NMEC Program Design with Missing Data, Zero Values, and Differing Meter Resolutions

Cory Read, EcoMetric Consulting, Boise, ID
Salil Gogte, EcoMetric Consulting, Exton, PA
Kyle McKenna, EcoMetric Consulting, Cedar Rapids, IA

ABSTRACT

Southern California Gas Company (SCG) initiated the Central Water Heater Multifamily Building Solution program (CWHMBS) in early 2017. The program design targets natural gas master-metered multifamily buildings built prior to 1984 and complies with the California Public Utilities Commission (CPUC) normalized metered energy consumption (NMEC) site-level program rules. Target measures for each site include central domestic hot water boiler, circulation loops, controls, and low flow showerheads and aerators. The program pays both pre-measure incentives based on initial savings estimates, and post-measurement incentives based on a 12-month measurement and verification period. The final verified site specific savings are determined using natural gas advanced metering infrastructure (AMI) hourly data reads.

Savings results have been mostly positive with initial whole facility savings ranging between negative savings up to 40% of baseline period program site consumption. Among the project sites, numerous data challenges arose during data preparation including several unique gas AMI challenges. Data challenge examples include strings of missing values, zero-valued gas reads, and differing meter data resolution ranging between three decimals to simple integer hourly reads.

For the analysis we walk through each highlighted data challenge and associated model goodness of fit metrics and energy savings impacts. Gas AMI data challenges negatively impacted model accuracy metrics which, for most sites, did not meet program defined thresholds while still passing overall savings uncertainty requirements.

Results provide a case study for dealing with data deficiencies for a whole-building NMEC gas program. Presentation insights will aid whole building, and meter data-based program planners and program implementers understand how different natural gas meter data issues affect model and savings results.

Introduction

Southern California Gas Company (SCG) initiated the Central Water Heater Multifamily Building Solution program (CWHMBS) in early 2017. The program was made possible by the 2015 California Assembly Bill (AB) 802[1] enabling high opportunity utility projects or programs (HOPPs) utility program submissions. Per SCG Advice Letter 4965-A[2] (Disposition, 2016), the final approved program design targets gas master-metered multifamily buildings built no later than 1984 with efficient central domestic hot water upgrades. Final approved program site measures included;

---

[1] https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201520160AB802

[2] "Disposition approving Advice Letter 4965-A (U 904 G), Southern California Gas Company High Opportunity Projects and Programs(HOPPs) – Central Water Heater Multifamily Building Solution (CWHMBS) Program", August 2, 2016, p. 2. https://tariff.socalgas.com/regulatory/tariffs/tm2/pdf/4965-A.pdf

- Central storage water heater or boilers
- Central water heater modulating temperature controllers
- Hot water system usage monitoring
- Low flow showerheads and faucet aerators
- Circulating demand pumps with controllers

The initial program rollout was limited to 20 enrolled SCG customer multifamily sites with installed hourly natural gas AMI. The piloted program design was a site-level normalized metered energy consumption (NMEC) design where all savings and associated incentives were to be estimated using actual measured hourly metered gas usage. Final savings were calculated by estimating changes between baseline pre-project and post-project periods. NMEC is a California Public Utilities Commission (CPUC) specific program category term. CPUC NMEC program guidelines are outlined in a program rulebook (California, 2020). CPUC regulated NMEC program designs are similar to pay for performance, strategic energy management, or other meter data based whole building efficiency designs. A specific NMEC program requirement is that programs produce weather normalized savings in additional to traditional "at the meter" or actual energy savings. Actual or meter based savings were calculated in addition to weather normalized savings. Actual savings provided the basis for incentive payments.

EcoMetric was contracted as an embedded M&V consultant to produce site specific claimable and weather normalized energy savings estimates. site-level energy savings and customer incentives were determined uniquely for each program site. Program participants authorized the release of a minimum 12 months of hourly reads prior to and after project installation. The M&V process required site- specific metered data quality control screens, weather merges and data feature creation, baseline and performance period model building.

In this paper, we outline the program M&V methodology, discuss gas AMI data challenges and chosen resolutions, and summarized data quality impacted baseline modeling impacts.

