Advanced Metering Infrastructure Billing Regression Study: Phase II

Final Report

June 5, 2020

Prepared for Pacific Gas and Electric,
San Diego Gas and Electric,
Southern California Edison, and
Southern California Gas Company

CALMAC ID: PGE0451.01

Table of Contents

Executive Summary

The Advanced Metering Infrastructure Customer Segmentation (AMICS) modeling approach has been extensively tested on residential HVAC programs in Phase I of the AMI Billing Regression study. This Phase II study has expanded this research to include a variety of commercial HVAC programs and the Gamma Wave of PG&E's residential Home Energy Reports program.

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 conditions differ from this average. 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 1. When applied to an entire program, the AMICS model provides separate savings estimates for each customer segment, which makes it a useful tool for targeting. Most other models provide one annualized kWh savings number. AMICS parses out the savings into individual hours and days by customer segment to pinpoint the conditions that produce savings.

Figure 1: AMICS Approach

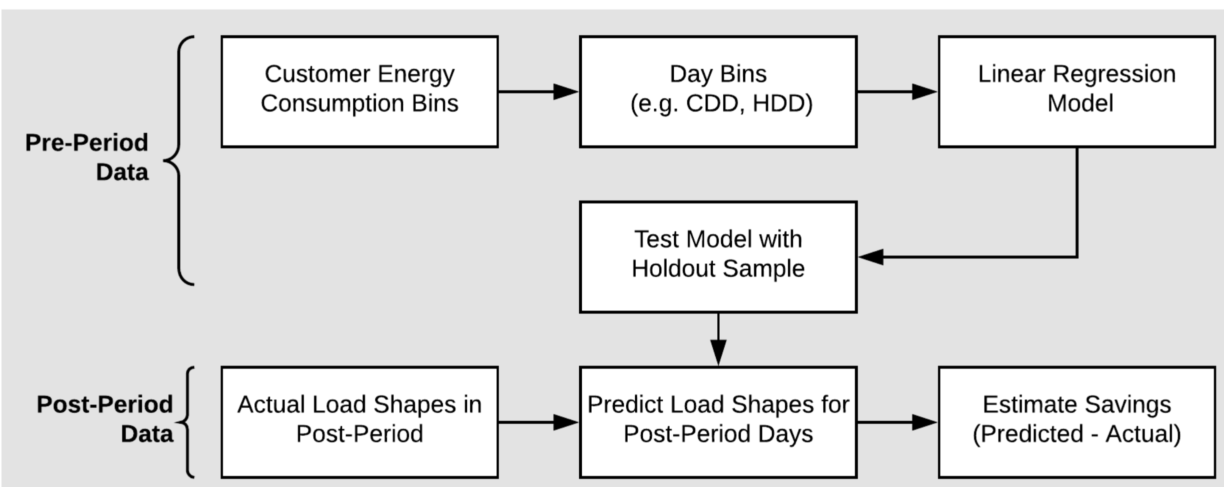


Table 1 provides a summary of the results from this Phase II research. We looked at two residential HVAC programs (the SCE Quality Installation and PG&E Quality Maintenance programs), the Gamma Wave of the PG&E Home Energy Reports program, four commercial HVAC programs offered by PG&E and SCE, and one commercial HVAC field data collection study. The holdout tests for each of these programs demonstrated that the AMICS model is able to produce reasonable load shape estimates, with prediction errors of less than 1 percent relative to the actual hourly energy usage of customers in the holdout samples. The AMICS model detected statistically significant savings for the SCE Quality Installation and PG&E Air Care Plus programs that were consistent with our expectations by season and time-of-day for improved air conditioning efficiency. While the AMICS model was not able to detect statistically significant savings at the program level for the remaining programs, our analysis by customer segment enabled us to identify a season and/or subset of customers (by baseline usage segment or industry) with savings.

Table 1: Summary of AMICS Results from Phase II

Program Type	IOU	Program Name	Holdout Error	Estimated Savings*	Savings Broken Out By
Residential HVAC	SCE	Quality Installation (QI)	-0.3%	6.0 ± 2.8%	Season, usage bin
	PG&E	Quality Maintenance (QM)	<0.1%	0.9 ± 1.8%	Season, usage & load shape bin
Home Energy Reports	PG&E	Home Energy Reports (Gamma Wave)	-0.1% control -0.4% treat	0.6 ± 1.0%	Season, load shape bin
Commercial HVAC	PG&E	Air Care Plus	-0.9%	3.9 ± 1.4%	Season, industry
		Quality Maintenance (CQM)		0.2 ± 2.2%	Season, industry
	SCE	Quality Maintenance (CQM)	1.0%	0.4 ± 1.5%	Season, industry
		Quality Installation (CQI)	0.3%	-0.5 ± 1.9%	Industry
	SCE	Field Data Collection Study	n/a	n/a	n/a

* Percentages represent kWh savings as a proportion of baseline kWh consumption.

Key findings:

- The AMICS model is able produce accurate load shape predictions for residential HVAC participants, households in the HERs treatment and control groups, and participants in each of the commercial HVAC programs.
- The estimated savings for the residential SCE QI program were consistent with our expectations by season and time-of-day for improved air conditioning efficiency.
- The AMICS segmentation of the residential PG&E QM program revealed that participants who were high energy users in the baseline period realized substantial energy savings from the program intervention.
- We found evidence of energy savings realized by the HERs treatment group above and beyond the natural changes observed in the control group, but these savings were not statistically significant at the program level.
- All of the commercial HVAC programs could benefit from improved targeting by business type.

1 Introduction

As electric utilities transition to advanced metering infrastructure (AMI), a greater amount and richer source of consumption data is becoming available to evaluators. A single customer's metered data at one-hour intervals translate to over 700 data points per month, providing an opportunity for evaluators to better understand the impacts that energy efficiency programs (and other factors) have on energy consumption during specific hours of the day, rather than a daily average derived from monthly data. A common concern among economists and other analysts working with monthly consumption data is that the aggregation conceals more than it reveals. The availability of short-interval meter data allows for potentially more accurate and robust models.

One of the key areas where AMI data have the potential to improve accuracy is in billing regression models used to estimate program savings—both for energy and demand impacts. Most of the literature to date has focused on using monthly consumption data, as these are typically all that have been available for estimating impacts at the program level. Other studies (particularly those regarding demand response programs) have utilized AMI data to estimate load shapes and demand impacts (Nexant's load impact evaluation for PG&E's SmartAC Program,[1] for example), but these models typically have been developed manually for each specific situation and therefore have not been practical for addressing a large number of customer types and time periods. Other works such as Hsiao et al.[2] provided an early application of the random coefficients model to energy efficiency, while Granderson et al.[3] have begun to look at developing AMI regression models in a more systematic fashion. None of these past studies, however, have presented a method for efficiently developing a large number of models that are tailored to a wide range of customer types and time periods that take full advantage of the information contained in the AMI data.

To explore how AMI data could be used in billing regression models, the California investor-owned utilities (IOUs) contracted with Evergreen Economics to conduct exploratory research using participant data from several residential HVAC programs.

---

[1] Nexant. 2014. *2013 Load Impact Evaluation for Pacific Gas and Electric Company's SmartAC Program.* Prepared for Pacific Gas and Electric.

[2] Hsiao, C., D. Mountain, M.W. Chan, K.Y. Tsui. 1989. "Modeling Ontario Regional Electricity System Demand Using a Mixed Fixed and Random Coefficients Approach." In *Regional Science and Urban Economics* Volume 19, Issue 4: 565-587.

[3] See for example: Granderson, J, PN Price, D. Jump, N. Addy, and M. Sohn. 2015. "Automated Measurement and Verification: Performance of Public Domain Whole-Building Electric Baseline Models." *Applied Energy* 144: 106-113. See also: Granderson, J., S. Touzani, C. Custodio, S. Fernandes, M. Sohn, and D. Jump. 2015. *Assessment of Automated Measurement and Verification (M&V) Methods*. Lawrence Berkeley National Laboratory, LBNL#-187225.

Phase I of this research was completed in 2016 and culminated in the development of a billing regression model that takes full advantage of AMI data granularity. This model is called the AMI Customer Segmentation (AMICS) model and proved to be very effective when tested using residential customer data.[4]

Phase I used two relatively small residential HVAC programs as testing grounds: Pacific Gas and Electric's residential Quality Maintenance (QM) program and Southern California Edison's residential Quality Installation (QI) program. This research demonstrated that the AMICS modeling approach produces similar results to a traditional fixed effects model at the program level, while providing valuable insights into the characteristics of customers and weather conditions that drive savings.[5]

To follow up on the promising results from the Phase I research, the IOUs contracted with Evergreen Economics in 2016 for Phase II of this study so that the AMICS model could be tested using a wider range of programs and customer types. The overarching goal of the Phase II research was to conduct a much more detailed investigation of the AMICS model and determine how well the modeling framework performed in a wider range of program applications. With this overarching goal in mind, the specific study objectives were to:

1. Refine the residential billing analysis methods using data from the same HVAC programs examined in Phase I;
2. Explore using the AMICS model to estimate savings for the Home Energy Reports (HERs) program;
3. Assess the AMICS model performance in evaluating commercial HVAC programs; and
4. Evaluate the AMICS model's potential capabilities for analyzing High Opportunity Programs and Projects, in regards to implementation of Assembly Bill (AB) 802.[6]

The steps for developing the AMICS model for each program are discussed in detail in the following section, but the basic steps are as follows:

1. Assign customers and weather conditions into distinct segments;

---

[4] In Phase I, this approach was referred to as the Random Coefficients Model (RCM), named for the specific type of regression. We have since re-branded the model to emphasize segmentation, the step that is unique to this approach.

[5] Evergreen Economics. 2016. *AMI Billing Regression Study Final Report*. Prepared for Southern California Edison. http://www.calmac.org/publications/AMI_Report_Volume_1_FINAL.pdf

[6] AB 802 directed the California Energy Commission to consider the overall reduction in normalized metered energy consumption (NMEC) in existing buildings as a measurement of energy savings.

2. Estimate average daily load shapes for each customer segment;
3. Use the load shape estimates from Step 2 to predict actual usage for a holdout sample of customers;
4. Calculate the prediction error and assess whether it meets the accuracy criteria established for the model; and
5. If the prediction error fails the accuracy test, repeat Steps 1 through 4 with different customer segments until the accuracy thresholds are met.

Throughout this report, we refer to differences in actual and predicted load shapes as "savings" that can be attributed to the program. For most of the programs examined in this study, however, there is no nonparticipant comparison or control group to include in the model that would help control for exogenous effects that might be impacting energy consumption. Consequently, the differences between actual and predicted energy use may be reflecting broader changes in the economy and not the result of any program actions. The results presented here should be interpreted with that caveat in mind. The exception to this is the HER program, where there is a nonparticipant control group available for our analysis.

The remainder of this report describes the AMICS modeling process in more detail, followed by different applications of the AMICS model to the different programs identified for this research. Table 2 summarizes the programs and available data for each program type.

Table 2: Data Sources Available for the AMICS Phase II Analysis

Program Type	IOU	Program Name	Number of Distinct Customers
Residential HVAC	PG&E	Quality Maintenance (QM)	31,615
	SCE	Quality Installation (QI)	2,119
Home Energy Reports	PG&E	Home Energy Reports (Gamma Wave)	152,292
Commercial HVAC	PG&E	Air Care Plus	1,503
		Quality Maintenance (CQM)	
	SCE	Early Retirement	5,059
		Quality Installation (CQI)	
		Quality Maintenance (CQM)	
		Quality Renovation (CQR)	
		Upstream HVAC	
		Field Data Collection Study	7

NOTE: This table provides the number of distinct customers with participation dates listed in the IOU program documentation. Some of these customers participated in multiple programs.

2 Methods

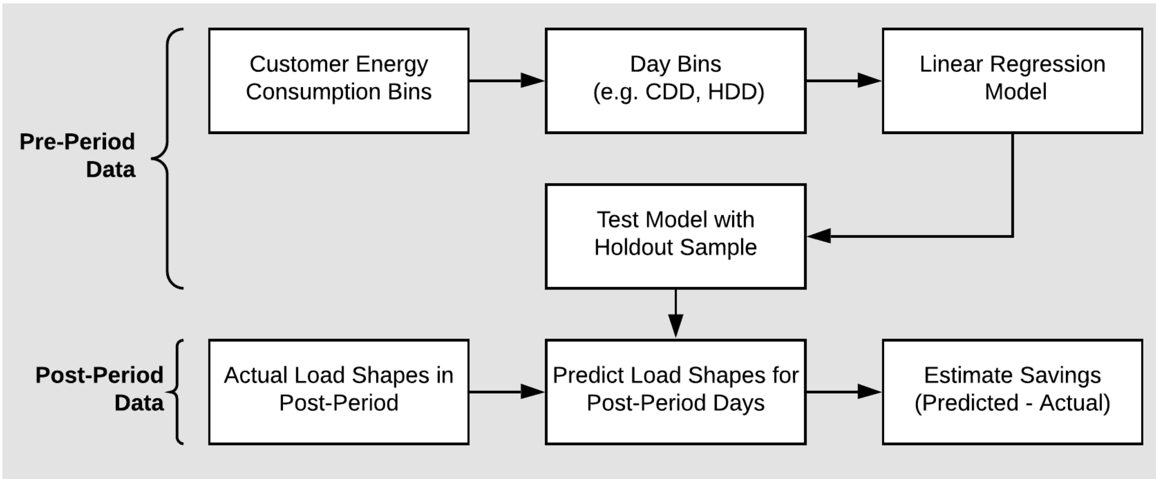
2.1 The AMICS Model

This section presents a general overview of the Advanced Metering Infrastructure Customer Segmentation (AMICS) modeling approach. Evergreen's previous research for SCE and PG&E has demonstrated that the AMICS model produces similar results to a traditional fixed effects billing regression at the *program* level. The real advantage of the AMICS model, however, is that the model provides very detailed and granular results that provide valuable insights into the characteristics of customers and weather conditions that drive savings. The AMICS model also provides detailed estimates of the time of day that the savings occurred.[7]

The AMICS approach produces a portfolio of daily energy use load shapes, representing how each customer uses energy across a wide range of different weather conditions. A unique step in the AMICS modeling approach is segmenting the AMI data into thousands of distinct segments (bins), as shown in Figure 2. Each bin contains interval energy use data for customers with similar energy usage patterns on days with specific weather conditions. 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. AMICS could easily be adapted to evaluate gas programs, providing the benefits of predictive power and energy savings by customer segment and day.

[7] The AMICS approach has been extensively tested and shown to accurately estimate energy savings for residential and commercial customers participating in HVAC programs, multifamily whole building retrofit programs, and home energy reports programs (both recipients and controls). In each study, repeated holdout testing was conducted to demonstrate the model's ability to make reasonable and consistent load shape predictions across the diverse sample of customers and days.

Figure 2: AMICS Approach



2.1.1 Segmentation

Much of the AMICS modeling process is devoted to categorizing customers based on the pre-period billing data (top half of Figure 2). The process used to develop the optimal customer segmentation scheme is described in more detail below.

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. Baseline energy usage – via fixed effects model
2. Daily energy usage
3. Load shape – via *k*-means clustering
4. Climate zone
5. Business type (e.g., retail, warehouse)
6. Building type
7. Individual

In most instances, a combination of two or three of these segmentation criteria are combined to create a series of small, more homogenous customer segments.

Baseline energy usage. One option for customer segmentation is to use a fixed effects regression model to estimate daily baseline electricity use for each home, while controlling for outside air temperature, as shown in Equation 1.[8] A characteristic of the fixed effects model is the estimation of a separate specific constant term (α_i) for each customer. This constant varies by customer site and accounts for time-invariant effects on consumption. In the model specification, the constant can be interpreted as site-specific baseline electricity consumption after controlling for variation in outside air temperature.

Equation 1: Fixed Effects Model Specification

$$kWh_{i,t} = \alpha_i + \beta_1 CDD_t + \beta_2 HDD_t + \varepsilon_{i,t}$$

Where:

$kWh_{i,t}$ = daily electricity consumption of customer i on day t of the pre-participation period

α_i = customer-specific fixed effect constant (i.e., estimated baseline consumption for customer i)

CDD, HDD = cooling and heating degree-days (base of 65°F)

$\beta_1, \beta_2, \ldots$ = coefficients estimated in the regression model

ε = random error term, assumed to be normally distributed

Once the constant term is estimated, the customers are ranked in ascending order of baseline energy use and assigned to one of 10 bins based on each home's weather normalized home usage in the pre-period. In this way, we group homes with similar energy consumption together. Each home group represents about 10 percent of total daily electricity (baseload) usage for the homes in our sample. Because of this, the number of homes in each bin varies, but the amount of daily kWh each bin represents is approximately the same.[9]

Daily energy usage. A separate binning process is used to capture differences in average daily energy usage, without removing the weather-sensitive component. This is simpler

---

[8] Before defining the customer segments and estimating any regression models, we removed days with fewer than 24 observations (one per hour) from the database to ensure that inclusion of incomplete days did not bias our estimates of energy consumption.

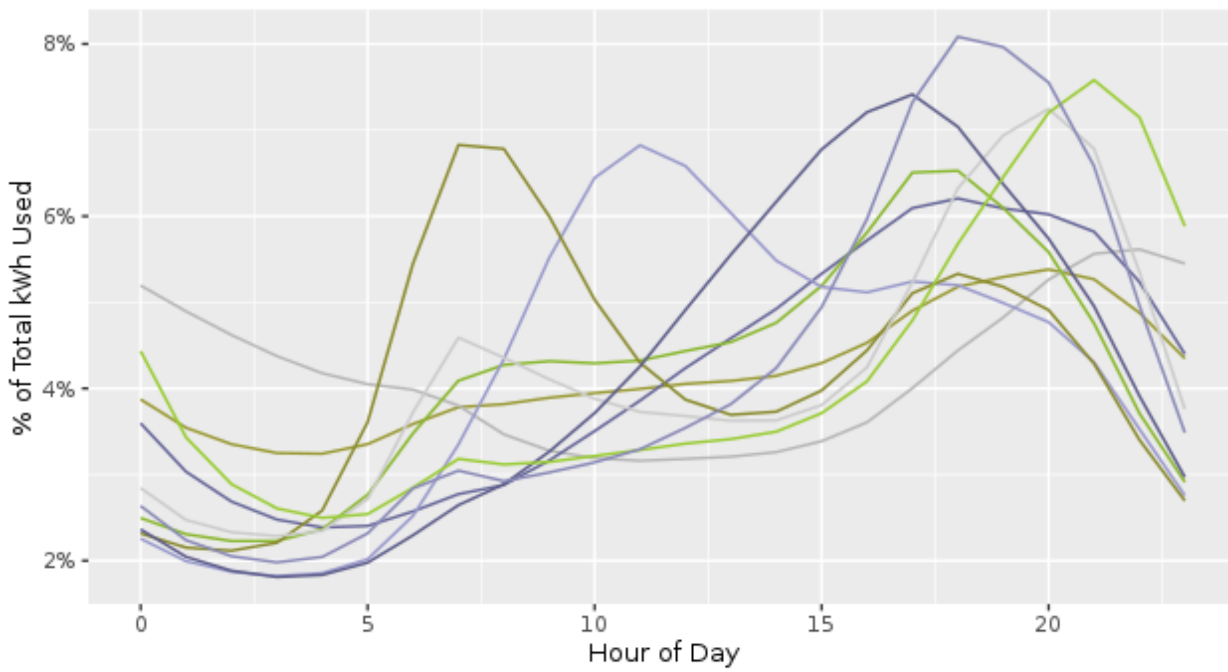
[9] Sites that are vacant in the pre-period due to long vacations, tenant turnover in rental properties, or other reasons will naturally fall into the lowest baseline usage group. If the site is not vacant during the post-period, the site's total usage will increase greatly and may mask program savings. The opposite is expected to be true as well. This is not a limitation of the binning procedure, but is a limitation on any analysis conducted with these data. If we had access to more information about these buildings (e.g., occupancy, owned vs. rental property, vacation vs. permanent residence), we would incorporate these elements into the binning procedure to limit any bias they may have on the resulting program savings estimates. To limit this potential for bias, we removed sites with extremely low consumption on the average day during the pre-period.

than the baseline approach, as it does not require any weather normalization. We assign customers to one of 10 bins by their average daily energy usage for the most recent full pre-period year, such that each bin represents 10 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 regular use of an inefficient air conditioning system, large buildings, and/or high occupancy buildings.

Load shape. The load shape bins are clusters of customers with similar hours of energy use. We used *k*-means clustering to identify the 10 unique clusters shown in Figure 3, each containing a subset of residential customers with similar load shapes during the pre-period. Cluster analysis is a machine-learning algorithm designed to detect patterns in data.[10] 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 occupant schedules, energy-intensive equipment, peak demand hours, and other factors.

[10] 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 3: Load Shape Clusters



Climate Zone. In programs where the 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).[11]

Building Type. For residential customers, building type can identify customers living in multi-family units and mobile homes from single-family detached units. Home size, occupant tenure, and vintage all contribute to energy usage and differ by building type. This van be a valuable option for customer segmentation when combined with another factor, such as load shape. The main limitation with building type for the residential programs is the data availability and variation within the participant population.[12]

---

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

[12] Building type was not utilized for any of the three residential programs in Phase II due to data limitations. All of the SCE QI participants had a building type listed as single-family; the vast majority (97%) of PG&E QM participants were in detached units, with a very small sample of shared-wall units (3%) and a single mobile home; building type was undefined for nearly all (99.4%) of the PG&E HERs participants.

Business Type. For commercial customers, a separate binning option is based on business type. The California investor-owned utilities (IOUs) collect and store information about their commercial and industrial customers based on the North American Industry Classification System (NAICS), which describes the primary business activity at the physical location associated with the utility account.[13] NAICS follows a hierarchical classification system, where the first two digits designate the sector, followed by digits designating the subsector, industry group, and then industry.[14] We defined broad industry groups using the first one to three digits of each six-digit NAICS code, as shown in Table 3. Businesses within each of these groups likely share some characteristics that are not directly represented in utility customer databases, such as operating schedules (e.g., seasonality in schools vs. retail) or primary end uses (e.g., lighting in retail vs. kitchen equipment in food service).

[13] We validated the NAICS codes of all SCE Commercial Quality Installation participants with manual lookups of each service address (i.e., distinct customer account and premise). While the NAICS provided by the utility were not flawless, there were some clear patterns in the types of businesses that were reclassified during this process. Of the 24 percent of valid NAICS codes that did not match on the first two digits, 52 percent were listed in SCE's data as lessors of real estate (531XXX). This was not surprising, as the utility classifications likely describe the property manager or lessor if they are responsible for paying the utility bill, rather than the business who occupies the building is actively using energy. Fortunately, the utility classification of lessors can still provide value in customer segmentation as it will isolate those business who are not responsible for their utility bill directly. As long as we segment businesses by a combination of NAICS and load shape, the differences in operating hours and major end uses will result in leased offices and leased retail spaces to be assigned to separate customer segments within the NAICS segment for lessors of real estate.

[14] The US Census Bureau's NAICS website can be found at https://www.census.gov/eos/www/naics/

Table 3: Business Type by Industry Classification

NAICS Code	Description
11****	Agriculture, forestry, fishing and hunting
22****	Utilities
23****	Construction
3*****	Manufacturing
42****	Wholesale trade
44****, 45****	Retail trade
48****, 49****	Transportation and warehousing
5*****	Information, finance and insurance, real estate, management; professional, scientific, and technical services
61****	Educational services
62****	Health care and social assistance
71****	Arts, entertainment, and recreation
721***	Accommodation
722***	Food services and drinking places
811***	Repair and maintenance (e.g., auto, household goods)
813***	Religious, civic, professional, and similar organizations
92****	Public administration
99	Undefined

When NAICS codes were not well defined across participants in a given program, we relied instead on customer segment, building type, or business type classifications provided in the program tracking data and/or utility customer databases.

Individual. Customer segment models can be sufficient to estimate savings in *individuals* 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. 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.

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.[15] These customer-days are assigned to a series of bins, each containing a range of six CDDs. This process is repeated to assign these same days to heating degree-day (HDD) bins, each containing a range of six HDDs. Segmenting days by their CDD and HDD in this manner before the regression explicitly incorporates temperature into our model.

To control for the differences in energy usage across days with the same weather conditions, we also binned by day type and season. Day type was typically defined by weekday versus weekend, with Saturday and Sunday assigned to day type 1.[16] The four seasonal bins are defined as winter (December-February), spring (March-May), summer (June-August), and fall (September-November).

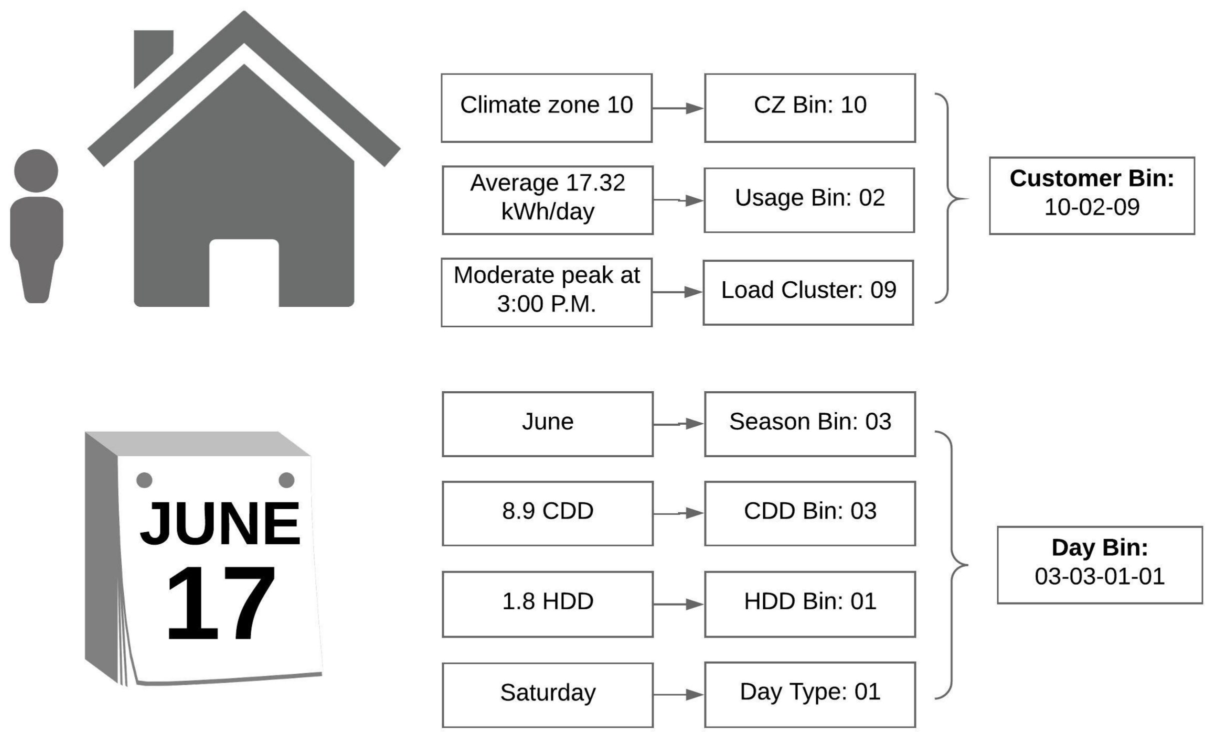
Composite Bins

Figure 4 provides an example of a single customer and day being binned. Each customer was assigned to just one customer bin, but because temperature and day type changes throughout the year, each customer has customer-days that were assigned to many different bins.

[15] 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.

[16] We explored variations in day type definitions, such as weekday versus weekend/holiday or seven distinct day-of-week bins. Unless these more granular day type bins resulted in significant reductions in model error, we opted for the simplicity of weekday versus weekend day types.

Figure 4: Customer-Day Segmentation Example



There are no rules for how many bins are required for the AMICS modeling approach to be successful. In general, fewer bins will increase the number of customers within each bin; as you increase the sample size for a model, you would generally expect to see tighter error bounds around the model predictions. However, this desire for large sample size in each bin must be balanced with a need to isolate outliers. If outliers with extremely high or variable energy usage are retained in bins with customers who have more consistent patterns of usage, then the model predictions for this entire customer segment will be influenced by the outlier and the variability will be reflected with wider error bounds. If the segmentation strategy successfully identifies this outlier as unlike all other customers, they will be assigned to a solo bin. A key to the AMICS approach is conducting trials with different segmentation strategies to compare the relative prediction error with holdout tests, as described in Section 2.1.3.

The customer-day segmentation process has the following benefits:

- Variation in CDD is controlled for in the bins so it does not need to be included as a variable in the model specification; the same is true for all other binning factors.
- Modeling customer-days allows us to exclude individual days with missing observations from the database. Rather than limiting the analysis to customers with

flawless data throughout the study period, we remove specific days with less than 24 complete hours of billing and weather data.

- Participants with no post-period observations are still useful when constructing models of the pre-period because they are simply a series of customer-days. These pre-period observations improve our ability to produce reasonable load shape predictions for other customers in the same segment that do have post-period observations. Later in the analysis, customers with no post-period observations are automatically excluded from the impact estimates.
- When binning annual observations by customer segment, weather, and day type, only one model is required. The output is generated at the bin level, so the model allows creation of load shapes and savings estimates for each bin (e.g., customer segment 20 on summer weekdays with 9-11 CDDs), group (e.g., annual savings for the customer segment with the highest energy usage during the peak period), or at the program level (i.e., annual savings across all customer segments), without the need to estimate additional models.

2.1.2 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 2) that contains a single dummy variable for each hour of the day.

Equation 2: AMICS OLS Regression Model

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

$$\begin{aligned} &Where: \\ kWh_{i,t} &= \text{Energy consumption, for customer in bin } i \text{ during hour } t \\ H00, H01, \ldots &= \text{Array of dummy variables (0,1) representing the hour of the day} \\ \beta_{0i}, \beta_{1i}, \ldots &= \text{Coefficients estimated by the model, for customers in bin } i \\ \varepsilon &= \text{Random error, assumed normally distributed} \end{aligned}$$

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 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).

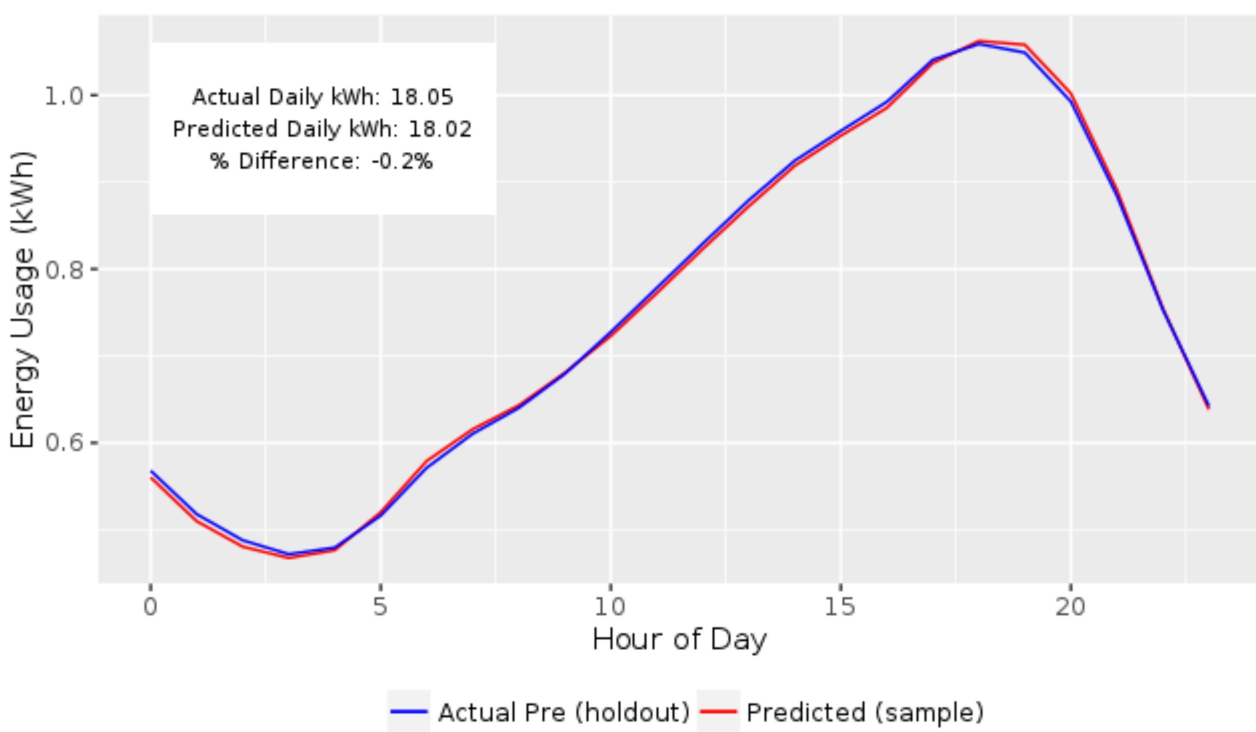
2.1.3 Validation: Holdout Tests

Load Shape Predictions

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 the customers in our database as a holdout sample, defining the bins and estimating the model using the remaining 70 percent of the customers, 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 holdout test are shown in Figure 5, comparing the predicted pre-period load shape from the model (red line) to the actual pre-period load shape for the comparison group from the holdout sample (blue line). When the model is performing well, the two lines will overlap. The holdout test relies exclusively on pre-period data so that any differences between the predicted and actual energy usage can be attributed to model error, and not to program savings. In this case, the model predictions track almost perfectly with the actual energy usage of the comparison group, with a total difference of only 0.2 percent on an average day.

Figure 5: Holdout Actual vs. Predicted Usage, Comparison Group in Pre-Period



If the holdout test reveals customer or day bins with high prediction error, we can adjust the binning criteria (e.g., the number of load shape clusters) to refine the segmentation and then repeat the holdout process to confirm improvement.[17] The iteration process continues with small variations to the AMICS binning criteria until the model prediction error stops showing significant improvement. If multiple binning strategies result in similarly low prediction errors, the simplest model is selected for ease of interpretation.

2.1.4 Predictions and Savings Estimation

Once we are confident that our AMICS model is sufficiently able to predict the pre-period consumption for the customers in our holdout sample, we re-estimate the model using the full sample (no holdout) to take advantage of all available data. We then use the model to predict what the load shapes would have looked like (for each customer segment on each day) in the post-period if the program intervention had not occurred. Finally, we compare the predicted load shapes to actual energy consumption over the same period to determine the total change in energy usage from the pre- to post-period, while controlling for any differences in weather and day type.

The AMICS model produces separate energy savings estimates for every day in the post-period. These can be aggregated to whatever level is requested by the utility, such as a single average annual savings by hour or the average energy savings which occurred during the utility's system peak demand (e.g., summer weekdays from 4-9pm). As long as interval data is available, peak demand savings can be estimated with the AMICS approach. However, our primary focus for this study was a demonstration of model fit, annualized hourly energy savings, and examples of patterns in energy savings by customer segment.

Computing Standard Errors

In the AMICS approach, we estimate individual regression models for thousands of customer-day segments, providing a kWh energy usage prediction for each hour.

Because the AMICS model is estimated using the pre-period data, we compute the relative variance for each hour of the day for each customer-day bin as the ratio of the variance to predicted hourly kWh usage. These relative variances are then applied to the post-period data to create confidence intervals for the model predictions of each hour of each customer-day in the post-period. With 24 hours per day and thousands of customer-day segments, we compute over 24,000 confidence intervals. For aggregated predictions, such as the annual and seasonal post-period load shapes, we use bootstrapping to estimate the

---

[17] 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.

relative variance for each hour, accounting for variation in the number of observations and relative kWh represented by each customer-day bin.

Any bias in the AMICS model predictions detected in the holdout validation test will be reflected in the error bounds on the predictions of post-period energy use and the corresponding savings estimates.

2.2 Database Creation

2.2.1 IOU Programs

The primary focus of this study was to test the capability of the AMICS approach across a wider range of customer types and efficiency programs than Phase I. To accomplish this, Evergreen received a sample of nearly 200,000 utility customers that participated in one or more IOU programs in recent years. Table 4 summarizes customer data available for this analysis.

Table 4: IOU Program Participant Data Sources

Program Type	IOU	Program Name	Number of Distinct Customers	Participation Dates
Residential HVAC	PG&E	Quality Maintenance (QM)	31,615	Jan 2015 to Jul 2016
	SCE	Quality Installation (QI)	2,119	Dec 2013 to May 2016
Home Energy Reports	PG&E	Gamma Wave	152,292	Nov 2011
Commercial HVAC	PG&E	Air Care Plus	1,503	Apr 2006 to Jul 2016
		Quality Maintenance (CQM)		
	SCE	Early Retirement	5,059	Feb 2010 to Aug 2016
		Quality Installation (CQI)		
		Quality Maintenance (CQM)		
		Quality Renovation (CQR)		
		Upstream HVAC		
		Field Data Collection Study	7	Aug 2016 to Feb 2017

NOTE: This table provides the number of distinct customers with participation dates listed in the IOU program documentation. Some of these customers participated in multiple programs.

The residential HVAC programs include PG&E's Quality Maintenance and SCE's Quality Installation programs. These programs were the focus of our Phase I research, for which the AMICS approach was originally developed. The updated AMICS model for these programs utilizes more recent program data to demonstrate improvements we have made to the customer segmentation process, along with improved graphics and model fit diagnostics.

The PG&E Home Energy Reports (HERs) program produces relatively small energy savings, but the opt-out randomized control trial (RCT) design makes it feasible to estimate net savings at the program level. We received a large sample of treatment and control group customers from the Gamma Wave of the HERs program, which includes a diverse population of households from all energy usage quartiles across PG&E's service territory. This HERs analysis adapts the AMICS approach for use in an RCT program evaluation, with control group matching and energy savings estimated with a difference-of-differences approach.

We focused on three commercial HVAC programs, including Air Care Plus, Quality Maintenance, and Quality Installation. These programs incentivize a variety of HVAC improvement activities including the purchase and installation of new efficient HVAC systems and/or maintenance contracts for system repairs and tuning. The remaining commercial HVAC programs (Early Retirement, Quality Renovation, and Upstream HVAC) had small participant populations with limited program tracking documentation. We decided to focus on the larger programs for this phase of commercial modeling.

In addition to these programs, Evergreen received data from SCE's recent Commercial HVAC Quality Installation field data collection study.[18] These data included a small sample of commercial customers with thorough on-site testing and metering of individual HVAC units, both before and after system upgrades. We obtained both HVAC metering and whole building AMI billing data for this sample to use as a case study of commercial load disaggregation with the AMICS approach.

