sites that began doubling in March, due a lack of perfect overlap between the billing cycles and the AMI data. To avoid this type of limitation, we recommend that anyone preparing data for use in baseline models for an NMEC program request billing data that span one month beyond the AMI interval data.

Figure 6~~Figure 6~~ provides the energy usage of customer A23 spanning the same time period as Figure 5~~Figure 5~~, after the adjustment factor was applied. As seen in this example, energy usage no longer shows a sudden increase on November 1, instead cycling between 45 and 80 kWh between October 15 and December 15 in predictable intervals by time-of-day.

**FIGURE 6: CUSTOMER A23 AFTER DATA CORRECTION**

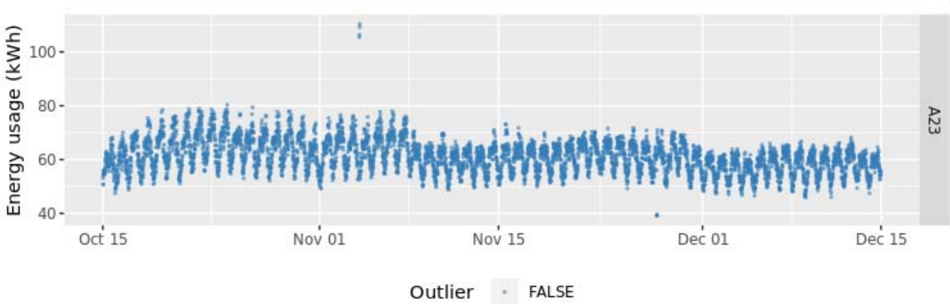
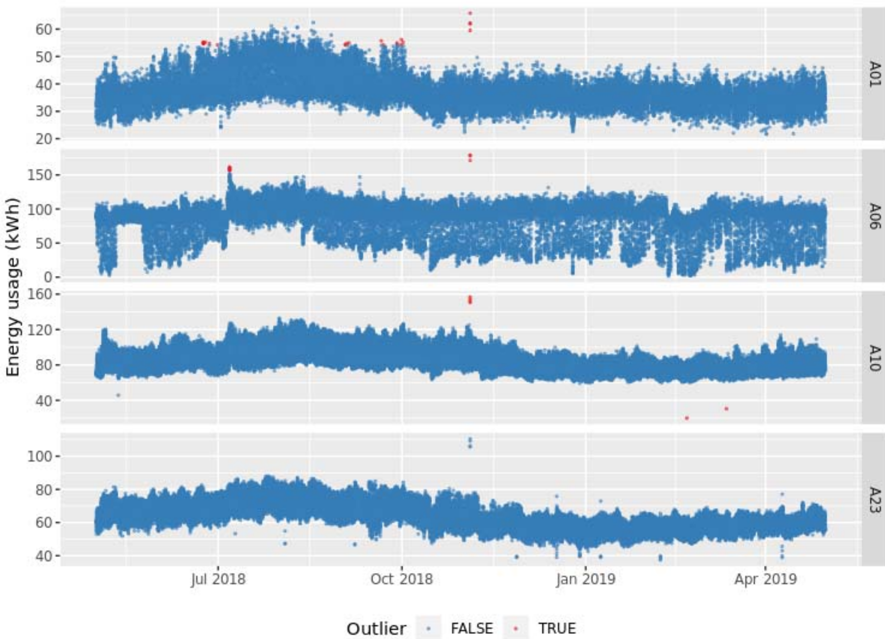



Figure 7~~Figure 7~~ shows the baseline energy usage of the same four participant sites from Figure 2~~Figure 2~~ with outliers indicated in red, after all adjustments to the outlier flags and data error correction. All four sites have a spike in energy usage on November 4, 2018, starting at 2:00 a.m., corresponding to the start of daylight savings time. To avoid any unnecessary bias, we decided to exclude the start and end of daylight savings (two days per year) from the baseline models.

**FIGURE 7: CUSTOMER ENERGY USAGE AFTER OUTLIER DETECTION AND ADJUSTMENT**




## NON-ROUTINE EVENT DETECTION

Evergreen considered a number of approaches to determine if a building's baseline (i.e., pre-intervention) data appear to be sufficiently well behaved to serve as a baseline for determining NMEC energy savings.[11] We empirically tested each approach against the AMI electricity usage data we received for the 39 participant sites. We believe there is much work still to be done in developing procedures for detecting NRE and screening AMI data to ensure they are sufficiently well behaved for calculating NMEC savings. Nevertheless, we believe the common statistical measures described below are useful indicators for quickly and efficiently identifying pre-period data series that are not suitable as baseline data for NMEC savings calculations.

### INDICATOR A – EXCESS KURTOSIS

Kurtosis is a unitless measure of the combined probability in the tails of a distribution. While it is a commonly included statistic in the summary output of regression analysis, it is generally ignored as it is not considered to provide meaningful information regarding the relationship between the dependent and

[11] Lawrence Berkeley National Laboratory (LBNL) has proposed some new NRE detection approaches. However, these diagnostics rely on energy usage data for both the baseline and reporting period for concurrent analysis. The Pre-Qualification Pilot design requires that we estimate baseline models prior to the program intervention. Hence, we needed to develop new methods for NRE detection that rely only on energy usage during the baseline year.

independent variables. Statisticians' understanding as to what kurtosis implies has evolved in recent years. Traditionally, kurtosis was considered a measure of the degree of sharpness in the peak of a distribution, with higher values indicating a more peaked distribution.

More recently, statisticians have come to understand that kurtosis provides information about the tails of the distribution. Higher values of kurtosis indicate that, relative to a normal distribution, higher variance in the empirical distribution is being caused by extreme values in one or both tails of the distribution. It is because kurtosis is a statistic that focuses on the tails of the distribution that we believe that kurtosis is a valuable indicator for determining whether a series of pre-period AMI data is suitable as a baseline for determining NMEC savings, as well as for determining whether a series of post-period data contains an NRE.

Excess kurtosis is computed as follows:[12]

**EQUATION 1: EXCESS KURTOSIS**

$$Excess\ Kurtosis = \frac{\sum_{i=1}^{n} \frac{(x_i - \bar{x})}{n}}{s^4} - 3$$

Where:

$n =$ number of observations
$x_i =$ the i-th AMI observation
$\bar{x} =$ mean of the AMI observations
$s =$ standard deviation of the AMI observations

Note that the standard deviation, s, is raised to the fourth power and represents the fourth moment of the distribution.[13] If the distribution is normally distributed, excess kurtosis is equal to zero.[14] A negative value of excess kurtosis indicates less probability in the tails than a normal distribution, which suggests a distribution with relatively few extreme values. A positive excess kurtosis indicates more probability in the tails than a normal distribution and is consistent with a distribution containing a relatively large number of extreme values.[15] While excess kurtosis is a unitless measure, the magnitude of the value indicates the extent to which the tails of the empirical distribution differ from a normal distribution.

[12] Excess kurtosis is simply kurtosis minus 3.

[13] Moments are quantitative measures defining the shape of a probability function. The zeroth moment is the total probability (100 percent), the first moment is divided by the expected value (i.e., the mean), the second moment is variance, the third moment is the skewness, and the fourth moment is the kurtosis.

[14] Excess kurtosis is often reported as “kurtosis"; however, as stated in footnote 12, it equals kurtosis minus 3.

[15] Heteroskedasticity is the term used to describe the phenomenon of unequal variance across a random variable's range of values. This metric is generally of interest (concern) to econometricians and other practitioners of regression analysis. While we did not examine heteroskedasticity in the energy usage data for these sites, it is likely that we would find the presence of heteroskedasticity in those buildings with high degrees of kurtosis. However, it is also likely that we would find many other buildings with “suitable” interval data that are also heteroskedastic. The next step would be to identify a specific pattern of heteroskedasticity that is “bad” versus other patterns that are “okay.” Our tests with excess kurtosis demonstrate that this metric can be used for simple and consistent detection of potential NREs during the baseline period.

#### INDICATOR B – RELATIVE VARIANCE AND "MODIFIED" RELATIVE VARIANCE

Relative variance is an easy-to-compute measure of the variation in a distribution relative to the mean of the distribution. It is computed by dividing the variance by the absolute value of the mean.

**EQUATION 2: RELATIVE VARIANCE**

$$RV = \frac{s^2}{|\bar{x}|}$$

Where:

$s^2 = variance$

$\bar{x} = mean$

Before computing relative variance, Evergreen aggregated 15-minute AMI data to hourly in order to diminish the impact of any extreme values and anomalous readings. We then computed three variables characterizing the expected hourly energy use and the range in hourly energy use for each day.