Methodology

Advice Letter Proposed Methodology

The initial program advice letter proposed using the Lawrence Berkley National Lab (LBNL) published Time of Week Temperature (TOW) model approach (Matthieu, 2011). The LBNL sponsored paper outlines a full methodology, but key method highlights include:

- **Ordinary Least Squares Regression (OLS)** Current CPUC NMEC rulebook guidelines do not call for a certain model type or structure. Guidelines only spell out that model formats are stated. The program advice letter called specifically for OLS regression methods.
- **Primary data prediction features were temperature and time of week** An indicator variable is created for each week hour creating 167 new dataset variables (168 hours in the week minus one variable to avoid collinearity issues). Temperature variables are added by creating linear spline variables that divide annual temperature ranges into buckets. The spline features incorporated into OLS regression approximate the curved hourly annual seasonal temperature.
- **Separate occupied and un-occupied time periods models** Baseline or performance period data is divided into occupied or unoccupied hours. Unoccupied hours are then modeled separately using only hourly temperature.

- **Normalized Metered Energy Consumption (NMEC)** Fully normalized savings estimates requires building both baseline and performance period models. The baseline model is built using pre-installation energy data while the performance prediction model is built to predict efficient energy post measure installation. The final baseline and performance period models are projected against normalized, long term weather conditions to create a weather normalized baseline and efficient performance period energy forecasts. The performance period model coefficients are also rerun using normalized weather conditions to create a weather normalized energy use forecast. Final weather normalized savings are the cumulative differences between the normalized baseline forecast and the normalized performance period energy use forecast.

Final Approved Methodology

The final approved program M&V approach differed slightly from the published TOWT approach. The targeted multi-family properties baseline datasets were not divisible into separate modeled occupied and unoccupied hours. Program participant buildings master metered gas data aggregated apartment level hourly natural gas usage. The aggregation masked discernable site specific occupancy trends. The published TOWT methodology introduces two primary data features for baseline and performance model building, temperature and time of week. The final M&V approach included additional non- TOWT method defined model variables including heating degree hour and holiday indictors. Table 1 lists the final prediction model variable set.

Table 1. Final baseline and performance model prediction variables

Variable	# of Variables
Temperature (linear splines)	6
Time of Week	167
Heating Degree Hour Moving Average (14 days)	1
Holiday indicator	1
Total # Variables	175

Other prediction variables were considered but ultimately were not included. Hot water usage is typically less correlated with weather than electric usage so looking for additional prediction variables when building baseline models is useful. Two primary data collection activities were investigated; 1) water usage, and 2) occupancy. Water bills were requested for several early sites with the hope that at least monthly bills would help adjust models seasonally. The majority of the water bills were bi-monthly resulting in only 6 unique water consumption data points across a single year. Bi-monthly bills masked seasonal transitions by averaging consumption across wider periods. One site was located in an area with monthly water bills and results were promising because water usage was correlated with gas usage. Because it took multiple customer contacts and several weeks to obtain the water bills while only a small number of the 20 program sites ultimately would have monthly water bills. Mastered metered water data also included irrigation and community pool usage so ultimately the effort was abandoned. Occupancy data was investigated for several sites via third party data sites[3] and also through participant data requests. Vacancy research revealed that a majority of sites during the program pilot had very constant vacancy rates with little month to month variability. Numerous sites participated in high consumer demand local and/or state rental assistance programs.

The coefficient of variation of the root mean square error CV(RMSE) (Equation 1)(Granderson, 2019) was the principle model variability goodness of fit metric. The CV(RMSE) describes how much

[3] E.g. Loopnet.com, costar.com, apartments.com, commercial real estate listings etc.

variation or randomness there is between the data and the model. CV(RMSE) is calculated by dividing the root-mean squared error (RMSE) by the period average hourly gas consumption. The program defined CV(RMSE) target threshold was less than 25% for each baseline and performance period model.

$$CV(RMSE) = \frac{1}{\bar{y}}\left[\frac{\sum(y_i - \hat{y})^2}{(n-p)}\right]^{1/2}$$

Equation 1. Model variance goodness of fit metric

Normalized bias error (NBE) (equation 2) (Granderson, 2019) detects if a model is more likely to miss above or below a prediction target value. NBE was the primary bias estimate with a program target range between -.5% and .5%.

$$NBE = 100 * \frac{\sum_i (E_i - \hat{E}_i)}{\sum_i E_i}$$

Equation 2. Model bias goodness of fit metric

The final model coefficient of variation (R^2) was also calculated for each site baseline and performance period model. The program target R^2 threshold was greater than 70%.