2.2.2 Customer Account and Billing Data

SCE and PG&E provided Evergreen with AMI hourly or 15-minute interval electricity billing data and account characteristics for each customer participating in one or more of the IOU programs listed in Table 4. The AMI billing data for this study contained nearly 5.2 billion observations from November 1, 2010, to October 31, 2017. For consistency across customers in the study, all 15-minute interval billing data were aggregated to the hourly

[18] National Comfort Institute Inc., Energy Solutions, and Solaris Technical LLC conducted this Commercial HVAC Quality Installation field data collection study to support SCE's program development.

level. These data capture energy usage patterns of each customer before, during, and after program participation.

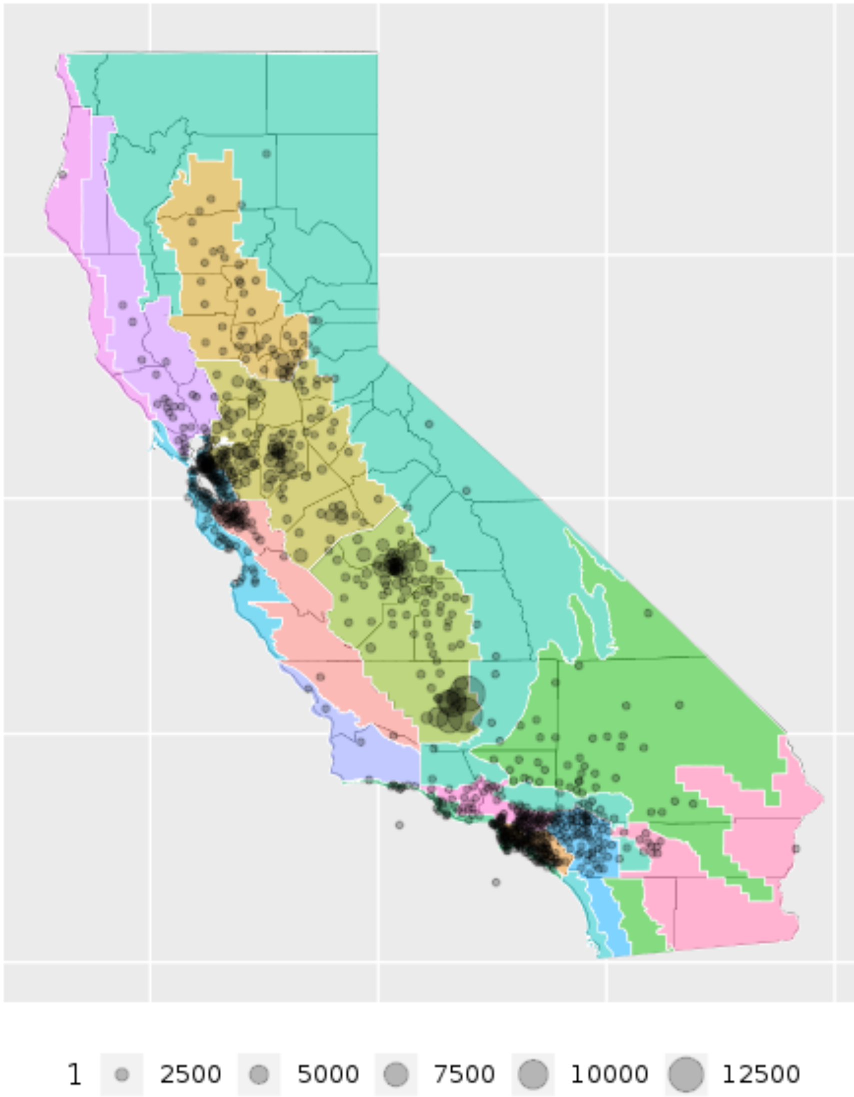
Many of these customers are likely duel-fuel with natural gas or other fuels besides electricity. As the original study was designed around a need for reliable, hourly interval load shape predictions for program evaluations, this phase of research was focused on electricity. Future work should extend this research to gas programs to demonstrate the applicability of the AMICS approach and this research for gas utilities.

2.2.3 Weather

Evergreen appended weather data obtained from the National Oceanic and Atmospheric Administration (NOAA) to develop a database that includes both AMI consumption information and hourly weather data for each customer 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.[19] 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. The distribution of participating customers across counties and climate zones is shown in Figure 6.

[19] *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

Figure 6: Participating Customers by County and Climate Zone



3 AMICS Applications

3.1 Residential HVAC

This section provides our analysis of two residential HVAC programs: PG&E's Quality Maintenance (QM) and SCE's Quality Installation (QI) programs. Evergreen originally developed the AMICS approach using prior participants of these two programs, described in the Phase I report finalized in February 2016. The new batch of participants we received for Phase II provided us with an opportunity to test and demonstrate improvements to customer segmentation and interpretation of the variation in savings across participants.

The SCE QI program provides a good candidate program for testing the AMICS model approach, with a participant pool of 2,126 customers and *ex ante* savings of over 5 percent, including new HVAC equipment and strict installation protocols that will improve the likelihood that savings will be detectable using a billing regression. The PG&E QM program has a larger population of nearly 30,000 participants but smaller *ex ante* savings (<5%) from HVAC testing and maintenance activities as needed. The savings from QM are not just smaller, but also more varied across customers than the QI program. This reduces the likelihood that we will be able to detect statistically significant savings with billing analysis.

We used similar filters and segmentation strategies for both residential HVAC programs. The holdout tests for both programs demonstrate that the AMICS model is able to produce accurate estimates of load shapes for participants of residential HVAC programs, accounting for the variation in load shapes across all four seasons. The AMICS model detected statistically significant savings for the QI program, consistent with our expectations by season and time-of-day for improved air conditioning efficiency. The QM program savings estimated by our model were not statistically significant during most hours, despite the tight error bounds around our predictions. However, the AMICS segmentation revealed a wide variation in energy savings across customer segments and weather conditions, with more substantial energy savings being realized by high energy users on days with low to moderate cooling loads.

3.1.1 SCE Quality Installation

Program Description

The SCE Quality Installation (QI) program is a California statewide program designed to achieve energy and demand savings through the installation of replacement split or packaged HVAC systems in accordance with industry standards.

Each home participating in the SCE QI program replaced an existing HVAC system[20] using an installation contractor who received additional training in quality installation practices through the program. The contractor is responsible for ensuring that the air conditioning unit is sized properly for the home and installing the new unit based on strict ENERGY STAR requirements, as well as connecting it to the ductwork/distribution system.[21] The quality installation theoretically improves cooling delivery (from reduced runtime and/or power draw) by preventing common installation problems that may cause the new unit to operate below its energy efficiency specification.

Note that the savings estimates from the AMICS model are measured as the difference in predicted and actual usage in the post-installation period. In the analysis presented here, the entire decrease in residential energy consumption is attributed to the QI program. This is consistent with the interpretation for any billing regression model used to estimate gross impacts.[22]

Database

SCE provided Evergreen with AMI billing data and account characteristics for the 2,126 distinct residential customers who participated in the QI program between December 2013 and May 2016. The SCE QI program data included HVAC technology type (central air conditioner vs. heat pump), installation date(s), SEER rating, size in tons, BTU per hour, *ex ante* gross energy and demand savings, climate zone, building type, and monthly rate schedule.

The AMI billing data for this study contained approximately 64 million hourly or 15-minute intervals from January 1, 2014 to September 23, 2016.[23] These data captured energy usage patterns of most households for a full year before their participation in the program.

We applied filters to exclude customers with:

- Net energy metering, such as onsite solar generation (n=341);
- No pre-period observations in the billing data (n=15); or
- Extreme changes from the pre- to post-periods of more than 150 percent or less than

[20] Eligible homes installed a new package or split system air conditioner or heat pump with a Seasonal Energy Efficiency Ratio (SEER) of 14 or greater.
[21] ANSI/ACCA 5 QI-2010: HVAC Quality Installation Specification
[22] It should also be noted that California (through Assembly Bill 802) is moving toward a new evaluation protocol where meter-based savings would be calculated; consequently, the existing equipment conditions would be used to measure savings. The method demonstrated here for the QI program is consistent with the Assembly Bill 802 approach.
[23] For consistency across customers in the study, all 15-minute interval billing data were aggregated to the hourly level before analysis.

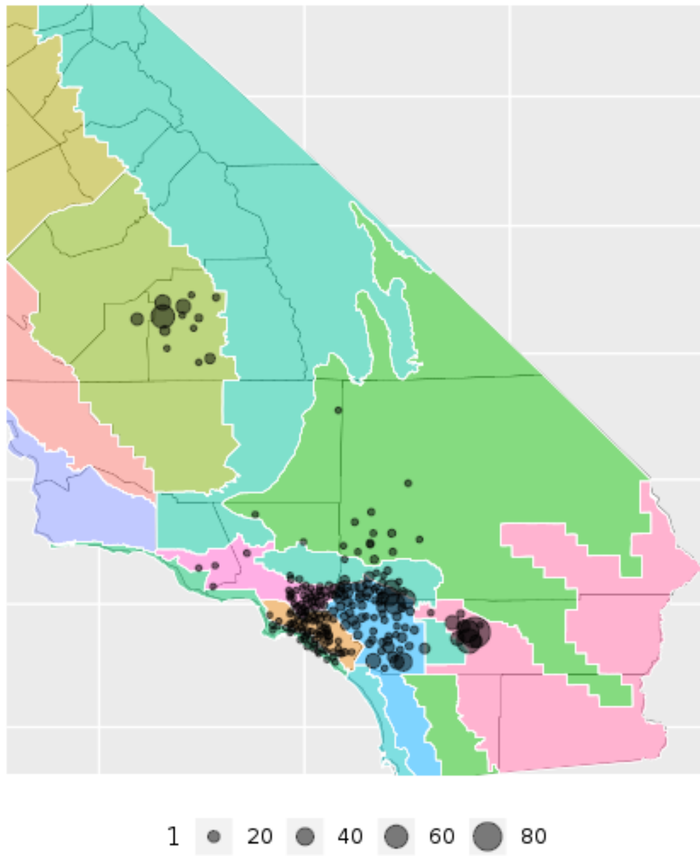
-66 percent (n=22).

These data screens reduced the analysis sample to 1,748 customers for the SCE QI program.

As shown in

Figure 7, this sample covers many different counties and climate zones, with a large number of participants concentrated in Los Angeles and surrounding cities. All of the participant dwellings (100%) were single-family homes. The homes ranged in size from 425 to 5,500 square feet, with an average of 1,900 square feet. Nearly one third (30%) of the participants installed HVAC equipment with a SEER rating of 18, but most of the installed equipment had a SEER rating of at least 16 (81%). One third (33%) were enrolled in one or more demand response programs, including 377 customers with direct load control switches that allow SCE to cycle their air conditioning equipment during summer peak events. One advantage of the AMICS modeling approach is that individual demand response event days can be excluded from the model, to avoid mis-attributing changes in peak load (i.e., savings) exhibited by participants to the HVAC program participation that are actually caused by demand response program participation.

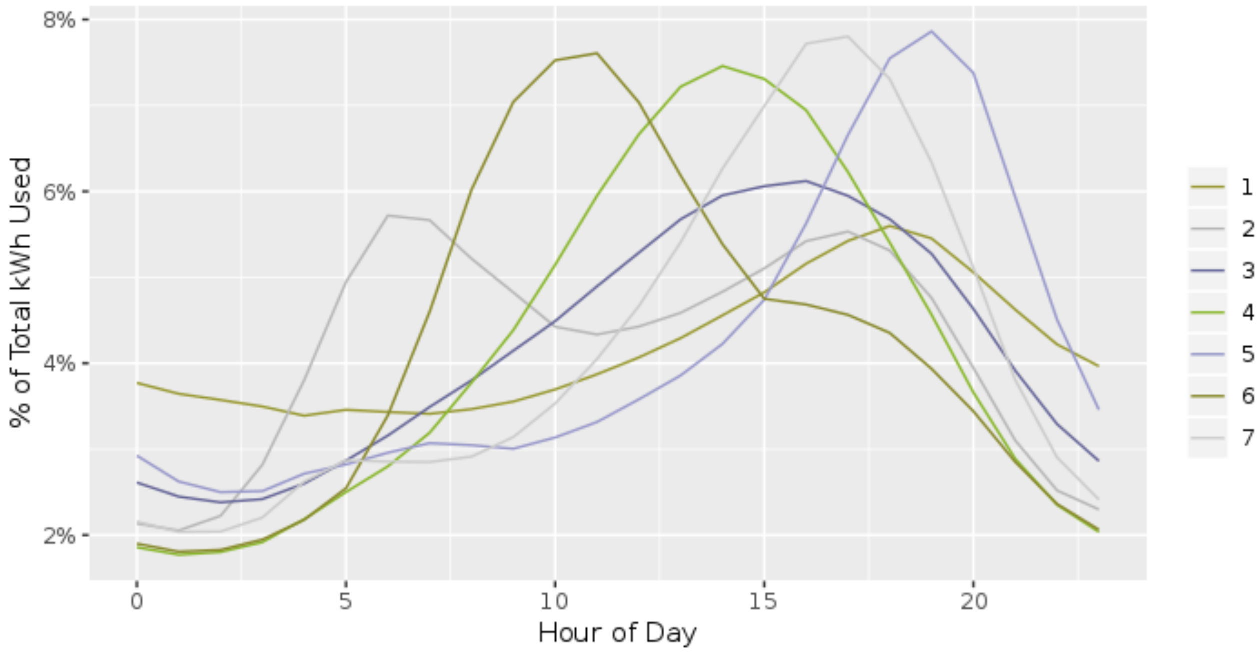
Figure 7: SCE Residential QI Participants by County and Climate Zone



Segmentation

For the SCE QI program, we defined customer segments with a combination of daily energy usage (magnitude) and normalized load shape (hours of use) in the pre-period. First, we assigned customers to one of seven bins by their average daily energy usage across the most recent pre-period year, with the highest energy usage bin containing the fewest customers. Next, we used *k*-means clustering to identify the seven unique clusters shown in Figure 8, each containing a subset of customers with similar load shapes during the pre-period. These load shapes exhibit the diversity in energy usage patterns we identified within the population of QI participants. Some of these participants' peak energy usage occurs as early as 6:00 a.m., while for others, peak energy usage occurs as late as 9:00 p.m.

Figure 8: SCE Residential QI Normalized Load Shape Bins



This segmentation approach defines 49 customer segments and 104 day bins, for a total of 5,243 distinct customer-day bins.[24]

Holdout Validation Tests

The results of one holdout test are shown in Figure 9, comparing the predicted pre-period load shape from the model (red line) to the actual pre-period load shape for the holdout sample (blue line). When the model is performing well, the two lines will overlap. The holdout test relies exclusively on pre-period data so that any differences between the predicted and actual energy usage can be attributed to model error, not to program savings. The model predictions track closely with the average actual load, slightly underestimating load during a few morning and evening hours. The AMICS model is able to reasonably predict the differences in the hourly load across all four seasons as shown in Figure 10, despite differences in weather conditions and schedules.

[24] The 49 customer segments are distinct combinations of 7 energy usage bins and 7 load shape clusters. The 104 day segments are comprised of 8 CDD bins, 8 HDD bins, 2 day types, and 4 seasons. Not all possible combinations of the customer and day segments were observed in the pre-period data.

Figure 9: SCE Residential QI Holdout Validation Test

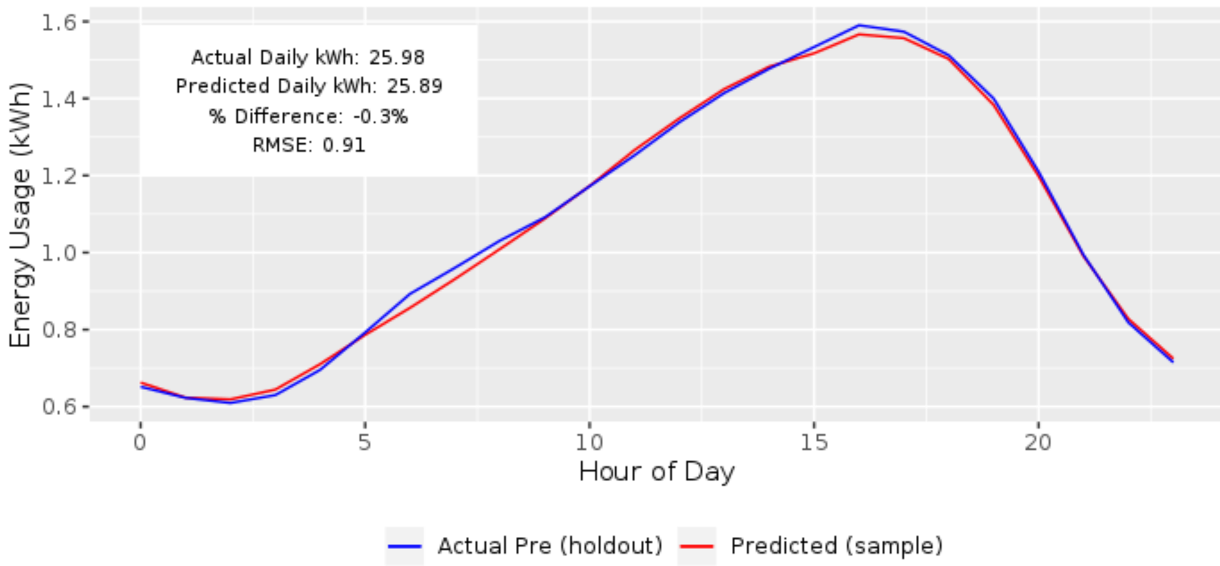
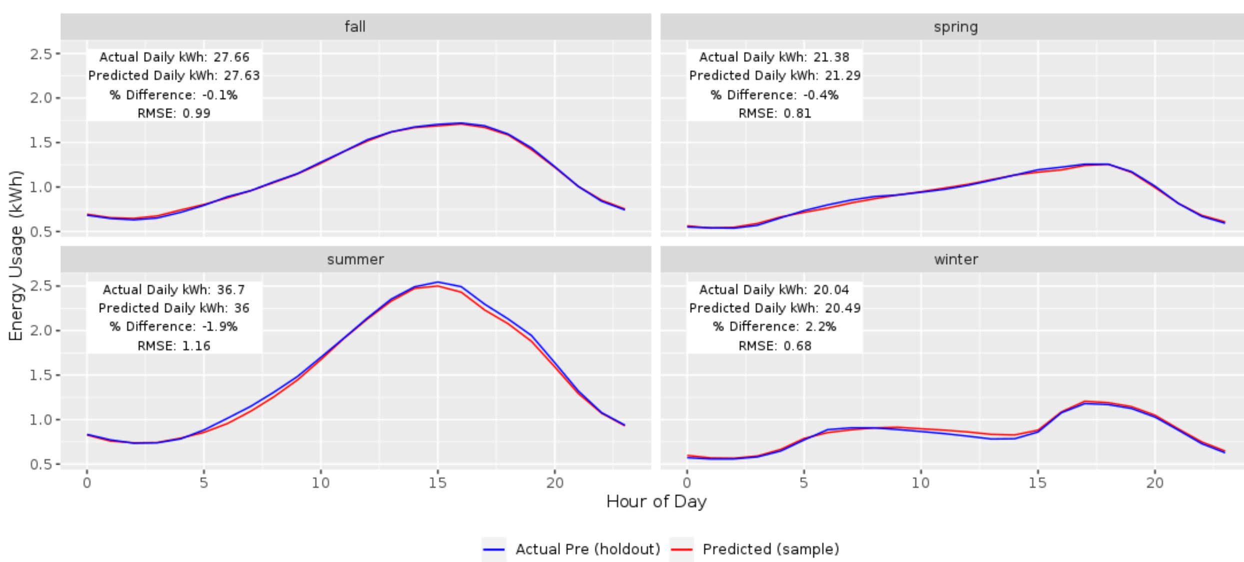


Figure 10: SCE Residential QI Holdout Validation Test by Season



Program Energy Savings

Figure 11 compares the post-period predicted load shape (red) with the actual post-period load shape (blue) across all residential QI participants in the database. This prediction is based on the pre-period energy consumption model and post-period weather data; it represents the expected load shape for these customers in absence of the program intervention. The error of each hourly prediction is depicted as a 95 percent confidence interval in the shaded area around each estimate. Whenever the actual post-period load shape (blue line) falls outside the predicted post-period load region (red area), this indicates that a statistically significant change was observed during that hour. In this case, the actual post-period load shape (blue) falls below the predicted load shape (red) during all hours, with substantial changes in the afternoon and early evening hours, when we would expect the majority of cooling load to occur. This is consistent with the seasonal load shapes in Figure 12, where we see a significant reduction in load during the summer and fall months, but only minor changes in the winter and spring.

Figure 11: SCE Residential QI Post-Period

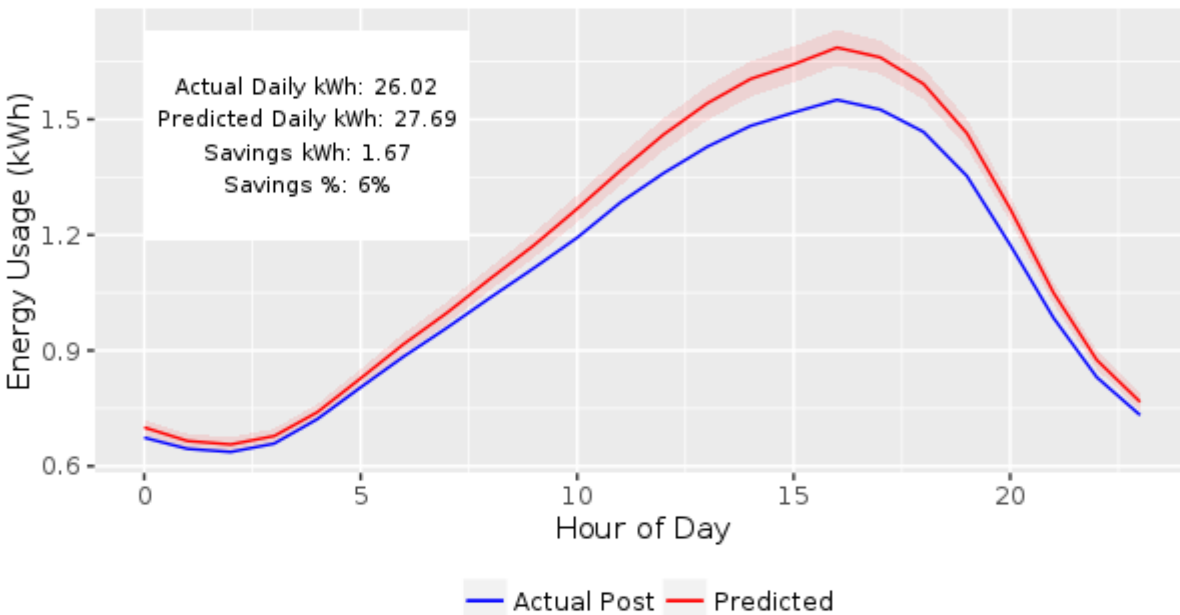


Figure 12: SCE Residential QI Post-Period by Season

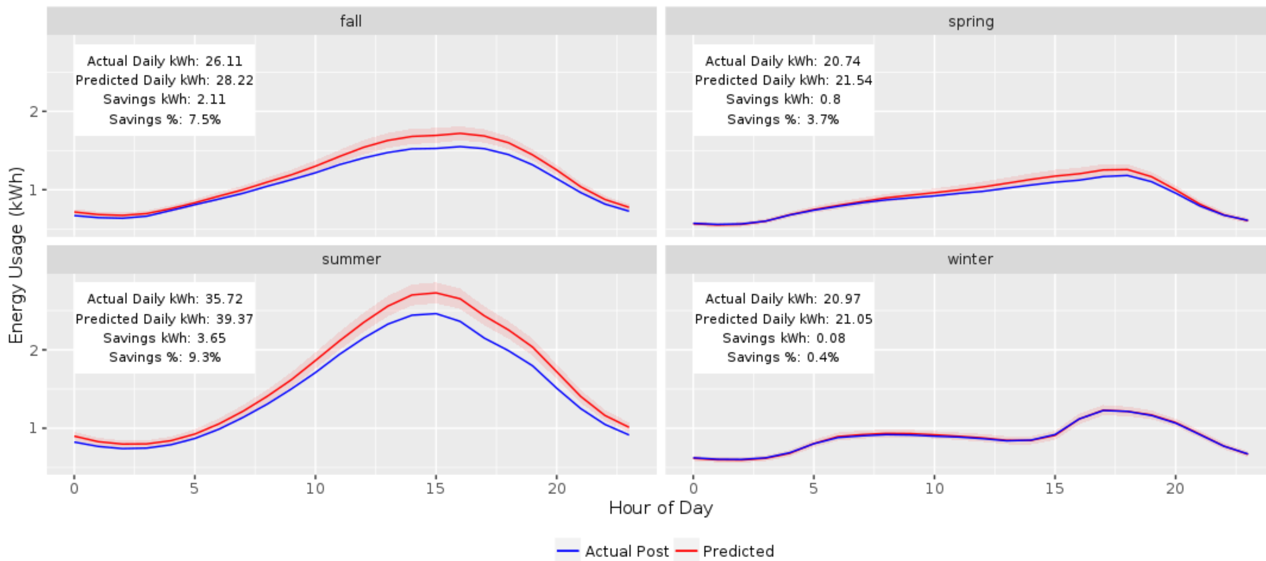


Figure 13 shows our estimated hourly kWh savings across the entire post-period, with error bars depicting 95 percent confidence intervals around each estimate. The SCE residential QI participants exhibited statistically significant energy savings during 23 of the 24 hours, with the exception being 4:00 a.m. Overall, we estimate that the SCE residential QI program resulted in energy savings of 1.67 kWh ± 0.77 per day (or 6.0% ± 2.8%).

Figure 13: SCE Residential QI Estimated Savings

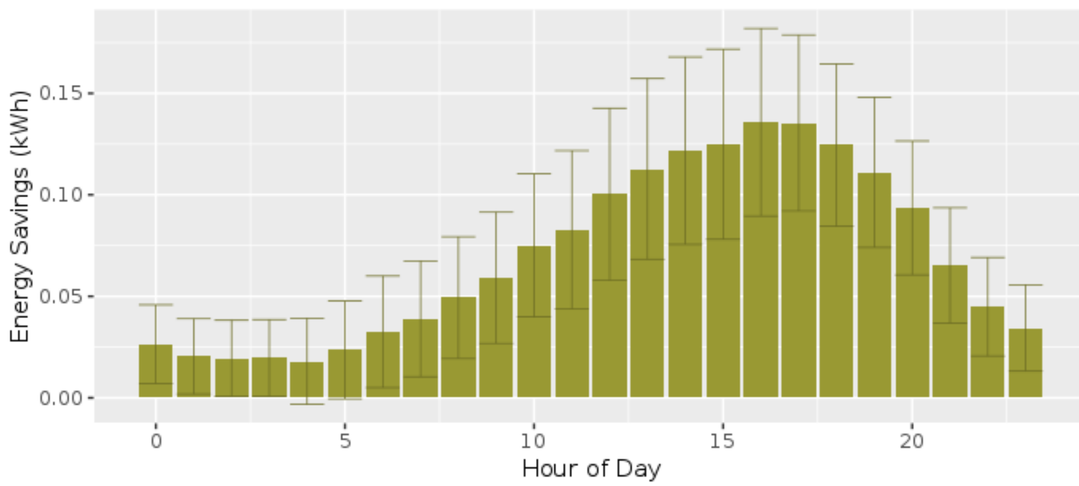
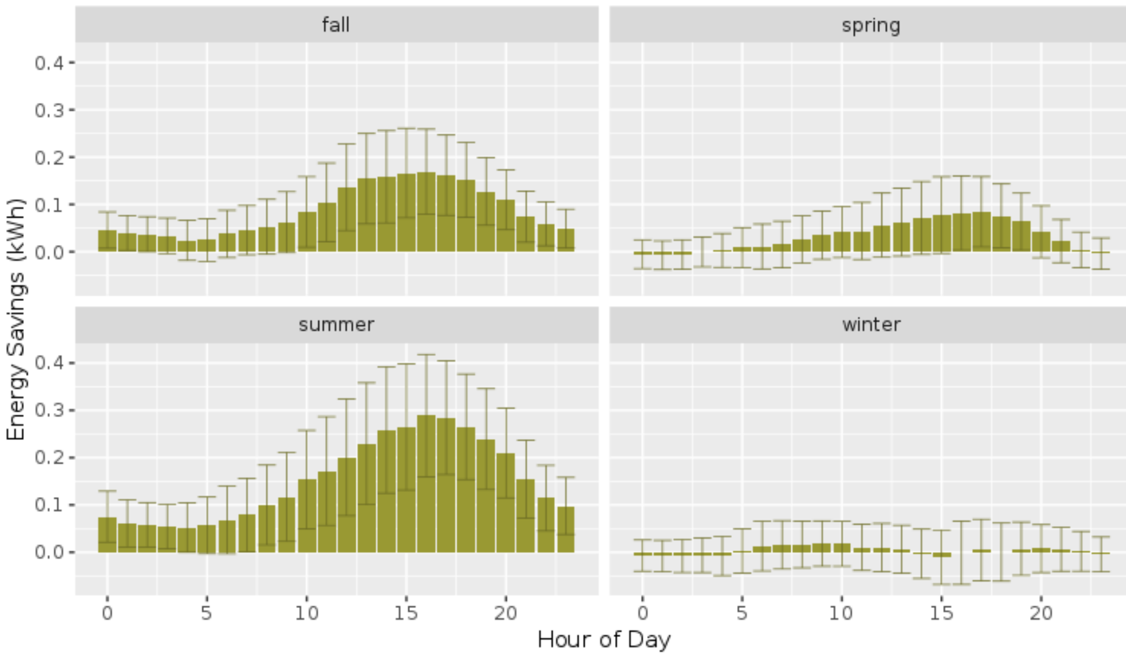


Figure 14 depicts the estimated savings in each of the four seasons. The AMICS model found positive and statistically significant energy savings in the summer and fall, with small and mostly insignificant savings in the winter and spring. These findings are consistent with our expectations since the vast majority of participants (97%) installed new central air conditioners, and only heat pumps (3%) would provide savings for both cooling and heating seasons.

Figure 14: SCE Residential QI Estimated Savings by Season



Savings by Segment

Figure 15 shows the average daily savings estimated by the AMICS model by pre-period energy usage and cooling load. The columns show the cooling load by CDD bin (hottest days on the right), and the rows show customers segmented by their average daily energy usage (lowest users on the bottom). We automatically color-coded the cells with the highest kWh savings in dark green and the lowest in dark orange (negative savings = increased usage); the yellow cells fall in the middle of this spectrum.

As this figure shows, customers with the lowest energy usage in the pre-period (rows 1 and 2) realized negative energy savings (i.e., increased energy usage) from the residential QI program across most observed levels of cooling load. One possible reason for this is that these low energy users did not use their air conditioning equipment before the program due to the high cost of operating inefficient cooling equipment. Hence, the residential QI program led them to increase space cooling and increase their overall energy usage on hot days. Fortunately, these increases were offset by substantial energy savings realized by moderate to high energy users (rows 4 through 7).

Figure 15: SCE Residential QI Energy Savings by Usage Bin and CDD Bin

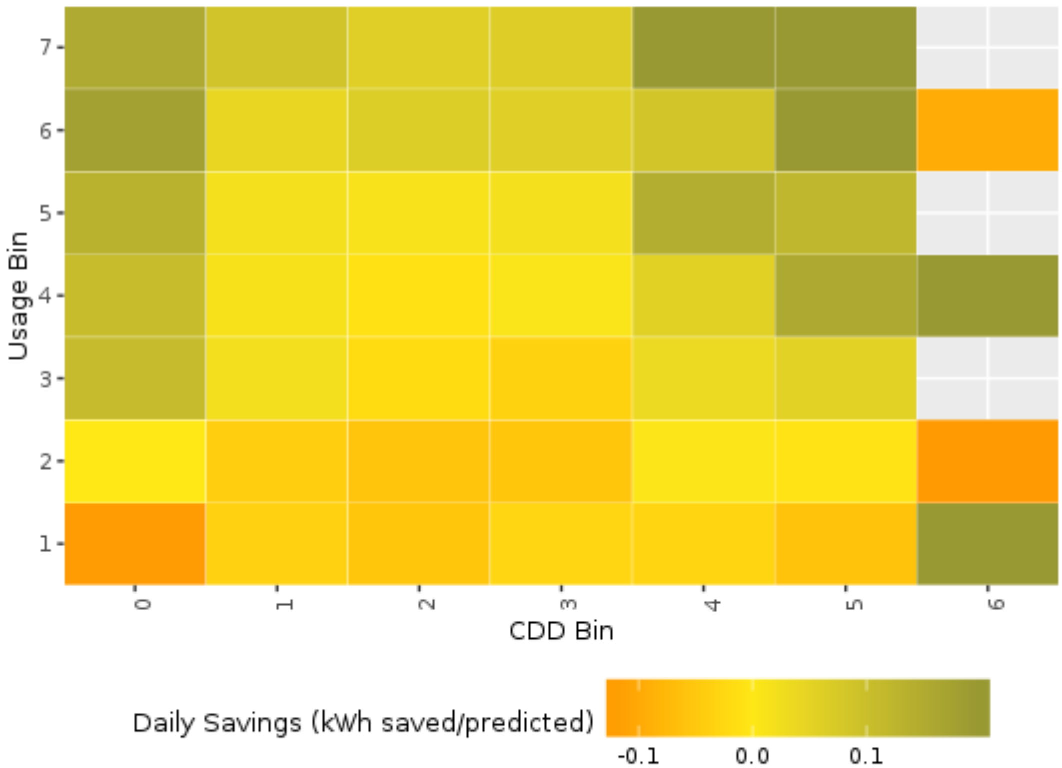


Figure 16 shows the average daily energy savings by pre-period load shape and cooling load. Corresponding to the load shapes shown in Figure 17, the customers with the flattest load shapes are at the bottom and the steepest are at the top. Independent of their daily energy usage in the baseline period, we see patterns in savings by the customers' load shape in the pre-period. Customers in load shape bin 02 had two peaks, one around 7:00 a.m. and another around 5:00 p.m.; this shape is common among homes with electric heating and low to moderate air conditioning usage. The heat table shows that customers in bin 02 realized positive energy savings (green) at all levels of cooling load, indicating that the program intervention reduced energy usage for days with cooling, heating, and ventilation needs. On the other end of the spectrum, customers in load shape bin 05 had negative savings (orange) during most weather conditions. This group's load shape reaches a peak at 7:00 p.m.; this is later than we would expect for a site with a high air conditioning load and suggests that HVAC is not the main driver of their peak usage. In general, we see that a customer's baseline load shape (and existing conditions) has an impact on the energy savings we can attribute to the program intervention.

Figure 16: SCE Residential QI Energy Savings by Load Shape and CDD Bin

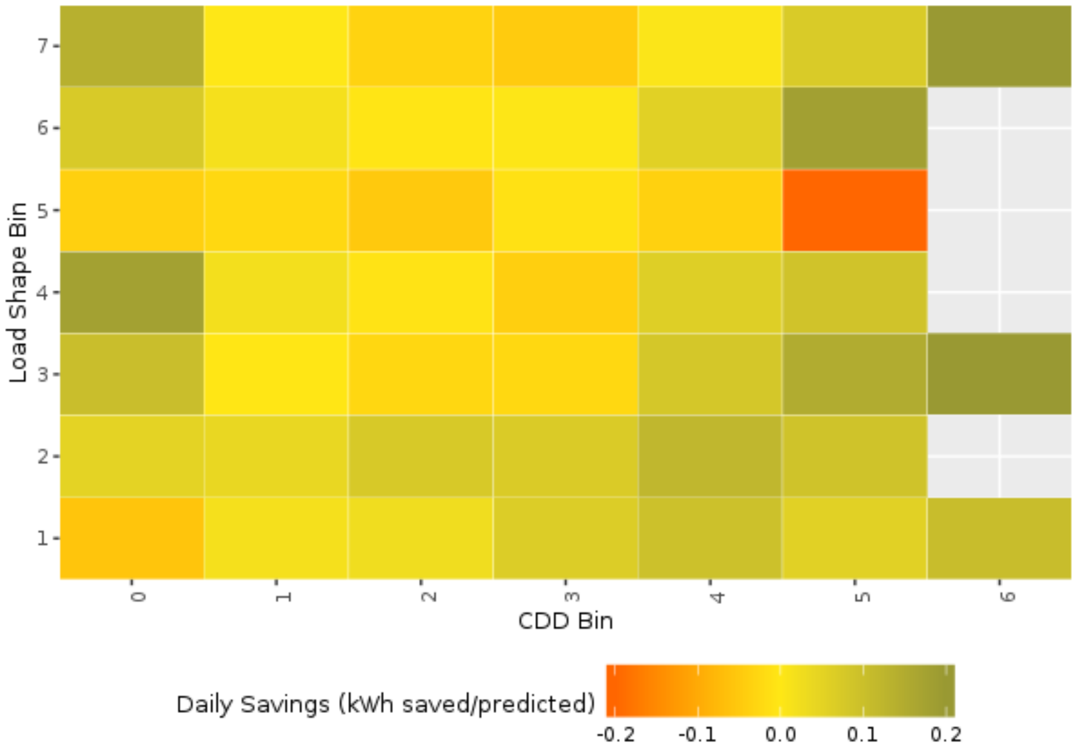
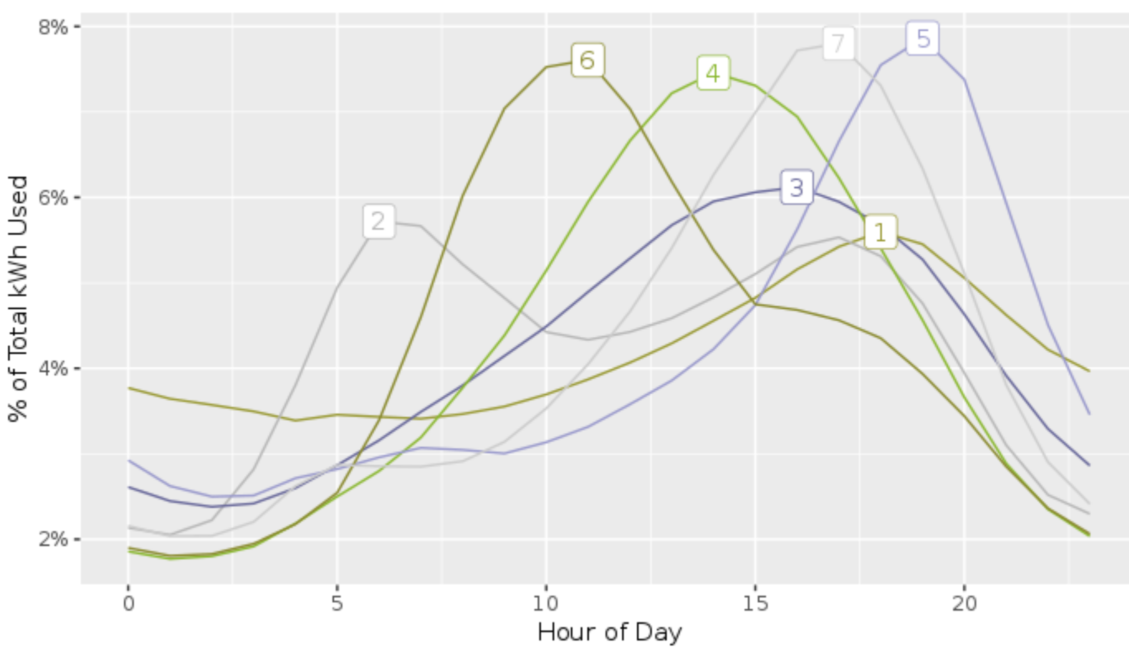


Figure 17: SCE Residential QI Load Shape Bins



Overall, the AMICS model estimated that the average energy savings attributable to the SCE residential QI program were 1.67 kWh ± 0.77 per day (or 6.0% ± 2.8%) with statistically significant savings in all but one hour of the day. The AMICS segmentation revealed some variation in energy savings across customer segments and weather conditions.