- Average hourly kWh use for each day of the pre-intervention period
- Minimum hourly kWh use for each day of the pre-intervention period
- Maximum hourly kWh use for each day of the pre-intervention period

Using these new variables, we computed the following three measures of relative variances:

**EQUATION 3: RELATIVE VARIANCE AND "MODIFIED" RELATIVE VARIANCE**

$$RV = \frac{S^2_{Avg\ hourly\ kWh}}{\bar{x}_{Avg\ hourly\ kWh}}$$

$$MRV_{Min} = \frac{S^2_{Min\ hourly\ kWh}}{\bar{x}_{Avg\ hourly\ kWh}}$$

$$MRV_{Max} = \frac{S^2_{Max\ hourly\ kWh}}{\bar{x}_{Avg\ hourly\ kWh}}$$

Where:

$RV =$ Relative variance of the average hourly kWh

$MRV_{Min} =$ Modified relative variance of the minimum hourly kWh

$MRV_{Max} =$ Modified relative variance of the maximum hourly kWh

$\bar{x}_{Avg\ hourly\ kWh} =$ Mean of the average hourly kWh

$S^2_{Avg\ hourly\ kWh} =$ Variance of the average hourly kWh

$S^2_{Min\ hourly\ kWh} =$ Variance of the minimum hourly kWh

$S^2_{Max\ hourly\ kWh} =$ Variance of the maximum hourly kWh

Note that each measure of relative variance relies on the same denominator—the mean of the average hourly kWh—which means that each measure of variation is relative to average hourly kWh.[16] Even for building with PV, we assume that hourly measures of kWh usage are non-negative and, therefore, the measures of relative variance are non-negative. Our expectation for "well behaved" AMI data is that the three measures will have the following relationship:[17]

**EQUATION 4: EXPECTED RELATIONSHIP BETWEEN VARIANCE METRICS**

$$MRV_{Min} < RV < MRV_{Max}$$

### BASELINE NRE DETECTION PROCESS

The following approach utilizes the kurtosis and relative variance indicators described above to quickly and efficiently identify data issues in the pre-intervention period that suggest the building's energy data are not suitable as baseline data for NMEC savings calculations. Please note that we present this as an approach *given what we have learned to date*. We recommend additional analysis and testing on a wider array of commercial buildings before developing and presenting a recommended approach.

**Step 1**: Compute excess kurtosis for baseline (i.e., pre-period) AMI interval usage, after correcting for known data errors.

**Step 2**: Evaluate kurtosis [*Note: we recommend additional testing using AMI data from other commercial buildings to evaluate the empirical usefulness of kurtosis and the optimal cut-off values for decision making.*]

- If kurtosis < -1.5 → Flag for visual review of interval data
- If -1.5 > kurtosis < 1.5 → Go to Step 5
- If kurtosis is > 1.5 → Go to Step 3

**Step 3**: Compute the three relative variance measures described above

**Step 4**: Evaluate relative variance measures [*Note: as a next stage, we recommend additional empirical review of relative variance and/or similar indicators using AMI data from other commercial buildings to evaluate the empirical usefulness of relative variance and the optimal cut-off values for decision making.*]

- If $MRV_{Min} > RV$ → Fail - Pre-intervention data not suitable for NMEC baseline
- If $MRV_{Min} \times 10 < RV$ → Fail - Pre-intervention data not suitable for NMEC baseline
- Else → Go to Step 5

**Step 5**: Proceed with baseline model estimation and assessing model fit.

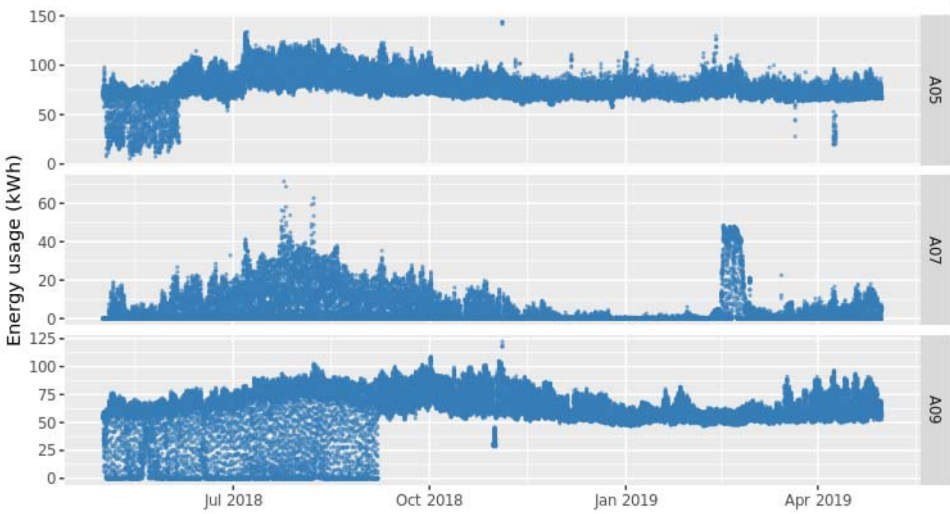
Following this relatively simple approach, **we would fail the following three buildings** shown in Figure 8. These three buildings failed in Step 2 (evaluate kurtosis) and were, therefore, further evaluated in Steps 3 and 4, where each failed one of the evaluation criteria of relative variance. Considering the sample size, we

[16] This is why we refer to the relative variance for minimum and maximum hourly kWh variables as "modified."

[17] Though we are still analyzing data, our assumption is that for well-behaved data, the size of the $MRV_{Max}$ is not more than ten times larger than the $MRV_{Min}$.

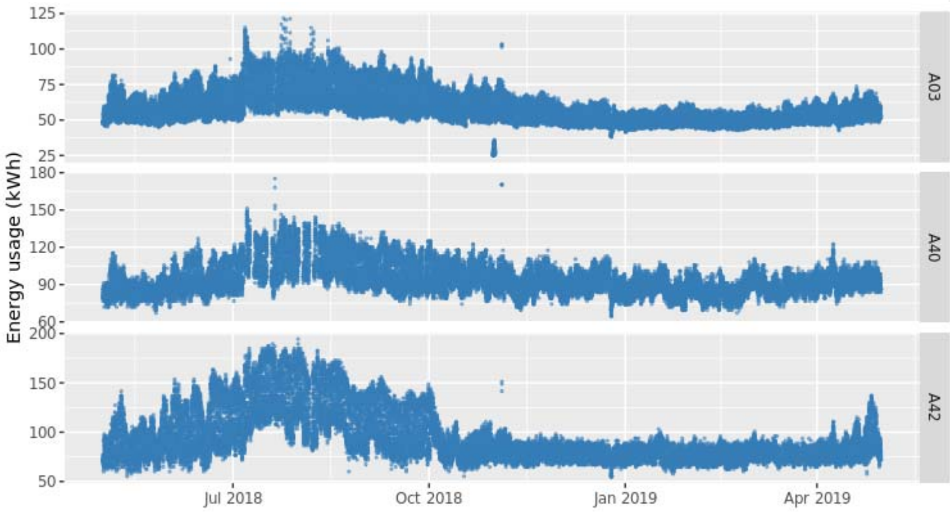
detected NREs in 8 percent of participant branches (n=3 of 39) and 11 percent of non-participant branches (n=9 of 85).

**FIGURE 8: PARTICIPANTS WITH BASELINE NRES**




Three additional buildings failed in Step 2 and were further evaluated in Steps 3 and 4. These three buildings passed the relative variance testing in Step 4, suggesting that the excess kurtosis produced a **false positive** indication of a baseline NRE. These buildings will move on to the next stage of model development.

**FIGURE 9: PARTICIPANTS WITH HIGH EXCESS KURTOSIS BUT NO BASELINE NRES**



The three sites with NREs shown in Figure 8 would require a follow-up conversation between the business managers and program staff. The goal of this outreach would be to learn what triggered the event, how long it lasted (temporary or permanent), and the likelihood it will happen in the reporting period. Many of these businesses have sophisticated energy management systems with historical records that will be useful for this type of diagnostic. If not, given the extreme nature of these events, it is likely that the branch manager would be able to recall what caused the change. Once we can explain the nature of the event, we can adjust the baseline model to better predict the future energy consumption at this branch.

## BASELINE MODEL DEVELOPMENT

To compare the options for baseline models of chain businesses, we used the AMI Customer Segmentation (AMICS) modeling approach to estimate load shapes for:

1. Individual sites;
2. Customer segments; and
3. Pooled chain businesses.

### TESTING PROCEDURES

First, we will review the baseline energy usage for each branch to determine which of the individual branches appear to be the best candidates for the proposed NMEC pilot program.

We will use these load shapes to select a subset of sites from each customer that appear to have the same operating hours, peak usage, and sensitivity to weather. These findings can be compared to the information provided to SCE by the business managers, who previously selected branches for participating, claiming that they are similar on all key characteristics. This independent validation of branch similarity will improve our confidence that these branches are similar enough to justify the use of a shared program intervention and M&V plan.

Next, we will estimate an AMICS model for each customer (i.e., chain business) as described in the following section, AMI Customer Segmentation Modeling Approach~~AMI Customer Segmentation Modeling Approach~~, including all of the branches that were identified as good candidates for NMEC and eligible for the program intervention in the previous step. We will consider three streamlined modeling approaches for each customer, including:

1. Individual sites – shared model specification and day bins, but separate load shape predictions for each branch;
2. Customer segments – grouping similar branches; and
3. Pooled businesses – model predictions are shared across all branches.

We will assess the relative error of each modeling approach based on the goodness of fit criteria that we have developed for our previous work with AMICS. This will include a cross validation exercise, utilizing a holdout group of days from the baseline period that are not used to develop the model, as well as all the baseline model fit criteria outlined in SCE's current NMEC guidelines. Most of the error thresholds are based on the metrics listed in SCE's current site-level NMEC procedures manual. We adjusted the fractional savings uncertainty (FSU) error threshold to reflect the CPUC's draft ruling for population-level NMEC issued October

**Formatted:** Font: 10 pt

2019, which is more stringent than SCE's threshold for site-level NMEC.[18] The formulas for each of these metrics can be found in the appendix (Model Error Metrics).