Final total savings uncertainty estimates were measured using the fractional savings uncertainty (FSU) metric (California, 2020) The FSU process is shown in Equations 3-5. FSU combines baseline and performance model uncertainty estimates which is then taken as a percent of total site-level estimated savings. The final FSU calculation is adjusted for autocorrelation bias because consecutive hourly energy use data are not fully independent. Even with autocorrelation adjustments, savings uncertainty including the root mean square error most likely underestimate the true model and savings uncertainty. This is a known industry issues for hourly and sub-hourly meter- based programs using RMSE based uncertainty approaches (Koran, et al.).

$$Uncertainty = t \times 1.26 \times CV(RMSE) \times \sqrt{\frac{n}{n'} * \left(1 + \frac{2}{n'}\right) * \frac{1}{m}}$$

Equation 3. Savings uncertainty

Where:

t = student's t-statistic for a given confidence level (90% in this analysis)
CV(RMSE) = Coefficient of variation of the root mean square error
m = number of points in the full period (8,760 for this analysis)
n = number of observations in model
n' = adjusted number of observations in the model, given by the equation:

$$n' = n \times \frac{1-\rho}{1+\rho}$$

Equation 4. FSU autocorrelation adjustment where ρ = the autocorrelation coefficient (the square root of the R^2 calculated for the correlation between the final model residuals and the residuals for the prior time period or previous hour)

The total uncertainty for each site is simply the square root of combined baseline and reporting periods' site-level uncertainty for that site (equation 5):

$$Uncertainty_{Total} = \sqrt{Uncert_{Pre}{}^2 + Uncert_{Post}{}^2}$$

Equation 5. Total savings uncertainty

Final Fractional Savings Uncertainty (FSU) equals the Uncertainty divided by the total final estimated savings.

Data Quality Issues

Pre-screening participant AMI data quality was not a program participation requirement so all identified data quality issues needed to be addressed or researched prior to final model building. Four common data quality issues were identified and investigated;

- zero value reads,
- poor temperature to gas usage correlation,
- low variability in gas usage, and
- reduce meter data resolutions (e.g. data only available in whole integers).

SCG staff reviewed each problem site data extracts ensuring data issues were not query errors. After confirming that the data was exported correctly summary metrics were created to help summarize the various data quality modeling impacts.

- **Percent zero reads** The metric assesses the percent of baseline or performance period hourly reads were zero valued meter reads. Days with all zero meter reads were ultimately removed before model building.
- **Temperature and natural gas usage correlation** The Pearson correlation[4] value was calculated between hourly gas meter values and actual temperatures. Several program sites showed little seasonality and weather relationship between hot water gas usage and temperature and temperature and gas usage correlation was generally low across all sites.
- **Gas meter read variability** The ratio of unique hourly gas values divided by the total number of baseline or performance hourly reads. Values closer to zero indicate less unique gas meter read values or reduced variability. Low variability in your prediction variable may impact model performance.
- **Gas decimal places** Gas AMI reads were delivered with three different levels of measurement resolution. No decimals, one decimal, and two decimal places. Gas data resolution varied according to the gas metering system installed at the site as well as how it was programmed.

Figure 1 visualizes program site #8 periods of zero value reads. Nearly 30% of Site #8 baseline data were consecutive zero value reads. Consecutive zero value meter reads were removed prior to model building.