Key findings:

- The model holdout test performed very well, with the AMICS model predicting energy use for the holdout sample within 1 percent error. This confirms findings from the Phase I research showing that the AMICS model performs well with residential customers, with reasonable load shape predictions for this type of program.
- Energy impact estimates are consistent with expectations and are occurring during the expected times of day (late afternoon/early evening).
- The AMICS model tracks well with the changes in load shapes across seasons. The seasonal impact estimates also conform to expectations, with larger savings shown in the summer and fall.

3.1.2 PG&E Quality Maintenance

Program Description

The PG&E Quality Maintenance (QM) program is part of a California statewide program designed to achieve energy and demand savings through assessment and optimization of existing residential HVAC units.[25] Participants are enrolled in an ongoing maintenance agreement with a qualifying contractor (who has received additional training through the program) who performs two maintenance calls per year—once in the pre-cooling season and once in the pre-heating season. During each visit, the contractor conducts a full ACCA Standard 4 HVAC System Assessment and then performs any required maintenance. Examples of these maintenance activities include blower motor retrofits, enhanced time delay relay, airflow correction, and refrigerant charge adjustment. These activities should improve cooling delivery (from reduced runtime and/or power draw) and thereby improve efficiency.

Database

PG&E provided Evergreen with AMI billing data and account characteristics for the 31,705 distinct residential customers who participated in the QM program between January 2015 and July 2016. The PG&E QM program data included household and program

[25] Eligible homes must have a central forced air conditioner or heat pump and be a single-family home or duplex.

participation information such as HVAC technology type, maintenance activity description by date(s), *ex ante* gross energy and demand savings, premise type, and rate schedule.

The AMI billing data for this study contained approximately 650 million hourly observations from January 1, 2014 to July 31, 2017.[26] These data captured energy usage patterns of most households a full year before their participation in the program.

We applied filters to exclude customers with:

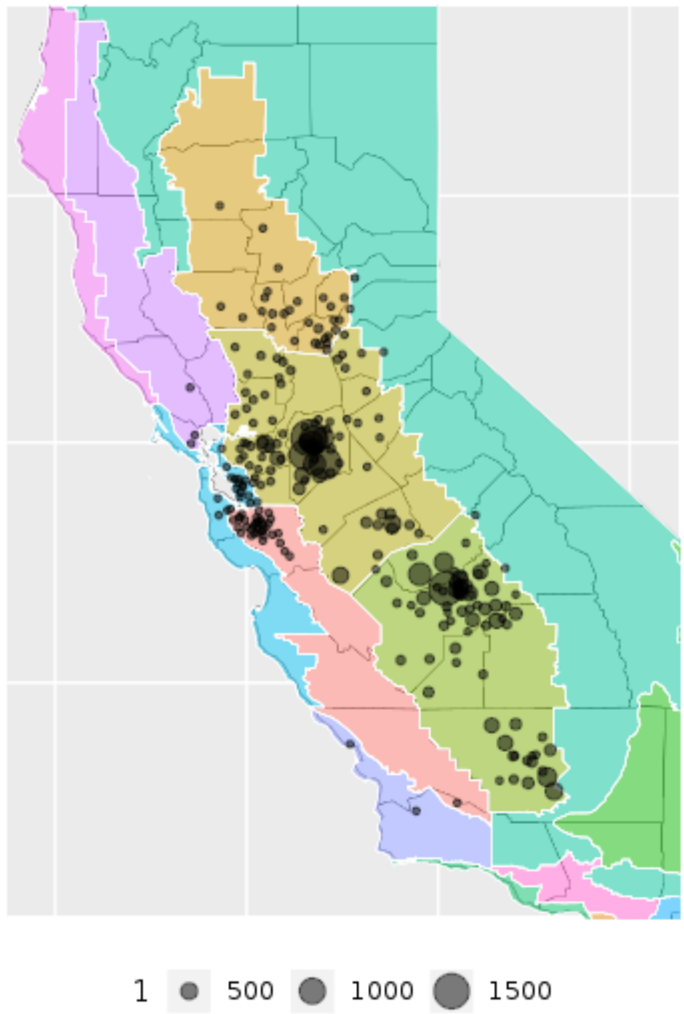
- No non-zero *ex ante* savings listed in the program documentation (n=2,053);[27]
- Net energy metering, such as onsite solar generation (n=2,342);
- No pre-period observations in the billing data (n=533); or
- Average energy usage in the pre- or post-period of less than 0.1 kWh or extreme changes from the pre- to post-periods of more than 150 percent or less than -66 percent (n=326).

These filters reduced the analysis dataset to 28,504 customers that participated in the QM program.

As shown in Figure 18, this sample was spread across many counties and CEC climate zones. The vast majority of participant dwellings (97%) were detached single-family homes. Nearly a quarter (18%) were enrolled in one or more demand response programs, including 4,166 customers with direct load control switches that allow PG&E to cycle their air conditioning equipment during summer peak events. Again, observations of energy usage on event days can be excluded to avoid double-counting peak load reductions (i.e., savings) that should be attributed to participation in a demand response program.

[26] For consistency across customers in the study, all 15-minute interval billing data were aggregated to the hourly level.
[27] Some QM program participants did not require any adjustments (i.e., tests revealed that their system did not need any maintenance); these participants were excluded from our post-period analysis because they would not have any energy savings attributable to the program.

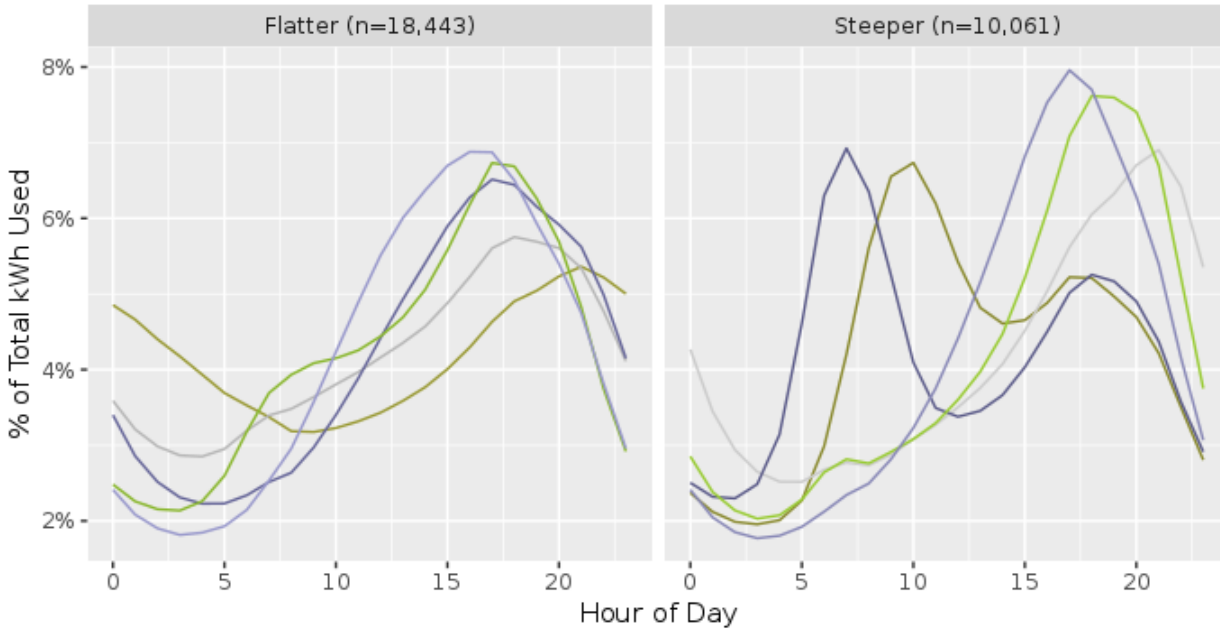
Figure 18: PG&E Residential QM Participants by County and Climate Zone



Segmentation

For the PG&E QM program, we defined customer segments with a combination of daily energy usage (magnitude) and normalized load shape (hours of use) in the pre-period. First, we assigned customers to one of 10 bins by their average daily energy usage across the most recent pre-period year, with the highest energy usage bin containing the fewest customers. Next, we used *k*-means clustering to identify the 10 unique clusters shown in Figure 19, each containing a subset of customers with similar load shapes during the pre-period. These load shapes exhibit the diversity in energy usage patterns we identified within the population of residential QM participants. Some customers have relatively flat load shapes with little change in energy usage throughout the day, while others ramp up to a steep peak in the morning or evening hours.

Figure 19: PG&E Residential QM Normalized Load Shape Bins



This segmentation approach defines 100 customer segments and 117 day bins for a total of 11,700 distinct customer-day bins.[28]

Holdout Validation Tests

The results of one holdout test are shown in Figure 20, comparing the predicted pre-period load shape from the model (red line) to the actual pre-period load shape for the holdout sample (blue line). The holdout test relies exclusively on pre-period data so that any differences between the predicted and actual energy usage can be attributed to model error, not to program savings. The model predictions track closely with the average actual load, with the two lines nearly indistinguishable. The AMICS model is able to accurately predict the differences in the hourly load across all four seasons for the PG&E residential QM program population, shown in Figure 21.

[28] The 100 customer segments are distinct combinations of 10 energy usage bins and 10 load shape clusters. The 117 day bins are comprised of 8 CDD bins, 10 HDD bins, 2 day types, and 4 seasons. Not all possible combinations of the customer and day segments were observed in the pre-period data.

Figure 20: PG&E Residential QM Holdout Test

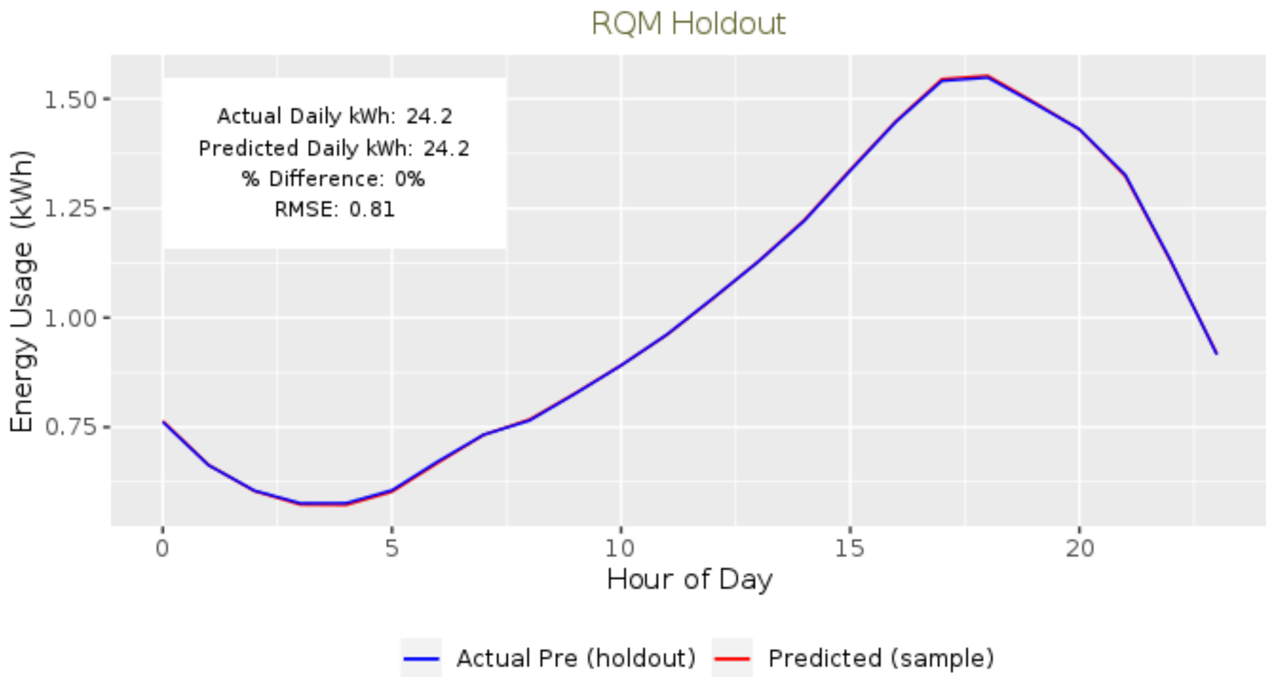
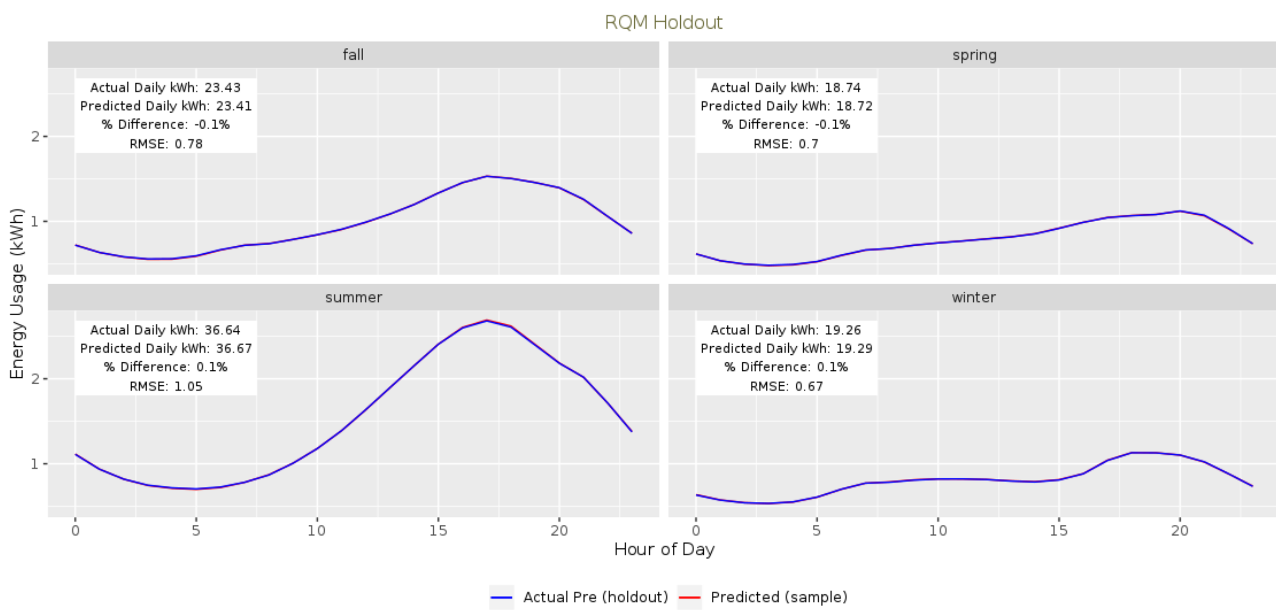


Figure 21: PG&E Residential QM Holdout Test, by Season



Program Energy Savings

Figure 22 and Figure 23 compare the post-period predicted load shape (red) with the actual post-period load shape (blue) across all residential QM program participants in the database. This prediction is based on the pre-period energy consumption model and post-period weather data; it represents the expected load shape for these customers in absence of the program intervention. The error of each hourly prediction is depicted as a 95 percent confidence interval in the shaded area around each estimate. Whenever the actual post-period load shape (blue line) falls outside the predicted post-period load region (red area), this indicates that a statistically significant change was observed during that hour. In this case, the actual post-period load shape (blue) falls just below the predicted load shape (red) during most hours, but the difference is minor.

Figure 22: PG&E Residential QM Actual versus Predicted Post-Period

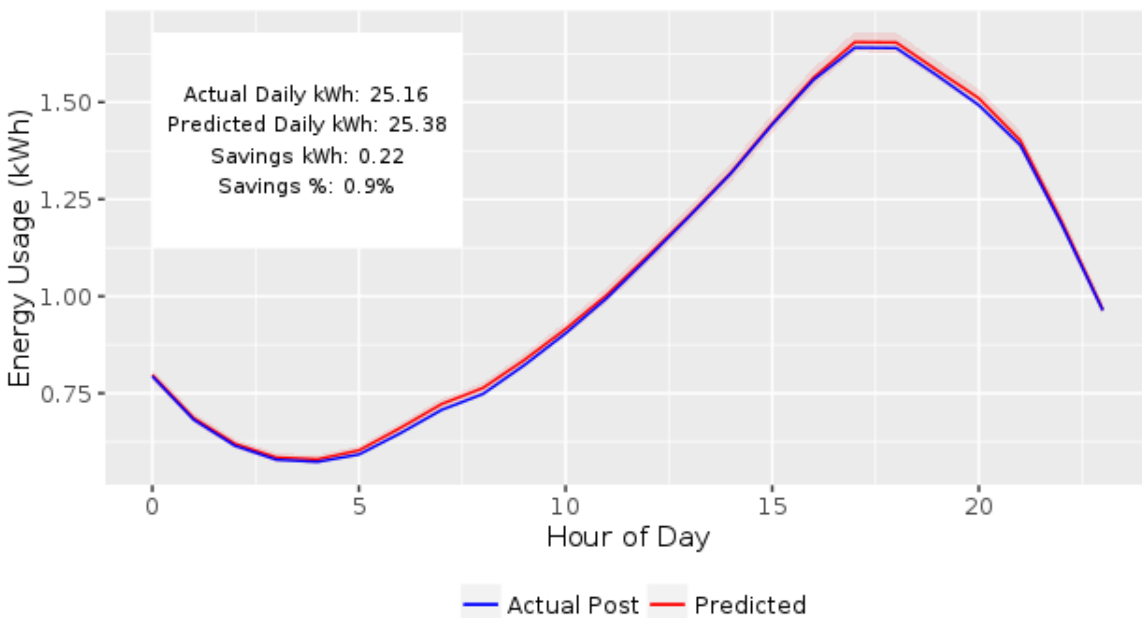


Figure 23: PG&E Residential QM Actual versus Predicted Post-Period, by Season

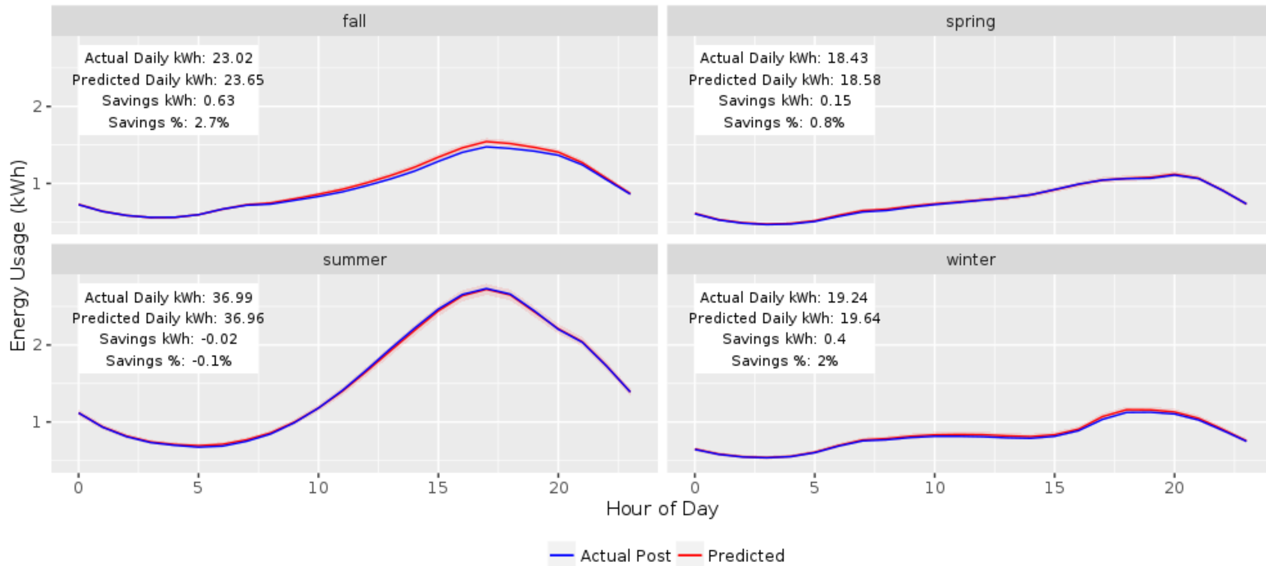


Figure 24 shows our estimated hourly kWh savings across the entire post-period, with error bars depicting 95 percent confidence intervals around each estimate. The PG&E residential QM program participants exhibited energy savings during all 24 hours, but these reductions in energy usage were only statistically significant during the morning hours of 6:00 a.m. to 8:00 a.m. Overall, we estimate that the PG&E QM program resulted in energy savings of 0.22 kWh ± 0.45 per day (or 0.9% ± 1.8%).

Figure 24: PG&E Residential QM Estimated Savings

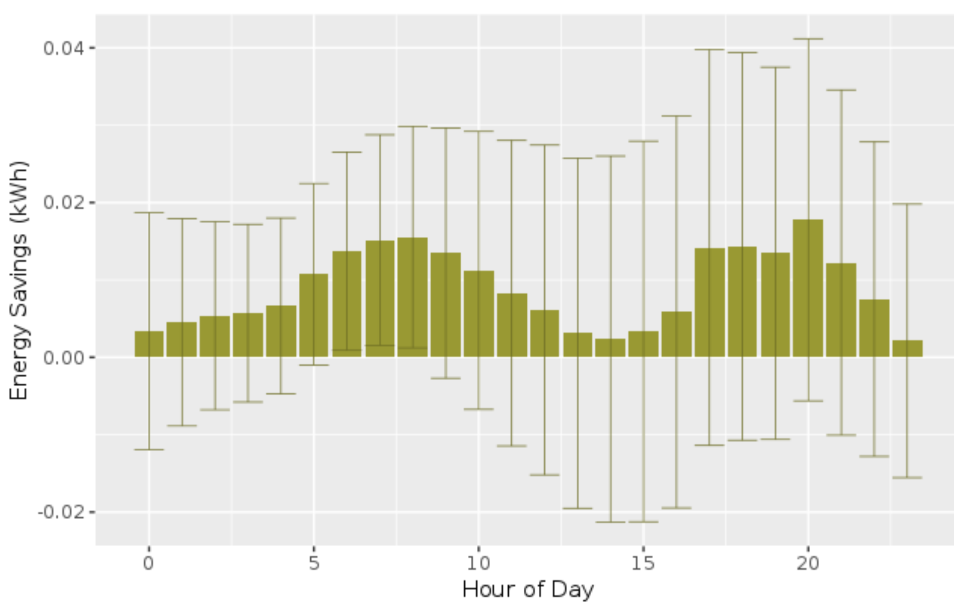
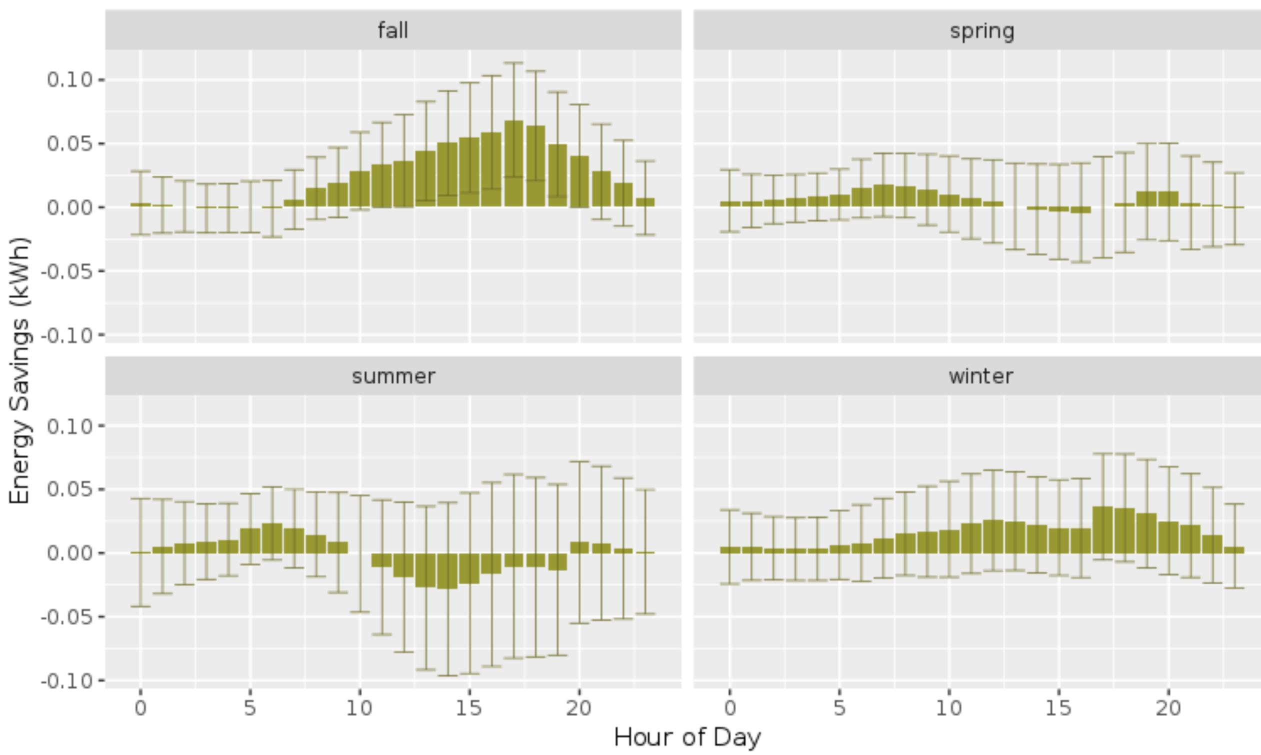


Figure 25 depicts the estimated savings in each of the four seasons. The AMICS model revealed an increase in energy usage (negative savings) during the peak hours of the summer. It is possible that the residential QM program's HVAC maintenance activities led participants to increase cooling during the summer to improve comfort (e.g., customers did not bother to use their cooling equipment as often when it was not functioning properly, leading to an increase in usage after the program intervention). Positive energy savings were detected in all other months, with the majority of savings occurring during peak hours of the fall.

Figure 25: PG&E Residential QM Estimated Savings by Season

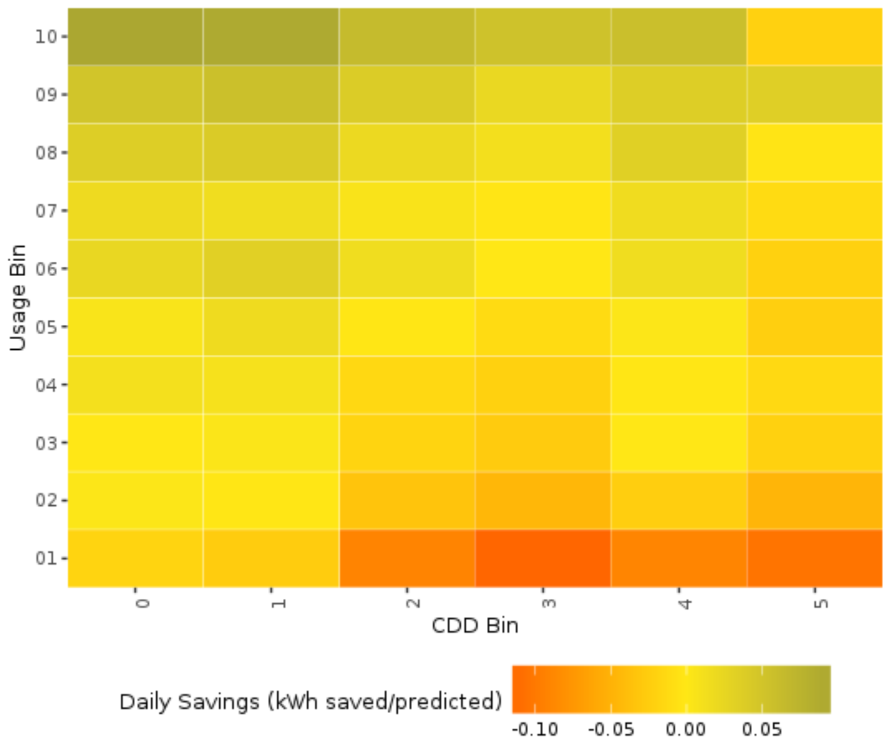


Savings by Segment

Figure 26 shows the average daily savings estimated by the AMICS model by pre-period energy usage and cooling load. The columns show the cooling load by CDD (hottest days on the right), and the rows show customers segmented by their average daily energy usage (lowest users on the bottom). We automatically color-coded the cells with the highest kWh savings in dark green and the lowest in dark orange (negative savings = increased usage); the yellow cells fall in the middle of this spectrum.

As this figure shows, customers with higher energy usage in the pre-period realized higher energy savings from the residential QM program across all observed levels of cooling load. Across all customer segments, participants realized greater energy savings on days with low to moderate cooling load (left), and most had negative savings on the hottest days (far right). This is consistent with the post-period seasonal load shapes, which indicated that the average participants increased their energy usage in the summer months and realized savings during the cooler months.

Figure 26: PG&E Residential QM Energy Savings by Usage Bin and CDD



Overall, the AMICS model estimated that the average energy savings attributable to the PG&E residential QM program were 0.22 kWh ± 0.45 per day (or 0.9% ± 1.8%), which is not statistically significant. The average hourly impact results are also statistically insignificant during most hours of the day. However, the AMICS segmentation revealed a wide variation in energy savings across customer segments and weather conditions, with more substantial energy savings being realized by high energy users on days with low to moderate cooling loads.

Key findings:

The small average savings provides an example of how more traditional billing regressions obscure impacts by aggregating everything to a single customer average using monthly billing data. With the AMICS model, even when the overall average program savings is not statistically significant, the detailed customer bins allow us to identify subsets of customers that are achieving significant savings. This information could be used by the IOUs to target similar customers for future program participation.

3.2 Home Energy Reports

There are three key benefits of a segmented modeling approach for estimating energy savings attributable to a Home Energy Reports program:

1. **Predictive power.** Creating separate predictions for each customer segment limits the variation across customers for which each model must account. Our previous research has demonstrated that segmentation is a useful tool to reduce error in load shape predictions, improving the predictive power of our models.
2. **Segmentation accomplishes the same goal as matching.** In some cases, randomized group assignment is not sufficient to produce balanced samples with similar energy usage patterns in the baseline period. Segmentation in the baseline period identifies and groups customers with similar load shapes, seasonality, and climate prior to any change in the program treatment. Performing difference-of-differences calculations within each customer segment improves the validity of our comparisons, focusing on the impact of any difference in the program treatment.
3. **Ease of distributional impact analysis.** The AMICS modeling approach creates separate model predictions and estimated post-period changes (i.e., energy savings) for each customer segment simultaneously. We do not simply provide the average treatment effect; instead, we expose the variation in treatment effects across program participants associated with key differences in the characteristics and energy usage patterns of these customers in the baseline period.

Program Description

The PG&E Home Energy Reports (HERs) program is part of a California statewide program designed to achieve energy and demand savings through customized reports with peer comparisons (energy usage by home type, size, and heating source) and customized tips for saving energy. The HERs program produces relatively small energy savings, but the opt-out randomized control trial (RCT) design makes it feasible to estimate net savings at the program level.

The program is designed as an RCT, utilizing a control group to estimate any natural change in energy consumption that occurs from the pre- to post-intervention period. The program impacts are estimated as the change in the treatment group, above and beyond any naturally occurring change exhibited by the control group. Note that for the HERs program, the existing conditions in the pre-implementation period are the appropriate baseline and therefore, no additional adjustments are needed to the baseline to calculate program impacts.

SCE's Gamma Wave customers have been receiving HERs (i.e., "being treated") since November of 2011.[29] The Gamma Wave includes residential customers from all energy usage quartiles and fuel types (i.e., electric only and duel-fuel). They received three initial monthly reports, followed by bi-monthly (standard frequency) or quarterly (reduced frequency) reports.

We selected the Gamma Wave for Phase II because it provides a large sample of residential customers across PG&E's entire service territory, stratified by usage quartile. Hence, all observations of the control group and the treatment group pre-period provide a diverse sample that is expected to be representative of the broader population of residential customers. In addition to testing whether AMICS can detect small energy savings, the HERs program provides an opportunity to study the benefits of segmentation with a large and diverse sample of households.

Database

PG&E provided Evergreen with account characteristics for the 152,292 distinct residential customers who were enrolled in the HERs Gamma Wave. The PG&E HERs program data included only the HERs wave assignment (i.e., gamma treatment vs. gamma control) and premise baseline territory (R, S, T, W, X).[30]

The AMI whole-home billing data for this study contained nearly 3.8 billion hourly observations from November 1, 2010 to October 31, 2013.[31] The treatment group received its first report on November 1, 2011; hence, these data captured energy usage patterns of most households for a full year of the pre-period and for two full years of the post-period.

We applied filters to exclude customers with:

- No pre-period observations in the billing data (n=4,222);
- Extreme changes from the pre- to post-periods of more than 150 percent or less than -66 percent (n=4,740); or
- Average energy usage in the pre- or post-period of less than 0.1 kWh (n=662).

---

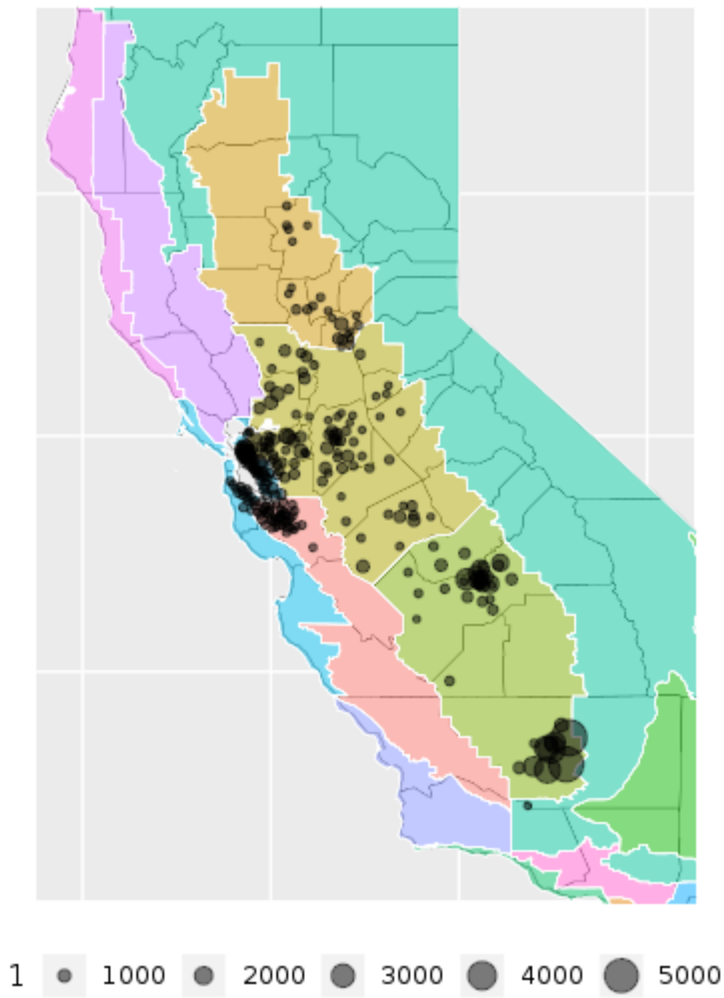
[29] The Gamma Wave refers to the third round of program implementation. It follows the original Alpha Wave pilot and the Beta Wave launched in August 2011 to the top quartile of energy users in the San Francisco Bay area. Each 'wave' is comprised of residential customers assigned by PG&E to a treatment or control group.

[30] PG&E baseline territories: https://www.pge.com/nots/rates/PGECZ_90Rev.pdf

[31] For consistency across customers in the study, all 15-minute interval billing data were aggregated to the hourly level.

As a result, the AMI analysis was limited to 142,668 customers, including 71,344 report recipients and 71,324 controls. As shown in Figure 27, these customers were spread across many counties and CEC climate zones.

Figure 27: PG&E HERs Participants by County and Climate Zone



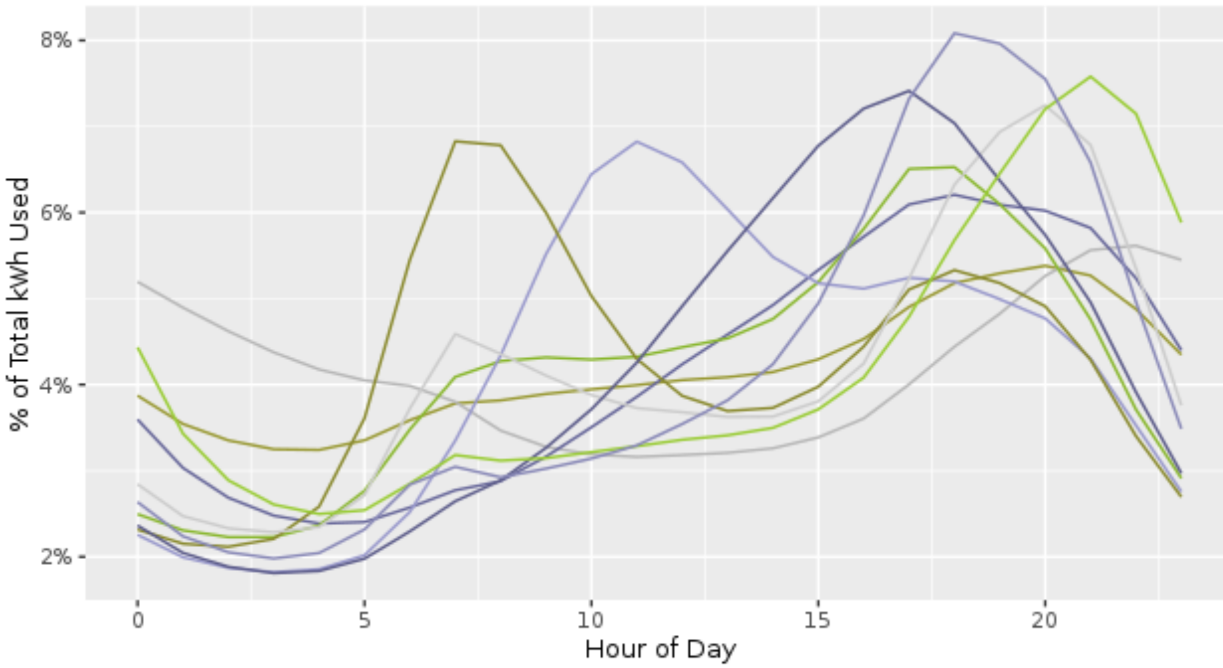
To reduce processing burden, we selected a random sample of 75,500 customers from PG&E's Gamma Wave, stratified by premise baseline (defined by PG&E) and group to ensure the sample had a representative assortment of report recipients and controls.