- Holdout test
  - NMBE
  - CV(RMSE)
- Full baseline model fit
  - NMBE<0.005%
  - CV(RMSE)<25%
  - FSU<25% at 90% confidence, bias adjusted
  - R-square>70% (preferably)

The final result of this task is a series of tables demonstrating the results of the pre-period holdout tests and full baseline model fit for each business and individual branch. The results for this sample of chain businesses will act as a proof-of-concept for the feasibility of streamlining NMEC baseline model development for the proposed pilot.

## AMI Customer Segmentation Modeling Approach

This section presents a general overview of the AMICS modeling approach. The AMICS approach estimates a separate load shape for each service account (i.e., distinct customer and service location) and day type, controlling for weather conditions and differences between weekdays and weekends.

The AMICS approach has been extensively tested on residential HVAC programs in Phase I of the AMI Billing Regression Study.[19] The ongoing Phase II study has expanded this research to include a variety of commercial HVAC programs and PG&E's residential Home Energy Reports.[20] We conducted a separate analysis of site level commercial HVAC savings for SCE in 2018 to demonstrate that the AMICS approach can be applied to individual commercial buildings and not just groups of program participants.[21] This study included a side-by-side comparison and cross validation exercise that found no significant difference in prediction error between AMICS and the Temperature and Time of Week (TTOW) modeling approach developed by Lawrence Berkeley National Laboratory (LBNL).

A key benefit of the AMICS model is avoiding over-reliance on 'average day' conditions. Most models essentially estimate the average load shape and then make a series of adjustments to that prediction depending on how the actual weather

[18] *Normalized Metered Energy Consumption Savings Procedures Manual (ET15SCE1130)*, December 2017 https://www.etcc-ca.com/reports/normalized-metered-energy-consumption-savings-procedures-manual

*Administrative Law Judge's Ruling for Issuing Draft Revised Rulebook for Normalized Metered Energy Consumption and Inviting Comments on Population-Level Rules, Measurement Methods and Calculation Software (Rulemaking 13-11-005)*, Valarie Kao, August 29, 2019.

http://docs.cpuc.ca.gov/SearchRes.aspx?docformat=ALL&docid=311581553

[19] *AMI Billing Regression Study*, Evergreen Economics, February 2016. http://calmac.org/publications/AMI_Report_Volume_1_FINAL.pdf

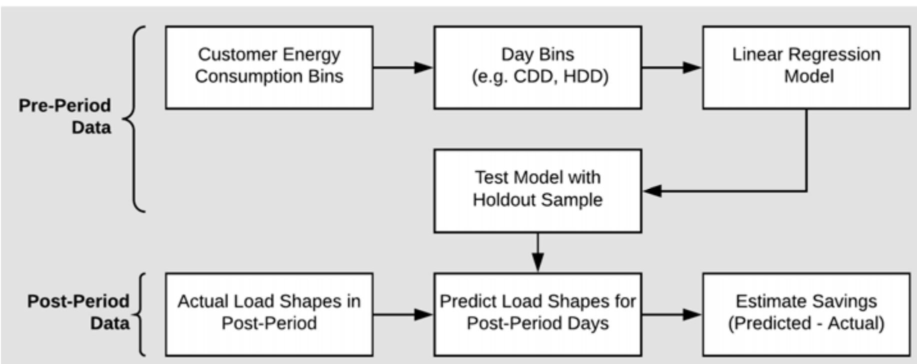
[20] *Advanced Metering Infrastructure Billing Regression Study: Phase II (Draft report)*, Evergreen Economics, August 2019.

[21] *AMI Analysis of Site Level Commercial HVAC Savings (ET17SCE1130)*, Evergreen Economics, July 2018.

conditions differ from this average. These models typically provide one annualized kWh savings number. AMICS parses out the savings into individual hours and days by segment to pinpoint the conditions that produce savings.

The AMICS approach uses segmentation to produce a portfolio of load shapes and then compares each day in the post-period against similar days in the pre-period, as shown in Figure 10~~Figure 10~~. When applied to an entire program, the AMICS model provides separate load shapes (and thus separate savings estimates) for each segment, which makes it a useful tool for targeting. Binning the data and then estimating separate regression models for each bin enables the overall model to control for a greater amount of the variation across both customers and weather conditions. This is not a proprietary “black box” method, but rather a series of simple linear regressions that are estimated with open source statistical software (R and PostgreSQL). Ultimately, the segmentation process reduces the prediction error for the load shape estimates, improving the predictive power of our models.

**FIGURE 10: AMICS APPROACH**




### CUSTOMER SEGMENTATION

Similar customers are modeled together, increasing the number of observations within each bin. The additional observations improve the model’s ability to separate out signals in energy usage from simple random noise. After modeling, the segments also provide insights into the characteristics of customers who are realizing the greatest energy savings from the program. In this way, customer segmentation can be an effective and meaningful process for evaluations focused on total program savings.

For this study, we explored a variety of customer segmentation techniques, including:

1. Individual
2. Chain business (parent company and/or business subtype)
3. Climate zone
4. Load shape – via *k*-means clustering
5. Daily energy usage

**Individual.** For programs with a small number of diverse customers, it is not always possible to construct meaningful customer segments that will consistently meet the normalized metered energy consumption (NMEC) error thresholds for a baseline model. This is likely for programs offering custom efficiency measures to commercial customers that are unique with respect to their building characteristics, operating hours, and economic activity. In these cases, each customer is assigned to their own bin, effectively constructing separate models for each individual customer. In this variation of the AMICS approach, we are no longer creating separate customer groups, but the segmentation of days (via weather conditions and day type) is still required.

**Chain business.** Customer segments comprised of all chains under the same parent company or chain business subtype can be sufficient to estimate savings for *individual sites* when the segments are constructed from a relatively homogenous target population (e.g., multifamily tenants) and/or a large number of customers with a full year of pre-period energy usage data. If the chains do not meet these criteria (homogeneity and large sample size), additional segmentation may be necessary to satisfy the NMEC error thresholds for baseline models. Chain subtype refers to the branch name, which may differ from the parent company name (e.g., market, superstore, corporate office, distribution center).

**Climate zone.** When participants cover a large geographic area, it can be beneficial to also segment by climate zone. The building climate zones defined by the California Energy Commission may help to control for differences in the typical climate (including temperature, humidity, and wind) as well as housing stock (e.g., building type, vintage, existing equipment).[22] Due to the small sample size, we aggregated climate zones into three categories: cool (climate zones 6-8), moderate (9), and hot (10, 13-16).[23]

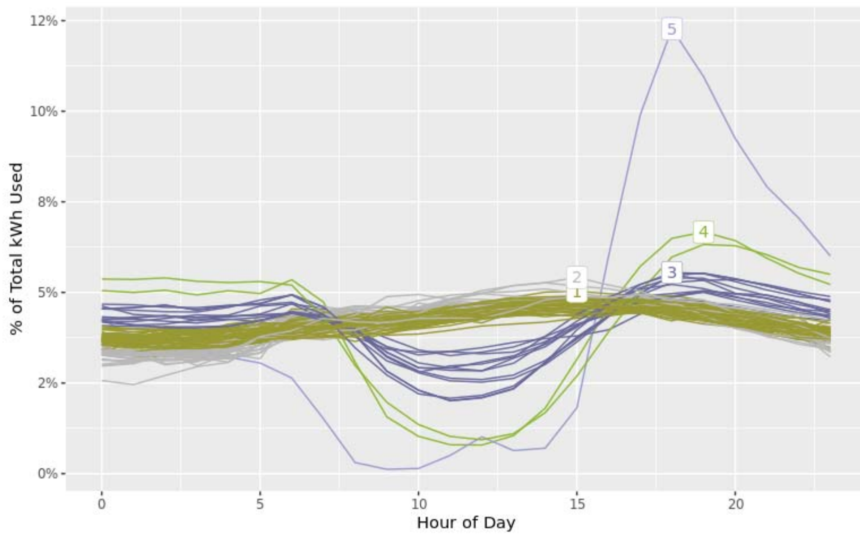
**Load shape.** The load shape bins are clusters of customers with similar hours of energy use. We used *k*-means clustering to identify the five unique clusters shown in Figure 11~~Figure 11~~, each containing a subset of customers with similar load shapes during the pre-period. Cluster analysis is a machine-learning algorithm designed to detect patterns in data.[24] In the AMICS application, the cluster analysis allows for identifying customers with similar load shapes and then grouping them together in the binning process. The benefit of cluster analysis is that similar customers are grouped automatically from the AMI data rather than relying on customer characteristics that are not typically tracked (or not regularly updated) in utility databases. Customers with similar energy usage on the average day (daily usage bin) can have drastically different load shapes. These load shape clusters help account for the differences in operating schedules, energy-intensive equipment, peak demand hours, and onsite solar generation.

[22] A description of the CEC climate zones can be found at https://ww2.energy.ca.gov/maps/renewable/building_climate_zones.html

[23] Our early model trials utilized individual climate zone bins instead of these three categories. However, the climate zone categories resulted in similar baseline model fit while also improving the overlap between customer bin assignments, which increased the number of participant customer segments with matched non-participant sites.

[24] The *k*-means clustering algorithm randomly assigns each customer's load shape to one of *k* clusters and then calculates the sum of the distance between each load shape and the centroid (i.e., average load) of the cluster to which it was assigned. Load shapes are then reassigned to the nearest cluster centroid, and the process is repeated until the variation within each cluster cannot be improved.