[4] https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

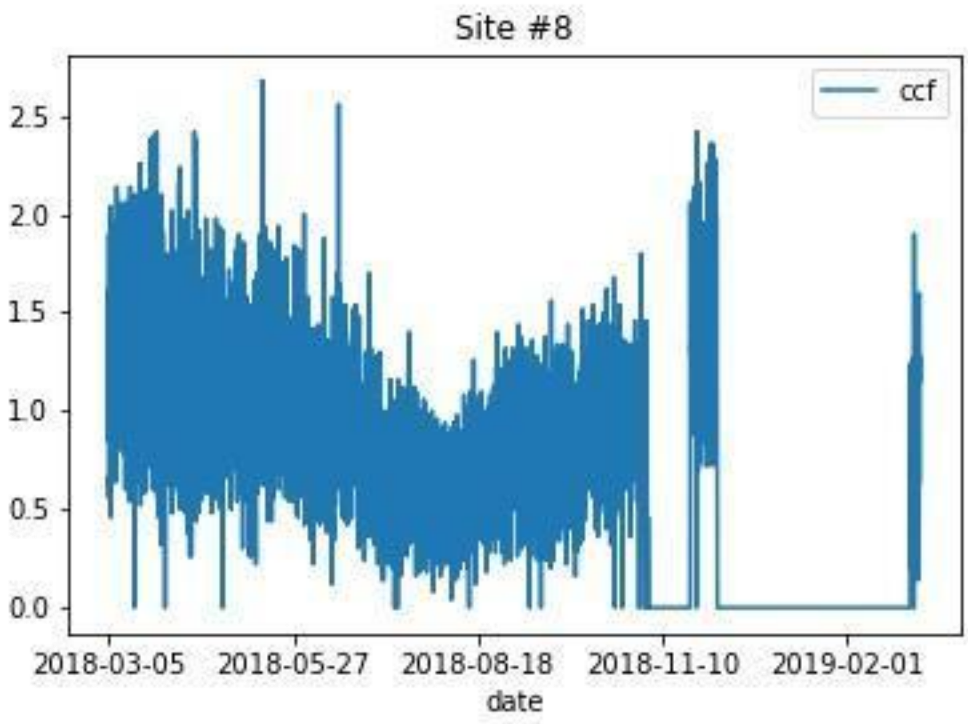

Figure 1. Multiple zero read value example

Figure 2 demonstrates poor hourly temperature and natural gas usage correlation. Gas usage still peaks seasonally in the figure, but the final hourly gas usage and temperature correlation was only -5.4% when the all site average was -26% (100 or -100% indicates perfect variable correlation). Hot water gas consumption usage and temperature were expected to be low in general across all program sites in part because of the regional mild climate.

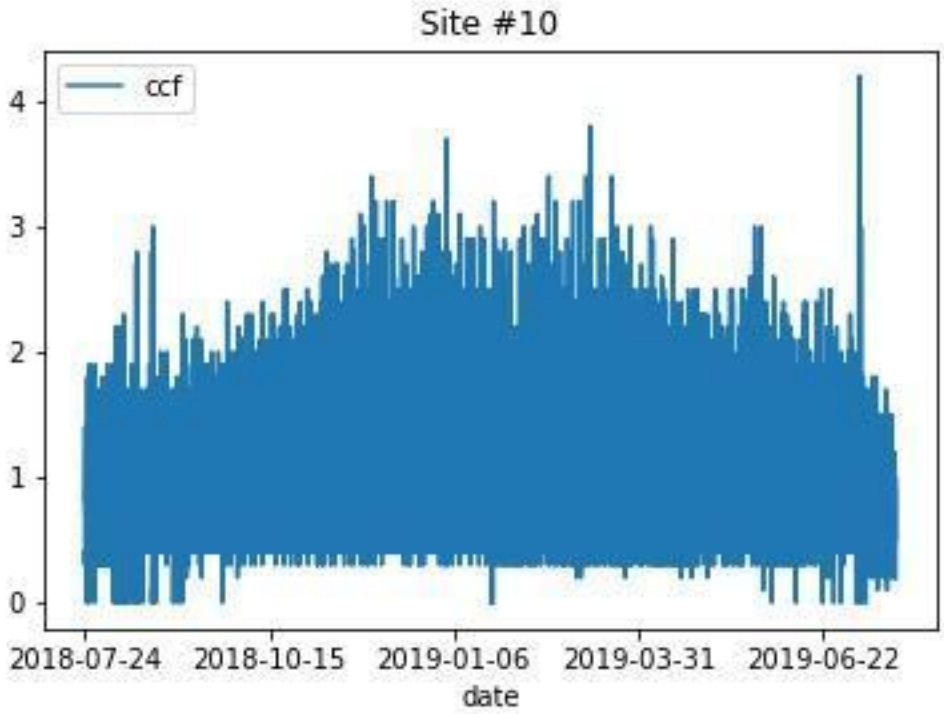

Figure 2. Low gas consumption and temperature correlation

Figure 3 displays low meter resolution output where the meter was programmed to only produce integer values. Hourly summer 2018 gas usage data oscillates between 0 and 1 depending on the hour of the day. The oscillation creates a block shape during summer 2018 hours and a very digital (squared shape over the whole year). No pre-model building data adjustments were made so the data was modeled as is.

