436 14th Street, Suite 1020, Oakland, CA 94612
Prepared by: Mike Maroney
Phone: (510)-473-8608
Email: mmaroney@trcsolutions.com

California Advanced Homes Program (CAHP) Billing Analysis

Final Report

May 31, 2019

Submitted To:

Pacific Gas and Electric Company (PG&E)
Josh Harmon
245 Market Street
San Francisco, CA 94111

TABLE OF CONTENTS

TABLE OF FIGURES

I. Acknowledgements

The California Investor Owned Utilities (IOUs) funded and supported the development of the Study. The Study has been carefully guided by Josh Harmon, the Pacific Gas & Electric (PG&E) project manager who provided oversight on this project.

PG&E provided energy consumption and cost data on homes built under the 2008, 2013 and 2016 Title 24 code. Additional data on projects (home size and number of bedrooms) was provided by Attom Data Solutions.

The study research plan and draft report was posted on CPUC's basecamp site. The project study plan and draft findings were reviewed by numerous stakeholders through a public review process. Written comments and follow-up conversations that followed from that process assisted the study teams in their efforts. While not all comments were incorporated into the study, these comments nevertheless helped broaden the purview of the study which strengthened the study findings and recommendations.

Mike Maroney was the study team lead, the primary author of this report and the project manager. Abhijeet Pande provided research oversight. Avani Goyal and Neil Perry provided data clustering, data cleaning and technical writing support. Stephen Wilson provided data management and encryption support. Margaret Pigman of Resource Refocus LLC conducted statistical analysis of the energy consumption and billing data included in this report.

2. Executive Summary

TRC (the consultant) in collaboration with Resource Refocus (the subcontractor) analyzed the impact of California Advanced Homes Program (CAHP) on metered energy use in participating homes compared with homes that did not take part in CAHP. CAHP pays incentives to builders to build homes that exceed the performance requirements of Title 24, Part 6. The consultant contributed data from a database of energy simulation models of CAHP homes built in PG&E territory during the 2008 and 2013 Title 24 code cycles. Separately, PG&E provided billing and hourly energy use data for all homes constructed during the 2008 and 2013 code cycles. The billing data for CAHP and non-CAHP homes was normalized by conditioned floor area (CFA). CAHP and non-CAHP homes were clustered so that all homes compared were within 30 miles of a cluster center—this clustering approach controlled for climate zone. The analysis only considered mixed fuel houses and houses that do not have solar photovoltaic (PV) panels.

CAHP participation shows the most significant performance improvement over non-CAHP homes in the hot climates (Fresno and Bakersfield), where CAHP homes have lower total energy consumption and more grid-friendly electricity load shapes (smaller peak demand and less steep afternoon demand increase) in summer. These clusters are also the places with the highest number of single-family new construction in the study. Normalized power demand (W/sf) on summer afternoons was lower in CAHP than non-CAHP homes in Fresno and Bakersfield–this has the potential to be a grid benefit beyond electricity and natural gas efficiency savings.

The data shows that the differences in energy consumption, bills ($), or load shapes between CAHP homes (sample group) and non-CAHP homes (control group), are small compared to the within-group variation; the same or similar asset yields a wide range of actual performance. After controlling for CFA, billing year, and vintage, CAHP homes show a 4.6% decrease in EUI compared to non-CAHP homes in Fresno and 15% in Bakersfield.