FIGURE 11: LOAD SHAPE CLUSTERS




**Daily energy usage.** A separate binning process is used to capture differences in average daily energy usage, without removing the weather-sensitive component. We assign customers to one of four bins by their average daily energy usage for the most recent full pre-period year, such that each bin represents 25 percent of the total kWh. The number of customers in each bin varies, with the highest energy usage bins containing the fewest customers. This binning strategy isolates customers who are atypical in terms of daily energy use, thereby reducing error in the model without removing these customers from the analysis. The last bin will include customers with the highest energy usage, such as those with large buildings or regular use of an inefficient air conditioning system.

### DAY SEGMENTATION

In addition to the segmentation schemes described above based on customer characteristics, each day of the study period is also categorized in terms of its weather, day type, and season.

The weather bins are created by calculating cooling degree-hours (CDH) for each hourly observation using a base temperature of 65 degrees Fahrenheit, and then taking the average of these hourly values to create a single cooling degree-day (CDD) value for each customer on each day (i.e., each "customer-day") in the study period.[25] These customer-days are assigned to a series of bins, each containing a

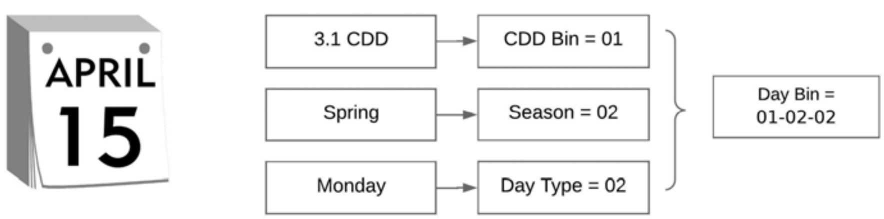
[25] A cooling degree-day (CDD) is a metric designed to measure the demand for energy required to maintain a comfortable temperature inside a building. It represents the number of degrees that the outdoor temperature exceeds an assumed baseline (in this case, 65°F), averaged across all hours in the day. By calculating this metric from hourly temperatures instead of daily averages, we can identify days that require some cooling during peak hours as well as heating in the early morning or evening.

range of three CDDs. Segmenting days by their CDD in this manner before the regression explicitly incorporates temperature into our model.[26]

To control for the differences in energy usage across days with the same weather conditions, we also binned by season and day type. The four seasonal bins are defined as winter (December-February), spring (March-May), summer (June-August), and fall (September-November). Every day of the week was assigned to its own bin; this helps control for any differences in operating hours and retail volume that vary throughout the week.

Figure 12~~Figure 12~~ provides an example of a single day being binned. Each customer was assigned to a single bin (with other branches of the same chain or as an individual), but because weather and day type change throughout the year, each customer has days that were assigned to many different bins.

**FIGURE 12: DAY SEGMENTATION EXAMPLE**




### REGRESSION MODEL

Once the data are segmented, the AMICS model approach involves estimating an ordinary least squares (OLS) regression model for each customer-day bin (Equation 5~~Equation 5~~) that contains a single dummy variable for each hour of the day.

**EQUATION 5: AMICS OLS REGRESSION MODEL**

$$kWh_{i,t} = \beta_{0i}H00_{i,t} + \beta_{1i}H01_{i,t} + \beta_{2i}H02_{i,t} + \cdots + \beta_{23i}H23_{i,t} + \varepsilon_{i,t}$$

Where:

$kWh_{i,t} =$ Energy consumption for a customer in bin $i$ during time interval $t$

$H00, H01, \ldots =$ Array of indicator variables (0,1) representing the hour of the day

$\beta_{0i}, \beta_{1i}, \ldots =$ Coefficients estimated by the model, for customers in bin $i$

$\varepsilon =$ Random error, assumed to be normally distributed

Unlike a traditional fixed effects regression model, which estimates a single set of slope coefficients for all customers and a constant term for each individual customer, the regression modeling approach employed by the AMICS model estimates a full

[26] This process was repeated to assign these same days to heating degree-day (HDD) bins, each containing a range of six HDDs. However, the holdout tests revealed that there was no improvement in model fit with the inclusion of HDD bins. It appears that the vast majority of these buildings use gas heat, and thus only the CDD term is necessary to explain the relationship between outdoor air temperature and electricity usage.

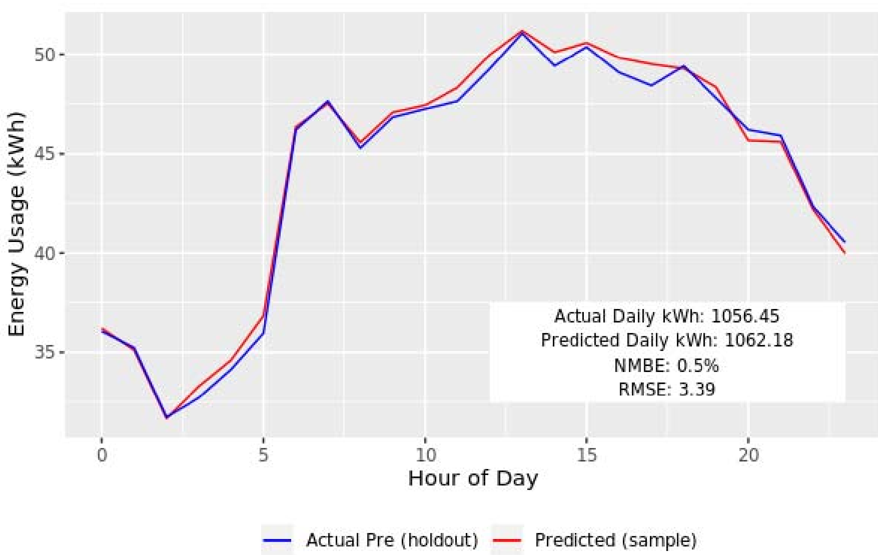
unique set of slope coefficient estimates for each customer segment (i.e., climate zone and load shape cluster) for each day bin (weather and day type).

### VALIDATION: HOLDOUT TESTS

To validate the AMICS model's ability to make accurate predictions, we conduct a holdout test (i.e., cross validation exercise) using only pre-period data. This involves randomly selecting 30 percent of weeks from each customer in our database as a holdout sample, defining the bins and estimating the model using the remaining 70 percent of the weeks, and then using the model results to predict energy usage for the holdout sample. This is sometimes referred to as a *cross-validation* exercise.

The results of one such holdout test are shown in Figure 13~~Figure 13~~, comparing the predicted pre-period load shape from the individual model (red line) to the actual pre-period load shape for the holdout (blue line) for a single holdout sample of site A20. When the model is preforming well, the two lines will overlap, and the error metrics will be minimized.

**FIGURE 13: EXAMPLE HOLDOUT SAMPLE IN BASELINE PERIOD, ACTUAL VS. PREDICTED USAGE (SITE A20)**



If the holdout test reveals segments with high prediction error, we can adjust the binning criteria (e.g., add day type bins) to refine the segmentation and then repeat the holdout process to confirm improvement.[27] The iteration process continues with small variations to the AMICS binning criteria until the model prediction error stops showing significant improvement. If multiple segmentation strategies result in similarly low prediction errors, the simplest model is selected for ease of interpretation.

## COMPARISON GROUP MATCHING

A matched comparison group is one option for estimating net savings for SCE's savings claim filings. We received a sample of non-participant branches within SCE's service territory for each of the three chain businesses. These differ from the participant branches that were hand-picked by the business manager for the proposed pilot. Our intention with this sample was to examine whether it would be feasible to utilize these sites as a comparison group to tease out any changes between the baseline and reporting period caused by external factors, rather than by program intervention.[28] Note that a comparison group is not necessary to determine payable savings, as performance payment calculations can be based on the NMEC gross savings with a static net-to-gross (NTG) adjustment.[29]

Segmentation in the baseline period identifies and groups customers with similar load shapes, seasonality, and climate prior to the program treatment. Performing difference-of-difference calculations within each customer segment in the reporting period will provide an estimate of net savings. This improves the validity of our comparisons, focusing on the impact of the program treatment, rather than all changes from the baseline to the reporting period.

We defined the customer segments based on our analysis of the characteristics and energy usage patterns of the 39 participant sites, as described in the previous chapter. Next, non-participants were assigned to these existing segments if they matched on all four binning criteria.

Table 6~~Table 6~~ shows the count of sites by chain after segmentation. Seventeen of the participant sites (n=3 from Grocery 1, n=7 from Grocery 2, and n=7 from Retail 1) were assigned to a customer segment that also contained at least one matched non-participant site. The 14 non-participants in these matched segments will act as a

[27] We consider a segmentation approach successful if the resulting AMICS model is able to separate patterns in energy usage from the simple random noise of individual observations, as measured by our holdout validation tests. This must be balanced with a need for easy interpretation, as the model results by customer segment will be used to provide insights into the characteristics of customers that were able to achieve the greatest energy savings.

[28] The proposed pilot uses an opt-in design, where participants are told about the program and decide whether or not they are interested in participating. Unfortunately, this introduces a concern about self-selection bias, where the customers that participate in the program are systematically different from those that do not choose to participate because they have additional motivation to pursue energy efficiency. In this case, the entire chain was offered an opportunity to participate and then representative from the chain business selected the 10-15 sites they believed would be a good fit for the program. The remaining non-participant branches provide our best estimate of how the participant branches would have continued to use energy, if it were not for the program intervention. While an opt-out, recruit-and-deny, or staggered participation program design may be more effective at controlling for self-selection bias, these carry additional risks such as customer confusion and dissatisfaction. A simpler option for assessing bias would be to ask follow-up questions of the chain representative to learn more about their motivations for participating and how they selected the 10-15 branches for participation.