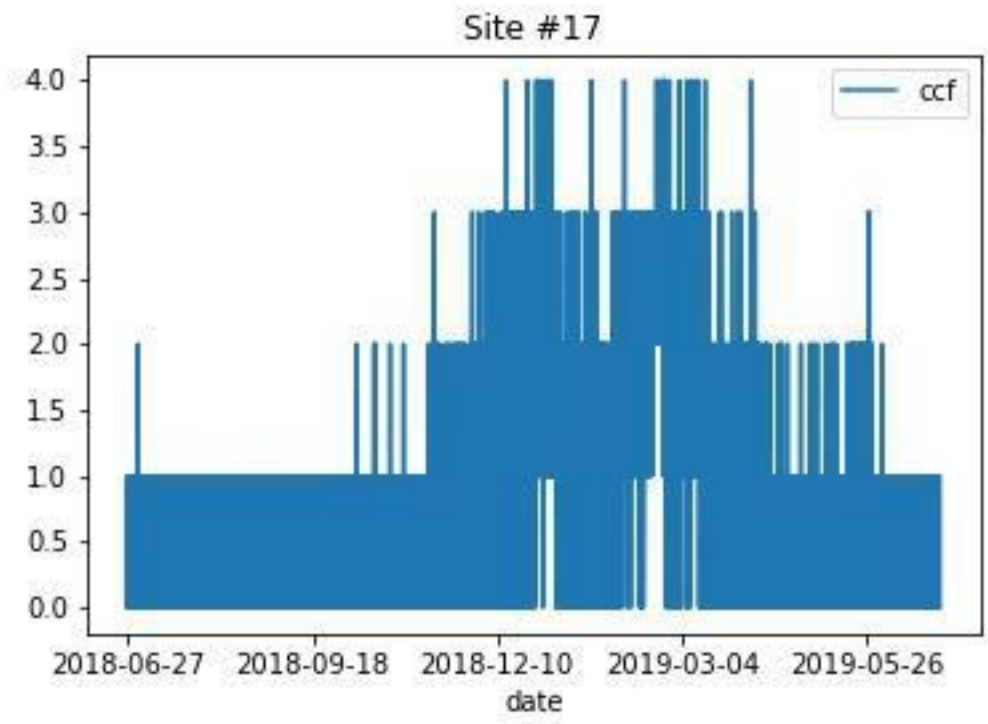

Figure 3. Low resolution meter hourly consumption

Model Results

A majority of all program sites did not meet program defined model goodness of fit metrics while 16 out of 20 sites still passed the overall model uncertainty (FSU) threshold. Three of the four non-compliant FSU sites were negative savings projects. Overall project annual energy savings as a percent of baseline usage was 8.7% when including the 3 negative savings projects. Energy savings as a percent of baseline usage without negative savings projects increased to 22%.

Table 4. # of program sites (n=20 total sites) not meeting model goodness of fit metrics and savings uncertainty thresholds

	# Sites Failing R^2 (< 70%)	# Sites Failing CV(RMSE) (> 25%)	# Sites Failing FSU (< 0 or > 50%)	Avg. Site Savings (% of Baseline usage)
Baseline	16	15	4	8.7%
Performance	16	14		

Poor model goodness of fit metrics including low R^2 and high CV(RMSE) values were not good indicators that projects would not pass final savings uncertainty (FSU) thresholds. Model goodness of fit metrics and model savings uncertainty formulas contain independent components so 100 percent alignment is not expected. Program defined model goodness of fit metrics should be expected to roughly align with savings uncertainty metrics. This is an area where further research should be conducted either through simulations or analysis of other similar model based program designs.

Lack of non-weather related variables and low weather and gas usage correlation likely contributed to poor final baseline and performance model fits. Only four baseline period and performance period sites met or exceeded the recommended 70% R^2 threshold. Low R^2 values indicate the degree that your prediction model variables are unable to explain gas usage or there are likely other unknown non-included model explanatory modeling variables. High percent zero meter reads, low temp-gas usage correlation, and low gas measurement variability were moderately correlated with higher (poor fit) CV(RMSE) values. Figure 4. (upper left) demonstrates ranked percent zero value reads (% data points removed) from high to low. Figure 4 (upper right) plots high to low temperature-gas correlation, and Figure 4 (bottom) shows increasing gas measurement variability all plotted with CV(RMSE) values on the right axis. Across each measure there is no consistent relationship data quality impact on model variability metrics and in some cases there are slight trends of model variability decreasing (increasing CV(RMSE)) as data quality metrics improve