Segmentation

For the HERs program, we defined customer segments with a combination of daily energy usage (magnitude) and normalized load shape (hours of use) in the pre-period. First, we assigned customers to one of 20 bins by their average daily energy usage across the most recent pre-period year, with the highest energy usage bin containing the fewest customers.

Next, we used *k*-means clustering to identify the 10 unique clusters shown in Figure 28, each containing a subset of customers with similar load shapes during the pre-period.

Figure 28: PG&E HERs Normalized Load Shape Bins



This segmentation approach defines 122 customer segments and 107 day bins, for a total of 13,054 distinct customer-day bins.[32]

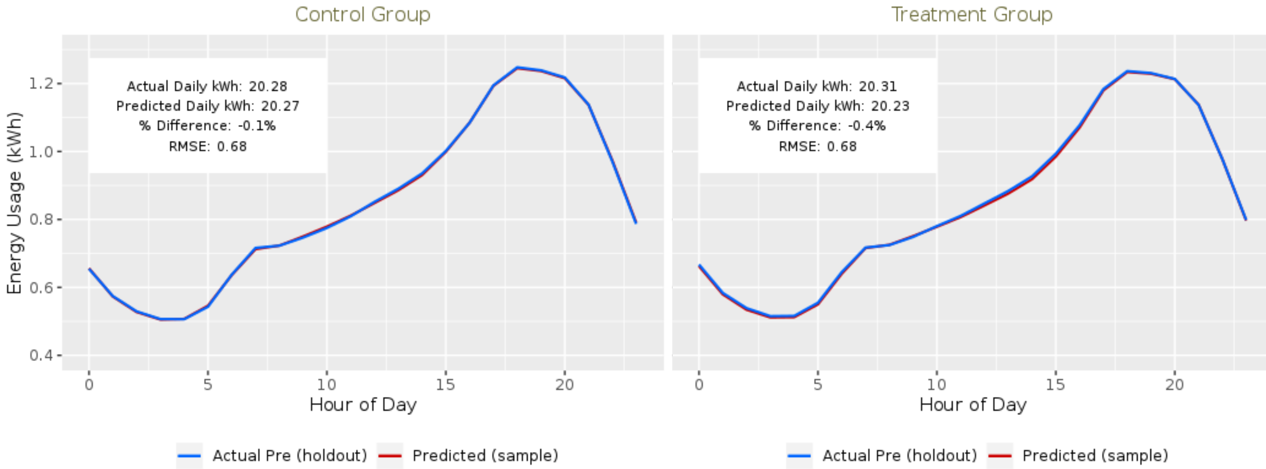
Holdout Validation Tests

The results of one holdout test are shown in Figure 29, comparing the predicted pre-period load shape from the model (red line) to the actual pre-period load shape for the holdout sample (blue line). When the model is performing well, the two lines will overlap. The holdout test relies exclusively on pre-period data so that any differences between the predicted and actual energy usage can be attributed to model error, not to program savings. The model predictions track closely with the average actual load, with a slight underestimation of energy usage in the treatment group during the afternoon hours and nearly perfect predictions for the control group.

---

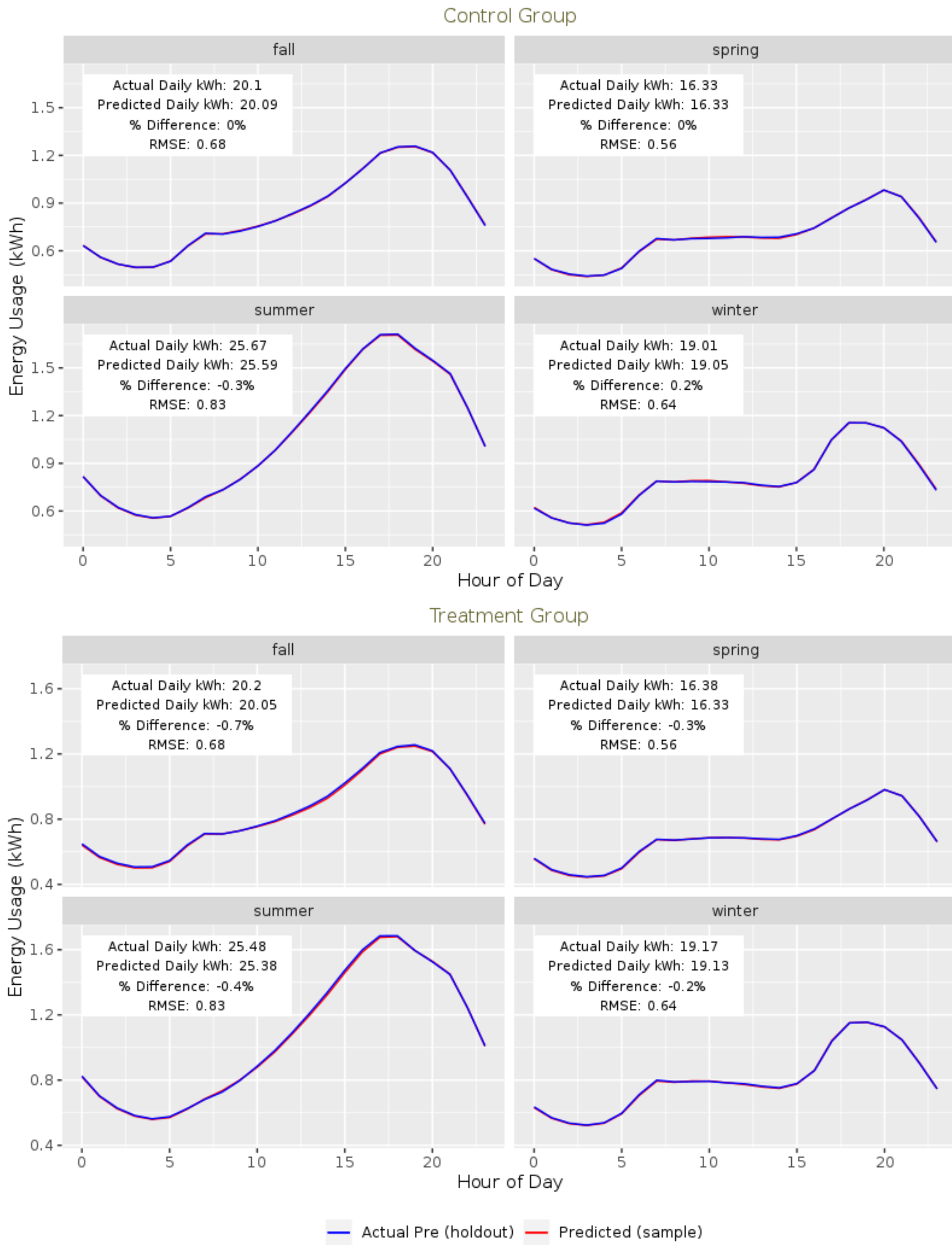
[32] The 122 customer segments are distinct combinations of 20 energy usage bins and 10 load shape clusters. The 107 day bins are comprised of 8 CDD bins, 9 HDD bins, 2 day types, and 4 seasons. Not all possible combinations of the customer and day segments were observed in the pre-period data.

Figure 29: PG&E HERs Holdout Test



As shown in Figure 30, the AMICS model is able to accurately predict the hourly load across all four seasons and two groups, within 1 percent of the actual load.

Figure 30: PG&E HERs Holdout Test by Season and Group



Program Energy Savings

This section provides our estimates for the energy savings experienced by customers that can be attributed to the HERs program, based on the AMICS model and a stratified random sample of recipient and control customers (n=75,500).

Figure 31 compares the post-period predicted load shape (red) with the actual post-period load shape (blue) across all customers in the HERs Gamma Wave sample. This prediction is based on the pre-period consumption model and post-period weather data; it represents the expected load shape for these customers in absence of the HERs program. The error of each hourly prediction is depicted as a 95 percent confidence interval in the shaded area around each estimate. The timeline depicted in these charts starts after the first Home Energy Report was mailed to the treatment group in November 2011 and spans two years of the post-period, allowing time for the reports to influence measure adoption and changes in behavior. Whenever the actual post-period load shape (blue line) falls outside the predicted post-period load region (red area), this indicates that a statistically significant change was observed during that hour.

The AMICS model detected reductions in the whole-building energy usage of the report recipients (i.e., treatment group), for a total reduction of 0.3 kWh per day, or 1.5 percent. Not all of the customers who received the HERs will have taken action to reduce their energy usage, but they all received the program mailings. The AMICS model also detected some reductions in energy usage by the control group of 0.18 kWh or 0.9 percent.

Figure 31: HERs Model Predictions vs. Actual Load in Post-Period

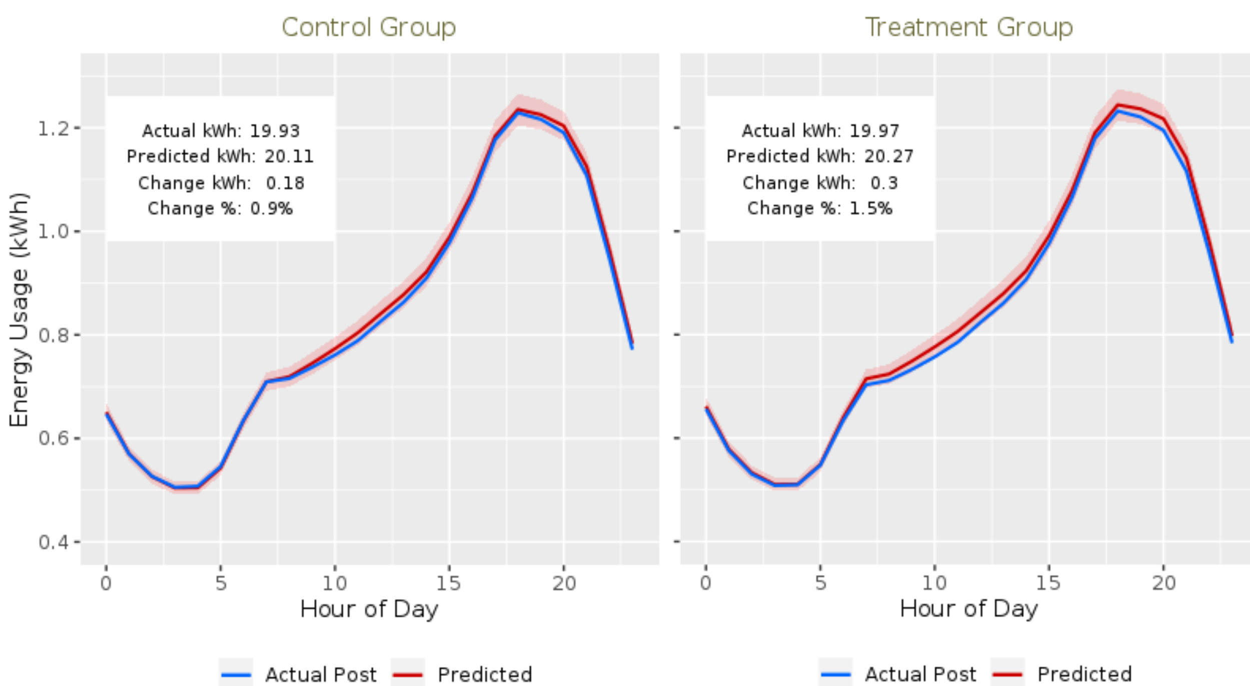


Figure 32 shows the estimated hourly change in kWh by group for a difference-of-differences estimation, with error bars depicting 95 percent confidence intervals around each estimate. The difference-in-differences between the control and treatment groups are performed within each customer-day segment and then weighted by the number of observations in the treatment group during the post-period. This helps to control for any known differences in the composition of customers and weather conditions in the control and treatment groups.[33]

The treatment group (in green) exhibited reductions in energy usage across all hours of the day. The control group (in orange) exhibited a similar pattern with the exception of slight increases in their energy usage from 3:00 a.m. to 5:00 a.m. During all 24 hours, the treatment group experienced greater reductions in energy usage than the control group, though the differences were not statistically significant. The treatment group experienced an overall reduction in energy usage of 0.30 kWh per day, while the control group decreased their usage by 0.18 kWh per day. As the control group provides our best estimate of the natural change that the treatment group would have experienced without the program, we estimate the total program impact to be a reduction of 0.12 kWh ± 0.21 per day (0.6% ± 1.0), as shown in Figure 33. The AMICS model does detect energy savings that we attribute to the HERs program treatment; however, these savings are not statistically significant at the program level.

[33] This comparison is restricted to customer and day segments that were observed in the post-period with both treatment and control households. This restriction is minor, as less than 1 percent of treatment customer days in the post-period had no similar customers or days in the control group during the post-period; unusual customers and days were automatically removed from the comparison. The post-period contains 4.23 million observations in 4,922 distinct customer-day segments, whereas this difference-in-differences comparison is based on 4.22 million observations from 4,023 customer-day segments (99.9% and 81.7%, respectively).

Figure 32: HERs Comparison of Post-Period Changes

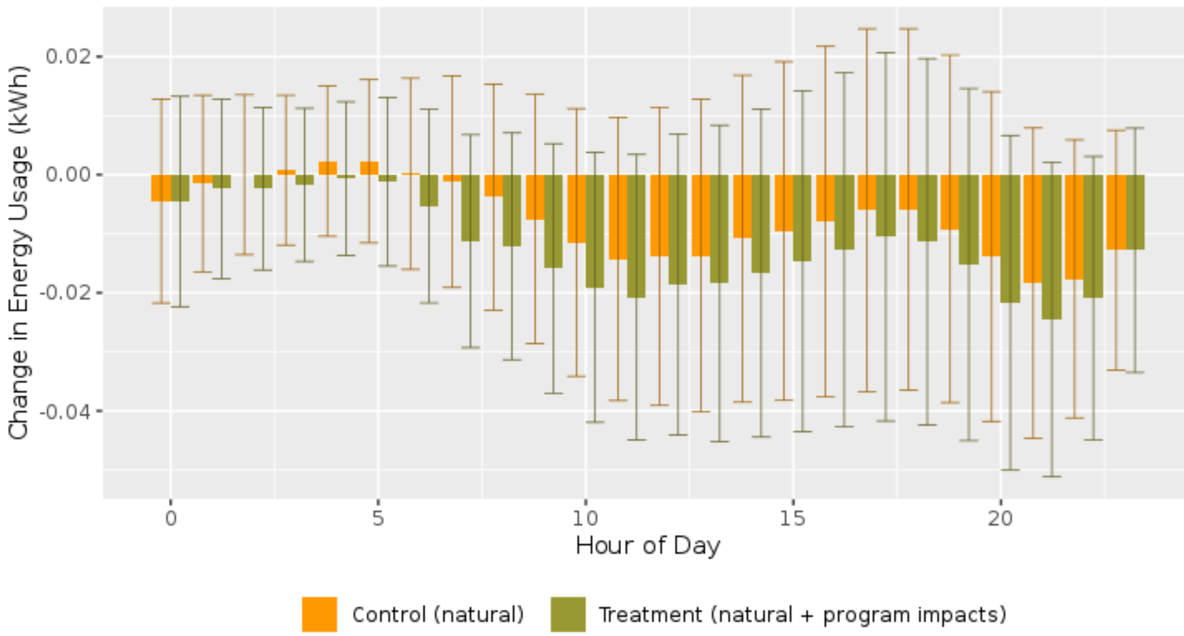


Figure 33: HERs Estimated Energy Savings

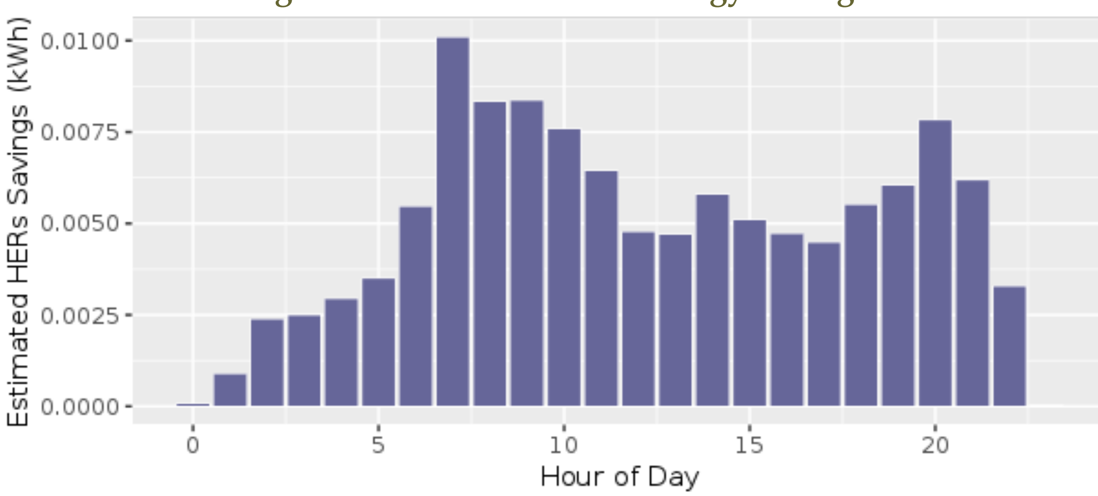


Figure 34 shows the estimated hourly change in kWh by group for a difference-of-differences estimation by season. The treatment group (in green) and control group (in orange) exhibited an increase in energy usage in the spring during the afternoon and evening hours. The treatment group experienced an overall increase in energy usage of 0.15 kWh per day in the spring months of the post-period, while the control group increased their usage by 0.26 kWh per day. As the control group provides our best estimate for how the treatment group would have continued to use energy in absence of the HERs program intervention, we estimate the HERs program led to energy savings of 0.11 kWh in the spring, as shown in Figure 35.

Figure 34: HERs Comparison of Post-Period Changes, by Season

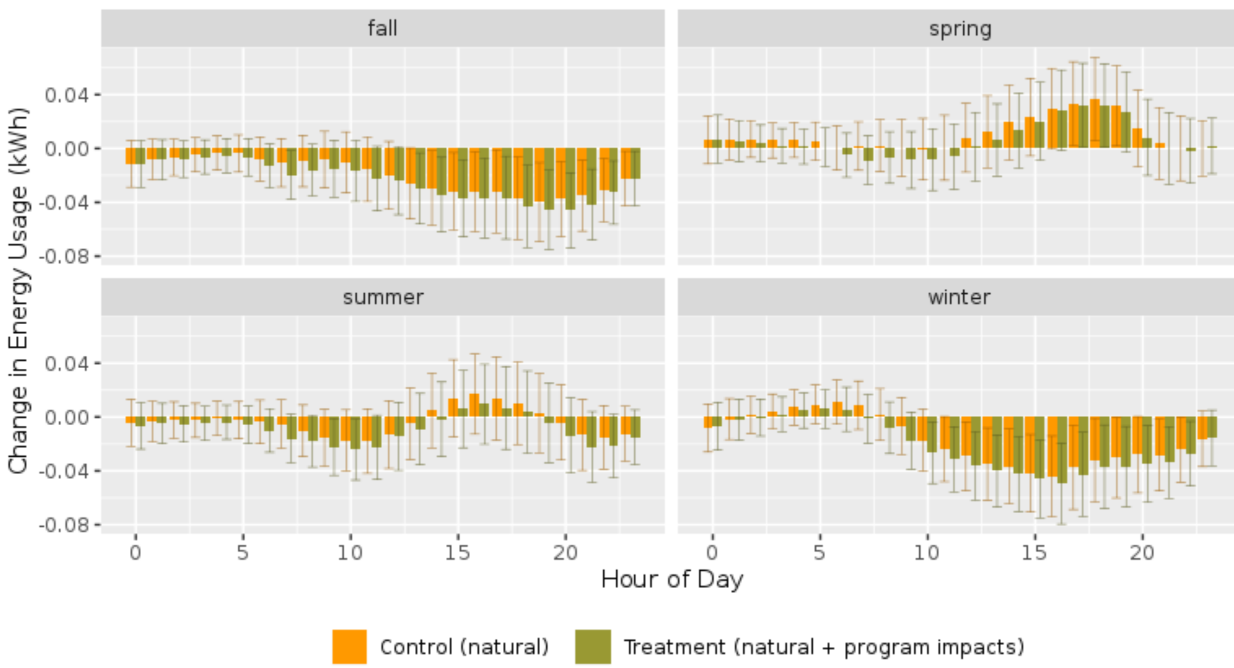
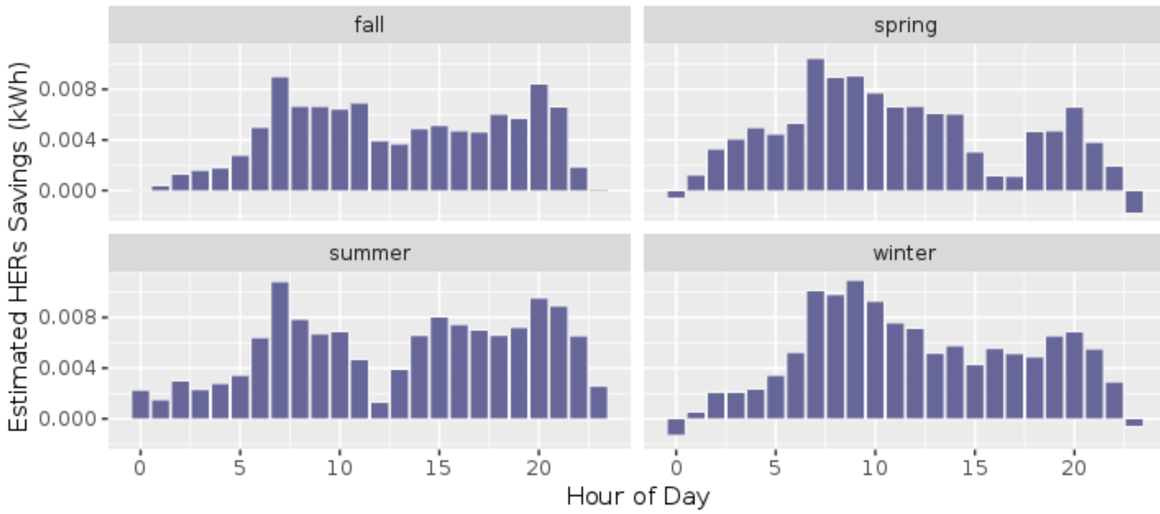


Figure 35: HERs Estimated Energy Savings, by Season



Savings by Segment

The segmentation in the AMICS model provides a unique opportunity to see the variation in program savings by customer segment and weather.

Figure 36 shows the average daily savings estimated by the AMICS model by season and cooling load. The columns show the four seasons, and the rows show customers segmented by their pre-period load shape. Corresponding to the load shapes shown in Figure 37, the customers with the flattest load shapes are at the bottom and the steepest are at the top. We automatically color-coded the cells with the highest kWh savings in dark green and the lowest in dark orange (negative savings = increased usage); the yellow cells fall in the middle of this spectrum.

As this figure shows, the HERs program impacts differ across the customer segments, with each segment exhibiting a different pattern of energy savings across the seasons. For example, the customers with the flattest pre-period load shape (load shape bin 01) had little to no savings in the summer or fall, but saved over 10 percent in the winter months. It is likely that these customers did not use much energy for cooling, and thus had less potential for savings in the summer months despite receiving the reports, whereas customers with the steepest pre-period load shape (load shape bin 10) realized savings across all four seasons.

Figure 36: PG&E HERs Energy Savings by Load Shape Bin and Season

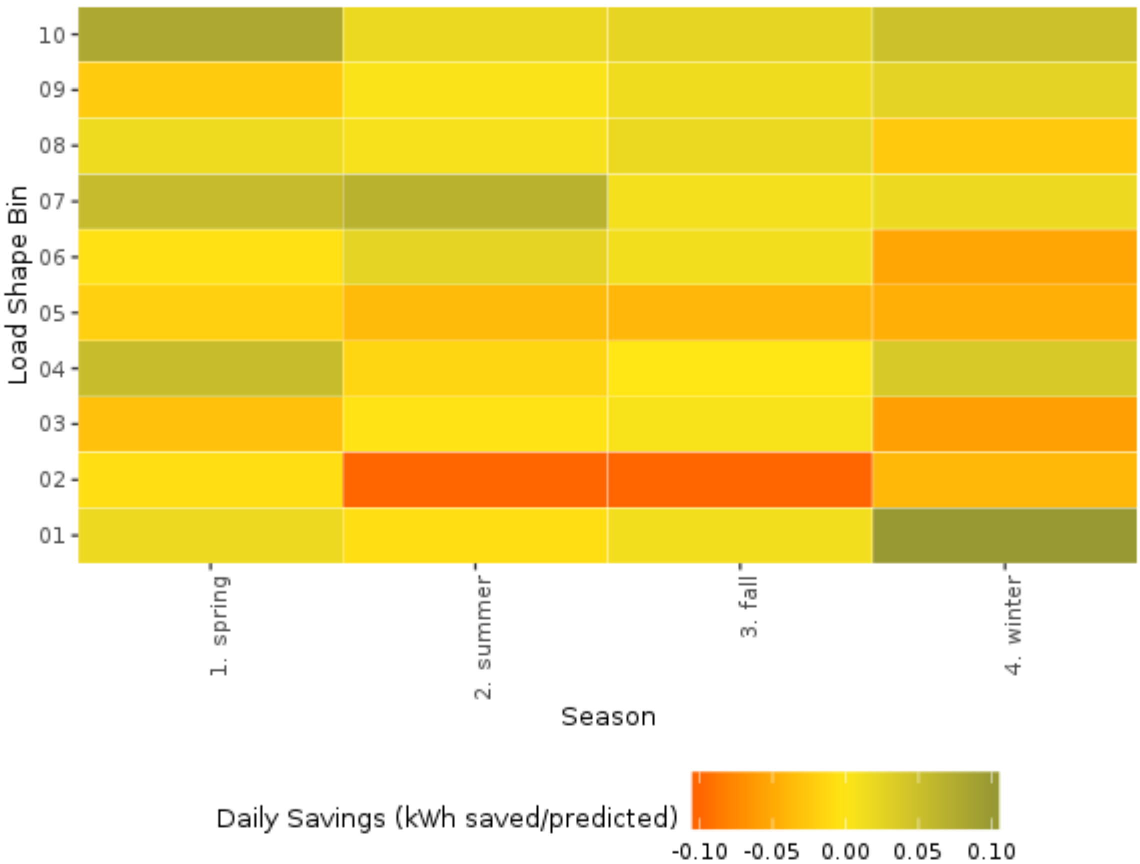
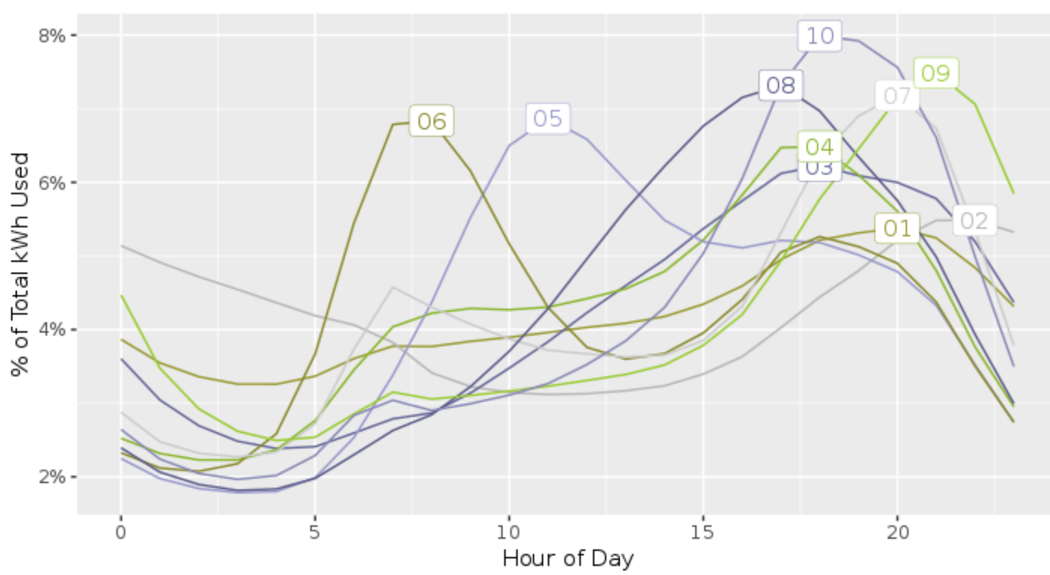


Figure 37: PG&E HERs Load Shape Bins



Overall, the AMICS model estimated that the average energy savings attributable to the HERs program treatment was 0.12 kWh per day, or 0.6 percent. These savings varied substantially across customer segments.

3.3 Commercial & Industrial HVAC

This section provides our analysis of multiple commercial HVAC programs. Commercial customers typically have greater variations in energy use across sites given the diversity of building types, end uses, and business activities relative to the residential sector. These programs provide an initial test of whether the AMICS model can be adapted to predict load shapes for commercial customers using only AMI billing data.

We used similar filters and segmentation strategies for each of the commercial HVAC programs. The holdout tests for each program demonstrate that the AMICS model is able to produce reasonable estimates of load shapes for participants of commercial HVAC programs, with predictions within 1 percent of the actual usage of the holdout sample. The AMICS model detected statistically significant savings for the PG&E Air Care Plus program, consistent with our expectations by season and time-of-day for improved air conditioning efficiency. The Commercial Quality Maintenance and Quality Installation program savings estimated by our model were not statistically significant during most hours at the program level, despite the tight error bounds around our predictions.

In each of the commercial programs, the AMICS segmentation revealed a wide variation in energy savings across customer segments. We found consistent energy savings attributed to HVAC interventions for participants in the retail sector, but these were offset (at least in part) by increases in energy usage attributed to participants in the manufacturing and health sectors. These findings suggest that the commercial HVAC programs could benefit from improved targeting.

3.3.1 PG&E Commercial Quality Maintenance and Air Care Plus

Program Description

The PG&E Commercial Quality Maintenance (CQM) program is part of a California statewide program designed to achieve energy and demand savings through assessment and optimization of existing commercial HVAC units.[34] Qualifying contractors have six months to repair the HVAC units to meet the program criteria or meet the ACCA 180 Standard before the customers receive incentive payments. CQM participants sign an

[34] Eligible buildings must have a commercial rooftop HVAC unit that weighs at least three tons. (or has a capacity greater than 3 tons). Most qualified buildings will have multiple rooftop HVAC units, and participants can choose to service all of their units or only a subset.

agreement with the contractor to receive quarterly system assessments and then perform any required maintenance.

The PG&E Air Care Plus program is similar but only requires a one-time HVAC maintenance visit with no industry standards to meet. Examples of the HVAC maintenance activities include refrigerant charge adjustment, coil cleaning, blower motor retrofits, enhanced time delay relay, airflow correction, and installation of a programmable thermostat. These activities should improve cooling delivery (from reduced runtime and/or power draw) and thereby improve efficiency.

Note that for the CQM and Air Care Plus programs, the existing conditions in the pre-participation period are the appropriate baseline and therefore, no additional adjustments are needed to the baseline to calculate program impacts. The savings estimates would benefit from utilizing a comparison group, however, which we were not able to explore in this analysis due to the data limitations discussed previously.

Database

PG&E provided Evergreen with AMI whole-building billing data and account characteristics for 1,503 distinct commercial customers that received incentives for HVAC maintenance between April 2006 and July 2016, including 1,205 CQM participants and 298 Air Care Plus participants. The PG&E program data included customer and program participation information such as HVAC technology type, maintenance activity description by date(s), *ex ante* gross energy and demand savings, building type, business NAICS code, and rate schedule.

The AMI billing data for this study contained approximately 125 million observations from November 6, 2011 to July 31, 2016.[35] These data captured energy usage patterns of most participating businesses for a full year before their program participation.

We applied filters to exclude customers with:

- No non-zero *ex ante* savings listed in the program documentation (n=255);[36]
- Net energy metering, such as onsite solar generation (n=67);
- No pre-period observations in the billing data or average energy usage in the pre- or post-period of less than 0.1 kWh or extreme changes from the pre- to post-

---

[35] For consistency across customers in the study, all 15-minute interval billing data were aggregated to the hourly level.

[36] Some participants did not require any adjustments (i.e., tests revealed that their system did not need any maintenance); these participants were excluded from our post-period analysis because they would not have any energy savings attributable to the program.

periods of more than 150 percent or less than -66 percent (n=172).

As a result, the AMI analysis was limited to 989 customers. As shown in Figure 38, this sample was spread across many counties and CEC climate zones. Half of the buildings (51%) were on rate schedules with demand charges, meaning their bill is impacted by their maximum demand (kW) in addition to their total energy usage (kWh) during each billing cycle. Over a third (36%) were enrolled in one or more demand response programs, with the potential to receive incentives for reducing their usage on event days.

Figure 38: PG&E Commercial HVAC Participants by County and Climate Zone

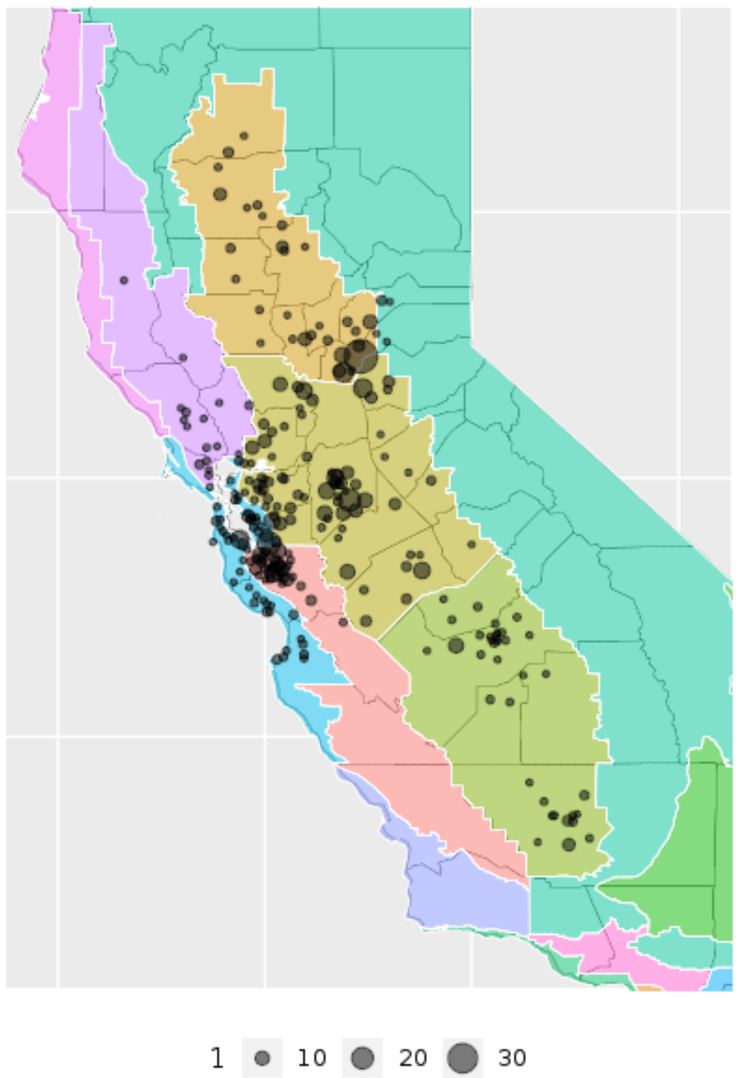


Table 5 shows the distribution of industries within the participant population. The top four sectors are the same across these two programs, but the distribution is different. Air Care Plus participants include proportionally more schools and restaurants, while the CQM program includes a higher proportion of offices (e.g., IT and finance).

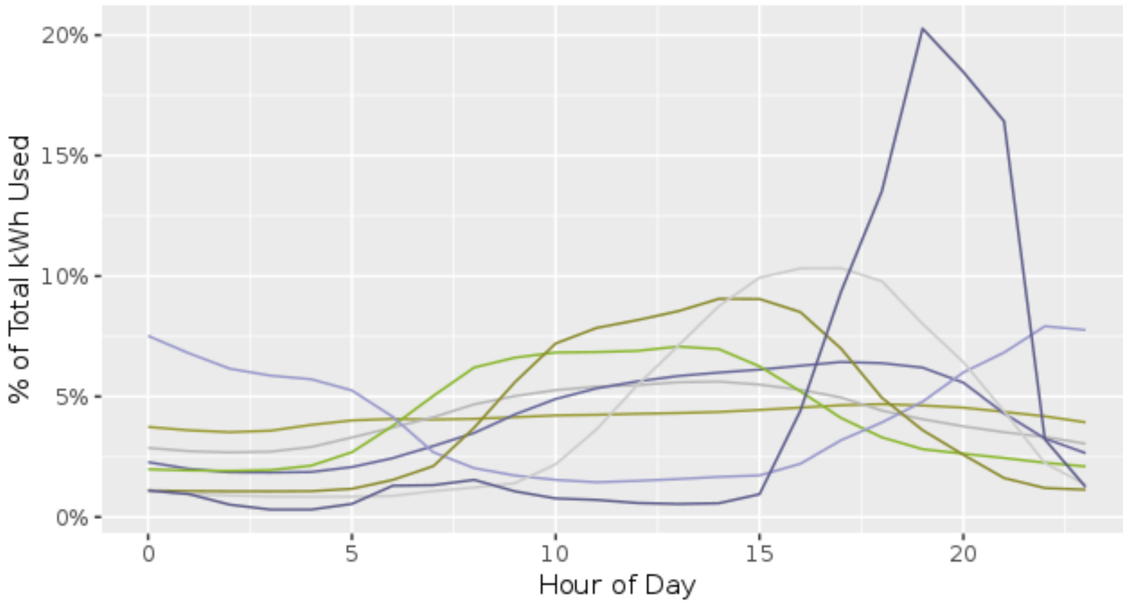
Table 5: PG&E Commercial HVAC Participants by Business Type

NAICS Codes	Description	Air Care Plus (n=148)	CQM (n=841)
5*****	Information, finance and insurance, real estate, management; professional, scientific, and technical services	9%	23%
44****, 45****	Retail trade	14%	16%
61****	Educational services	24%	10%
722***	Food services and drinking places	20%	7%
62****	Health care and social assistance	5%	7%
71****	Arts, entertainment, and recreation	7%	7%
813***	Religious, civic, professional, and similar organizations	6%	6%
-	Other	10%	14%
-	Undefined	5%	10%

Segmentation

For these programs, we defined customer segments with a combination of daily energy usage (magnitude), normalized load shape (hours of use), and business type in the pre-period. First, we assigned customers to one of 20 bins by their average daily energy usage across the most recent pre-period year, with the highest energy usage bin containing the fewest customers. Next, we used *k*-means clustering to identify the eight unique clusters shown in Figure 39, each containing a subset of customers with similar load shapes during the pre-period. Lastly, we used the NAICS codes contained in the utility customer information system to identify 20 distinct business types, describing their sector and primary business activity.