[29] The CPUC's draft NMEC rulebook released August 2019 lists the following definition: "Payable savings are the savings determined via the method and calculation software described in a program's M&V Plan which constitute the basis of payments between the Program Administrator and Implementer(s). Payable savings determinations may differ from claimable savings in that payable savings may account differently for net-to-gross determinations, nonroutine events and outliers, and/or other similar considerations."

comparison group, providing an estimate for how the participants would have continued to use energy in the absence of the program intervention. Another eight participants were assigned to segments with other participant sites but did not match to any of the non-participants. The remaining 14 participants were assigned to solo segments, meaning that they were too unique to match with any other participant or non-participant sites.

**TABLE 6: SEGMENTATION STATUS OF SITES BY BUSINESS CHAIN**

| SEGMENTATION STATUS | BUSINESS CHAIN | N PARTICIPANT BRANCHES | N NON-PARTICIPANT BRANCHES |
|---|---|---|---|
| **Matched Segments** | Grocery 1 | 3 | 2 |
| | Grocery 2 | 7 | 4 |
| | Retail 1 | 7 | 8 |
| | **Subtotal** | **17** | **14** |
| **Segmented within Group** | Grocery 1 | 5 | 2 |
| | Grocery 2 | 3 | 0 |
| | Retail 1 | 0 | 23 |
| | **Subtotal** | **8** | **25** |
| **Solo Segments** | Grocery 1 | 5 | 9 |
| | Grocery 2 | 2 | 0 |
| | Retail 1 | 7 | 11 |
| | **Subtotal** | **14** | **20** |
| **Total** | | **39** | **59*** |

* The remaining 38 non-participant branches were excluded from the segmentation phase due to extreme non-routine events (NREs) or when they did not fall within the expected range for one or more binning criteria. For example, the Grocery 1 participant sites are from climate zones 9-15; we excluded all Grocery 1 non-participant sites located in climate zones 6 or 8.

Direct 1:1 comparison group matching is only possible with a large sample size and limited variation across customers. Our sample of 1:2 was insufficient to identify good matches for most of the participant sites. We tested a variety of segmentation methods, attempting to increase the number of participants that were assigned to a matched segment with non-participants. Unfortunately, these broad customer segments with increased overlap came at the cost of worsening the model fit and predictive power of our baseline models. The segments utilized in this report attempt to balance these two priorities.

Going forward, we recommend that the pilot measurement and verification plan propose one of the following options:

1. Start with a larger sample of non-participant sites (e.g., all non-participant branches) to increase the likelihood that good matches will be found for each participant.
2. Create synthetic comparison customers from the existing non-participant sample, combining energy consumption data from multiple non-participant sites that match on some, but not all, segmentation criteria. Note, this option would require further testing.

3. Utilize the DEER net-to-gross ratio of 0.95 for non-residential NMEC programs with a combination of measures or other methods to estimate program attribution.[30]

Future projects would benefit from a larger sample size. We could implement a power analysis to determine sample size for a more rigorous and comprehensive demonstration. However, to conduct the power analysis, we would need to know or have a good estimate of the full extent of variability in electricity usage of the population of non-residential buildings to which our analytical approach would be applied. Then, we would need data on a random sample of such buildings—for both the baseline and reporting periods. To date, we have been pragmatic in our approach in that we analyze all the available data. For example, the power analysis may indicate that we need a sample of 500 sites to meet an 0.80 level of power, but if there are only 200 branches in SCE's service territory, then the analysis will naturally be limited to 200 sites.

At this point, we can only speculate on which comparison group design—a larger sample with direct matching or synthetic comparison sites—would provide a better control for detecting impacts. Our hypothesis is that, given the uniqueness of each non-residential site, an individual comparison site will not be adequate in most instances. Instead, we hypothesize that a synthetic comparison "site" composed of multiple comparison buildings will better serve the role as a comparison because the impact of extreme random behavior of any one comparison site will be minimized. Further testing is necessary to determine an appropriate protocol for developing a synthetic comparison group. There are certainly instances where suitable comparison sites are not available, and we believe protocols can be developed for determining when that is the case. In these instances, a synthetic comparison customer (option 2) or alternate method (option 3) would be necessary to estimate net savings.

[30] *Resolution E-4952*, Public Utilities Commission of the State of California: Energy Division, October 11, 2018 http://www.deeresources.com/files/DEER2020/download/Resolution%20E-4952.PDF

# Findings

In this section, we provide the holdout test results and full baseline model error metrics for each model. The three main variations we tested were:

1. Individual sites – shared model specification and day bins, but separate load shape predictions for each branch;
2. Customer segments – predictions shared across similar branches; and
3. Pooled businesses – model predictions are shared across all branches.

Our assessment of baseline model fit is based on the following metrics. These are based on SCE's current site-level NMEC procedures manual and the CPUC draft ruling issued in August 2019, which includes a draft rulebook for population-level NMEC prepared by CPUC Energy Division staff.[31] Whenever the two rules disagreed, we utilized the more stringent error threshold. The formulas for each of these metrics can be found in the appendix (Model Error Metrics).

- **NMBE<0.005%** – The normalized mean bias error (NMBE) measures the average difference between the model prediction and actual metered energy usage. NMBE is a directional measurement; a negative NMBE indicates that the model underestimated the site's actual energy usage.
- **CV(RMSE)<25%** – The coefficient of variation of the root mean square error, CV(RMSE), measures the model's prediction error across the entire sample and is focused on the distance between the actual and predicted energy usage (not the direction).[32]
- **FSU<25% at 90% confidence, bias corrected** – The fractional savings uncertainty (FSU) is a measurement of whether the baseline model is sufficiently accurate to detect the estimated program savings (in this case, we assumed a minimum savings of 10%).[33] The "bias correction" refers to an adjustment in the error statistic that is made to account for the impact of correlated residuals (which can make model error appear unrealistically small); this correction is necessary for all daily or hourly models.
- **R-square>0.7 (preferable)** – The coefficient of determination (R-square) measures how well the independent variables predict variation in the dependent variable—in this case, how well the binning procedure and hourly indicator variables explain the variation in interval energy usage (kWh).[34]

---

[31] *Normalized Metered Energy Consumption Savings Procedures Manual (ET15SCE1130)*, December 2017
https://www.etcc-ca.com/reports/normalized-metered-energy-consumption-savings-procedures-manual

*Administrative Law Judge's Ruling for Issuing Draft Revised Rulebook for Normalized Metered Energy Consumption and Inviting Comments on Population-Level Rules, Measurement Methods and Calculation Software (Rulemaking 13-11-005)*, Valarie Kao, August 29, 2019.
http://docs.cpuc.ca.gov/SearchRes.aspx?DocFormat=ALL&DocID=311581553

[32] The NMBE can appear near zero when overestimations are consistently balancing out underestimations to create an accurate prediction on average. The CV(RMSE) does not measure direction (i.e., consistent bias), but focuses on the magnitude of the prediction error.

[33] When the FSU of our baseline models fall well below the error threshold of 25 percent, this suggests that we would be able to detect even smaller energy savings (<10%).

[34] Unlike the other error metrics, an R-square value below the threshold of 0.7 does not disqualify a baseline model from use in the baseline period. Low R-square values suggest that additional independent variables should be tested. If no additional variables are feasible (due to limitations in the data), then a lower R-square value may be accepted, provided that the thresholds for the other three error metrics are met.

There are no formal thresholds for the holdout test error metrics. High holdout error will sometimes occur if, by random chance, the 70 percent of weeks utilized in the sample model differ systematically from the remaining 30 percent of weeks in the holdout sample. Generally, we would hope to see the holdout:

- NMBE within ±1% and
- CV(RMSE)<25%.

In each of the model fit tables (starting with Table 7~~Table 7~~), we have provided these two metrics for the holdout test and all four error metrics for the full baseline model. Values are shown in red if they fail to meet the thresholds listed above.

## POOLED BASELINE

This section starts with the highest level of aggregation, where a single model is estimated for each chain. Note that pooled models can be accurate predictors of the average load shape across all branches in the business chain, but they will not necessarily create accurate predictions for a single individual branch. Hence, this option is only appropriate if the savings claims and performance payments are made for each chain, rather than for individual branches within the chain.

Table 7~~Table 7~~ shows the model fit statistics for baseline models pooled by business chain. This version includes all 39 participant branches and 62 non-participant branches (i.e., the subset with a full baseline year and no NREs) with the same parent company. The full baseline models for both grocery chains meet all four model fit criteria. These models are accurate enough to detect energy savings of less than 10 percent. The retail chain's baseline model has a high CV(RMSE) of 45 percent, well above the threshold of 25 percent. The combination of a low NMBE and high CV(RMSE) suggests that the predictions for the retail chain include many overestimations that are balancing out underestimations on other branches or other days. This balance results in an accurate average prediction and thus a low NMBE. The CV(RMSE) does not measure direction of error (i.e., consistent bias) and instead focuses on the magnitude of the prediction error. Hence, the high CV(RMSE) tells us that there are many branches or days with large variances between the predicted and actual energy usage.