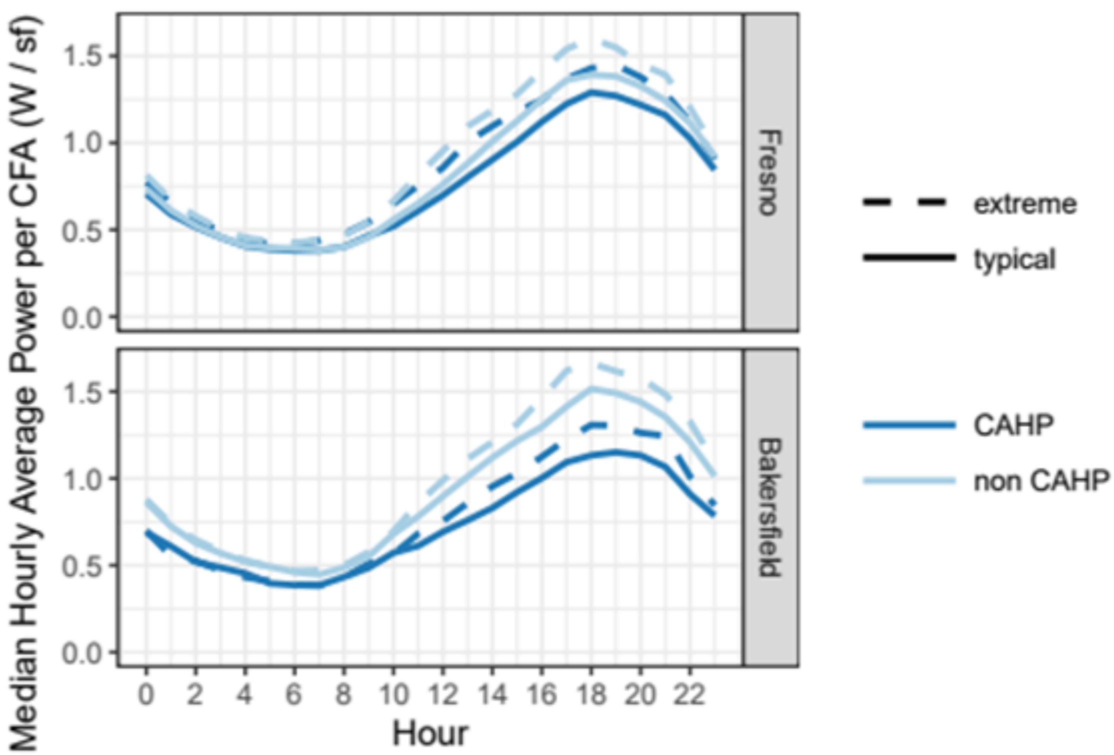

Figure 1. Median load profiles for Fresno and Bakersfield

CAHP participation is a statistically significant predictor of annual utility cost[1] in five of the seven cluster cities after controlling for billing year, CFA and the year of occupancy—however the study team was not able to identify if utility costs are influenced by other factors such as differences in appliances that tend to be gas versus electric. To investigate this result, the subcontractor performed additional analysis (see Section 5-Results) to parse total EUI savings by fuel-type. The study team suspects that fuel-type breakdown by system could further clarify why, for example, San Francisco homes data shows relatively little variation in EUI and significant variation in cost.

1 These savings estimates are based on a regression model predicting log(CFA-normalized bills) using CFA, billing year, vintage, and CAHP participation as the independent variables. Weather is accounted for by using a separate regression models for each cluster.

3. STUDY OVERVIEW AND GOALS

The California Advanced Homes Program (CAHP) pays incentives to builders to build homes that exceed the performance requirements of Title 24, Part 6. CAHP homes have the potential to use less energy than homes built strictly to code due to the improvements made to the built asset (envelope, systems) through CAHP that promotes efficiencies at least 10% better than code. However, this performance potential is based solely on the results of an energy simulation model (Energy Pro and Micropas were the simulation models used during 2008 and 2013 code cycles, CBECC-Res for 2016 code cycle).

PG&E commissioned this study to assess actual energy performance of occupied CAHP homes compared with performance of homes not participating in CAHP, built in the same geographic cluster (30-mile radius), normalizing results by conditioned floor area (CFA). The consultant (TRC—the implementer for CAHP) and subcontractor (Resource Refocus—the research group conducting billing analysis) documented the study results of the actual energy use of CAHP homes relative to homes constructed without program influence.