Figure 39: PG&E Commercial HVAC Normalized Load Shape Bins



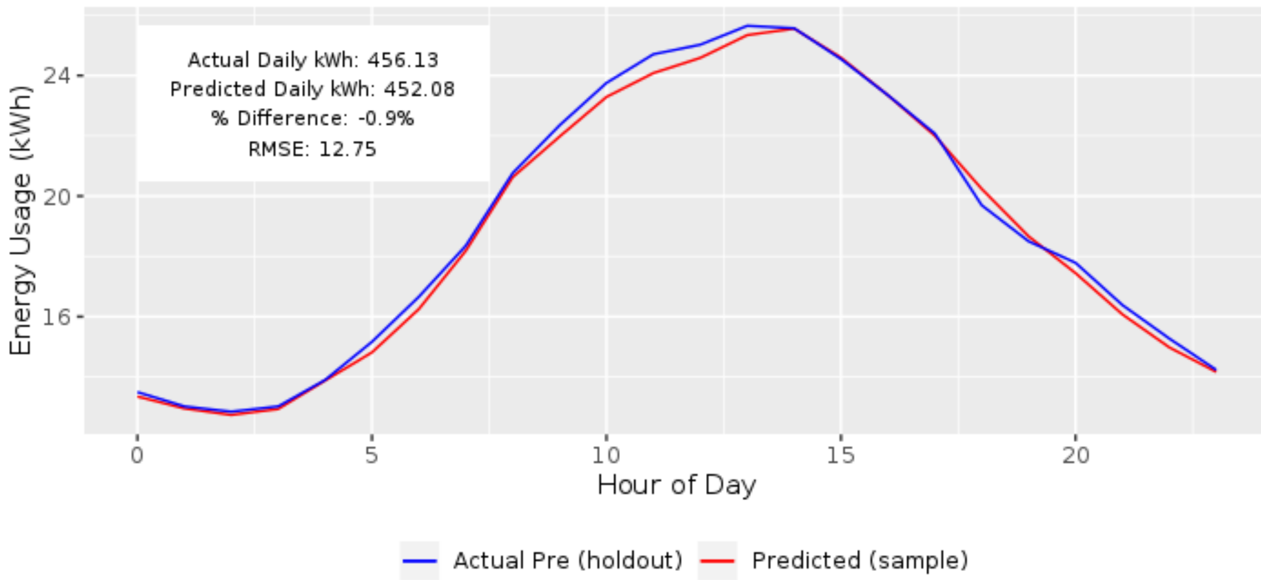
This segmentation approach defines 395 customer segments and 52 day bins, for a total of 20,540 distinct customer-day bins.[37]

Holdout Validation Tests

The results of one holdout test are shown in Figure 40, comparing the predicted pre-period load shape from the model (red line) to the actual pre-period load shape for the holdout sample (blue line). As a reminder, the holdout sample was randomly drawn from the population of program participants with AMI data, with a variety of business types. When the model is performing well, the two lines will overlap. The holdout test relies exclusively on pre-period data so that any differences between the predicted and actual energy usage can be attributed to model error, not to program savings. The model predictions track closely with the average actual load, with an overall difference of around 1 percent.

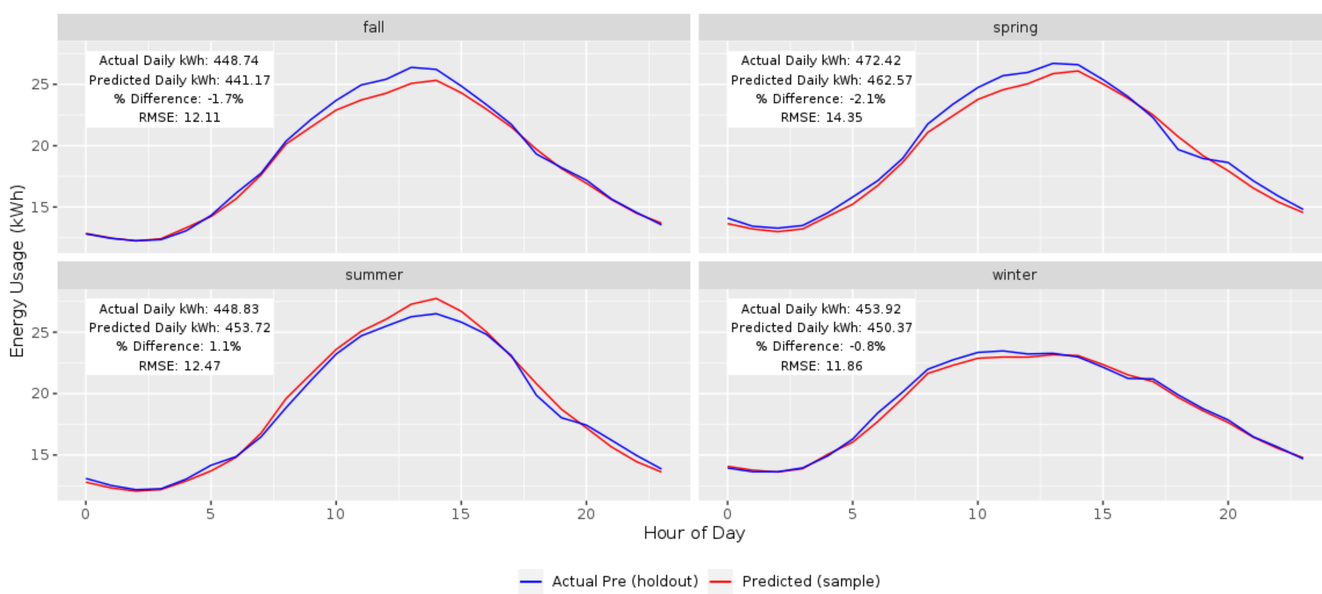
[37] The 395 customer segments are distinct combinations of 20 energy usage bins, 8 load shape clusters, and 26 industry groups. The 52 day bins are comprised of 9 CDD bins, 12 HDD bins, and 2 day types (but not season). Not all possible combinations of the customer and day segments were observed in the pre-period data.

Figure 40: PG&E Commercial HVAC Holdout Test



As shown in Figure 41, the AMICS model predictions were less accurate for the holdout sample during specific seasons, with slight overestimations during summer afternoons and underestimations during these hours of the spring and fall.

Figure 41: PG&E Commercial HVAC Holdout Test, by Season



Program Energy Savings

Figure 42 compares the post-period predicted load shape (red) with the actual post-period load shape (blue) across all CQM and Air Care Plus participants in the database. These predictions are based on the pre-period energy consumption model and post-period weather data; they represent the expected load shape for these customers in absence of the program intervention. The error of each hourly prediction is depicted as a 95 percent confidence interval in the shaded area around each estimate. Whenever the actual post-period load shape (blue line) falls outside the predicted post-period load region (red area), this indicates that a statistically significant change was observed during that hour.

The actual and predicted load shapes of CQM participants are nearly indistinguishable, indicating that there were no significant savings. For the Air Care Plus participants, the actual post-period load shape (blue) falls below the predicted load shape (red) during early morning and afternoon hours, with overall savings of 3.9 percent.

Figure 42: PG&E CQM/ACP in Post-Period

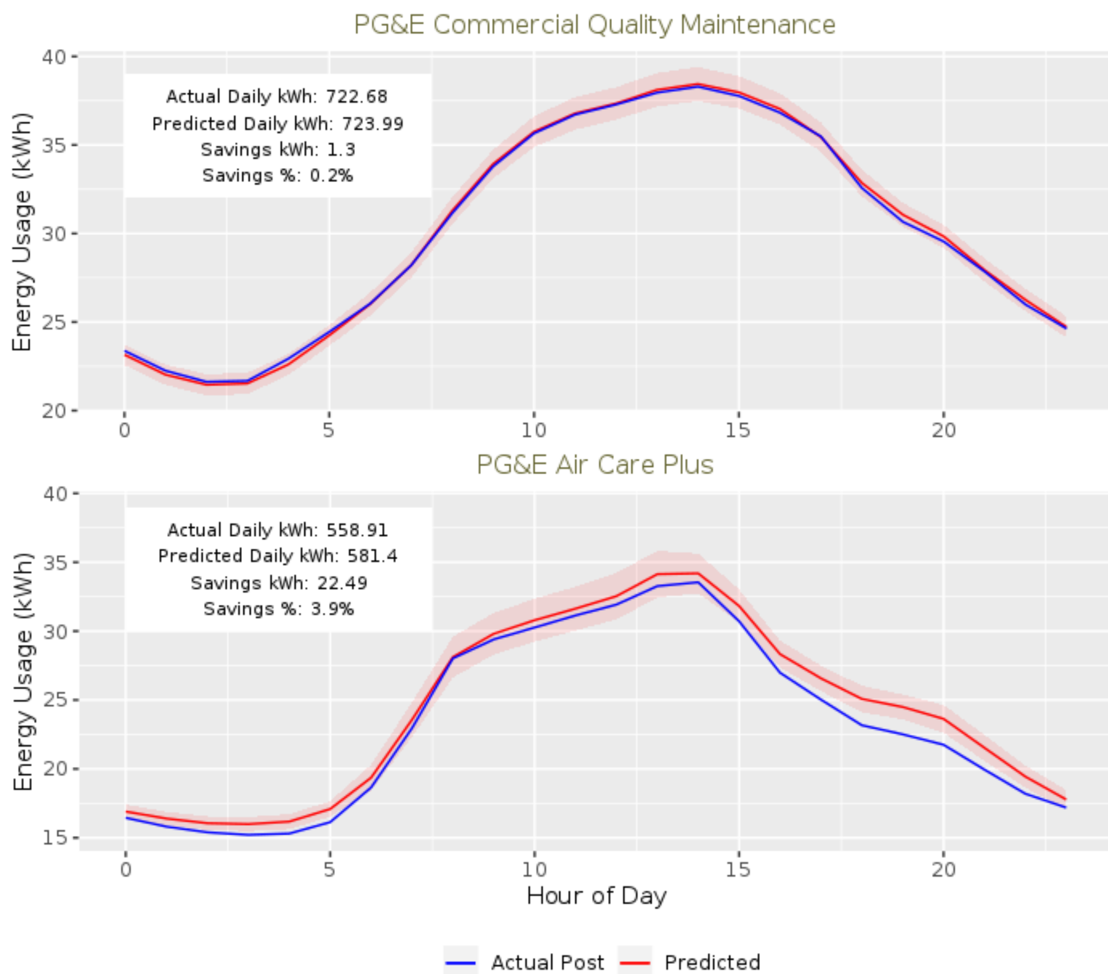


Figure 43 shows our estimated hourly kWh savings across the entire post-period, with error bars depicting 95 percent confidence intervals around each estimate by program. The PG&E CQM participants realized very limited energy savings, and none of the reductions in energy usage were statistically significant. The PG&E Air Care Plus participants realized significant energy savings from 1:00 a.m. to 5:00 a.m. and then from 4:00 p.m. to 11:00 p.m., including the system peak period. Overall, we estimate that the PG&E CQM program produced energy savings of 1.3 kWh ± 16.4 per day (or 0.2% ± 2.2%) while the Air Care Plus program produced savings of 22.49 kWh ± 9.64 per day (or 3.9% ± 1.4%).

Figure 43: PG&E Commercial HVAC Estimated Savings

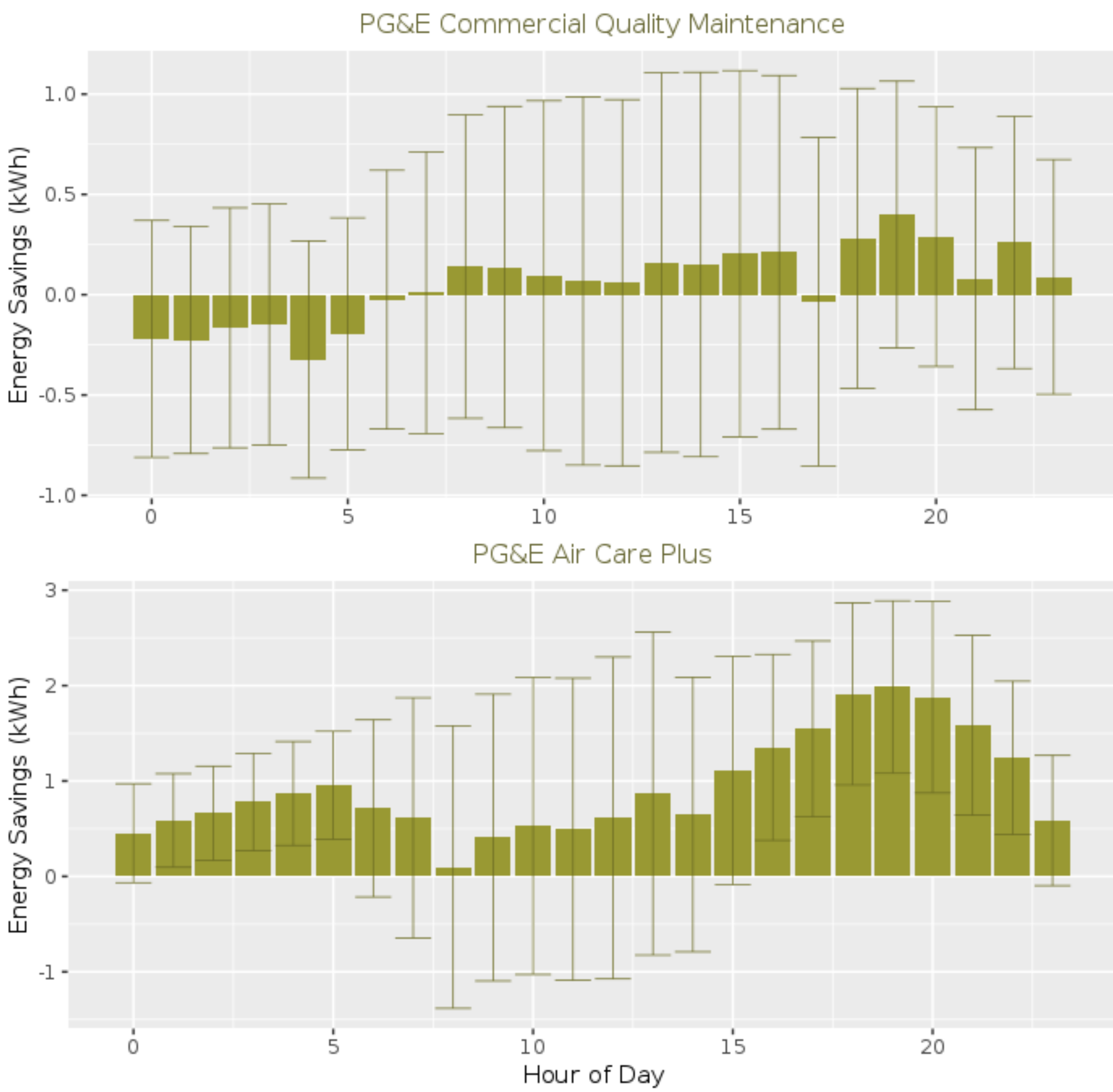
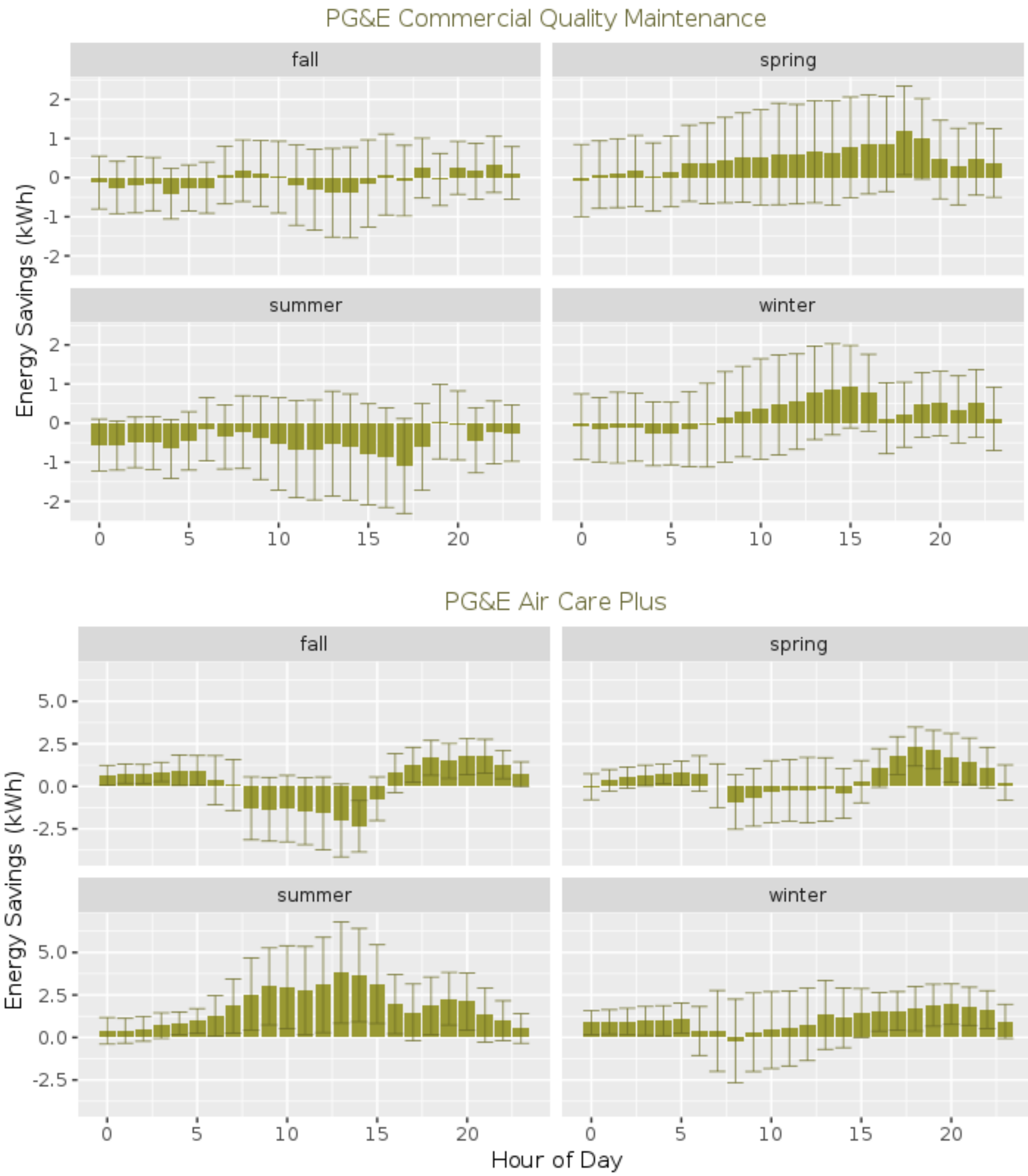


Figure 44 depicts the estimated savings in each of the four seasons. The AMICS model found positive, though statistically insignificant, savings among CQM participants during the spring and winter. We found positive and statistically significant energy savings among Air Care Plus participants during peak hours across all four seasons, with the most substantial savings on summer days.

Figure 44: PG&E Commercial HVAC Estimated Savings by Season



Savings by Segment

Table 6 shows the estimated energy savings for the four most common sectors among participants. We see a wide range of estimated energy savings by kWh and proportion of baseline consumption (%) across customer industries within each program. This can be explained, in part, by the variation in *ex ante* gross savings, which accounts for the HVAC size, type, and specific maintenance activities performed. However, there likely are additional differences in HVAC controls, operating schedules, and existing conditions that contribute to the differences in energy savings realized by participants from each business type.

Table 6: PG&E Commercial HVAC Energy Savings by Business Type

	Estimated Savings				*Ex Ante* Gross Savings			
	Air Care		CQM		Air Care		CQM	
Business Type	kWh	%	kWh	%	kWh	%	kWh	%
Information, finance and insurance, real estate, management; professional, scientific, and technical services	83.5	8%	-5.8	-1%	114.9	12%	73.0	9%
Retail trade	57.8	14%	10.4	2%	142.2	35%	47.1	7%
Educational services	32.2	4%	-3.3	-1%	67.8	7%	110.6	25%
Food services and drinking places	19.0	6%	-20.1	-3%	17.8	6%	18.5	3%

Note: Percentages represent kWh savings as a proportion of baseline kWh consumption.

3.3.2 SCE Commercial Quality Maintenance

Program Description

The SCE Commercial Quality Maintenance (CQM) program is part of a California statewide program designed to achieve energy and demand savings through assessment and optimization of existing HVAC units in commercial and industrial buildings.[38] SCE CQM participants sign a maintenance agreement with a qualified contractor to receive regular system assessments and any maintenance required to maintain HVAC performance and meet the ACCA 180 Standard. Note that for the SCE CQM program, the existing conditions in the pre-participation period are the appropriate baseline and

[38] Most qualified buildings will have multiple rooftop HVAC units, and participants can choose to service all of their units or only a subset.

therefore, no additional adjustments are needed to the baseline to calculate program impacts.

Database

SCE provided Evergreen with AMI whole-building billing data and account characteristics for 998 distinct commercial customers that received incentives for HVAC maintenance between February 2014 and October 2015. The SCE CQM program data included customer and program participation information such as maintenance activity description by date(s), *ex ante* gross energy and demand savings, HVAC size (tons), building type, vintage, facility size (square footage), business NAICS code, and rate schedule.

The AMI billing data for this study contained 23 million observations from January 1, 2013 to September 31, 2016.[39] These data captured energy usage patterns of most participating businesses for a full year before their program participation.

We applied filters to exclude customers with:

- Net energy metering, such as onsite solar generation (n=19);
- No pre-period observations in the billing data (n=42);
- No non-zero *ex ante* savings listed in the program documentation (n=508, retained only pre-period observations for the baseline model);[40] or
- Average energy usage in the pre- or post-period of less than 0.1 kWh or extreme changes from the pre- to post-periods of more than 150 percent or less than -66 percent (n=5).

As a result, the AMI analysis was limited to 932 customers. As shown in Figure 45, this sample was concentrated around Los Angeles, with fewer participants in the inland counties and mountainous climate zones. Nearly all of these buildings (99%) were on rate schedules with time-of-use and demand charges, meaning their bill is impacted by their energy usage (kWh) during peak hours and maximum demand (kW) in addition to their total energy usage (kWh) during each billing cycle. Almost one half (43%) were enrolled in a demand response program, with 34 percent enrolled in a direct load control program that allows SCE to cycle their HVAC or other connected equipment to reduce usage on event days.

---

[39] For consistency across customers in the study, all 15-minute interval billing data were aggregated to the hourly level.

[40] Some participants did not require any adjustments (i.e., tests revealed that their system did not need any maintenance); these participants were excluded from our post-period analysis because they would not have any energy savings attributable to the program.

Figure 45: SCE CQM Participants by County and Climate Zone

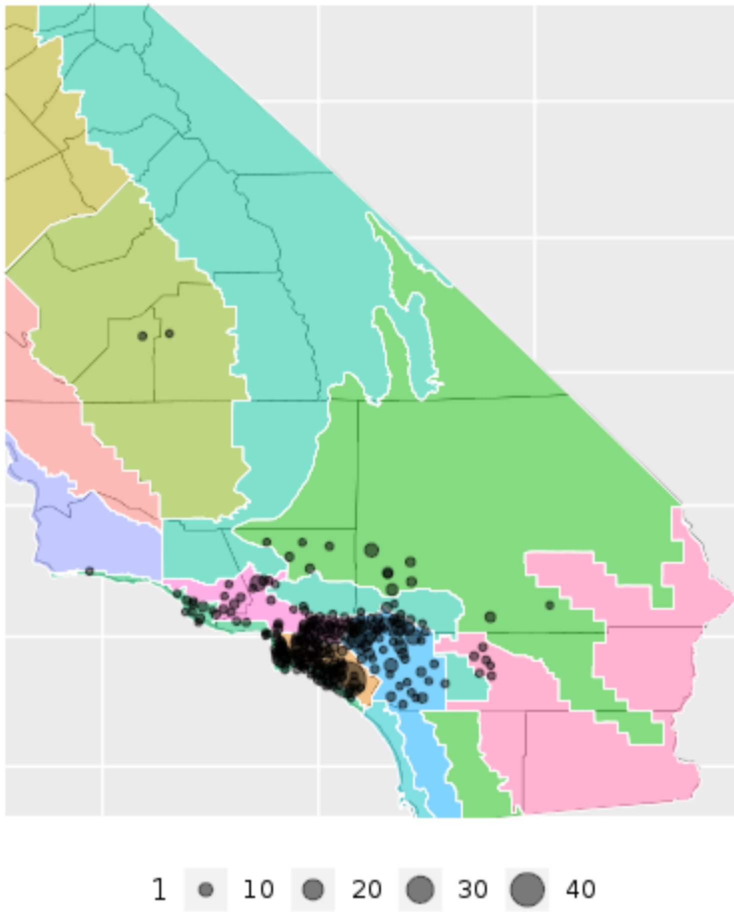


Table 7 shows the distribution of business types within the participant population; these were determined using a combination of NAICS codes (i.e., industry) and building types listed in SCE's customer database. The retail trade and grocery sectors make up more than half of all participants in SCE's CQM program. The top five sectors are the same as for the PG&E CQM and Air Care Plus programs.

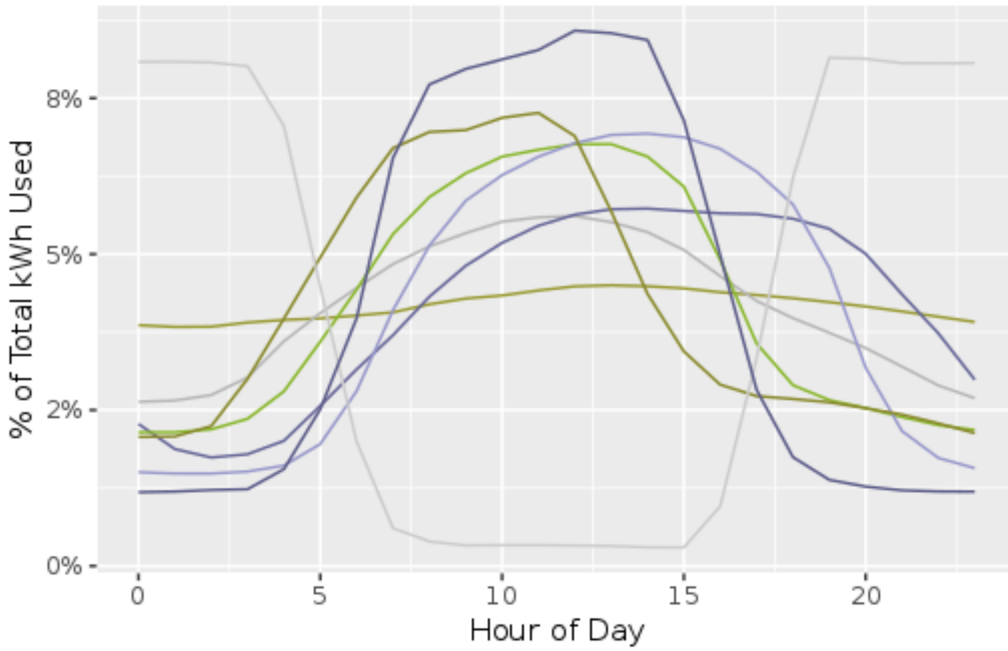
Table 7: SCE CQM Participants by Business Type

Business Type	Relevant NAICS Codes	CQM Participants
Retail trade	44****, 45****	33%
Grocery	445**	21%
Office (e.g., insurance, management)	5*****	13%
Restaurant	722***	7%
Education	61****	7%
Manufacturing	3*****	5%
Wholesale trade	42****	4%
Health care and social assistance	62***	4%
Religious, civic, and professional associations	813***	2%
Construction	23****	2%
Other	-	2%

Segmentation

For the SCE CQM program, we defined customer segments with a combination of daily energy usage (magnitude), normalized load shape (hours of use), and business type. First, we assigned customers to one of 20 bins by their average daily energy usage across the most recent pre-period year, with the highest energy usage bin containing the fewest customers. Next, we used *k*-means clustering to identify the eight unique clusters shown in Figure 46, each containing a subset of customers with similar load shapes during the pre-period. The U-shaped load group includes 10 buildings that are not on a net energy metering (NEM) rate schedule, but have a load shape consistent with interconnected generation or storage. Lastly, we used the building type and NAICS codes contained in the utility customer information system to create 11 distinct business types based on their building type and industry.

Figure 46: SCE CQM Load Shape Clusters



This segmentation approach defines 235 customer segments and 44 day bins, for a total of 10,340 distinct customer-day bins.[41]

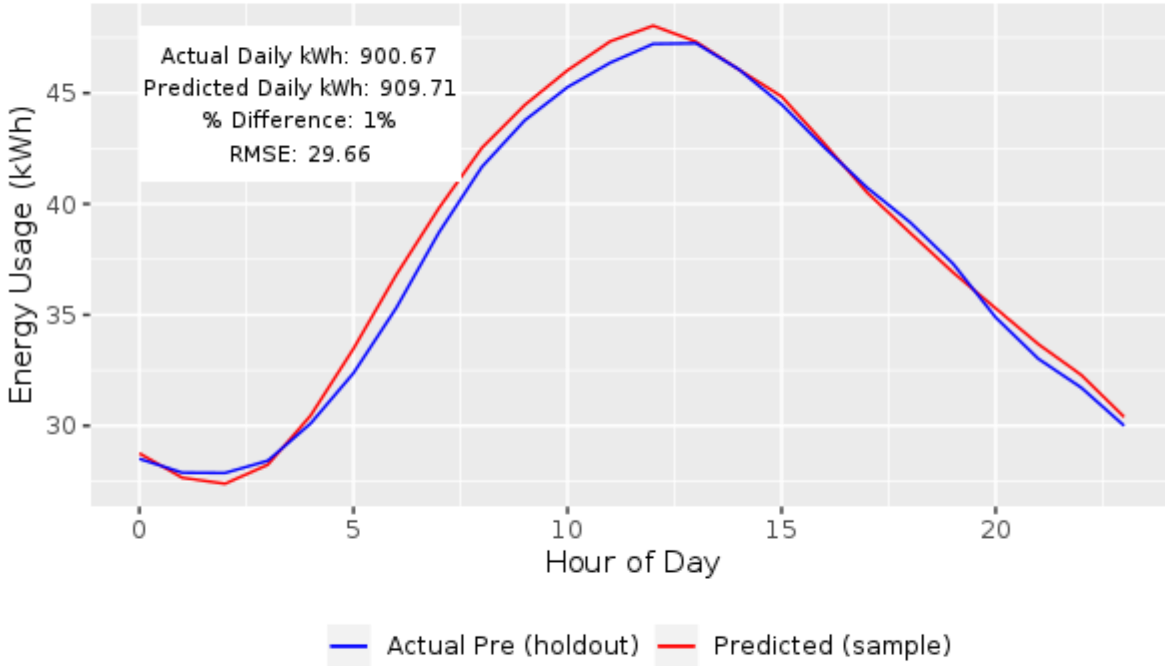
Holdout Validation Tests

The results of one holdout test are shown in Figure 47, comparing the predicted pre-period load shape from the model (red line) to the actual pre-period load shape for the holdout sample (blue line). When the model is performing well, the two lines will overlap. The holdout test relies exclusively on pre-period data so that any differences between the predicted and actual energy usage can be attributed to model error, not to program savings. The model predictions track with the average actual load, with an overall difference of 1 percent. The predictions are more accurate during the afternoon and evening, around the system peak.

---

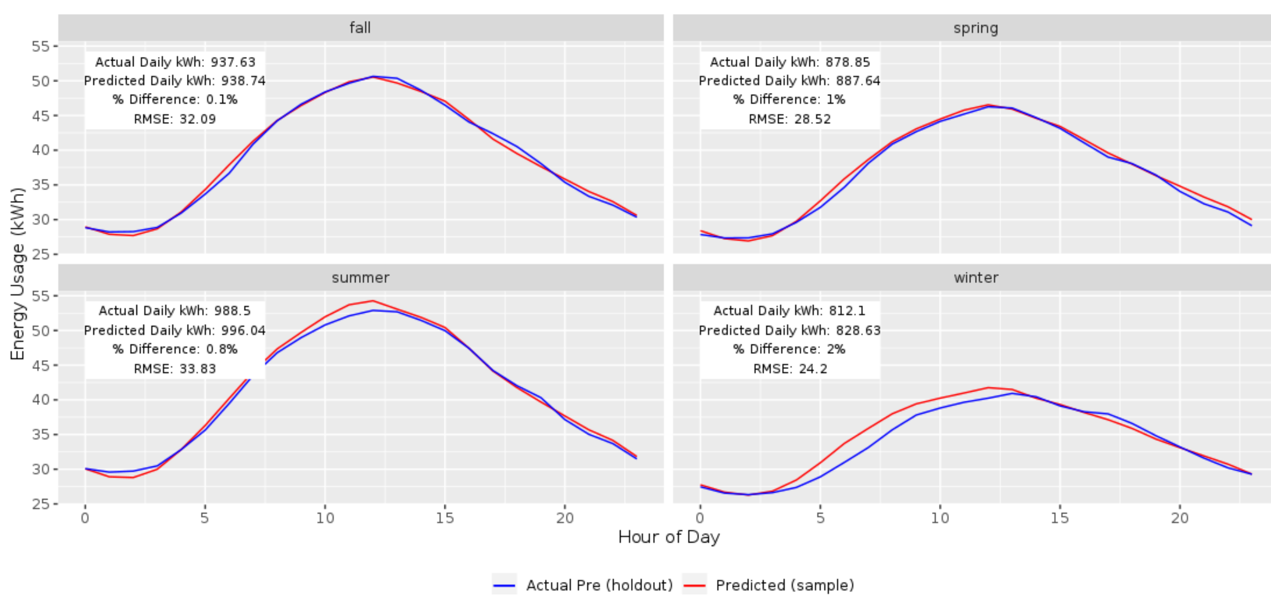
[41] The 235 customer segments are distinct combinations of 20 energy usage bins, 8 load shape clusters, and 11 business types. The 44 day bins are comprised of 10 CDD bins, 9 HDD bins, and 2 day types (but not season). Not all possible combinations of the customer and day segments were observed in the pre-period data.

Figure 47: SCE CQM Holdout Test



As shown in Figure 48, the AMICS model predictions were less accurate for the holdout sample during individual seasons, with overestimations of the summer usage from 10:00 a.m. to 1:00 p.m. and of the winter usage from 5:00 a.m. to 1:00 p.m. Even in the winter, the predictions are within 2 percent of the actual energy usage at sites in the holdout sample.

Figure 48: SCE CQM Holdout Test, by Season



Program Energy Savings

Figure 49 compares the post-period predicted load shape (red) with the actual post-period load shape (blue) across all SCE CQM participants in the database with records of HVAC maintenance.[42] This prediction is based on the pre-period baseline energy consumption model and post-period weather data; it represents the expected load shape for these customers in absence of the program intervention. The error of each hourly prediction is depicted as a 95 percent confidence interval in the shaded area around each estimate. Whenever the actual post-period load shape (blue line) falls outside the predicted post-period load region (red area), this indicates that a statistically significant change was observed during that hour. In this case, the actual post-period load shape (blue) falls just below the predicted load shape (red) during the afternoon and evening hours, but the difference is minor.

Figure 49: SCE CQM in Post-Period

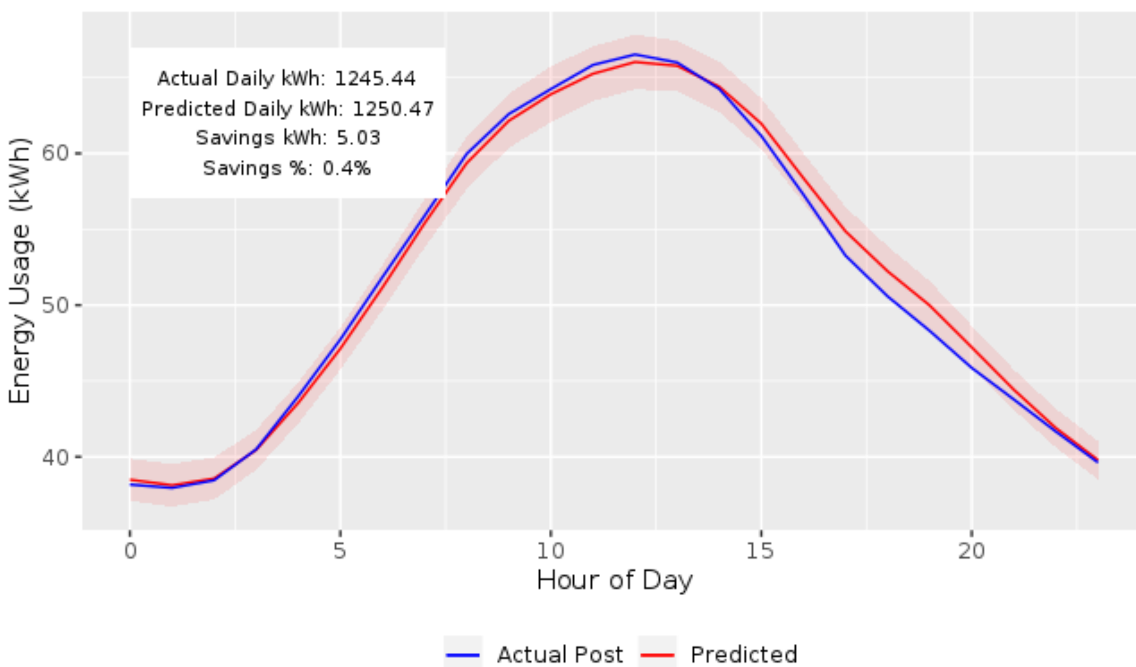


Figure 50 shows our estimated hourly kWh savings across the entire post-period, with error bars depicting 95 percent confidence intervals around each estimate. The SCE CQM participants exhibited energy savings during all 24 hours of the day, but these savings were only statistically significant during the evening hours of 5:00 p.m. to 7:00 p.m.

[42] Some participants did not require any adjustments (i.e., tests revealed that their system did not need any maintenance); these participants were excluded from our post-period analysis because they do not have any energy savings attributable to the program.

Overall, we estimate that the SCE CQM program led to energy savings of 5.03 kWh ± 19.56 per day (or 0.4% ± 1.5%).