**TABLE 7: POOLED CHAIN BASELINE MODEL FIT WITH ALL COMPARISON BRANCHES**

| | | Holdout Test | | Full Baseline | | | |
|---|---|---|---|---|---|---|---|
| Business Chain | N Branches | HO NMBE | HO CV(RMSE) | NMBE | CV(RMSE) | R-Square | FSU |
| Grocery 1 | 26 | **-2.27%** | **27.1%** | <0.001% | 22.2% | 0.954 | 3.32% |
| Grocery 2 | 16 | **1.38%** | 15.7% | <0.001% | 13.7% | 0.982 | 1.44% |
| Retail 1 | 56 | 0.66% | **45.9%** | <0.001% | **44.7%** | 0.839 | 9.87% |

Table 8~~Table 8~~ shows another variation on the pooled model that splits each business chain into separate models by subtype. While the subtypes have the same parent company, it would be reasonable to see consistent differences across chain subtypes and less variation within these groups (e.g., retail hours, product offerings, and end use equipment). The same 39 participants and 62 non-participant branches were included in this variation. Grocery 1 split into two subtypes, with improved model fit for the 14 branches in group A (reduction in CV(RMSE), increase in R-square and FSU), and worse model fit for the 11 branches in group B. Splitting by subtype helps to reveal the subset of branches with greater variation in energy usage during the baseline year, which may not be adequately represented in the original pooled model.

**Table 8: Pooled Chain Subtype Baseline Model Fit with All Comparison Branches**

| | | | Holdout Test | | Full Baseline | | | |
|---|---|---|---|---|---|---|---|---|
| Business Chain | Subtype | N Branches | HO NMBE | HO CV(RMSE) | NMBE | CV(RMSE) | R-Square | FSU |
| Grocery 1 | A | 14 | 0.58% | 20.7% | <0.001% | 16.7% | 0.973 | 2.2% |
| | B | 11 | **-2.97%** | **31.0%** | <0.001% | **26.0%** | 0.938 | 10.6% |
| | C | 1 | **-1.75%** | 9.4% | <0.001% | 6.3% | 0.996 | 1.6% |
| Grocery 2 | A | 16 | **1.38%** | 15.7% | <0.001% | 13.7% | 0.982 | 1.4% |
| Retail 1 | A | 37 | **1.73%** | **36.8%** | <0.001% | **35.1%** | 0.894 | 8.1% |
| | B | 16 | -0.74% | **53.3%** | <0.001% | **49.2%** | 0.814 | 17.9% |
| | C | 3 | -0.76% | 15.1% | <0.001% | 11.8% | 0.987 | 3.1% |

Table 9~~Table 9~~ provides the last variation of the pooled model, which restricts the non-participant sample to the subset of branches that exhibit similar energy usage patterns and characteristics as the 39 participant branches (based on the customer segmentation criteria described in the previous section). This filter reduced the non-participant sample from 62 to 27 branches (44%). While this was a significant filter, it would be unfair to include a more diverse population of non-participant branches in the baseline models than what is observed in the participants, as the non-participants are intended to provide our best estimate for how the participant branches would continue to use energy in absence of the program intervention. As shown in Table 9~~Table 9~~, the pooled chain subtype model with selected comparison branches has improved CV(RMSE) and R-square for Grocery 1 A and B and Retail 1 A chains. However, the CV(RMSE) and FSU for Retail 1 B both worsened (i.e., increased). The Retail 1B non-participant sample was very diverse, with only 2 of 11 branches exhibiting similar energy usage patterns and characteristics as the 5 participant branches. It is likely that this group will need further segmentation to create accurate model predictions.

**Table 9: Pooled Chain Subtype Baseline Model Fit with Matched Comparison Branches**

| | | | Holdout Test | | Full Baseline | | | |
|---|---|---|---|---|---|---|---|---|
| Business Chain | Subtype | N Branches | HO NMBE | HO CV(RMSE) | NMBE | CV(RMSE) | R-Square | FSU |
| Grocery 1 | A | 10 | **1.16%** | 19.7% | <0.001 % | 13.1% | 0.983 | 1.3% |
| | B | 9 | **3.26%** | 20.9% | <0.001 % | 18.0% | 0.969 | 6.0% |
| Grocery 2 | A | 16 | **1.38%** | 15.7% | <0.001 % | 13.7% | 0.982 | 1.4% |
| Retail 1 | A | 24 | **1.73%** | 23.2% | <0.001 % | 20.4% | 0.961 | 3.7% |
| | B | 7 | **-1.00%** | **63.8%** | <0.001 % | **57.3%** | 0.767 | **32.4%** |

## SEGMENTED BASELINE

The rationale behind customer segmentation is that we can further divide the chains into groups of customers that are similar enough to justify sharing a single set of predictions. Segmenting the customers and estimating separate regression models for each segment enables the overall model to control for a greater amount of the variation across both customers and weather conditions, improving baseline model fit. Compared to individual site-level models, segmented models are based on more observations (a full baseline year from every branch in the segment). The additional observations can, at least in theory, improve the model's ability to separate out signals in energy usage from simple random noise. After modeling, the segments also provide insights into the characteristics of customers who are realizing the greatest energy savings from the program. Savings claims and performance payments can be based on each customer segment or aggregated back up to the full business chain.

For this model variation, we segmented customers on:

1. **Chain business** by parent company and business subtype;
2. **Climate zone** to separate cool, moderate, and hot regions;
3. **Load shape** hours of energy use, via *k*-means clustering; and
4. **Annual energy usage**.

Table 10~~Table 9~~ provides the baseline model fit for each of the models by customer segment. Every one of these segmented models meets all the model fit criteria, indicating that these models would be adequate for NMEC savings estimation. However, 14 of the 39 participant branches (36%) are not shown in this table because they were assigned to their own customer segment; that is, they did not match with any other participant or non-participant branch on all four segmentation criteria. This is common in segmentation of commercial buildings, due to a wide variation across buildings (e.g., building size and composition, end use equipment) relative to more homogenous populations such as residential customers. See the earlier section **Error! Reference source not found.**~~Comparison Group Matching~~ for additional detail on the results of segmentation for matching between participants and non-participants. Individualized customer segments were excluded from Table 10~~Table 10~~ because they have the same model fit as the individual models provided later in Table 11~~Table 10~~.

**TABLE 10: SEGMENTED BASELINE MODEL FIT AMONG MATCHED BRANCHES**

| | | | | Holdout Test | | Full Baseline | | | |
|---|---|---|---|---|---|---|---|---|---|
| Business Chain | Subtype | Segment ID | N Branches | HO NMBE | HO CV(RMSE) | NMBE | CV(RMSE) | R-Square | FSU |
| Grocery 1 | A | 1221 | 3 | **1.02%** | 10.78% | <0.001% | 7.80% | 0.994 | 1.75% |
| | | 1222 | 3 | **1.23%** | 13.75% | <0.001% | 10.53% | 0.989 | 3.02% |
| | | 1232 | 2 | **1.09%** | 7.22% | <0.001% | 10.22% | 0.990 | 0.79% |
| | B | 6221 | 2 | -0.04% | 6.81% | <0.010% | 5.62% | 0.997 | 1.03% |
| | | 6231 | 2 | 0.60% | 6.57% | <0.001% | 5.12% | 0.997 | 1.17% |
| Grocery 2 | A | 4112 | 3 | 0.69% | 9.68% | <0.001% | 8.73% | 0.993 | 1.39% |
| | | 4211 | 3 | **2.02%** | 13.33% | <0.001% | 10.22% | 0.990 | 1.55% |
| | | 4212 | 8 | **1.05%** | 16.14% | <0.001% | 13.74% | 0.982 | 2.03% |
| Retail 1 | A | 5131 | 4 | **-1.01%** | 10.54% | <0.001% | 9.40% | 0.991 | 2.60% |
| | | 5141 | 4 | 0.51% | 11.38% | <0.001% | 9.36% | 0.991 | 2.83% |
| | | 5232 | 2 | 0.68% | 21.68% | <0.001% | 15.72% | 0.978 | 6.58% |
| | | 5233 | 2 | **3.33%** | 22.40% | <0.001% | 13.29% | 0.984 | 3.13% |
| | | 5241 | 5 | **1.35%** | 16.91% | <0.001% | 13.04% | 0.984 | 3.64% |
| | | 5341 | 2 | 0.10% | 12.66% | <0.001% | 10.12% | 0.990 | 4.63% |
| | | 5343 | 2 | **1.58%** | 16.53% | <0.001% | 13.29% | 0.983 | 3.52% |
| | B | 7141 | 2 | -0.71% | 10.86% | <0.001% | 10.33% | 0.990 | 5.27% |

## INDIVIDUAL BASELINE

Individual, site-level models are the simplest to explain and often produce strong model predictions for the commercial sites. While segmented models have the advantage of a higher number of observations, this does not guarantee an improvement in model fit. In fact, the benefits of segmentation are only realized when the segments are successfully identifying customers who are sufficiently similar to help explain patterns of energy usage in each other. If the branches have any systematic differences that are not controlled for in the regression model specification, the segmented model will have higher error (i.e., worse model fit) than an individual model.