The goals of the study were to:

- Compare the metered energy use of CAHP homes vs non-CAHP homes in PG&E territory while controlling for factors such as climate zone and home size
- Evaluate energy use reductions beyond first year of occupation
- Compare the actual energy use data to the estimates by Title 24 models for CAHP homes
- Support future program improvements to reduce post occupancy energy consumption in program-treated homes
- Develop load shapes of electricity consumption for typical and extreme conditions for summer and winter periods

4. Study Approach and Methodology

The consultant prepared the official research plan and schedule for the analysis. As part of this task and under collaboration requirements from D. 10-04-029[1], the consultant submitted the research plan to review by ZNE/RNC PCG members, California Public Utilities Commission (CPUC) Energy Division (ED) staff, and public stakeholders on December 11, 2017. For the purposes of this project, the Draft Research plan was posted to Basecamp and EnergyDataWeb ("the PDA") by the PG&E Project Manager for review and comment by all stakeholders. The research plan was approved by PG&E and posted on PDA and Basecamp for public review. The research plan received limited comments from stakeholders and the final version was uploaded to EnergyDataWeb after approval by the PG&E Project Manager.

The consultant requested billing data from PG&E, as well as authorization from PG&E to use Title 24 modeling data already in the consultant's possession as implementers of CAHP since 2011. The subcontractor conducted all data analysis and performance comparison between CAHP modeling data and billing data. The subcontractor investigated the relative billed energy use of occupied CAHP homes in comparison to occupied non-CAHP homes (controlling for relevant details). This report includes the subcontractor's un-edited analysis findings (Appendix D). The subcontractor completed all data analysis of this project without direct involvement or direction from the consultant to minimize bias. The consultant as the CAHP implementer did not influence the results and findings of the analysis completed by Resource Refocus.

4.1 Data Collection

The consultant collected CAHP project data from an internal program implementation database. The consultant requested CAHP and non-CAHP project billing and usage data from PG&E.

4.1.1 CAHP Data Collection

The consultant collated the following CAHP details from their existing program database to be shared with the Subcontractor for analysis and comparison:

- Code cycle (2008, 2013)
- Number of bedrooms
- Incentive amount
- Year built
- House size (conditioned floor area)
- Climate zone
- Address
- Title 24 performance modeling results – including compliance percent above code, CAHP Score, and simulated site-energy use by end-use

The consultant reviewed the scope and scale of the program database to begin to cluster homes by climate zone and location. The consultant found 20,751 CAHP homes were built over the period of analysis, spread across 205 zip codes and 140 cities. Figure 2 shows the breakdown of the 20,751 homes across CFA bin and

1 http://www.cpuc.ca.gov/General.aspx?id=4795

climate zone. Climate zones 4, 12 and 13 and home size bins 2000-2500 square feet (sf) and 2500-3000 sf have the highest density of CAHP homes.

CFA vs climate zone	1	2	3	4	5	11	12	13	16	No Data	Total
0-500			1				2	50	3		56
500-1000	25	1	2				37				65
1000-1500		11	28				66	62			167
1500-2000		39	95	265	43	294	958	800	7		2,501
2000-2500		45	178	216		642	1,731	839	1		3,652
2500-3000	1	169	228	1,129	92	586	4,318	2,548			9,071
3000-3500	60	2	43	298		517	1,278	831			3,029
3500-4000				44		279	334	80			737
4000-5000			42	21		91	343	9			506
5000-6000				11		1	4	1			17
>6000			2	2							4
No Data			88	52		203	252	286		65	946
Total	86	267	707	2,038	135	2,613	9,323	5,506	11	65	20,751

Figure 2. CAHP project database statistics by CFA and Climate Zone

4.1.2 PG&E Billing Data Collection

Data Request

The consultant submitted data requests to PG&E to determine data availability and to inform initial sample groups. The consultant requested the following data for all single family homes built on or after January 1, 2011 (see Appendix A – Data Request Memo and Appendix B – Data Request Memo Follow-up for the data requests submitted to PG&E):

- Billing data from the time that the first customer moved in, for current and past PG&E customers, which contains 15-minute (or hourly, if 15-minute data is not available) interval data of kWh consumption as well as the customer's rate plan
- If the house is PV integrated and net-metered, individual consumption and generation data is requested in addition to the net-metered data
- Daily (or monthly) therms consumption data as well as the customer's rate plan

- Utility bill data in dollars

PG&E Data Received

PG&E provided two datasets: Billing and Usage. Billing data contained monthly billing information including total kWh, therms and the cost (dollars) for that billing period. Usage data contained electric and gas information—kWh on 15-minute interval and therms consumption on daily interval. The PG&E billing dataset comprised 3,668,062 records for 78,176 unique UUIDs. This translates to 77,844 unique addresses across 495 zip codes and four climate zones (2, 3, 4, and 13).