Figure 50: SCE CQM Estimated Savings

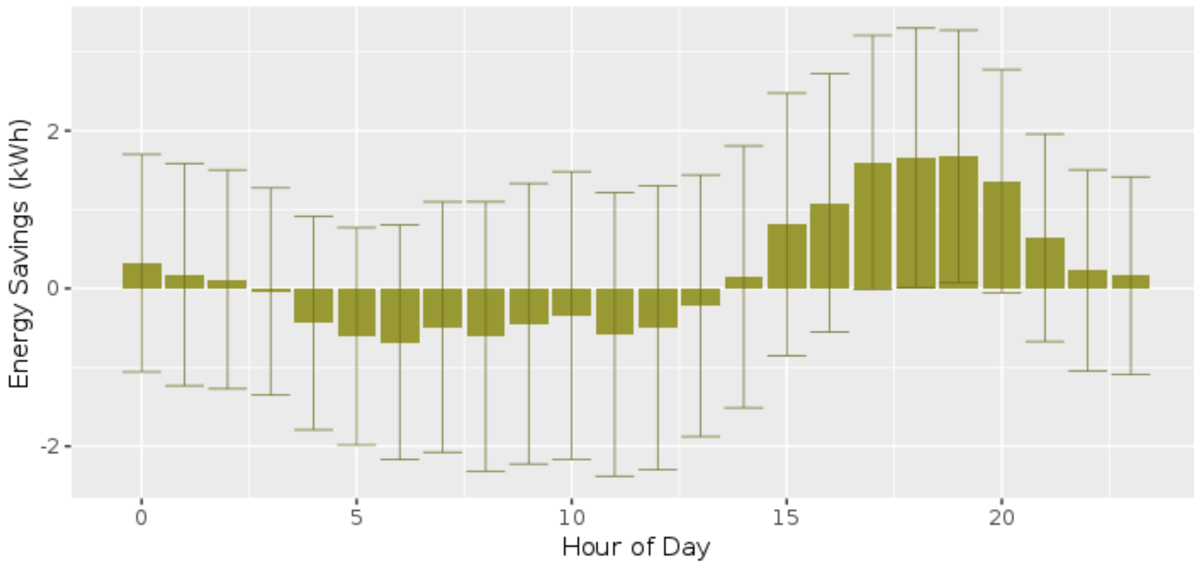
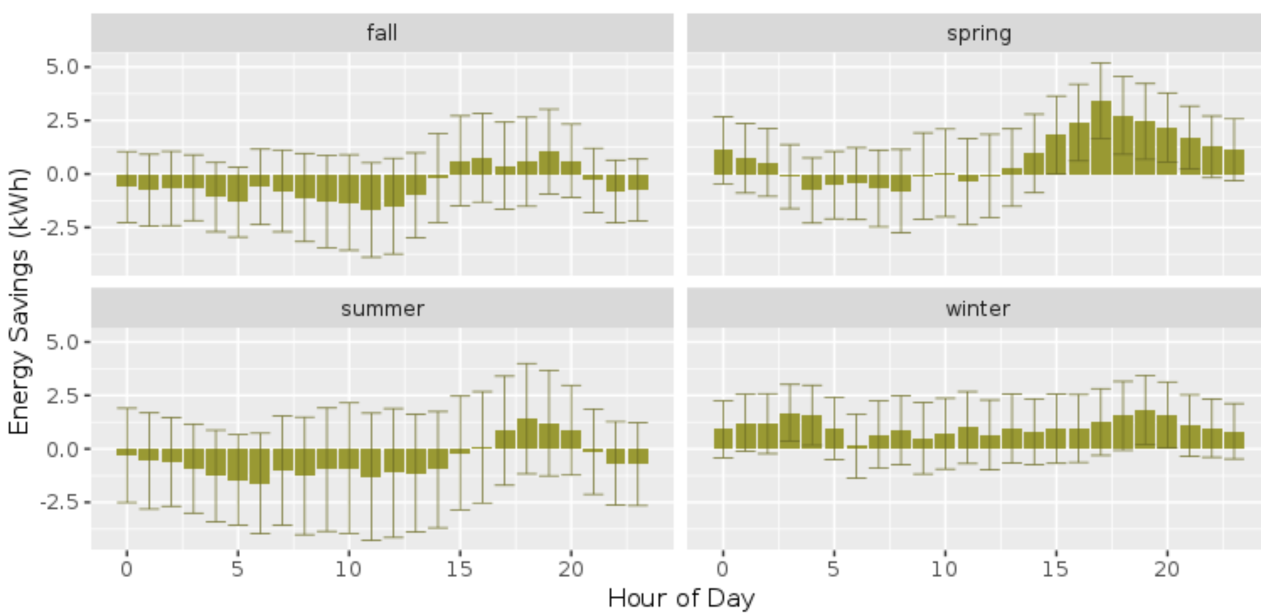


Figure 51 depicts the estimated savings in each of the four seasons. The AMICS model found positive and statistically significant energy savings in the spring, with small and mostly insignificant savings in the winter.

Figure 51: SCE CQM Estimated Savings by Season



Savings by Segment

The segmentation in the AMICS model provides a unique opportunity to see the variation in program savings by customer segment and weather.

Table 8 shows the AMICS estimate of energy savings by business type among participant sites with non-zero *ex ante* savings that were observed in the post-period. We see a wide range of estimated energy savings by kWh and proportion of baseline consumption (%) across sectors within each program. This is due in part to differences in HVAC maintenance activities, reflected in the *ex ante* gross savings but also differences in savings potential related to the HVAC capacity, controls, and operating schedules across sectors. Three business types exhibited negative savings, increasing their energy usage in the post-period; this will offset some of the positive energy savings realized by the other business types when viewed at the program level. This information can then be used by program implementors to target and focus recruitment efforts on the business types with the highest potential for savings (i.e., based on estimate savings of prior participants rather than ex ante assumptions). Additionally, this could prompt a process evaluation with a focus on why the current offering is not working well businesses in the construction, health care, or manufacturing industries.

Table 8: SCE CQM Gross Energy Savings by Business Type

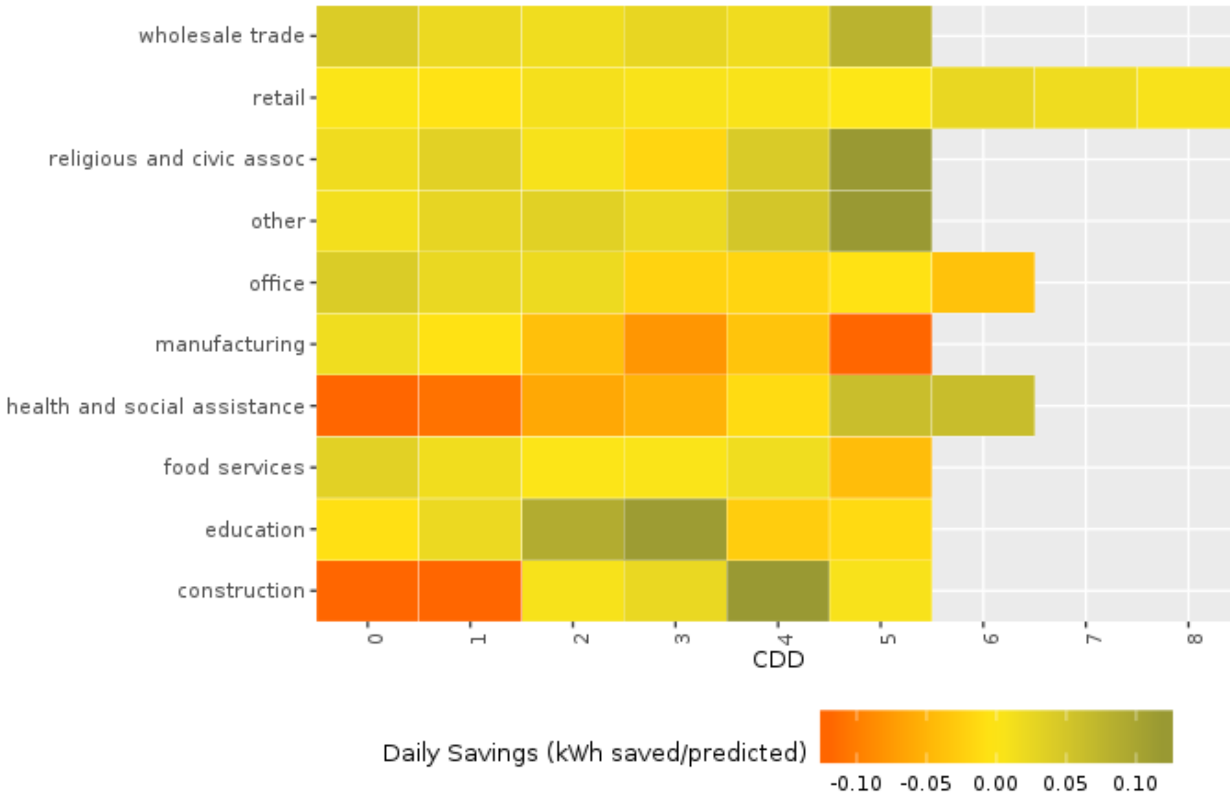
Business Type	CQM Participants	N Sites in Post-Period	Estimated Savings kWh	Estimated Savings %	Ex Ante Savings kWh	Ex Ante Savings %
Retail trade	33%	238	2.8	0.4%	10.88	1.3%
Office (e.g., insurance, management)	13%	67	33.2	2.2%	26.57	2.0%
Education	7%	38	30.2	2.7%	39.15	3.3%
Manufacturing	5%	33	-57.0	-1.3%	35.73	0.8%
Wholesale trade	4%	16	21.3	2.9%	25.08	1.6%
Restaurant	7%	12	57.4	2.0%	17.15	0.8%
Health care and social assistance	4%	9	-69.2	-8.9%	21.04	2.8%
Religious, civic, and professional associations	2%	8	17.5	1.8%	37.85	3.7%
Construction	2%	6	-93.0	-11.1%	13.47	1.7%
Other	2%	8	29.3	2.7%	27.47	2.7%

Note: Percentages represent kWh savings as a proportion of baseline kWh consumption.

Figure 52 shows the average daily savings estimated by the AMICS model by business type and cooling load. The rows show customers segmented by business type, and the columns show the cooling load by CDD (hottest days on the right). We automatically color-coded the cells with the highest kWh savings in dark green and the lowest in dark orange (negative savings = increased usage); the yellow cells fall in the middle of this spectrum.

Participants in the health and social assistance industry realized some positive energy savings (green) on days with high cooling load (on the right), but those savings were offset by negative savings (orange) on days with little to no need for cooling (on the left). Participants in the retail trade realized much lower savings (light yellow-green) that were stable across days with low, moderate, and high cooling loads.

Figure 52: SCE CQM Energy Savings by Business Type and CDD



3.3.3 SCE Commercial Quality Installation

Program Description

This application of the AMICS model used a sample of commercial customers that participated in SCE's Commercial Quality Installation (CQI) program. The CQI program requires installation of higher efficiency HVAC units in addition to quality installation procedures that help ensure that the units are installed and operating in a manner that maximizes their efficiency. The AMICS model was used to estimate savings for participants relative to a pre-installation baseline, as discussed previously. The sample did not include a comparison group of similar HVAC installations, so we were not able to parse the estimated savings into the portion attributable to the new efficient HVAC equipment versus the quality installation processes.

Database

SCE provided Evergreen with AMI whole-building billing data and account characteristics for 1,972 distinct commercial and industrial customers that received incentives for installation of an efficient HVAC unit by a qualified contractor between February 2014 and May 2016. The SCE program data included customer and program participation information such as *ex ante* gross energy and demand savings, HVAC size (tons), building type, business NAICS code, and rate schedule.

The AMI billing data for this study contained 26 million observations from January 1, 2014 to September 31, 2016.[43]

We applied filters to exclude customers with:

- No pre-period observations in the billing data (n=353);
- Net energy metering, such as onsite solar generation (n=94);
- Average daily use greater than 40,000 kWh (n=13);[44] or
- Extreme changes from the pre- to post-periods of more than 150 percent or less than -66 percent (n=172).

As a result, the AMI analysis was limited to 1,340 customers. As shown in Figure 53, this sample covers a wide geographic region, with participants along the coast and in the mountainous climate zones. Nearly all of these buildings (99%) were on rate schedules with time-of-use and demand charges, meaning their bills are impacted by their energy usage (kWh) during peak hours and maximum demand (kW) in addition to their total energy usage (kWh) during each billing cycle. Eighteen percent were enrolled in a demand response program, with 92 enrolled in a direct load control program that allows SCE to cycle their HVAC or other connected equipment to reduce usage on event days.

[43] For consistency across customers in the study, all 15-minute interval billing data were aggregated to the hourly level.
44 These very large customers have too much variation in their load shapes, and their inclusion in the model would have led to less accurate forecasts for the remaining commercial customers. They should be modeled individually.

Figure 53: SCE CQI Participants by County and Climate Zone

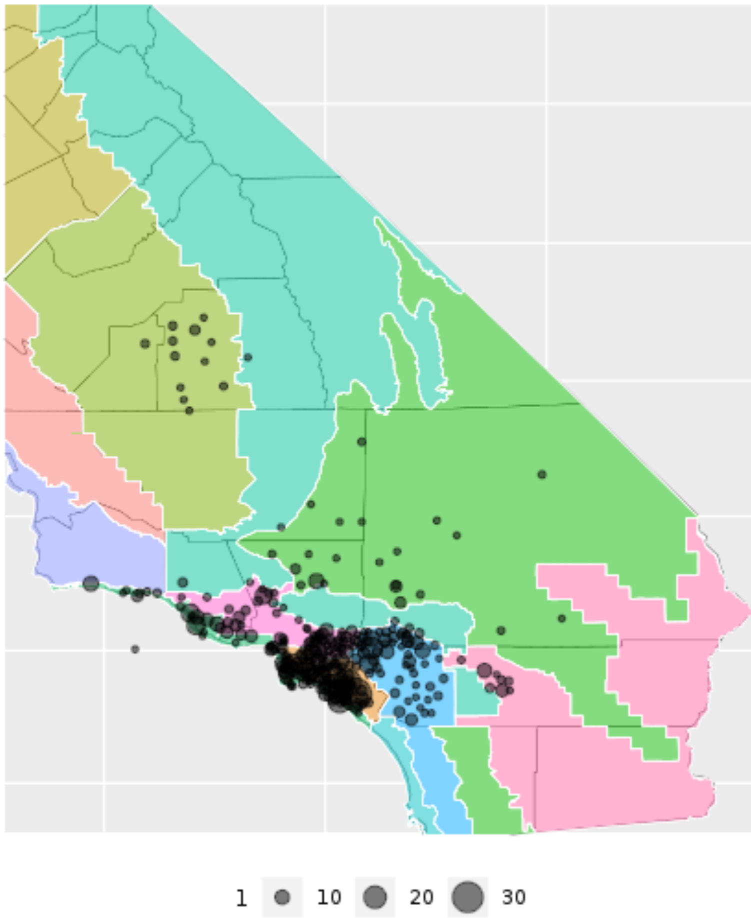


Table 9 shows the distribution of business types within the participant population; these were determined using the customer segment listed in SCE's database. We have provided a column with relevant NAICS codes to aid comparisons across programs. Offices, retail, and schools comprise half of all participants in SCE's CQI program.

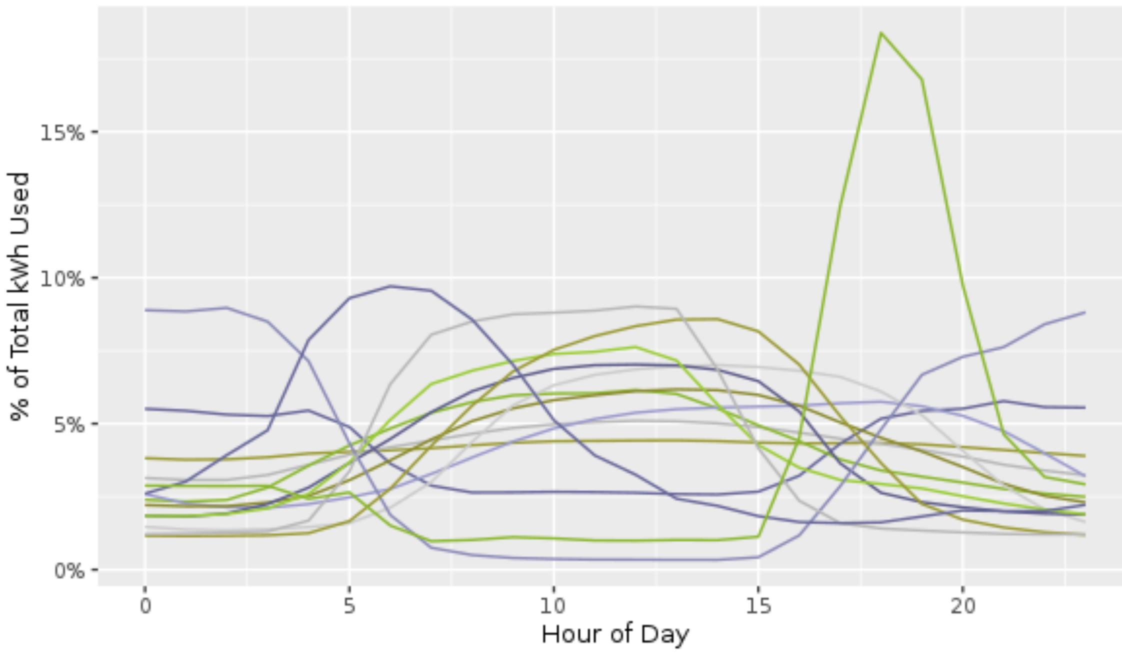
Table 9: SCE CQI Participants by Business Type

Business Type	Relevant NAICS Code	CQI Participants
Office buildings	5*****	29%
Retail stores	44****, 45****	13%
Other commercial	51****, 71****, 81****	11%
Schools	6111**	10%
Restaurants	722***	9%
Other industrial	3*****	7%
Warehouses	42****	4%
Builders	23****	3%
Hospitals/medical facilities	62****	2%
Colleges & universities	6112**, 6113**	2%
Food stores & refrigerated warehouses	445***	2%
Other	-	10%

Segmentation

For this program, we defined customer segments with a combination of daily energy usage (magnitude), normalized load shape (hours of use), and business type. First, we assigned customers to one of 25 bins by their average daily energy usage across the most recent pre-period year, with the highest energy usage bin containing the fewest customers. Next, we used *k*-means clustering to identify 15 unique clusters, each containing a subset of customers with similar load shapes during the pre-period. Figure 54 shows the 14 clusters with more than one site. Lastly, we used the business segment contained in the utility customer information system to create 21 distinct groups describing the primary business activity.

Figure 54: SCE CQI Load Shape Clusters



This segmentation approach defined 510 customer segments and 45 day bins, for a total of 22,950 distinct customer-day bins.[45]

Holdout Validation Tests

The results of one holdout test are shown in Figure 55 and Figure 56, comparing the predicted pre-period load shape from the model (red line) to the actual pre-period load shape for the holdout sample (blue line). When the model is performing well, the two lines will overlap. The holdout test relies exclusively on pre-period data so that any differences between the predicted and actual energy usage can be attributed to model error, not to program savings. The final customer segmentation approach resulted in a model that closely matched the holdout sample, with an overall prediction error of only 0.8 percent. For the seasonal models, the CQI model prediction error was around 1 percent in three seasons, with higher error in the winter months. The differences in prediction error across seasons will be evident in the width of our error bounds by season in the post-intervention period.

---

[45] The 510 customer segments are distinct combinations of 25 energy usage bins, 15 load shape clusters, and 21 business types. The 45 day bins are comprised of 10 CDD bins, 9 HDD bins, and 2 day types (but not season). Not all possible combinations of the customer and day segments were observed in the pre-period data.

Figure 55: SCE CQI Holdout Test

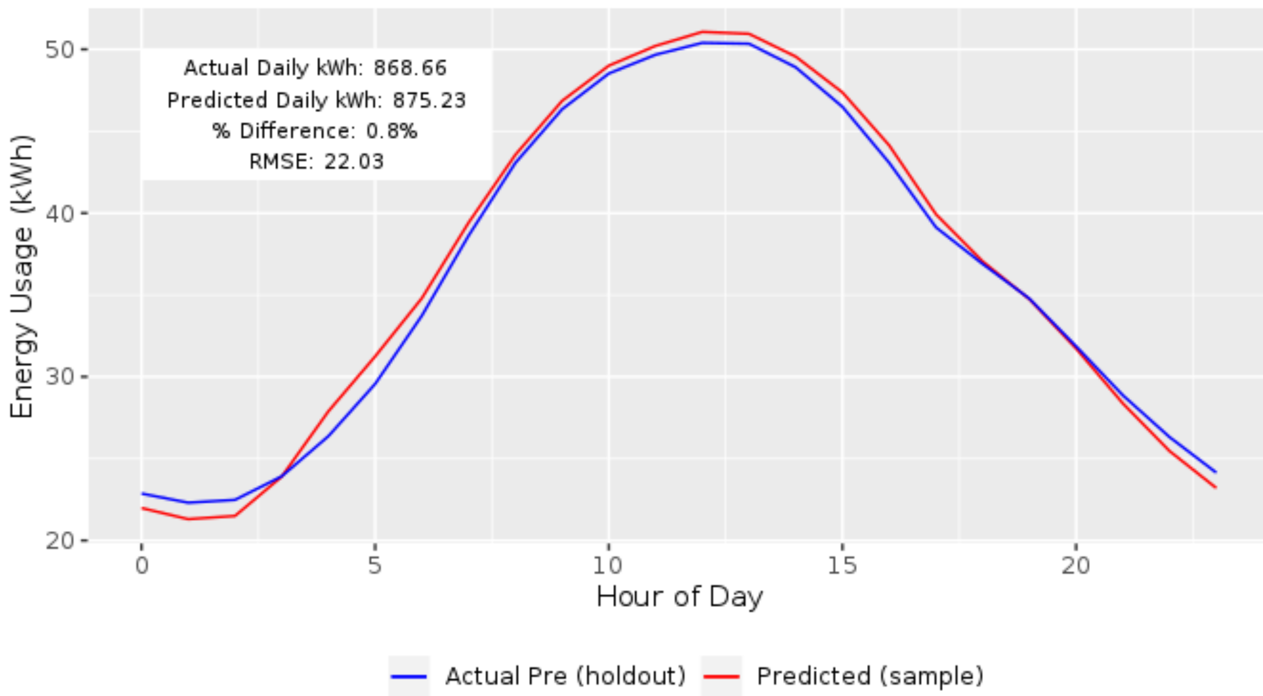
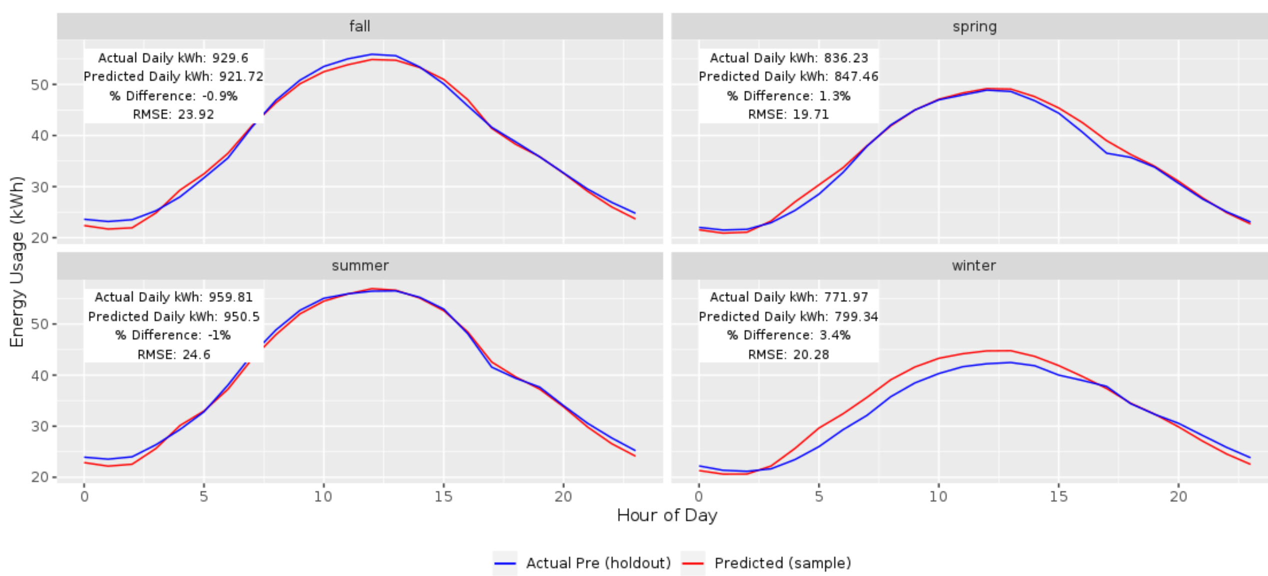


Figure 56: SCE CQI Holdout Test, by Season



Program Energy Savings

Figure 57 shows the overall average daily impacts in the post-period. The average daily model predicts savings of -0.5 percent of consumption, or an increase of approximately 11.2 kWh per day on average (4,088 kWh per year).

Figure 57: SCE CQI in Post-Period

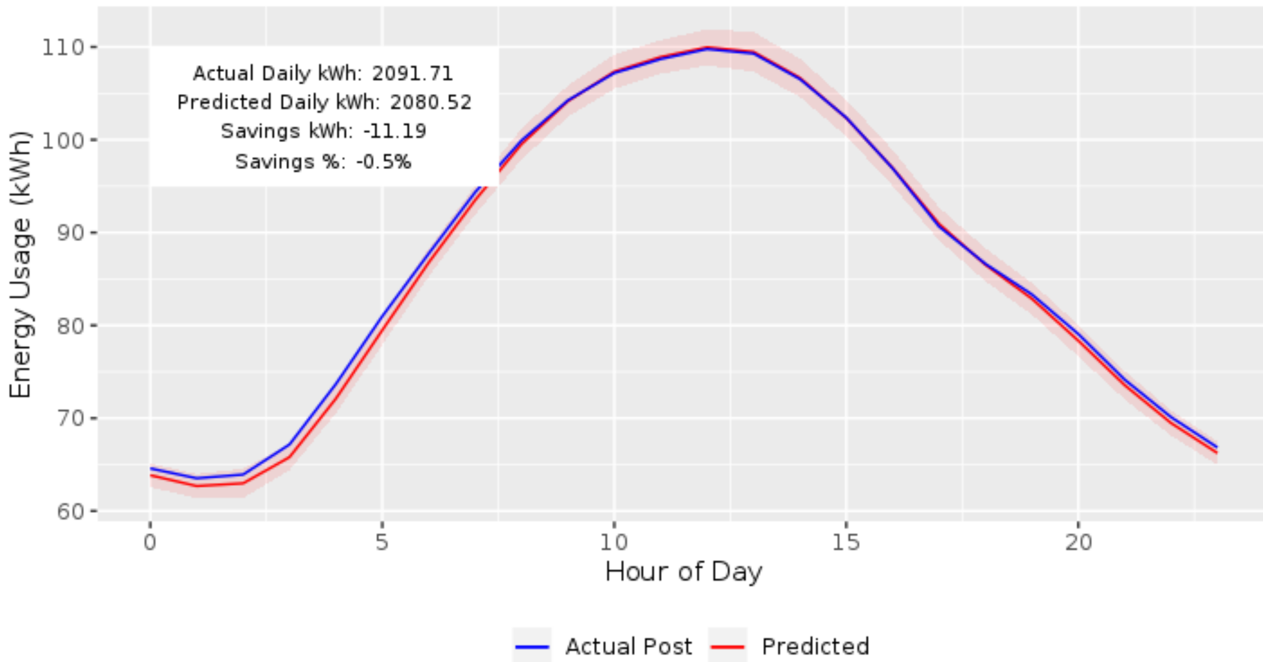


Figure 58 provides the post-period load shapes for the four most common business types. Note the variation across groups in total daily kWh energy usage, load shape, and reductions in energy usage. The largest savings were observed in retail stores, with average savings of 78 kWh per day or 28,434 kWh per year. The relatively high kWh savings estimate is reflective of the large retail customers in the CQI sample, which have an average annual consumption of 14,964 kWh. The savings estimate on a percentage basis (6.2%) appears reasonable given these customers' sizes.

Figure 58: SCE CQI in Post-Period by Business Type

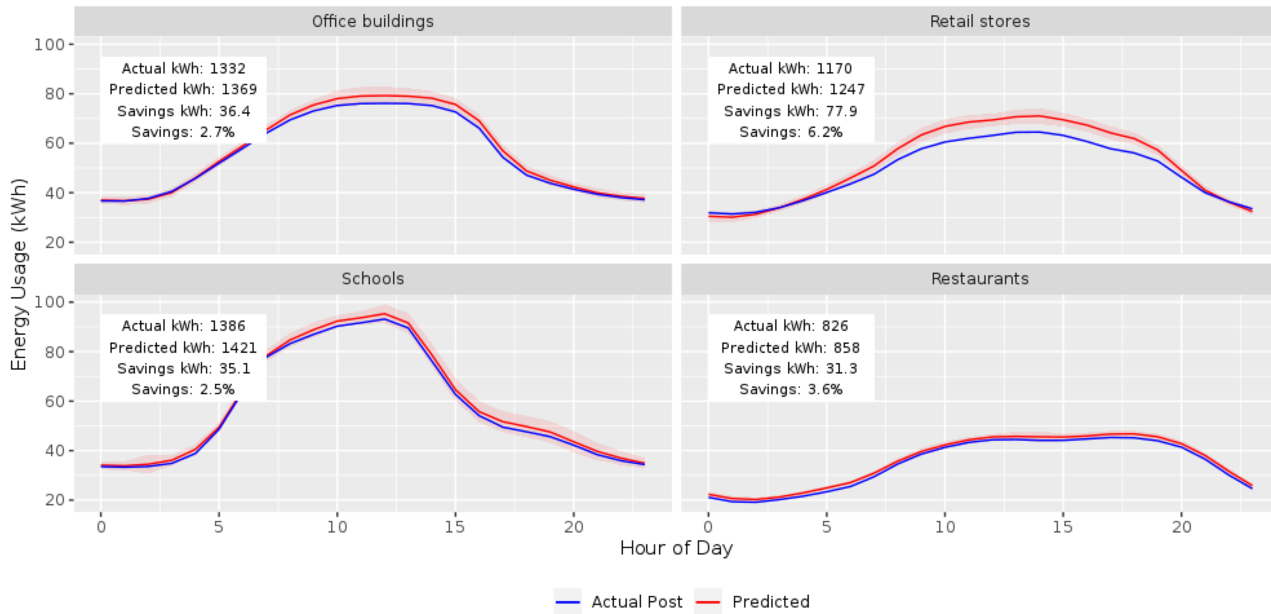


Figure 59 shows our estimated hourly kWh savings across the entire post-period, with error bars depicting 95 percent confidence intervals around each estimate, and Figure 60 shows the savings broken out by business type. Across all participants, the SCE CQI program led to insignificant energy savings during the midday from 9:00 a.m. to 5:00 p.m., but these savings were offset by increases in energy usage during the morning and evening hours. Overall, we estimate that the SCE CQI program resulted in energy savings of -11.19 kWh ± 40.16 per day (or -0.5% ± 1.9%).

The savings by business type in Figure 60 show large and statistically significant savings in retail stores and office buildings. Restaurants and schools have savings during most hours of the day, suggesting an improvement in baseline energy usage, but not all of these savings were statistically significant. Participants with other business types (e.g., industrial, warehouses, hospitals) exhibited a significant increase in energy usage during most hours.

Figure 59: SCE CQI Estimated Savings

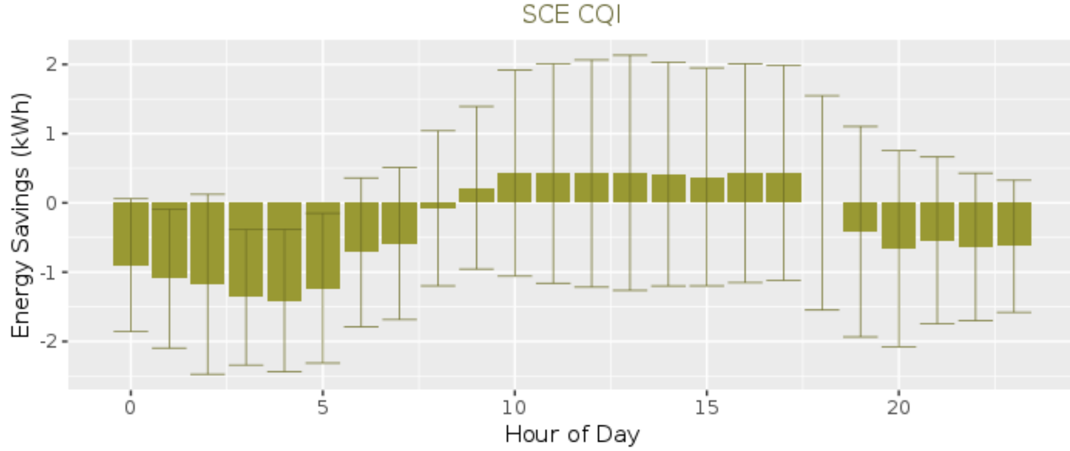
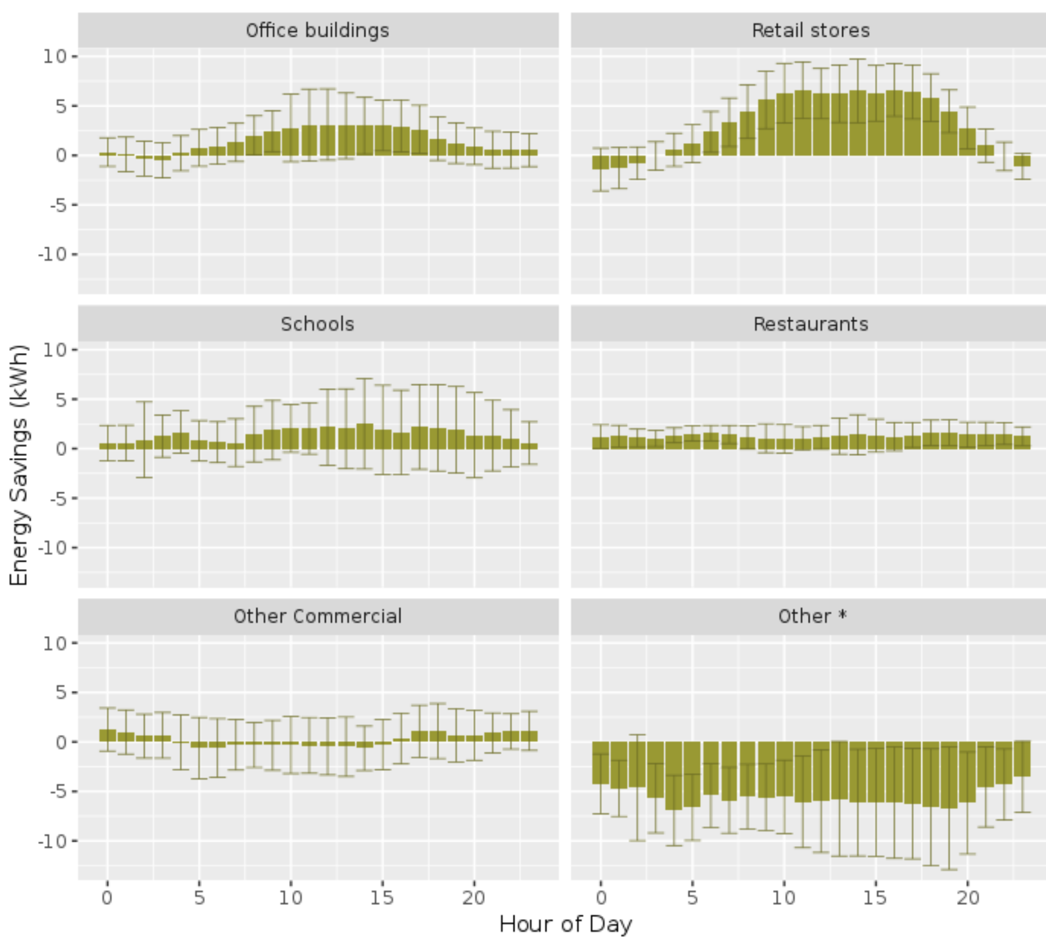


Figure 60: SCE CQI Estimated Savings by Business Type



* The "Other" category includes industrial, warehouses, builders, hospitals/medical facilities, colleges and universities, food stores and refrigerated warehouses, and other business types.

Savings by Segment

Table 10 shows the AMICS estimate of energy savings by business type among participating customers that were observed in the post-period. We see a wide range of estimated energy savings by kWh and proportion of baseline consumption (%) across business types within the SCE CQI program. This is due in part to differences in HVAC size, capacity, and efficiency relative to existing equipment, reflected in the *ex ante* gross savings (with inherent assumptions from the underlying workpapers). Five of the business types exhibited positive savings, reducing their energy usage after the program intervention. These were also the five most common business types within the CQI program, with over 100 participants in each. All other business types exhibited negative savings, increasing their energy usage in the post-period; this offset most of the positive energy savings realized by the other sectors when viewed at the program level.

Table 10: SCE CQI Gross Energy Savings by Business Type

Business Type	CQI Participants	N Sites in Post-Period	Estimated Savings		*Ex Ante* Savings	
			kWh	%	kWh	%
Office buildings	29%	290	36.4	2.7%	74.8	5.5%
Retail stores	13%	161	77.9	6.2%	30.9	2.5%
Schools	10%	129	35.1	2.5%	40.2	2.8%
Other commercial	11%	127	5.7	0.4%	36.2	2.6%
Restaurants	9%	110	31.3	3.7%	4.5	0.5%
Other industrial	7%	51	-45.9	-1.7%	10.3	0.4%
Warehouses	4%	42	-234	-8.4%	37.1	1.3%
Hospitals/medical facilities	2%	25	-575	-6.7%	44.8	0.5%
Colleges & universities	2%	23	-218	-5.0%	132	3.0%
Other	10%	166	-86.9	-2.0%	50.3	1.1%

Note: Percentages represent kWh savings as a proportion of baseline kWh consumption.

3.3.4 Metered Commercial HVAC

In addition to applying the AMICS approach to a variety of commercial programs at a whole building level, we also used HVAC sub-metering data to better understand the degree to which the AMICS approach is capable of predicting changes in HVAC usage and disaggregating HVAC usage from whole building data. As part of Evergreen Economics' analysis of SCE's Comprehensive Value Chain HVAC (CVC-HVAC) program as a High Opportunity Programs or Projects (HOPPs) offering, we demonstrated that the

AMICS model was able to estimate energy savings for individual non-residential sites while meeting the precision requirements established for NMEC projects in most cases. In addition to whole building data for these sites, we received HVAC sub-metered data for seven customers that participated in the SCE Field Data Collection Study. Since these seven customers have both whole building and HVAC sub-metered data available, they provide a unique opportunity to test how the AMICS model can estimate HVAC loads for commercial customers.