Table 11~~Table 11~~ provides the baseline model fit for each of the models by site. Nearly all the individual models for participant sites (n=38 of 39) meet the error thresholds for all four metrics of model fit. The one exception is Retail 1 site A07, which has an extremely high CV(RMSE) and FSU. This is not too surprising because A07 was identified in the section on Non-Routine Event Detection~~Non-Routine Event Detection~~ as one of the sites with a significant temporary event during the baseline year, which will impact our ability to create a reasonable prediction of energy usage for the reporting period (after the program intervention).

**Formatted:** Font: 10 pt

Sites A05 and A09 from Retail 1 also have non-routine events (NREs), but these were permanent shifts in baseline energy usage. The seasonal binning criteria in the day segmentation of the AMICS model creates separate predictions for each season. This means that an NRE in winter will only impact the predictions for other days in the winter. Hence, the full baseline models for A05 and A09 were still sufficiently accurate to meet the NMEC error thresholds for a site-level model. However, these

NREs should still be investigated to confirm that the NRE is permanent (rather than temporary or seasonal), prior to estimating savings in the reporting (post) period.

**Table 11: Individual Baseline Model Fit by Site**

| Business Chain | Subtype | Site ID | Holdout Test: HO NMBE | Holdout Test: HO CV(RMSE) | Full Baseline: NMBE | Full Baseline: CV(RMSE) | Full Baseline: R-Square | FSU |
|---|---|---|---|---|---|---|---|---|
| Grocery 1 | A | A31 | **1.41%** | 6.75% | <0.001% | 12.63% | 0.985 | 1.12% |
| | | A32 | **-1.57%** | 6.20% | <0.001% | 4.60% | 0.998 | 1.23% |
| | | A33 | **1.16%** | 6.76% | <0.001% | 5.07% | 0.997 | 0.89% |
| | | A34 | 0.04% | 7.03% | <0.001% | 4.51% | 0.998 | 1.14% |
| | | A35 | 0.09% | 5.96% | <0.001% | 4.51% | 0.998 | 0.85% |
| | | A36 | 0.16% | 6.17% | <0.001% | 4.59% | 0.998 | 1.02% |
| | | A37 | **-2.32%** | 10.44% | <0.001% | 7.13% | 0.995 | 2.15% |
| | | A38 | **2.47%** | 8.74% | <0.001% | 5.77% | 0.997 | 1.86% |
| | B | A08 | -0.73% | 4.42% | <0.001% | 3.55% | 0.999 | 0.92% |
| | | A23 | 0.00% | 7.67% | <0.001% | 4.81% | 0.998 | 1.51% |
| | | A24 | **1.51%** | 5.06% | <0.001% | 3.98% | 0.998 | 0.92% |
| | | A26 | **1.66%** | 9.67% | <0.001% | 6.83% | 0.995 | 2.32% |
| | | A28 | **1.22%** | 14.02% | <0.001% | 9.65% | 0.991 | 3.46% |
| Grocery 2 | A | A01 | **1.04%** | 8.89% | <0.001% | 6.58% | 0.996 | 1.08% |
| | | A11 | -0.23% | 8.03% | <0.001% | 5.57% | 0.997 | 1.03% |
| | | A12 | -0.32% | 9.66% | <0.001% | 7.21% | 0.995 | 1.59% |
| | | A13 | 0.06% | 9.20% | <0.001% | 6.51% | 0.996 | 1.59% |
| | | A15 | 0.45% | 8.38% | <0.001% | 5.88% | 0.997 | 1.36% |
| | | A17 | 0.65% | 9.01% | <0.001% | 7.56% | 0.994 | 1.18% |
| | | A19 | 0.07% | 8.12% | <0.001% | 6.27% | 0.996 | 1.15% |
| | | A20 | -0.31% | 7.42% | <0.001% | 5.88% | 0.997 | 1.04% |
| | | A21 | 0.54% | 7.65% | <0.001% | 5.93% | 0.997 | 1.22% |
| | | A22 | **1.31%** | 8.73% | <0.001% | 6.49% | 0.996 | 0.90% |
| | | A27 | 0.32% | 9.50% | <0.001% | 6.67% | 0.996 | 1.62% |
| | | A30 | **1.01%** | 9.80% | <0.001% | 7.69% | 0.994 | 1.08% |
| Retail 1 | A | A04 | **3.28%** | 22.31% | <0.001% | 13.25% | 0.985 | 4.42% |
| | | A05 | 0.24% | 13.83% | <0.001% | 8.80% | 0.992 | 3.75% |
| | | A06 | **3.24%** | 18.14% | <0.001% | 13.67% | 0.983 | 4.93% |
| | | A10 | -0.16% | 6.73% | <0.001% | 4.36% | 0.998 | 1.10% |
| | | A29 | **4.32%** | 14.30% | <0.001% | 8.61% | 0.993 | 4.28% |
| | | A39 | -0.13% | 5.70% | <0.001% | 4.44% | 0.998 | 1.21% |
| | | A40 | -0.26% | 7.70% | <0.001% | 5.45% | 0.997 | 1.71% |
| | | A41 | 0.02% | 10.33% | <0.001% | 6.21% | 0.996 | 2.28% |
| | | A42 | **3.03%** | 9.85% | <0.001% | 6.49% | 0.996 | 2.01% |
| | B | A02 | **4.50%** | **29.00%** | <0.001% | 19.53% | 0.971 | 5.74% |

| | | | Holdout Test | | Full Baseline | | | |
|---|---|---|---|---|---|---|---|---|
| Business Chain | Subtype | Site ID | HO NMBE | HO CV(RMSE) | NMBE | CV(RMSE) | R-Square | FSU |
| | | A03 | 0.54% | 6.89% | <0.001% | 5.10% | 0.997 | 1.41% |
| | | A07 | **-135.65%** | **425.28%** | <0.001% | **131.59%** | **0.655** | **102.34%** |
| | | A09 | **-1.20%** | 19.35% | <0.001% | 15.97% | 0.978 | 9.01% |
| | | A25 | **-2.00%** | 8.90% | <0.001% | 6.37% | 0.996 | 2.75% |

## Summary of Models

Table 12~~Table 12~~ provides a summary of the baseline model fit by chain and model variation. The individual baseline models met all the NMEC error thresholds for all but one participant site. This suggests that individual models provide the most accurate predictions for the largest number of commercial customers in our sample. However, it is worth noting that the segmented and pooled models have sufficient accuracy to meet NMEC thresholds for both grocery chains. Only the retail chain required these individual models. If the proposed program does not commit to site-level performance payments or savings claims, it would be reasonable to proceed with a segmented or pooled model under a population-level approach, with the caveat that the program would need to target businesses that are relatively homogenous, such as grocery store chains. Additional detail with results by site is available in the appendix (Site-Level Results).

**Table 12: Baseline Model Fit Summary**

| Business Chain | Subtype | N Part Branches | N Comp Branches | Pooled (all) | Pooled (of matched) | Segmented[1] | Individual[2] |
|---|---|---|---|---|---|---|---|
| Grocery 1 | A | 8 | 2 | Pass | Pass | Pass (n=4) | Pass (all) |
| | B | 5 | 4 | | Pass | Pass (n=2) | Pass (all) |
| Grocery 2 | A | 12 | 4 | Pass | Pass | Pass (n=10) | Pass (all) |
| Retail 1 | A | 9 | 15 | **Fail** | Pass | Pass (n=7) | Pass (all) |
| | B | 5 | 2 | | **Fail** | N/A | Pass (n=4) + **Fail (n=1)** |

1. N refers to number of participant branches that were successfully segmented. The remaining branches were too unique for segmentation. These sites would need a larger non-participant sample to draw from or utilize an individual site-level model.
2. Only a single participant branch failed to meet the NMEC criteria for a successful individual site-level baseline model.

# Conclusions and Recommendations

This section provides the study conclusions and our recommendations for implementation of the proposed normalized metered energy consumption (NMEC) pilot. These should help address most of the limitations and barriers we faced during the analysis. Many of these recommendations can be applied to other NMEC program designs targeting commercial businesses.

## Pre-Screening

The program implementers should identify concurrent program participation and any non-routine events in the baseline year of energy consumption. Additional data collection during the application phase will be required to proceed with program participation (as these data will be necessary to produce accurate savings estimates).

- Net energy metering (NEM) should not disqualify sites from participation, provided that any onsite generation is interconnected prior to the start of the baseline year.
- Collect details on all third-party implemented demand response contract (3P DRC) programs.
- Use monthly billing records to identify and correct errors in interval kWh data.
- Use excess kurtosis and relative variance metrics to identify sites with potential non-routine events (NREs) in the baseline period that could impact NMEC savings.
- Ask customers to explain baseline NREs or exclude these sites from participation.

## Comparison Group

While a matched comparison group of non-participant branches is feasible, this will require a much larger sample or synthetic comparison customers to ensure a match for every participant branch.

- The participant population was too diverse to ensure a match for every branch, given our limited non-participant sample (2:1). A larger sample would have improved our chances of finding a good match for each participant branch more appropriate. Additional non-participant branches of these three chains do exist within SCE's service territory, but the data were not available in time for this study.
- Another option is to create synthetic comparison customers from a composite of multiple non-participants. A composite may resemble the participant branches more precisely than any one individual non-participant branch.
- A comparison group is only one option for estimating net savings and program attribution. Program implementers may want to consider the labor cost and administrative burden of each option.

## Baseline Models

Individual baseline models consistently provide the most accurate predictions. Pooled and segmented models may be considered for populations that are relatively homogenous, such as grocery chains.