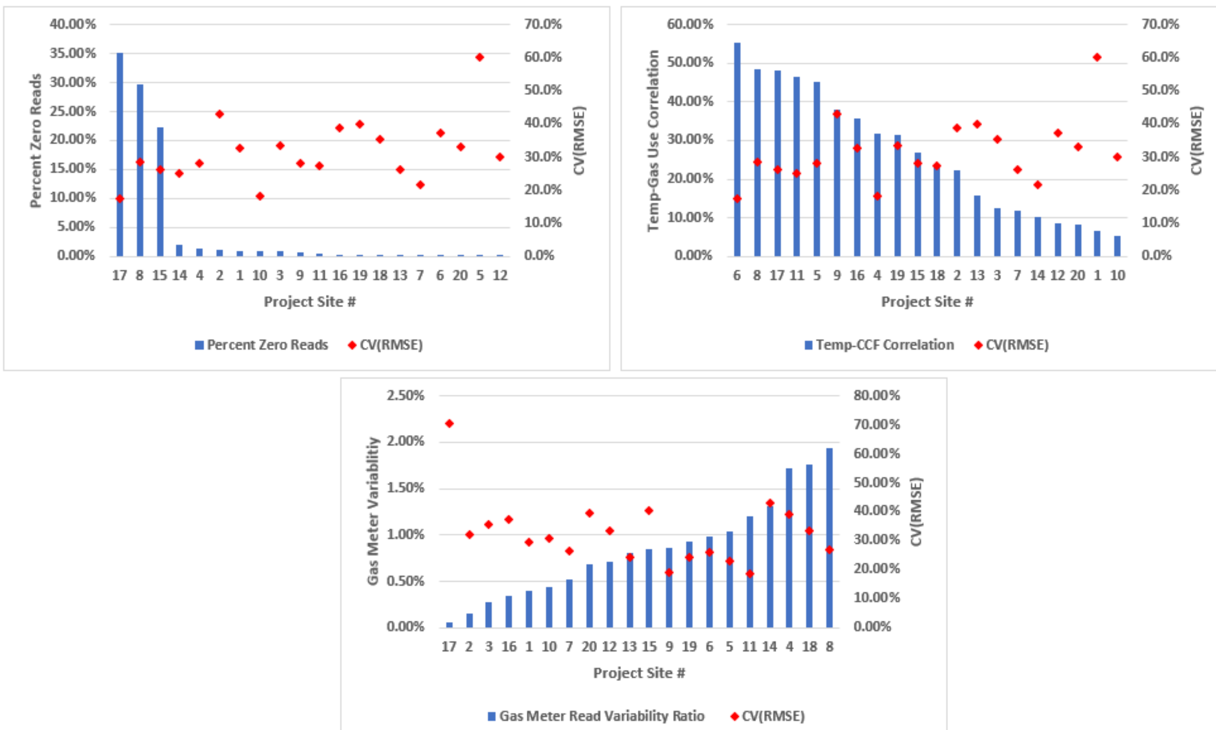

Figure 4. Baseline model percent zero meter reads (top left), temperature-gas use correlation (top right), and gas read variability (bottom) and CV(RMSE) values

Alternative Modeling Approaches

Two post project retrospective modeling approaches were testing with the four poorest model fit program sites. The first method was to test an array of alternative model algorithms apart from OLS regression and the second was to attempt daily models instead of the program prescriptive hourly models. Table 5 lists the tested algorithms which include high/low bias and high/low variance structures along with a variety of tunable model parameters (e.g. machine learning hyperparameters) to fine tune model final model fits. Low bias algorithms tend to be less restrictive in their form. Low variance model types tend to have less uncertainty with lower variability, but are often associated with higher bias. Several of the algorithms have been researched by LBNL (Touzani, 2019). Python based scikit-learn modules were used for all testing. Algorithm documentation can be viewed at;

https://scikit-learn.org/stable/supervised_learning.html. Alternative and flexible model type testing only moderately improved the four poorest performing program site baseline period CV(RMSE) values. Advanced alternative modeling could not overcome poor weather fit, missing data, and/or low meter read variability.

Table 5. Tested alternative model algorithms

Model	Model type	Bias	Variance	Learning Parameters
Linear Regression (OLS)	linear	high	low	None
Lasso Regression (Lasso)	linear	high	low	1
Ridge Regression (Ridge)	linear	high	low	1
Elastic Net (EN)	non-linear	low	high	4
Ada Boost (AB)	non-linear	low	low	2
Gradient Boosting (GBM)	non-linear	low	low	4
Random Forest (RF)	non-linear	low	low	5
Extra Trees Regression (ET)	non-linear	low	high	5
MLP Regressor (NN)	non-linear	high	low	4

Daily models significantly improved (lowered) baseline model CV(RMSE) values for the four tested sites. Energy savings estimates were consistent between hourly and baseline models excluding site #17 where the model change was dramatic and the underlying data issues more severe (see Figure 3). Daily modeling options should be considered to address program data variability and/or quality concerns.