Database

All seven customers shown in Table 11 completed HVAC upgrade projects incentivized by SCE and were part of the Field Data Collection study.[46] This study provided a small sample of commercial customers with thorough on-site testing and metering of individual HVAC units, both before and after system upgrades. For this analysis, we focused exclusively on the period prior to the system upgrades. Table 11 summarizes the data we were provided with for each site. While we were provided with at least five months of pre-period whole building AMI data for each customer, only a portion of that time also had HVAC sub-metering—usually the summer months. Overall, we were able to analyze at least 100 pre-period days for each customer where there were both whole building AMI interval data and sub-metered HVAC usage data.

Table 11: Overview of Field Data Collection

Customer	Business Type	Pre-period AMI Coverage	HVAC Metering Data Overlap	N Days in Pre-Period with Data
01	Restaurant	<6 months	Apr-Aug 2016	107
02	Restaurant	<6 months	May-Aug 2016	114
03	Restaurant	<6 months	May-Aug 2016	107
04	Restaurant	<6 months	Jun-Sept 2016	109
05	Restaurant	<6 months	Jun-Sept 2016	105
06	Office	6-12 months	Jun-Oct 2016	124
07	Office	6-12 months	Aug-Dec 2016	110

Note that most of the selected customers have less than six months of pre-period AMI energy usage data. Some evaluations require 12 month of pre-period data to ensure that model predictions are based on energy usage data that cover the continuum of seasons

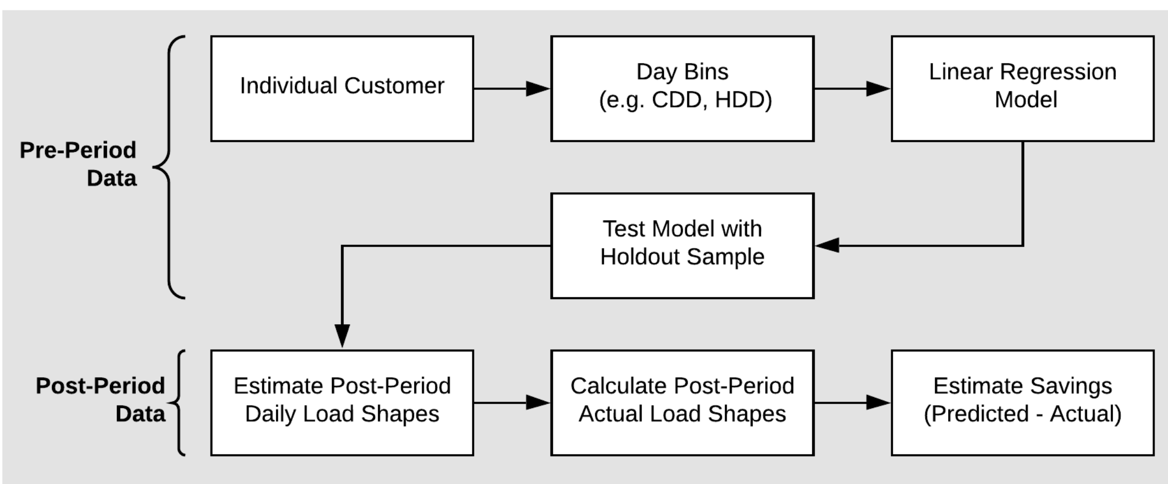
[46] National Comfort Institute Inc., Energy Solutions, and Solaris Technical LLC conducted this commercial quality installation data collection study to support SCE's program development.

and weather conditions. Additionally, because traditional billing regression only uses monthly data, a longer pre-period was needed to maximize the amount of data available for the model. With AMI data, however, a pre-period of six months is likely sufficient as long as it includes days with significant heating and cooling opportunities.

AMI Analysis Methods

As part of Evergreen's initial proof-of-concept for SCE's HOPPs CVC-HVAC pilot program, our standard AMICS modeling approach was applied to the whole building AMI data for the seven participating customers, with several minor program specific adjustments to the approach. The HOPPs CVC-HVAC pilot program offers custom HVAC retrofit measures with pay-for-performance contractor incentives. In this type of program design, it is necessary to estimate the realized savings of each *individual* participant, including some customers with limited pre-period data. Given the uniqueness of commercial customers with respect to building characteristics and economic activity, this application of the AMICS modeling approach excludes the customer segmentation component. Instead, customers are modeled individually. In this model variant, shown in Figure 61, the AMICS approach is not defining *customer segments*, but is modeling each individual customer's interval data on its own, based on characteristics of the days observed.

Figure 61: AMICS Approach



Similar to the approach for other programs, every day of the study period was binned in terms of its weather and day type. The weather bins were 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. These customer-days were assigned to a series of bins, each containing a range of five CDDs. This process was repeated to assign days to heating degree-day (HDD) bins, again using a base temperature of 65 degrees Fahrenheit. To help control for the differences in energy usage across days of the week with the same weather conditions, we binned the days by type. Weekends were assigned to day type 1, and weekdays were assigned to day type 0. To aid in our ability to make direct comparisons between HVAC and whole building data, the day bins from the whole building analysis were used to define the day bins for the HVAC data as well.

Holdout validation was conducted for the whole building model. Figure 62 compares the predicted pre-period load shape from the whole building model (red line) to the actual pre-period load shape for the holdout (blue line) for a single site and holdout sample. As was the case across multiple sites, the AMICS approach was able to effectively model whole building data from each site.

Figure 62: Example Holdout Sample in Pre-Period, Actual vs. Predicted Whole Building Usage

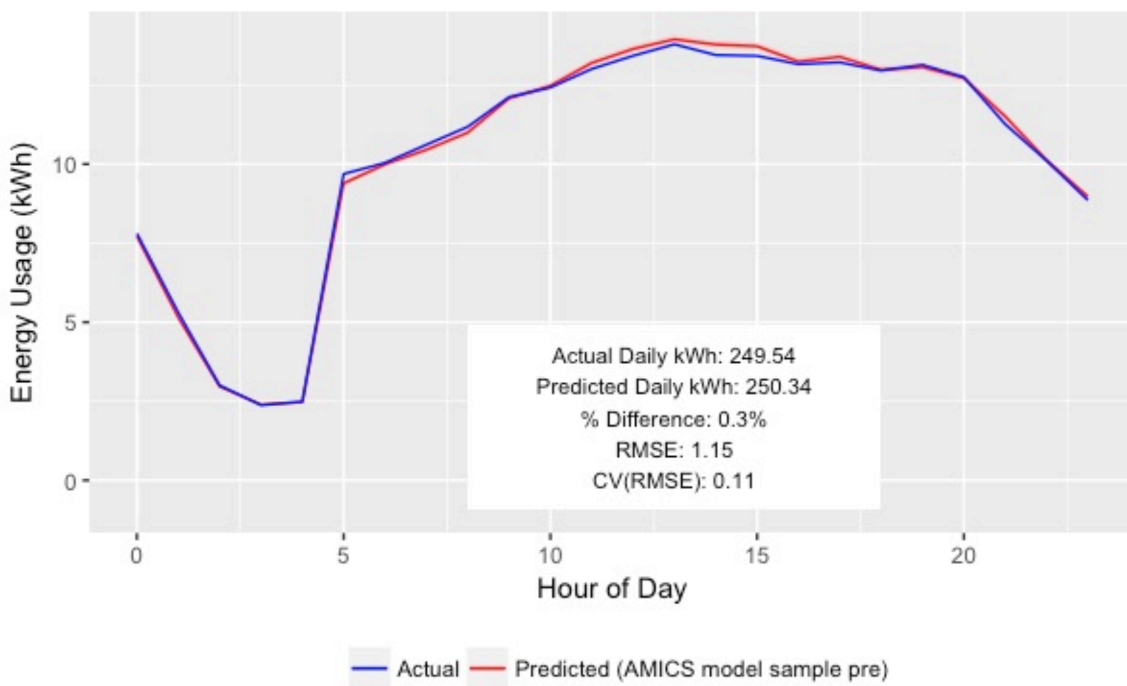


Table 12 summarizes the ability of AMICS to model whole building data at each of the seven sites for the entire pre-period using all three of the SCE NMEC error threshold metrics. The normalized mean bias error (NMBE) measures the average difference between the model prediction and actual metered energy usage. NMBE is a directional measurement; a negative NMBE indicates that the model underestimated the site's actual energy usage. 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).[47] When provided with all pre-period observations (i.e., no holdout days), our model meets the SCE NMEC error thresholds for the majority of sites in the sample.[48]

Table 12: Pre-Period Baseline Model Results

Customer	AMICS			Do Models Meet NMEC Criteria?
	NMBE	CV(RMSE)	R-Sq	
01	-9.6E-18	12%	81%	Yes
02	-3.9E-17	10%	88%	Yes
03	-7.5E-18	11%	92%	Yes
04	-3.6E-17	10%	65%	No, R^2<70%
05	-3.4E-17	9%	55%	No, R^2<70%
06	-4.7E-17	17%	92%	Yes
07	1.9E-18	16%	91%	Yes

Weather Sensitivity Analysis Methods

With the same day bins applied to the whole building and HVAC data of each customer, we compared the weather-sensitive component of whole building usage and the weather-sensitive component of HVAC usage under the hypothesis that the two would be closely related. We calculated the weather-sensitive component of usage (either whole building or HVAC) in two steps:

1. We estimated the hourly non-weather-sensitive component of usage (baseline) by determining average hourly usage on neutral days (CDD bin = 0 and HDD bin = 0). These values were calculated separately for weekdays and weekends to maintain day-type differences.

[47] The NMBE can appear near zero when overestimations are consistently balancing out underestimations to create an accurate average prediction. The CV(RMSE) does not measure direction (i.e., consistent bias), but focuses on the magnitude of the prediction error.
[48] Current SCE NMEC guidelines require NMBE<0.005%, CV(RMSE)<25%, and R-Square>70% for models with 12 months of pre-period data. When the models do not meet these error thresholds, further analysis and/or customer follow-up is suggested to improve the models prior to estimating energy savings realized in the post-period.

2. We then subtracted the average hourly baseline usage from the actual usage of each hour to determine how much of an hour's usage was weather sensitive. This calculation was completed independently for whole building data, AMICS estimates, and the HVAC sub-metered data.

Figure 63 shows an example from Customer 03 of how weather sensitivity was calculated for whole building usage at an individual site on an individual day. The black line indicates the average hourly usage across all neutral days in the pre-period as predicted by the AMICS model, i.e., baseline usage. The blue line indicates the actual usage at the site on a warm day in June 2016. The blue-shaded area is the difference between the actual usage and the average usage on all neutral days and is our estimate of weather-sensitive usage on this day for this site at the whole building level. As expected, the figure shows increased usage during peak cooling hours on this day.

Figure 63: Example of Weather-Sensitive Whole Building Calculation

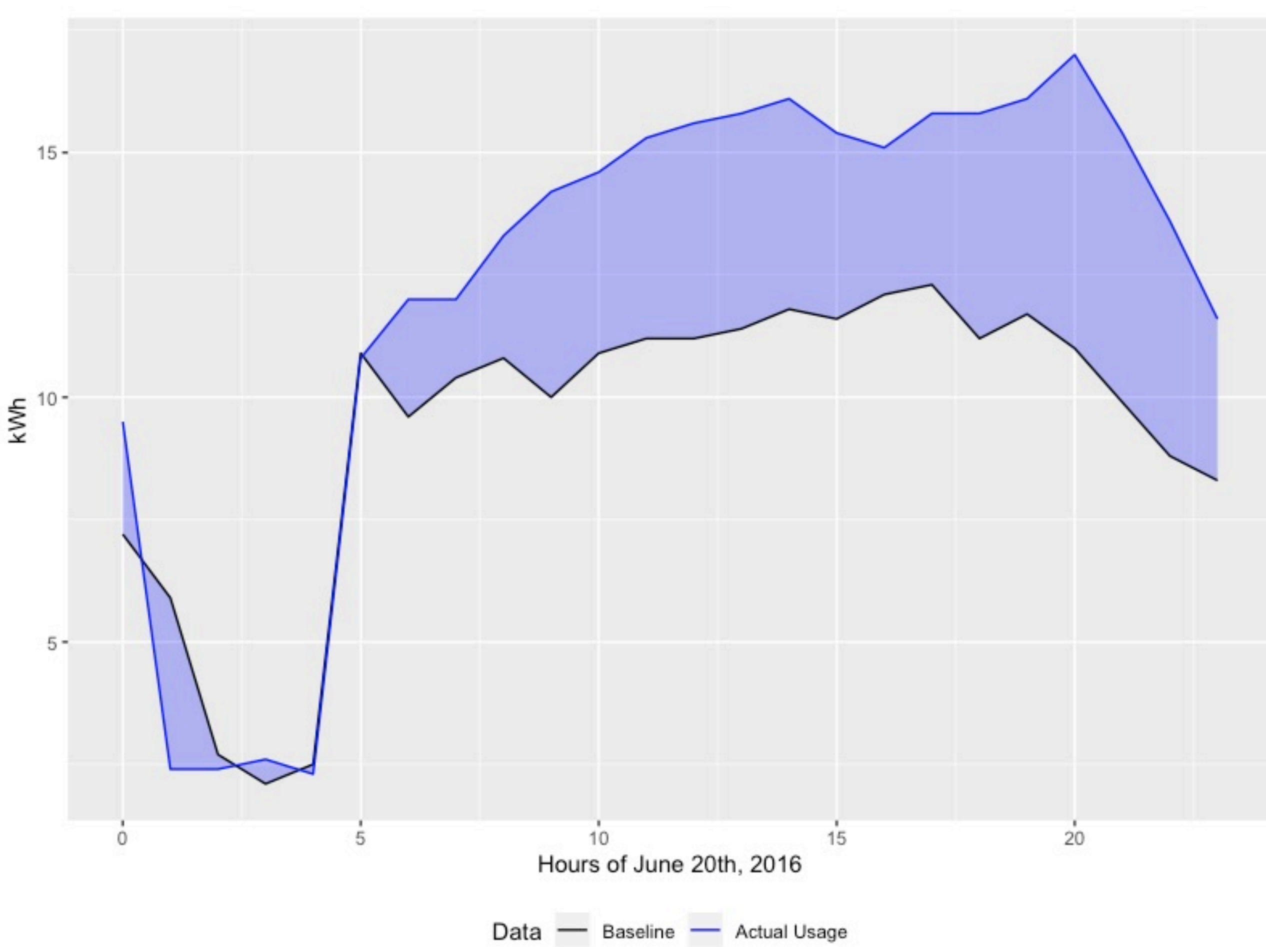


Figure 64 shows an example of how weather sensitivity was calculated for HVAC usage for the same site and the same day as shown in Figure 63. For this figure, the blue line indicates the average hourly HVAC usage across all neutral days in the pre-period, and the green line indicates the actual HVAC usage at the site on a warm day in June 2016. The green-shaded area is the difference between the actual usage and the average usage on all neutral days and is our estimate of weather-sensitive HVAC usage on this day for this site.

This calculation was repeated for every site and for every hour with HVAC data to create hourly weather-sensitive estimates. From these estimates, we created site-specific average hourly weather-sensitive load shapes. Using NMBE and CV(RMSE), we compared these load shapes to evaluate how accurately the weather-sensitive component of whole building usage predicted weather-sensitive HVAC usage.

Figure 64: Example of Weather-Sensitive HVAC Calculation

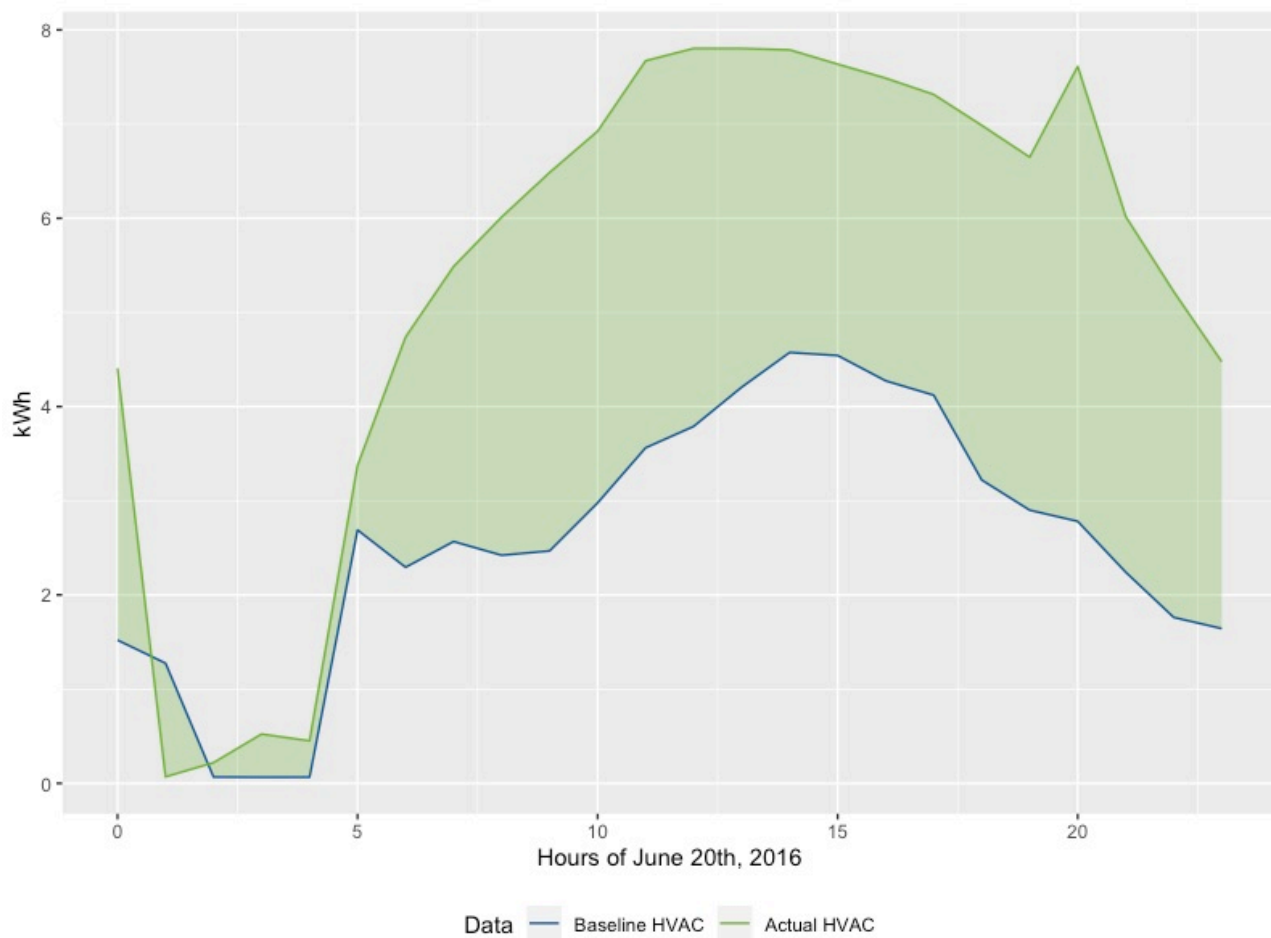


Figure 64 also shows that HVAC usage has a non-zero baseline; that is, even on temperate days, the HVAC system is still used. This non-weather-sensitive HVAC usage (non-WS HVAC) is critical to this analysis, because it prevents us from being able to directly estimate HVAC usage from whole building usage. This is because the model includes the HVAC baseline usage in the whole building baseline calculation. For example, there is no way of disaggregating the HVAC baseline in Figure 64 from the whole building baseline in Figure 63 without HVAC sub-metering because there are no days (regardless of weather) that the HVAC is not in use. However, this does not prevent us from evaluating how weather-sensitive changes in HVAC usage (WS HVAC) compare to weather-sensitive changes in whole building usage.

Weather Sensitivity Findings

Given that AMICS was effective at simulating whole building load shapes, the capability of the model to account for weather sensitivity can be evaluated with the goal of understanding how well the whole building model can account for weather-sensitive changes in HVAC usage. The results from this analysis can help determine on a site-by-site basis the degree to which AMICS is capable of disaggregating HVAC usage from whole building data.

Figure 65 shows an example comparison between weather-sensitive changes in HVAC usage and AMICS-predicted weather-sensitive changes in whole building usage for a single day. In this figure, the blue line corresponds to the blue shaded area in Figure 64 and represents the weather-sensitive change in HVAC for Customer 03 on a warm day in June 2016. Similar to the calculation in Figure 63, the red line represents the expected weather-sensitive change in whole building usage based on AMICS predictions for the same day. For this particular day for this particular customer, weather-sensitive changes in HVAC energy usage mapped closely with weather-sensitive changes in AMICS predicted usage.

Figure 65: Example of Weather-Sensitive Comparison on an Individual Day

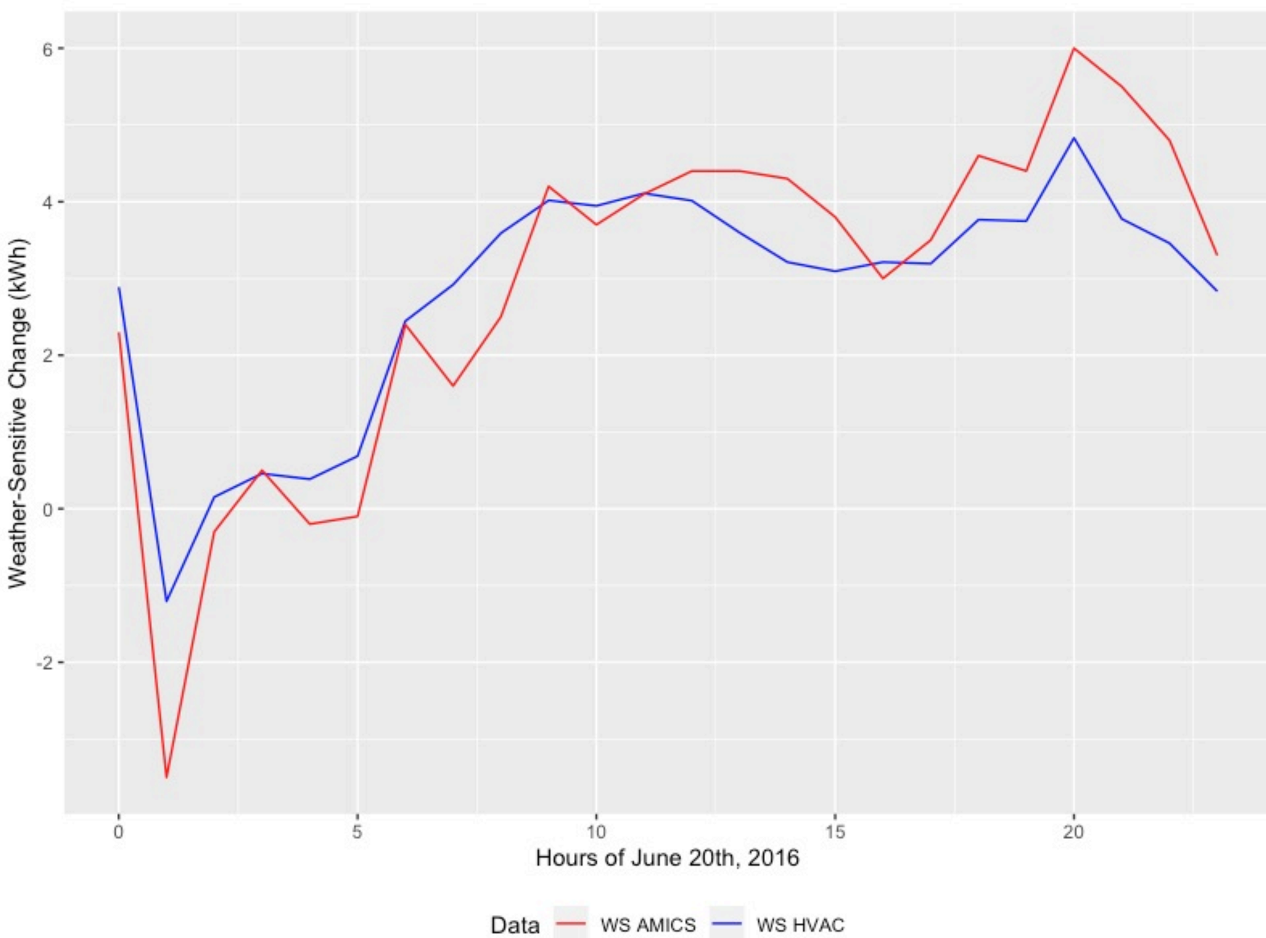


Figure 66 shows an example, from Customer 03, of how weather-sensitive changes in HVAC usage compared to weather-sensitive changes in the AMICS-modeled usage on average across the entire period with HVAC data. While weather-sensitive changes in modeled whole building usage do not perfectly align with weather-sensitive changes in HVAC usage, the general pattern of peak increases in usage during peak cooling hours is shared between both datasets. Weather-sensitive changes in whole building usage not attributed to weather-sensitive changes in HVAC usage could be attributed to other changes in customer usage that are correlated with weather. For example, given that Customer 03 is a fast food restaurant, the peak hour increases in weather-sensitive whole building usage could be attributed to increased business during those hours on warmer days.

Figure 66: Example of Weather-Sensitive Comparison on Average Day

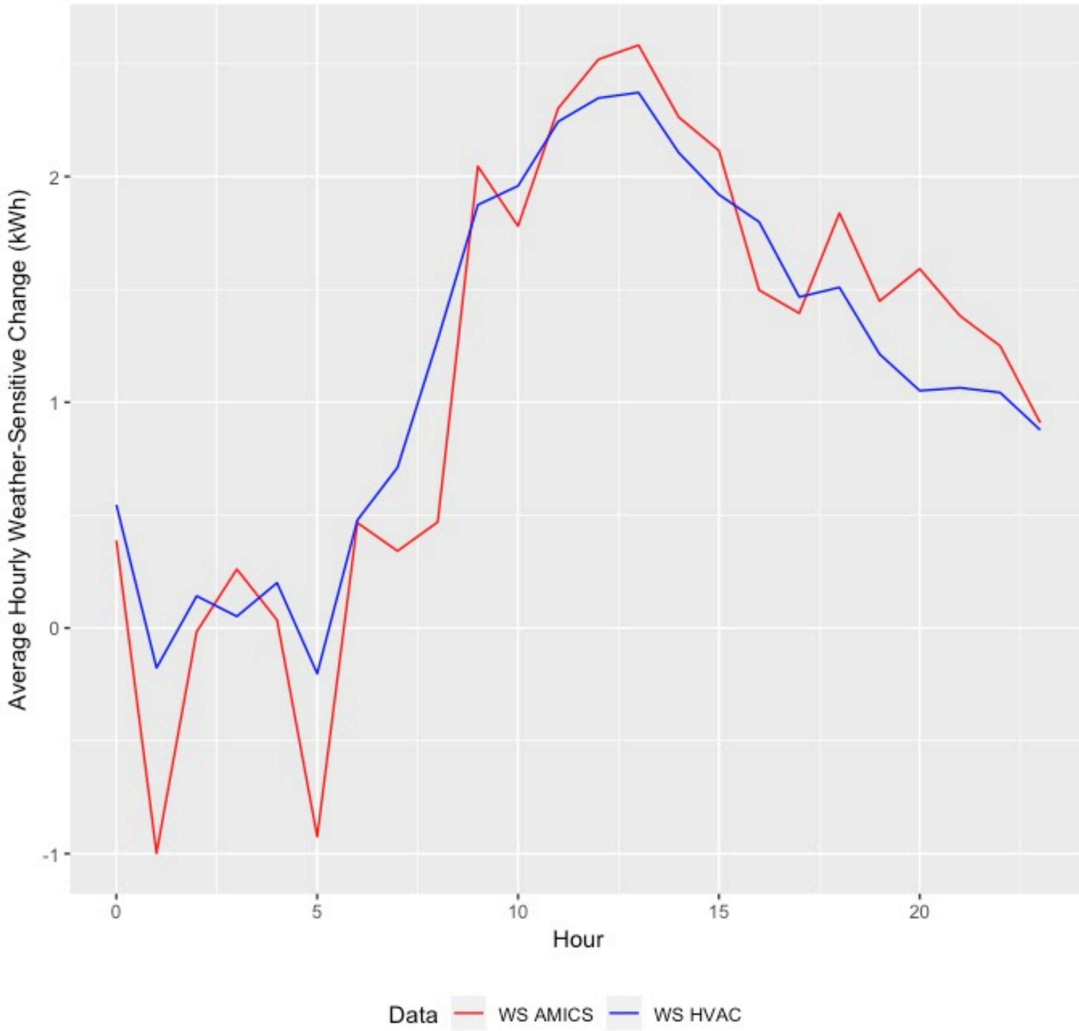


Table 13 summarizes the ability of the AMICS model to predict weather-sensitive changes in HVAC usage across all sites. For these metrics, the weather-sensitive component of HVAC usage (actual HVAC usage minus average HVAC usage on neutral days) was compared to the weather-sensitive component of the whole building AMICS model (modeled usage minus average usage on neutral days).

Comparing HVAC usage and whole building usage can generally be expected to have a higher level of error than when comparing modeled whole building usage against actual whole building usage. One reason is that variation in non-HVAC usage (which is not being metered independently) may be correlated with weather. This weather-sensitive non-HVAC usage (WS non-HVAC) shows up in the weather-sensitive whole building estimates, but is not caused by weather-sensitive changes in HVAC usage, leading to error.

As a result, the expected error in a successful model may be higher than in a model that only uses whole building data. While a successful whole building model is expected to have an NMBE of +/-5 percent and a CV(RMSE) of 25 percent, a comparison of weather-sensitive HVAC usage and weather-sensitive whole building usage will have higher error.

The predictability of weather-sensitive HVAC usage varied widely between sites. In particular, Customer 06 and Customer 07 had very high error, while Customer 01, Customer 03, and Customer 05 had relatively low error given the datasets being compared. The causes of variation in error between sites, especially the cases of Customer 06 and Customer 07, will be discussed in more depth later, but outside of those sites, changes in HVAC usage were generally predictable from whole building usage, especially in terms of kWh.

Table 13: AMICS Capability of Disaggregating HVAC from Whole Building

Customer	WS HVAC	WS AMICS	NMBE	CV(RMSE)
01	0.55	0.50	-9%	98%
02	0.78	0.88	14%	101%
03	1.16	1.12	-3%	61%
04	0.53	0.60	13%	63%
05	0.22	0.22	0%	97%
06	0.37	0.59	59%	114%
07	0.30	1.00	239%	526%

As described previously, one possible source of error is changes in whole building usage that are correlated with weather but are not caused by changes in HVAC usage. Leveraging the whole building usage data, as well as the HVAC sub-metering, we can impute these weather-sensitive non-HVAC usage levels on a site-by-site basis. Figure 67 shows an example of the average hourly usage components for Customer 03. For reference, in this figure:

- WS HVAC (green) corresponds to WS HVAC in Figure 66,
- Baseline non-HVAC (light blue) corresponds to Baseline in Figure 63,
- And, Baseline HVAC (dark blue) corresponds to Baseline HVAC in Figure 64.

The remaining component, WS non-HVAC, is imputed as the total average hourly actual usage minus the previous three components. While WS non-HVAC is generally the smallest component of usage, its size (especially relative to WS HVAC) can cause errors when using whole building data to estimate weather-sensitive changes in HVAC. This is because all weather-sensitive changes in whole building usage are assumed to be caused by changes in HVAC, when, in fact, some are not.

Figure 67: Example of Usage Components

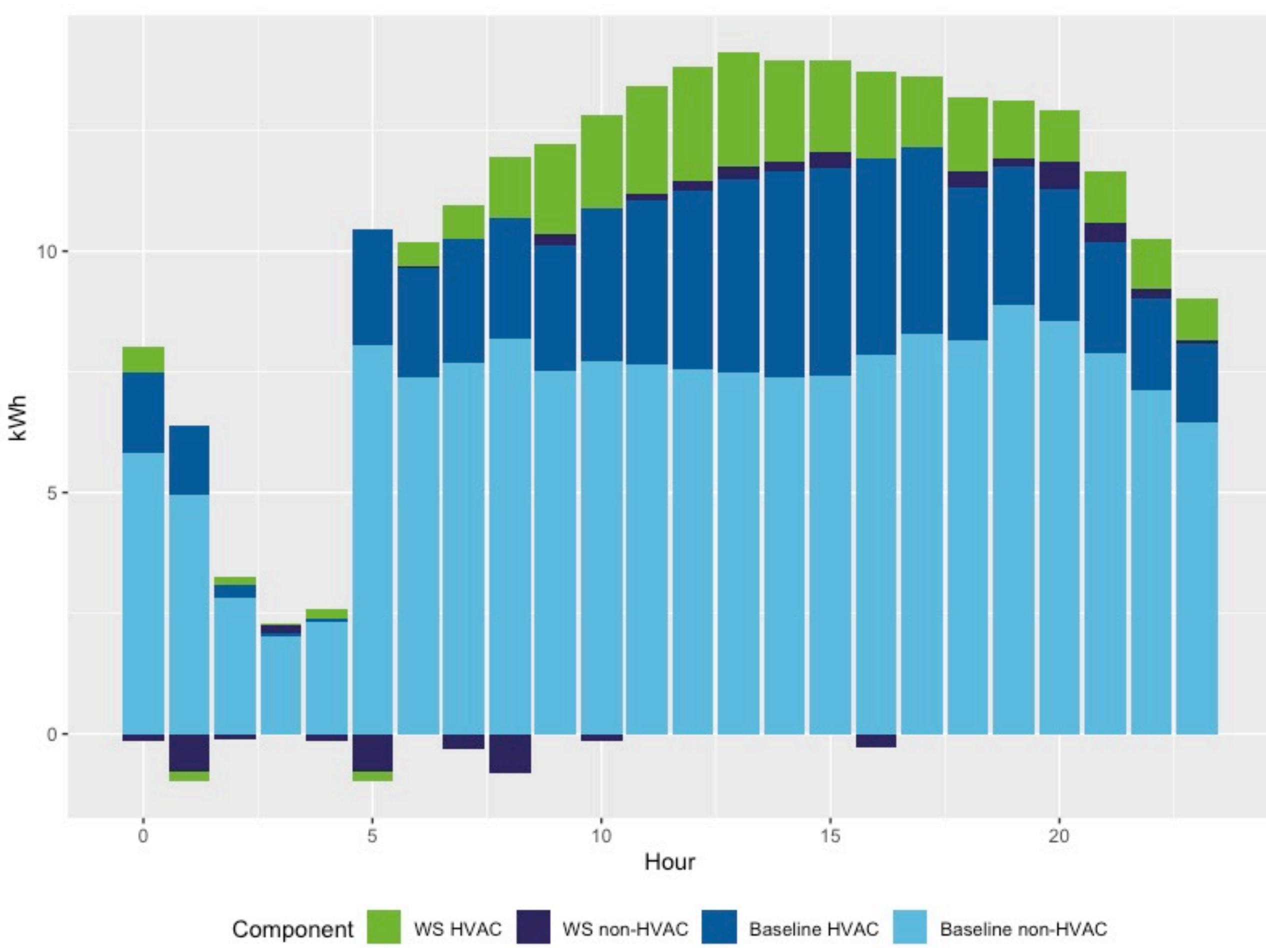


Table 14 summarizes the usage components on an average daily basis at each site. WS non-HVAC varied between sites, and the weather-sensitive AMICS estimates tended to perform better (especially in terms of NMBE) at sites where WS non-HVAC was small relative to WS HVAC. In particular, Customer 07, which performed the worst, had the largest average WS non-HVAC value. In fact, based on the available data, weather-sensitive changes in non-HVAC usage were actually larger than weather-sensitive changes in HVAC for Customer 07.

Table 14: Summary of Usage Components

Customer	Baseline non-HVAC	Baseline HVAC	WS HVAC	WS non-HVAC	WS non-HVAC as % of WS HVAC
01	9.79	3.06	0.55	0.06	11%
02	9.25	2.96	0.78	0.26	33%
03	9.43	2.55	1.16	-0.01	-1%
04	5.50	1.03	0.53	0.12	23%
05	6.05	1.41	0.22	0.05	23%
06	4.17	1.10	0.37	0.28	76%
07	15.07	1.88	0.30	0.37	123%

Conclusions

While sub-metered HVAC will always be the most effective measurement of HVAC load, we have demonstrated that the AMICS approach can estimate weather-sensitive changes in HVAC usage under certain circumstances. For specific customers in this study, like Customer 03, the AMICS approach created meaningful baseline usage estimates, from which we were able to make relatively accurate predictions of weather-sensitive changes in HVAC usage.

Table 15 summarizes the ability of AMICS to estimate weather-sensitive changes in HVAC usage in addition to some key site characteristics. The best predictor of our ability to successfully model the impact of weather on HVAC usage was the amount of weather-sensitive non-HVAC usage at a site. While the most successful sites had a moderate amount of weather-sensitive non-HVAC usage, the two sites with the highest model error, Customer 06 and Customer 07, had weather-sensitive non-HVAC usage that was nearly as great or greater than their weather-sensitive HVAC usage. While both of these sites are offices, another factor correlated with weather-sensitive non-HVAC usage was the amount of conditioned square feet per HVAC ton at each site. Therefore, a possible predictor of weather-sensitive non-HVAC usage (and by extension our ability to predict weather-sensitive changes in HVAC usage), is the size of business space relative to the size of HVAC used to cool that space (i.e., ratio of conditioned square-footage to HVAC tons). That is, the amount of weather-sensitive non-HVAC usage that can occur is greater when the amount of HVAC is small relative to the floor space of the business.

Table 15: Summary of Characteristics and Results

Customer	Business Type	HVAC Data Season	WS non-HVAC as % of WS HVAC	Sq ft per HVAC Ton	NMBE	CV(RMSE)
01	Restaurant	Summer	11%	90	-9%	98%
02	Restaurant	Summer	33%	140	14%	101%
03	Restaurant	Summer	-1%	160	-3%	61%
04	Restaurant	Summer	23%	100	13%	63%
05	Restaurant	Summer	23%	96	0%	97%
06	Office	Summer	76%	210	59%	114%
07	Office	Fall	123%	318	239%	526%

While using the AMICS approach to estimate weather-sensitive changes in HVAC usage has been effective in some cases, especially for certain types of commercial sites, further research is needed to evaluate the accuracy of these predictions prior to use as a baseline. One particular area of research that would improve this approach is developing a better understanding of the causes of weather-sensitive changes in non-HVAC usage (WS non-HVAC). This might include additional study on the load of typical commercial end-uses, particularly in office buildings, and how these loads react to changes in weather. Generally, a better understanding of the causes of WS non-HVAC and how to predict it would make for stronger estimates of weather-sensitive HVAC usage when using whole building data.