- Individual, site-level, baseline models were consistently the most accurate. These can be aggregated to produce savings for the chain. Unless further research is conducted to identify sectors that will consistently be a good fit for a pooled or segmented model, we suggest that the proposed pilot program utilize site-level NMEC.
- Pooled and segmented models can provide sufficient accuracy to satisfy SCE's requirements for an NMEC baseline model. However, this application should be limited to relatively homogenous groups of businesses, such as grocery chains. For non-residential programs, pooled and segmented models are not necessarily any cheaper to estimate than site-level models, as the majority of the labor hours are devoted to data preparation and documentation.
- Both grocery chains had less variation across sites and within sites than the retail chain.

# Appendices

## MODEL ERROR METRICS

This section provides definitions of each model error metric listed in the report.

**EQUATION 6: MODEL GOODNESS OF FIT METRICS**

$$NMBE = \frac{\sum_{i=1}^{n} y_i - \hat{y}_i}{\sum_{i=1}^{n} y_i}$$

$$CV(RMSE) = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}}}{\bar{y}}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Where:

$NMBE =$ normalized mean bias error

$CV(RMSE) =$ Coefficient of variation of the root mean square error

$R^2 =$ coefficient of determination

$y_i =$ measured energy usage at time interval *i*

$\bar{y} =$ average of *n* measurements of energy usage

$\hat{y}_i =$ predicted energy usage at time interval *i*

$n =$ number of data points

$p =$ number of parameters in the model

**EQUATION 7: FRACTIONAL SAVINGS UNCERTAINTY FOR DAILY OR HOURLY MODELS**

$$FSU = \frac{y \times t` \times CV`}{F}\sqrt{\left(1 + \frac{2}{n`}\right)\frac{1}{m}}$$

Where:

$FSU =$ fractional savings uncertainty,

$y =$ 1.26 for hourly values and $-0.00024m^2 + 0.03535m + 1.00286$ for daily

$F =$ savings expressed as a fraction of the baseline period energy use

$=$ 0.10, minimum expected savings for projects approved through the pilot

$\rho =$ autocorrelation coefficient of residuals at lag 1

$n =$ number of measurement periods in the baseline period

$m =$ number of measurement periods in the reporting period

$n` = n \times (1-\rho)/(1+\rho)$

= corrected number of points or measurement periods in the baseline

$t`$ = corrected t-statistic for the specified confidence interval and degrees of freedom $(df = n` - p)$

$CV`$ = corrected CV(RMSE), replacing $n$ with $n`$

## SITE-LEVEL RESULTS

Table 13~~Table 13~~ provides a high-level summary of the baseline model fit by site, along with notes about additional data needs, such as third-party program participation details (i.e., measure type and EUL) and non-routine events identified during the baseline period.

**TABLE 13: PRE-QUALIFICATION SUMMARY BY SITE**

| | | | | Baseline Model Fit | | | | |
|---|---|---|---|---|---|---|---|---|
| **Business Chain** | **Subtype** | **Site ID** | **Bin ID** | **Pooled** | **Segmented** | **Individual** | **Additional Data Needs** | **Matched Comp?** |
| | | A32 | | | | PASS | - | |
| | | A33 | 1221 | | PASS | PASS | | |
| | | A38 | | | | PASS | | |
| | A | A34 | 1222 | PASS | PASS | PASS | | TRUE |
| | | A37 | | | | PASS | | TRUE |
| | | A35 | 1231 | | PASS | PASS | | |
| Grocery 1 | | A31 | 1232 | | PASS | PASS | | TRUE |
| | | A36 | 1321 | | PASS | PASS | | |
| | | A08 | 6141 | | PASS | PASS | | |
| | | A28 | 6223 | | PASS | PASS | | |
| | B | A23 | | PASS | | PASS | | |
| | | A24 | 6231 | | PASS | PASS | | |
| | | A26 | | | | PASS | | |
| | | A13 | 4111 | | PASS | PASS | LCR NRG participation | |
| | | A15 | | | | PASS | LCR NRG participation | |
| | | A17 | 4112 | | PASS | PASS | LCR NRG participation | |
| | | A21 | | | | PASS | LCR NRG participation | |
| | | A01 | 4211 | | PASS | PASS | | TRUE |
| Grocery 2 | A | A22 | | PASS | | PASS | | TRUE |
| | | A11 | | | | PASS | | TRUE |
| | | A12 | | | | PASS | | TRUE |
| | | A20 | 4212 | | PASS | PASS | | TRUE |
| | | A27 | | | | PASS | | TRUE |
| | | A30 | | | | PASS | | TRUE |
| | | A19 | 4312 | | PASS | PASS | LCR NRG participation | |

| Business Chain | Subtype | Site ID | Bin ID | Baseline Model Fit: Pooled | Segmented | Individual | Additional Data Needs | Matched Comp? |
|---|---|---|---|---|---|---|---|---|
| Retail 1 | A | A10 | 5141 | PASS | PASS | PASS | LCR Stem participation | TRUE |
| | | A40 | | | | PASS | LCR Stem participation | TRUE |
| | | A06 | 5143 | | PASS | PASS | | |
| | | A05 | 5231 | | PASS | PASS | Baseload NRE starting 06/06/18 | |
| | | A04 | 5233 | | PASS | PASS | | TRUE |
| | | A29 | 5241 | | PASS | PASS | | TRUE |
| | | A41 | | | | PASS | | TRUE |
| | | A42 | | | | PASS | | TRUE |
| | | A39 | 5131 | | PASS | PASS | | TRUE |
| | B | A25 | 7111 | FAIL | PASS | PASS | | |
| | | A09 | 7123 | | PASS | PASS | Baseload NRE starting 09/07/18 | |
| | | A02 | 7214 | | PASS | PASS | | |
| | | A07 | 7215 | | FAIL | FAIL | Temporary NRE from 02/15/19 to 02/28/19 | |
| | | A03 | 7221 | | PASS | PASS | | |

Table 14 provides the fractional savings uncertainty (FSU) for each site and model. Unlike the FSU provided in the summary tables in the Pooled Baseline and Segmented Baseline subsections of the Findings section of this report, this table shows the savings uncertainty if the chain pooled or segmented models were used to estimate savings for an individual site. The purpose of these models was to provide the best possible predictions for the average site under the average conditions, not necessarily to provide accurate predictions for individual sites within the population. The intent of this table is to emphasize the reduction in savings uncertainty for individual models versus segmented or pooled. While all three model variations may pass the minimum model fit criteria, models with lower FSU will be capable of identifying even smaller changes in energy usage during the reporting period (e.g., energy savings<10%).

**Table 14: Fractional Savings Uncertainty by Site and Baseline Model**

| Business Chain | Subtype | Site ID | Pooled | Segmented | Individual |
|---|---|---|---|---|---|
| Grocery 1 | A | A31 | 5.8% | 1.1% | 1.1% |
| | | A32 | 2.5% | 2.5% | 1.2% |
| | | A33 | 4.1% | 2.4% | 0.9% |
| | | A34 | 3.3% | 4.0% | 1.1% |
| | | A35 | 3.0% | 0.8% | 0.8% |
| | | A36 | 3.2% | 1.0% | 1.0% |
| | | A37 | 8.1% | 4.9% | 2.1% |
| | | A38 | 4.9% | 3.4% | 1.9% |
| | B | A08 | 23.9% | 0.9% | 0.9% |

| Business Chain | Subtype | Site ID | Pooled | Segmented | Individual |
|---|---|---|---|---|---|
| | | A23 | 3.7% | 2.0% | 1.5% |
| | | A24 | 3.5% | 1.3% | 0.9% |
| | | A26 | 14.3% | 2.3% | 2.3% |
| | | A28 | 17.0% | 3.5% | 3.5% |
| Grocery 2 | A | A01 | 4.4% | 2.2% | 1.1% |
| | | A11 | 4.8% | 4.0% | 1.0% |
| | | A12 | 3.3% | 3.3% | 1.6% |
| | | A13 | 7.0% | 1.6% | 1.6% |
| | | A15 | 3.2% | 2.0% | 1.4% |
| | | A17 | 2.3% | 2.0% | 1.2% |
| | | A19 | 2.6% | 1.2% | 1.2% |
| | | A20 | 6.7% | 6.2% | 1.0% |
| | | A21 | 3.3% | 2.3% | 1.2% |
| | | A22 | 2.4% | 2.3% | 0.9% |
| | | A27 | 5.3% | 4.7% | 1.6% |
| | | A30 | 2.3% | 2.0% | 1.1% |
| Retail 1 | A | A04 | **27.0%** | 4.4% | 4.4% |
| | | A05 | 10.5% | 3.8% | 3.8% |
| | | A06 | 20.0% | 4.9% | 4.9% |
| | | A10 | 7.6% | 2.9% | 1.1% |
| | | A29 | 9.3% | 6.7% | 4.3% |
| | | A39 | 4.8% | 2.5% | 1.2% |
| | | A40 | 6.4% | 3.7% | 1.7% |
| | | A41 | 15.9% | 8.5% | 2.3% |
| | | A42 | 5.4% | 6.0% | 2.0% |
| | B | A02 | **28.5%** | 5.7% | 5.7% |
| | | A03 | **48.9%** | 1.4% | 1.4% |
| | | A07 | **98.3%** | **102.3%** | **102.3%** |
| | | A09 | **93.2%** | 9.0% | 9.0% |
| | | A25 | **62.8%** | 2.7% | 2.7% |