Table 6. Daily vs. hourly baseline CV(RMSE) values for poor model fit sites

Program Site #	Hourly CV(RMSE)	Model Type	Daily Energy CV(RMSE)
17	70%	Gradient Boosting	24%
14	43%	Ridge Regression	15%
15	40%	Ridge Regression	23%
20	40%	Random Forest	11%

Conclusions

Data oriented gas energy efficiency programs will expand as AMI is rolled out across additional service territories allowing for expanded data driven program designs. Even with substantial data quality issues such as high zero reads and low temperature-gas usage correlation in mild climates program rules and designs can help offset data quality issues. Program designs should allow and encourage exploration of available modeling variables apart from weather. Examples may include building level water usage for hot water end uses or production oriented variables for industrial gas programs. Additional non-weather data collection requires time and potential interactions with program participants to collect data.

Program identified gas AMI program data quality issues included high percent zero meter reads, low temp-gas usage correlation, and low gas measurement variability. Each identified data quality issue was correlated with higher (poor model fit) CV(RMSE) values. Weather model variables alone for gas hot water usage modeling were not sufficient for most program sites to pass model goodness of fit metrics. Missing and unknown variables likely contributed to poor model goodness of fit metrics. Poor model goodness of fit metrics were not predictors of final savings uncertainty. Clear high savings estimates for most projects resulted in lower FSU values even when almost all program sites failed one or more model goodness of fit threshold. The program model metrics and uncertainty thresholds were consistent with industry standards so this is an area where more investigation is needed either through simulation or analyzing other program results.

Alternative tested model approaches apart from the prescribed program approach had mixed results. Machine learning based model structures were not able to improve model goodness of fit metrics for tested poor fit sites and overcome underlying data quality issues. Daily aggregated models seemed to smooth hour to hour data issues while improving baseline model goodness of metrics for tested program sites. Final daily model energy savings estimates were within 1% of hourly model savings for all but one tested site. Daily model flexibility should be considered as a backup option for programs where hourly results are not mandatory and especially where there is a chance for inconsistent data quality

References

California Public Utilities Commission (CPUC), Energy Division. (2020) Rulebook for Program and Projects based on Normalized Metered Energy Consumption, v2.0.

Disposition approving Advice Letter 4965-A (U 904 G), Southern California Gas Company High Opportunity Projects and Programs(HOPPs) – Central Water Heater Multifamily Building Solution (CWHMBS) Program" (2016), p. 2. https://tariff.socalgas.com/regulatory/tariffs/tm2/pdf/4965-A.pdf

Granderson, J., P. Gruendling, P. Torok, P. Jacobs, N. Gandhi. (2019) Site-Level NMEC Technical Guidance: Program M&V Plans Utilizing Normalized Metered Energy Consumption Savings Estimation, v2.0. LBNL for the CPUC Energy Division

Granderson, J., S. Touzani, C. Custodio, M.D. Sohn, D. Jump, and S. Fernandes. (2016) Accuracy of automated measurement and verification (M&V) techniques for energy savings in commercial buildings. *Applied Energy*, *173*: 296-308.

Koran, B., E. Boyer, M. Khawaja, J. Rushton, J. Stewart. (2017) A Comparison of Approaches to Estimating the Time Aggregated Uncertainty of Savings Estimated from Meter Data, IEPEC-2017 Proceedings

Matthieu, J.L., P.N. Price, S. Kiliccote, and M.A. Piette (2011) "Quantifying Changes in Building Electricity Use, With Application to Demand Response," IEEE Transactions on Smart Grid, 6 2:507-518, 2011.

Touzani, S., J. Granderson, D. Jump, and D. Rebello. (2019). "Evaluation of methods to assess the uncertainty in estimated energy savings." Energy and Buildings, 193: 216-225.

Touzani, S., J. Granderson, and S. Fernandes. (2018) "Gradient boosting machine for modeling the energy consumption of commercial buildings". Energy and Buildings 158, pp.1533-1543