Another area for further research is HVAC baseline usage. Our ability to predict total HVAC usage from whole building usage was limited by the existence of non-zero baseline HVAC usage (non-WS HVAC), where the power draw can not be explained by a need for heating or cooling. If additional HVAC submetering data becomes available, we suggest that researchers examine the relationship between commercial HVAC load shape and weather. Another method for estimating HVAC baseline usage would be to combine whole building AMI data with data associated with the HVAC controls such as HVAC runtime, temperature set points, and indoor temperature readings. For example, the difference in whole building usage between when indoor air temperature is at a relative minimum compared to when indoor air temperature is at a relative maximum could be used to predict the minimum/baseline amount of HVAC usage. This information, combined with estimates of weather-sensitive changes in HVAC (as discussed in this section), could be used to create total HVAC usage estimates without any HVAC sub-metering. While theoretically possible, further study comparing actual HVAC baseline usage and predicted HVAC baseline usage would be required.

4 Conclusions

This research has demonstrated that the AMICS model performs well when predicting hourly energy usage across the population of participants in the residential and commercial programs we analyzed. While the AMICS model was not always able to detect statistically significant savings at the program level, our customer segmentation was able to identify a subset of customers (by baseline usage or industry) with significant savings that were hidden in aggregate, suggesting a need for targeting.

Below, we present the high level conclusions as they relate to the key research objectives, which are provided in bold.

1. **Refine the residential billing analysis methods using data from the same HVAC programs we examined in Phase I.**

In Phase I, we segmented customers by their baseline electricity consumption after controlling for weather using a simple fixed effects model. In Phase II, we refined our approach, segmenting customers with a combination of their average daily usage (kWh) and load shape (hours of use). We used *k*-means clustering to identify groups of customers with similar load shapes automatically from the AMI data, rather than relying on customer characteristics that are not typically tracked (or not regularly updated) in the utility databases. Customers with similar energy usage on the average day can have drastically different load shapes. The load shape clusters help account for the remaining differences in occupant schedules, energy-intensive equipment, peak demand hours, and other factors.

The holdout tests for both residential HVAC programs (PG&E's Quality Maintenance and SCE's Quality Installation programs) demonstrate that the AMICS model is able to produce accurate estimates of load shapes for participants of residential HVAC programs, accounting for the variation in load shapes across all four seasons. This confirms findings from the Phase I research showing that the AMICS model performs well with residential customers.

The AMICS model detected statistically significant savings for SCE's QI program, consistent with our expectations by season and time-of-day for improved air conditioning efficiency. PG&E's QM program had a larger population of nearly 30,000 participants but very small *ex ante* savings (<5%) that can be difficult to detect with billing analysis. The QM program savings estimated by our model were not statistically significant during most hours, despite the tight error bounds around our predictions. However, the AMICS segmentation revealed a wide variation in energy savings across customer segments and weather conditions, with more substantial energy savings being realized by high energy users on days with low to moderate cooling loads. This information could be used by the IOUs to target similar customers for future program participation.

Key Findings:

- The AMICS model is able to produce accurate load shape predictions for residential HVAC participants.
- The estimated savings for SCE's QI program were consistent with our expectations by season and time-of-day for improved air conditioning efficiency.
- The AMICS segmentation of PG&E's QM program revealed that participants who were high energy users in the baseline period realized significant energy savings from the program intervention.

2. Explore using the AMICS model to estimate savings for the Home Energy Reports (HERs) program—including the control group of non-participant residential customers.

The holdout test provided evidence that the AMICS model is able to accurately predict the hourly load of both the treatment and control groups within 1 percent across all four seasons.

Overall, we estimated that the average energy savings attributable to the Gamma Wave of PG&E’s Home Energy Reports was 0.12 kWh per day, or 0.6 percent. This estimate is lower than prior evaluations, which attributed 1.2 to 1.9 percent savings to the HERs program in the Gamma Wave; though the difference was not statistically significant. The AMICS model contributes three key benefits for the HERs program:

1. **Hourly intervals.** The existing evaluations relied on energy usage data in monthly or daily intervals. We used hourly intervals to provide more information about the hours when savings occur. Smaller time intervals can provide more information, but this comes at the cost of increasing random noise, for which the model must account.
2. **Segmentation improves matching.** In some cases, randomized group assignment is not sufficient to produce balanced samples with similar energy usage patterns in the baseline period. Our customer segmentation in the baseline period identifies and groups customers with similar load shapes, seasonality, and climate prior to any change in the program treatment. Performing difference-of-differences calculations within each customer segment improves the validity of our comparisons, focusing on the impact of the program treatment.

3. **Ease of distributional impact analysis.** The AMICS modeling approach creates separate model predictions and estimated post-period changes (i.e., energy savings) for each customer segment simultaneously. We do not simply provide the average treatment effect; instead, we expose the variation in program impacts across participants associated with key differences in the characteristics and energy usage patterns of these customers in the baseline period.

Key Findings:

- The AMICS model was able produce accurate load shape predictions for households in the HERs treatment and control groups.
- We found evidence of energy savings realized by the HERs treatment group above and beyond the natural changes observed in the control group, but these savings were not statistically significant at the program level.

3. Adapt the AMICS model as a potential tool to evaluate commercial and industrial HVAC programs, and assess the AMICS model's potential capabilities for analyzing High Opportunity Programs and Projects, in regard to implementation of Assembly Bill (AB) 802.

When constructing customer segments for commercial and industrial customers, we discovered that it was necessary to consider the business type, not just energy usage. For instance, a large office building and elementary school may have similar energy usage (kWh), operating hours, and peak energy usage; however, they will still differ in their seasonality and distribution of energy usage by end use (e.g., HVAC, cooking). Further categorizing customers by their NAICS code, building type, or utility segment improved the prediction error for the holdout sample and led to tighter error bounds around our estimates.

The holdout tests for each program demonstrated that the AMICS model is able to produce reasonable estimates of load shapes for participants of commercial HVAC programs, with predictions within 1 percent of the actual usage of the holdout sample. The AMICS model detected statistically significant savings for PG&E's Air Care Plus program, consistent with our expectations by season and time-of-day for improved air conditioning efficiency. The Commercial Quality Maintenance and Quality Installation program savings estimated by our model were not statistically significant during most hours at the program level, despite the tight error bounds around our predictions.

In each of the commercial programs, the AMICS segmentation revealed a wide variation in energy savings across customer segments. We found consistent energy savings attributed to HVAC interventions for participants in the retail sector, but these were offset (at least in part) by increases in energy usage attributed to participants in the manufacturing and health sectors. These findings suggest that the commercial HVAC programs could benefit from improved targeting.

Key Findings:

- The AMICS model was able produce accurate load shape predictions for participants in each of the commercial HVAC programs.
- All of the commercial HVAC programs could benefit from improved targeting by business type.

5 Appendix

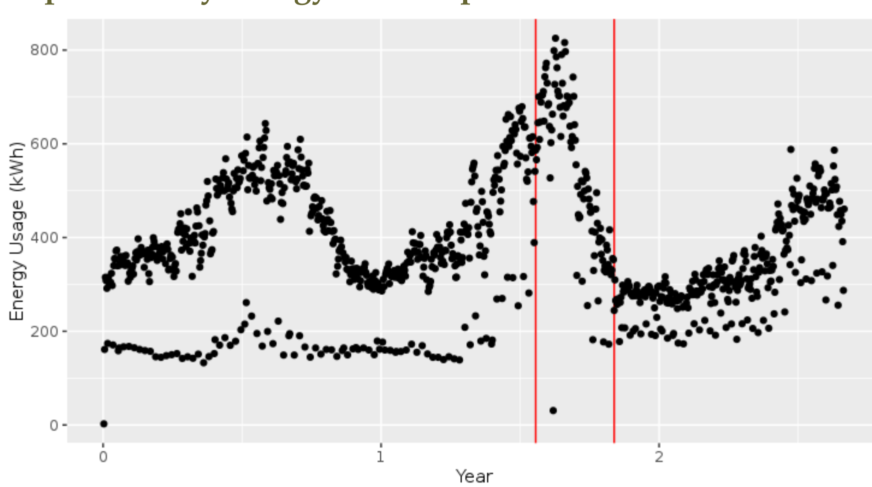
5.1 Non-Routine Events

During the analysis, we identified some commercial customers with substantial changes in energy usage that occurred outside the program retrofit period. In some cases, these could be explained by tenant turnover (in buildings with leased space), or on-site construction or other major renovations. These types of non-routine events (NREs) can obscure energy savings derived from a meter-based approach if these events are not identified, confirmed, and measured (or estimated), and if model predictions are not adjusted in the baseline period.

Motivated (in part) by the introduction of site-level normalized metered energy consumption (NMEC) programs, the evaluation community is currently working on developing rigorous and transparent approaches for NRE detection and adjustment. The following figures provide some examples of daily kWh energy usage in individual buildings that were screened during the creation of the participant pool for the billing analysis. In each figure, the red lines indicate the first and last program intervention date on record. The days to the left of the red line are the pre-period, between the lines is the installation/intervention period (if applicable), and days to the right are the post-period.

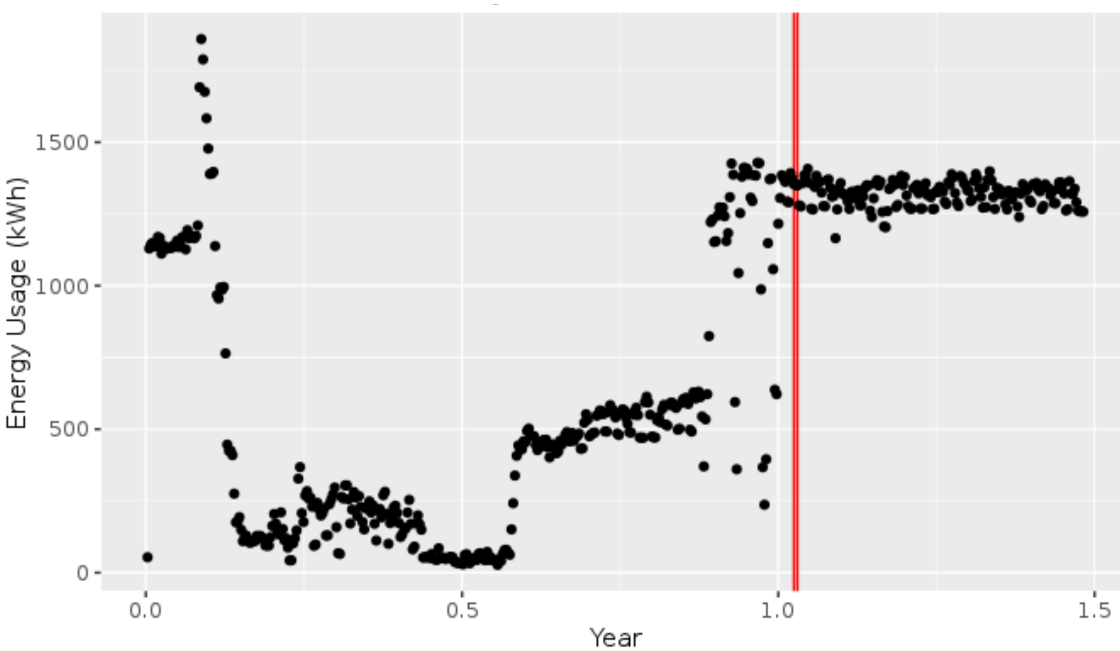
Figure 68 shows a well-behaved building, with consistent patterns in energy usage across days and fluctuations across months that appear consistent with seasonality during the baseline period, a short term increase in energy usage during the program intervention, and then some minor but fairly consistent changes in the patterns of energy usage maintained throughout the post-period.

Figure 68: Example of Daily Energy Consumption Over Time with a Seasonal Baseline



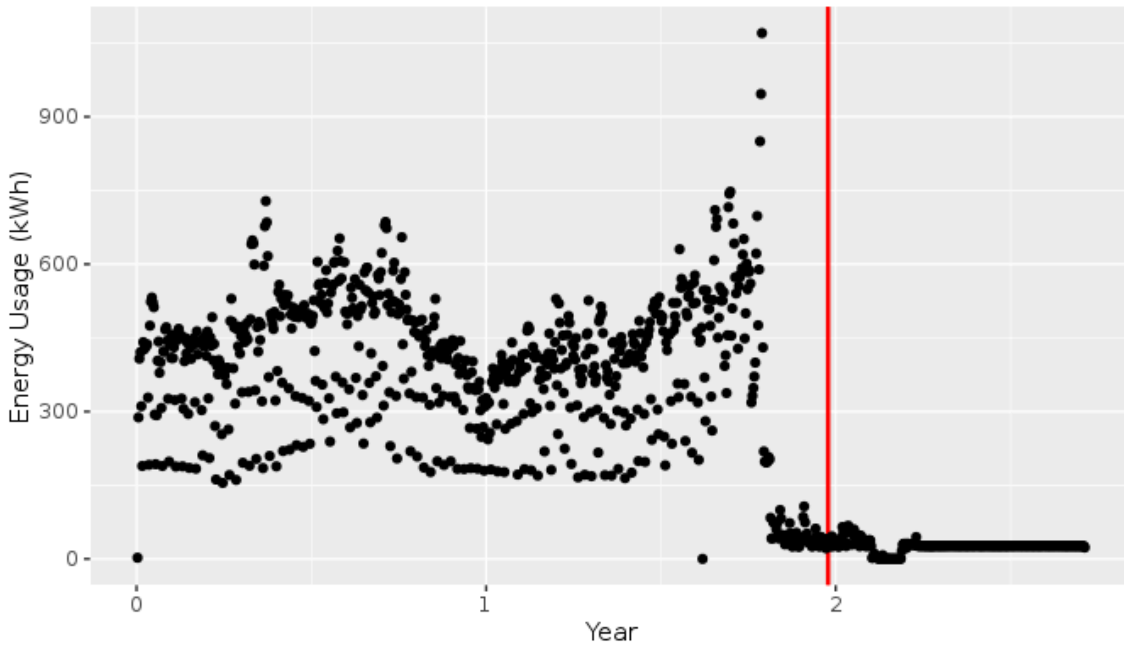
In Figure 69, we see a customer with relatively stable energy consumption that shifts to a different tier of energy usage every five months or so. A model relying on this full year of pre-period energy usage would provide an unrealistically low prediction for this building in the post-period, likely underestimating energy savings attributable to the program. This pattern has appeared in large office and industrial buildings with multiple tenants, where not all units are consistently occupied. This building's inconsistent energy usage would pose unique challenges for pre-post billing analysis (thereby disqualifying them from an NMEC program) unless additional data could be collected to control for changes in building occupancy and operating hours.

Figure 69: Example of Daily Energy Consumption Over Time With Tiered Baseline



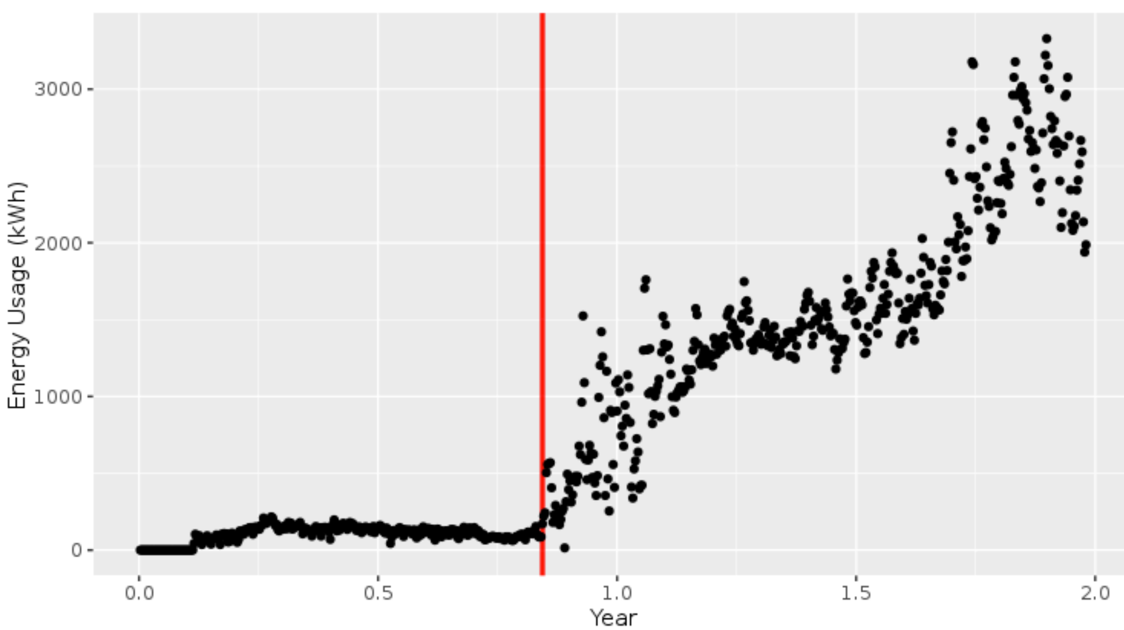
In Figure 70, we see a consistent range of kWh energy usage for one and a half years in the pre-period. However, a few months before the program intervention, there is a dramatic drop in energy usage that is maintained throughout the post-period. It appears that there were additional changes to the building operation prior to program participation. In this case, a model of the pre-period would provide an unrealistically high prediction for this building in the post-period, likely overestimating energy savings, if additional data are not collected. Unlike the previous example, this type of issue cannot be avoided with pre-period screening of program applicants (i.e., analysis to confirm consistency in energy savings prior to participation); *ex post* data collection would be necessary to identify, explain, and then correct for this event.

Figure 70: Example of Daily Energy Consumption Over Time With Sudden Drop



In this last example, Figure 71, we see a dramatic increase in energy usage. It is likely that this building was newly constructed or not fully occupied until after the program intervention date.

Figure 71: Example of Daily Energy Consumption Over Time With Increase



Even if the cause of inconsistent energy usage could be determined from the energy usage data alone, adjusting the model predictions to account for these changes in individual sites was not within the scope of this study. We applied filters to exclude customers with extreme changes from the pre- to post-period. Additional research is needed to develop algorithms for consistent NRE detection and adjustment.

5.2 Related AMICS Publications

This section provides a brief summary of additional research, program evaluations, and conference proceedings that have been published in the past few years related to the development and applications of the AMICS approach.

5.2.1 AMI Billing Regression Study (AMI Phase I)

Report prepared for Southern California Edison on behalf of SCE, PG&E, SDG&E, and SoCalGas on February 23, 2016. Calmac study ID: SCE0383.01.[49]

Southern California Edison (on behalf of the four California investor-owned utilities [IOUs]) hired Evergreen Economics to conduct a study of how traditional billing regression analysis tools could be adapted for use with advanced metering infrastructure (AMI) data. Correctly understanding and leveraging the great wealth of information provided by AMI data (in addition to developing methods for systematically processing very large amounts of customer billing data) can revolutionize how energy efficiency programs are evaluated.

This study presented a new approach—the AMI Customer Segmentation (AMICS) model—that allows savings estimates to be tailored more closely to individual customer characteristics. This is accomplished by first grouping customer consumption data into different categories based on energy use and weather conditions. Separate billing regression models (patterned after the random coefficients model specification) are then estimated for each usage/weather category, which allows for separate load shape predictions for very specific customer types.

In this study, the AMICS model specification was tested using data from two HVAC efficiency programs in California: Southern California Edison's HVAC Quality Installation (QI) program and Pacific Gas and Electric's HVAC Quality Maintenance (QM) program. Both of these programs had samples of over 1,000 customers and involved analyzing AMI billing data in 1-hour increments covering multiple years.

[49] http://calmac.org/publications/AMI_Report_Volume_1_FINAL.pdf

Using the AMI data from both programs, average daily load shapes were calculated for specific day types (weekday, weekend, seasonal) and used to estimate energy savings. When estimated load shapes were compared against a holdout sample of customers, the AMICS model performed extremely well; load shape predictions were within 1 percent of the actual load shapes for the holdout sample. Energy savings estimates for these programs ranged from 4 percent to 7 percent of annual energy use for the QM and QI programs respectively, which was consistent with the original savings expectations for these programs. Most of the savings occurred during peak hours and in summer months (as would be expected), which provided additional support for the model specification.

The AMICS modeling approach was also used to estimate the HVAC load using the Residential Building Stock Assessment Metering (RBSAM) dataset, as this was the only dataset available that contained both whole house and HVAC metered data. There was a small sample of homes (n=61 for homes with central heating or cooling) within the RBSAM that could be used to test how well the model could predict just the HVAC end use. Using this sample, the AMICS prediction was within about 1 percent of actual HVAC load on a daily basis.

In addition to producing accurate baseline models and impact estimates, the automated categorization and AMICS modeling processes developed by Evergreen allow for separate savings estimates and load shapes to be developed easily for a variety of different situations (e.g., time of day, day types, and seasons), rather than providing a simple average annual savings estimate. The AMICS model also provides an opportunity to develop customer-specific predictions of energy use and potential savings associated with various efficiency programs, thus empowering utilities to target the most beneficial programs to each customer.

5.2.2 A Smart Approach to Analyzing Smart Meter Data

Paper presented at the 2016 American Council for an Energy-Efficient Economy (ACEEE) Summer Study on Energy Efficiency in Buildings in Pacific Grove, CA.[50]

The wealth of information contained within AMI data offers great promise to utilities in designing and understanding the impacts of energy efficiency and demand-side management programs. At the same time, fully capturing the information contained within AMI data is challenging due to the sheer volume of data.

In this paper, we discussed some of the methods we employed for a recently completed research project for the California IOUs in which we employed a random coefficients model to estimate more than 1,000 unique load shapes, each representing one of 20

[50] https://aceee.org/files/proceedings/2016/data/papers/12_626.pdf

different home-bins on one of more than 50 different combinations of cooling degree-days (CDD) and heating degree-days (HDD). Unlike standard methods of regression analysis, which fit a single line through a scatter of data, the random coefficients model fits a unique regression line to each load shape while simultaneously accounting for correlations in energy use across all load shapes.

We began this paper by discussing the abundance of data generated by AMI—are all these data too much of a good thing? We then briefly discussed the fixed-effects model for estimating a billing regression and then presented the random coefficient model. We concluded with a discussion of potential applications for the random coefficient model with respect to AMI data.

5.2.3 Random Walk to Savings: A New Modeling Approach Using a Random Coefficients Model and AMI Data

Paper presented at the 2016 International Energy Policy & Programme Evaluation Conference (IEPPEC) in Amsterdam, Netherlands.[51]

This paper presented a new energy savings estimation approach, one that provides accurate impact estimates by taking full advantage of hourly AMI data. This approach differs from traditional methods in that it automatically develops a large number of customer-specific regressions covering a wider range of customer types, weather conditions, and time periods. The approach uses a type of hierarchical linear model—the random coefficients model—that allows savings estimates to be tailored more closely to individual customer characteristics. This is accomplished by first grouping customer consumption data into different categories based on energy use and weather conditions. Separate models are then estimated for each usage/weather category, which allows for separate load shape predictions for very specific customer types.

The random coefficients model specification was tested using data from two HVAC efficiency programs in California. Using participant and AMI data from both of these programs, average daily load shapes were calculated for specific day types (weekday, weekend, seasonal) and used to estimate program impacts. When estimated load shapes were compared against a holdout sample of customers, the random coefficients model performed extremely well; load shape estimates were within 1 percent of the holdout sample. Energy savings estimates for these programs ranged from 4 to 7 percent of annual energy use, which was consistent with expectations. Besides producing accurate impact estimates, the automated categorization and modeling processes allow for separate savings estimates and load shapes to be developed easily for a variety of situations.

---

[51] https://www.iepec.org/wp-content/uploads/2018/04/Paper-Grover.pdf

5.2.4 Take It From the Top! An Innovative Approach to Residential and Commercial Program Savings Estimation Using AMI Data

Paper presented at the 2017 International Energy Program Evaluation Conference (IEPEC) in Baltimore, MD.[52]

This paper presented a new energy savings estimation approach—referred to as the AMI Customer Segmentation (AMICS) model—that provides accurate impact estimates by taking full advantage of hourly AMI data. This approach differs from more traditional methods in that it automatically develops a large number of customer-specific regressions covering a wider range of customer types, weather conditions and time periods. The approach uses a type of hierarchical linear model—the random coefficients model—that allows savings estimates to be tailored more closely to individual customer characteristics. This is accomplished by first grouping customer consumption data into different categories based on energy use and weather conditions. Separate models are then estimated for each usage/weather category, which allows for separate load shape predictions for very specific customer types.

The AMICS model specification was tested using data from both residential and commercial HVAC efficiency programs in California. Using participant and AMI data from both of these programs, average daily load shapes were calculated for specific day types (weekday, weekend, seasonal) and used to estimate program impacts. When estimated load shapes were compared against a holdout sample of customers, the random coefficients model performed extremely well; load shape estimates were within 1 percent of the holdout sample. Besides producing accurate estimates of energy use, the automated categorization and modeling processes allow for separate savings estimates and load shapes to be developed easily for a variety of situations.

5.2.5 Taking Control: Using AMI Data to Estimate Impacts from Peer Comparison Programs

Poster presented at the 2017 International Energy Program Evaluation Conference (IEPEC) in Baltimore, MD.[53]

Peer comparison programs that utilize a randomly selected control group of customers are a popular way for achieving savings, with impact estimates typically ranging between 1 to 3 percent of annual usage. This poster presented results from a new AMI customer

[52] https://www.iepec.org/wp-content/uploads/2018/02/2017paper_grover_cornwell_monohon_helvoigt-1.pdf
[53] https://www.iepec.org/wp-content/uploads/2017/08/Cornwell_IEPEC2017_Poster.pdf

segmentation (AMICS) modeling approach that can easily provide more granular program savings estimates than the traditional fixed effects model.

The AMICS model process starts with automated customer segmentation, grouping customers by their total annual energy consumption, load shape (i.e. hours of use), and the weather conditions they experience. These groupings remove a substantial amount of uncertainty from the model by reducing the variation in energy usage across customers. Separate models are then simultaneously estimated for each usage/weather category, producing separate predictions for each.

The AMICS modeling approach was previously used to estimate impacts from two residential HVAC programs. In these earlier studies, the model generated load shapes for over 1,000 different combinations of customer types and weather conditions. When estimated load shapes were compared against a holdout sample of customers, the model performed extremely well; hourly energy load estimates were within 1 percent of the actual load for the holdout sample.

Given the promising results from the earlier study, the random coefficients model presents an exciting opportunity for estimating energy savings for a peer comparison program. To test this, the model is being used to estimate impacts for customers who began receiving home energy reports from Pacific Gas and Electric in late 2011. The AMI data contains over 3.5 billion hourly observations of treatment and control customers from 2010-2013. Our analysis starts with the customer segmentation process to group customers (both treatment and control) with similar energy usage. Then, the model simultaneously estimates thousands of load shapes, one for each usage/weather category. We then use the control group to calculate the difference-in-differences for each category.

This poster was of interest to researchers wishing to use AMI data to estimate savings for a peer comparison program, as well as those seeking to understand the underlying customer segments that are the largest contributors to overall program savings.

5.2.6 Cultural Factors in Energy Use Patterns of Multifamily Tenants: EPIC AMI and Load Shape Development

Report prepared for the California Energy Commission in February 2018 by TRC Engineers. Evergreen Economics acted as the interval data analyst for this study under contract with Pacific Gas and Electric.[54]

This project used the AMICS model to predict load shapes and estimate energy savings from hourly interval energy usage data of each tenant residing in a building that

[54] https://ww2.energy.ca.gov/2018publications/CEC-500-2018-004/CEC-500-2018-004.pdf

participated in PG&E's Multifamily Upgrade program, funded by California's Electric Program Investment Charge (EPIC). For this study, the model provided valuable insights into the characteristics of customers and days that drive program savings, including additional analysis with survey data to identify demographic and cultural attributes that are strong predictors of energy usage.

More traditional regression models will focus on the average energy usage and average program impact across all customers in the program. One major benefit of the AMICS model comes from the fact that customers are only modeled with other members of their customer segment. The midday peak users have a separate model and program savings estimate from the evening peak users. The evening peak users experienced greater savings on hotter days whereas the afternoon peak users actually increased their usage on the hottest days.

5.2.7 AMI Analysis of Site Level Commercial HVAC Savings

Report prepared for Southern California Edison on July 19, 2018. It was developed as part of SCE's Emerging Technologies Program, under internal project number ET17SCE1130.

The goal of the study was to demonstrate that the AMICS modeling approach is able to quantify interval energy savings for individual non-residential sites and is capable of meeting the requirements of normalized metered energy consumption (NMEC) analysis.

In the case of programs like SCE's HOPPs CVC-HVAC program, billing analysis must provide savings estimates for each individual participant. Given the small number of diverse commercial customers expected to participate in this program, we believed it would be unlikely that we could construct meaningful customer segments to consistently meet the NMEC error thresholds. Instead, we assigned each customer to their own bin, effectively constructing separate models for each individual commercial customer. In this variation of the AMICS approach, we were no longer creating customer segments, but the segmentation of days (via weather conditions and day type) was still required.

A key benefit of the AMICS model is avoiding over-reliance on the average day. Models like Temperature and Time of Week (TTOW) essentially estimate the average load shape and then make a series of adjustments to that prediction depending on how the actual weather conditions differ from this average. 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.

Table 16: Comparison of the AMICS and TTOW Models

	AMICS	TTOW
Developer	Evergreen Economics	Lawrence Berkeley National Laboratory
Prediction interval	Hourly or 15-minute	Hourly or 15-minute
Temperature dependence	Non-linear. Portfolio of predicted load shapes, one for each set of distinct weather conditions (combo of CDD and HDD).	Piecewise linear. Coefficients capture the average incremental energy usage for a series of component temperatures. Estimated separately for occupied vs. unoccupied periods.
Time dependence	Time-of-day adjustments. Estimated separately for weekdays and weekends.	Time of week adjustments.
For each…	Customer segment or individual	Individual

NOTE: Small modifications to each method could bring the two closer into alignment, but we chose to keep them distinct to emphasize the current design and relative strength of each.

Our cross validation exercise (i.e., pre-period holdout tests) did not find any significant differences in the prediction error between these two modeling approaches. We believe that the AMICS and TTOW approaches are both well suited for AMI analysis of residential and commercial customers, and choosing one model over the other will not significantly affect the analytical results.

5.2.8 M&V 2.0: Leveraging Machine Learning to Improve Energy Savings Estimates

Paper presented at the 2018 American Council for an Energy Efficient Economy (ACEEE) Summer Study on Energy Efficiency in Buildings in Pacific Grove, CA.[55]

Access to advanced metering infrastructure (AMI) data has given rise to many opportunities to improve upon the way in which the energy efficiency industry understands end-use customers, measure energy savings and inform estimates of energy savings potential. These emerging methods promise much more reliable savings estimates and predictions at much lower cost.

This paper describes how AMI data have been used to predict usage patterns for customer and weather segments, based on innovative modeling efforts. The study team developed a

[55] https://aceee.org/files/proceedings/2018/node_modules/pdfjs-dist-viewer-min/build/minified/web/viewer.html?file=../../../../../assets/attachments/0194_0286_000152.pdf

new approach—the AMI Customer Segmentation (AMICS) model—that reveals the variability in energy savings across customer groups and weather conditions. A key step in this modeling approach is using machine-learning algorithms to identify similar load shapes and segment the AMI data into thousands of distinct bins. Each bin contains customers with similar energy usage patterns on days with similar characteristics. Separate billing regression models are then estimated for each customer/weather segment, creating thousands of distinct load shape predictions.

The paper also describes new efforts to incorporate new data—residential end use and circuit-level interval data collected as part of two residential monitoring studies—along with more granular 15-minute AMI data to further refine the AMICS model to more effectively utilize AMI data to estimate energy efficiency program savings.

This paper will be useful for energy efficiency professionals who are interested in understanding the potential of AMI data to improve upon the ability to understand energy savings opportunities, with more reliability and much lower cost than traditional methods.

5.2.9 SCE Smart Thermostat Impact Analysis

Final report prepared for Southern California Edison on December 7, 2018.

SCE contracted with Evergreen Economics in 2018 to estimate the energy savings of the smart (communicating) thermostats installed in residential single-family homes, relative to existing conditions, likely manual or programmable thermostats.

This study used a convenience sample of customers who enrolled in SCE's Rush Hour Rewards (RHR) demand response program as of July 2018.[56] This demand response program offers customers incentives to purchase a qualifying energy–efficient smart thermostat that enables them to reduce their energy usage during peak demand events, in exchange for ongoing bill credits. Consequently, the RHR participants are all residential customers with smart thermostats that SCE had already identified in its service territory, but they do not necessarily represent the broader population of SCE customers with smart thermostats.

For this study, we conducted a two-pronged modeling approach to billing analysis that was designed to make the most of the available data while still producing seasonal smart thermostat energy savings estimates (i.e., daily kWh savings). In both models, we used a comparison group of future RHR program participants to help control for any natural changes in energy consumption over the study period that should not be attributed to the

[56] This program is now referred to as the Smart Energy Program (SEP).

smart thermostats. We identified and excluded RHR demand response event days to avoid double counting savings caused by event participation.

In the first model, we estimated smart thermostat impacts based on daily kWh energy usage data and a fixed effects billing regression. Next, we used the larger AMI billing database with hourly interval kWh energy usage to estimate savings using the AMICS model.

Both the daily kWh fixed-effects regression and hourly interval AMICS models found statistically significant increases in average energy use that we attribute to the installation of smart thermostats by RHR participants. However, the AMICS analysis by customer segment revealed a high variation in energy savings across households. Customers with low energy usage (kWh) in the pre-installation period substantially increased their energy usage after installing the smart thermostat; these offset the energy savings achieved by more moderate energy users—leading to an overall increase, or lack of savings.

5.2.10 EE Savings from Optimized Connected Thermostats

Final report prepared for Emerging Products (EP) group at Southern California Edison on December 10, 2018. It was developed as part of SCE's Emerging Technologies Program, under internal project number ET17SCE8010.

The manufacturer administered the optimization algorithm to a randomly selected group of customers in SCE's service territory who already have a connected thermostat. Each customer in the treatment group received a message on their thermostat prompting them to opt into the project, allowing the manufacturer's algorithm to adjust their temperature set points and thereby reduce their home's cooling load. The manufacturer also maintained a control group in the service territory that was not given an option to participate. This control group made it possible to estimate the incremental savings from the messaging treatment and set point changes, over the existing conditions of having a connected smart thermostat.

SCE contracted with Evergreen Economics to conduct an independent validation of the manufacturer's calculation of the energy savings attributable to the optimization implemented in 2018.

Evergreen used the AMICS model to estimate program impacts on whole home AMI interval energy use for a sample of participants that could be identified through a web survey. These savings estimates were then compared to estimates of the project's impact on actual HVAC runtimes, following the analysis methods suggested by the thermostat manufacturer. The kWh and percent energy savings varied across each of these methods, but most suggested that there were energy savings attributable to the optimization project.

The differences between the estimates were not statistically significant, due at least in part to the small sample size.

5.2.11 Predictions with Restrictions: C&I Metered Energy Consumption

Paper presented at the 2019 International Energy Program Evaluation Conference (IEPEC) in Denver, CO.[57]

This paper presents the results of a case study that compares and contrasts the Advanced Metering Infrastructure Customer Segmentation (AMICS) and Temperature and Time of Week (TTOW) models to estimate daily electricity load shapes for a sample of 10 businesses that completed an HVAC retrofit project between 2015 and 2017. This sample covers a wide range of business types, operating schedules, and variability in load shapes. Both the AMICS and TTOW models were designed for the purpose of using AMI interval data to predict whole building hourly or sub-hourly energy usage, while accounting for the impacts of outdoor temperatures and weekly operating schedules. To assess the relative accuracy of these two modeling approaches, we conducted a cross validation using a series of randomized pre-period holdout tests for each site in our sample.

This paper builds on existing research (AMI load shape analysis and prediction error diagnostics) and offers new insights for the next generation of programs. This paper will be of interest to evaluators, policymakers, and program implementers who are choosing between multiple industry-accepted methods for estimating savings for individual buildings and developing new evaluation policies to realize the potential normalized metered energy consumption (NMEC) benefits. The NMEC measurement and verification approach offers the opportunity for program implementers to gain more real-time realized savings feedback.

5.2.12 When Are Smart Thermostats a Smart Investment?

Paper presented at the 2019 International Energy Program Evaluation Conference (IEPEC) in Denver, CO.[58]

This paper presents the results of two separate studies estimating the energy efficiency and load impacts of smart thermostats in Southern Californian homes. These two studies provide a unique and robust exploration into the variation in energy savings across customer segments. Our findings suggest a need for targeting and education to increase energy savings and improve the cost-effectiveness of smart thermostat programs. The results and recommendations will be valuable for a wide audience, as the potential for

[57] https://www.iepec.org/2019_proceedings/index.html#/paper/event-data/046-pdf

[58] https://www.iepec.org/2019_proceedings/index.html#/paper/event-data/118-pdf

smart devices is increasingly of interest for utilities to enhance program offerings and optimize customer experience.

The first study focuses on 26,000 customers who received incentives for a smart thermostat that enabled them to participate in demand response events. We excluded event days to focus on the energy efficiency of these new smart thermostats relative to existing conditions. The second study estimated the impact of a temperature set point adjustment algorithm offered to around 63,000 connected smart thermostats across the utility's service territory. The algorithm reduces energy usage by making small improvements to thermostat settings, thereby reducing HVAC runtimes. The analysis was comprised of a simple fixed effects regression and a more complex AMICS model with hourly interval kWh, both using comparison groups to control for any natural changes that should not be attributed to the smart thermostats.

5.2.13 SCE NMEC Pre-Qualification Pilot Feasibility Study

Final report prepared for Emerging Products (EP) group at Southern California Edison on December 31, 2019. It was developed as part of SCE's Emerging Technologies Program, under internal project number ET19SCE7010.

SCE contracted with Evergreen Economics to conduct AMI data analysis to determine if a streamlined normalized metered energy consumption (NMEC) approach might be feasible to estimate energy savings for multiple business branch locations across a single business entity.

In this initial proof-of-concept study, Evergreen utilized a modified AMICS modeling approach and pre-screening algorithm to develop baseline models of energy consumption for three commercial chains, including two grocery chains and one retail chain, with 39 proposed participants. We utilized all current NMEC requirements and guidelines for assessing model fit. We also tested whether a matched comparison group could be extracted from the remaining branches (i.e., non-participants) from these three chains to estimate net savings in the post-period. Since the pilot has not begun implementation, this study did not cover performance payment calculations or savings claim estimates.

The individual baseline models met all of the NMEC model fit criteria for the vast majority of participant sites (n=38/39).[59] 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

59 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.

model prior to program intervention. The pooled and segmented baseline models of both grocery chains met all of the NMEC model fit criteria, but the retail chain did not.

Conclusions and 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.