Additionally, PG&E provided a match file to convert the UUID to physical address for the entire database. The consultant used an energy direction flag in the interval data to identify homes representing energy flow back to the grid. PG&E marks the energy consumption with either ‘D’ (for delivered energy to the customer) or ‘R” (for received energy from the customer). The consultant thus marked all the accounts with ‘R’ flag to be net metered. All sites with PV generation were omitted from analysis so that the low consumption or potentially negative energy consumption due to PV doesn’t skew the analysis. The billing data did not facilitate an accurate way of identifying the actual energy consumption separately from the PV generation. Hence, all addresses with installed PV were removed from the further analysis.

Sensitive Data Handling

The PG&E usage data involved sensitive information such as customer’s addresses and their energy usage. To maintain customer confidentiality, the consultant transferred the data through secure file transfer protocol (sFTP) from PG&E’s Axway sFTP service. The consultant downloaded the data to review the contents and completeness of the provided information. The usage information was mapped to UUID to maintain anonymity and the consultant converted the UUID to the physical addresses. The consultant reviewed usage and the addresses data in separate files. The consultant encrypted the files with a password, to prohibit access by an unintended user. The consultant deleted the usage and billing data from the initial computer immediately after the initial review. The consultant shared the template of data received with the subcontractor to inform them of the data structure and contents.

The contractor established a cloud location to conduct further analysis on this data. The consultant considered two cloud-based solutions—Amazon Web Services (AWS) and Azure—and ultimately selected AWS as prices were similar and our IT team has prior experience with the platform. In co-ordination with IT, the AWS account was set up for 200 GB of data storage and the consultant uploaded all data to this secure cloud-based machine and shared access with the subcontractor. This AWS solution was secure and provided a platform for easy collaboration between contractor and subcontractor.

Demographics Data

In addition to billing and usage information, the consultant requested demographics-related information (e.g. non-CAHP home floor area, home sales data, resident age, resident occupation etc.). PG&E did not have access to building characteristic (or occupant demographic) information for the comparison sample group. The research plan projected that the consultant could access data sources such as county assessor database and the Experian dataset and that these sources could inform income level and education level. The consultant explored other resources to obtain the building characteristic information, beginning with the county assessor data, but not all counties had made the data publicly available on their websites. Real estate websites like Experian and Zillow contain property related information, but the technical challenges of creating an API to access the data, and the data access restrictions from Zillow ultimately deemed this data source too costly to pursue. The consultant also explored accessing the CalCERTS database (which hosts data about all new construction HERS testing), but it does not provide physical addresses to third parties and thus the database was not useful for this study. The consultant found a data source provided by a firm Attom Data and secured agreement to retrieve building characteristics data for the required addresses. Attom Data provided conditioned floor area (sf), number of bedrooms and number of bathrooms for a list of >10,000 addresses that the consultant identified as the control sample for analysis.

4.2 Sampling Approach

The homes in this study group were constructed across two cycles of the Title 24, Part 6, residential energy code: 2008 and 2013. Each energy code is more stringent than the prior code in terms of permissible energy use.

The consultant categorized the CAHP homes into different sample groups based on climate zone and local cluster centers (<30 miles radius). The consultant prepared a sample group of non-CAHP homes constructed during similar years as the CAHP sample, and within a local radius (after initial analysis the consultant used 30 miles) of a comparable CAHP home cluster. The consultant binned all homes in the non-CAHP homes sample group by home size (sf).

The comparison of CAHP and non-CAHP homes performance was completed within each of the individual groups on an aggregate level.

4.2.1 Clustering

The consultant team clustered the CAHP and non-CAHP homes on their climate zone and distances on a city level. The consultant used only ~50,000 addresses from the PG&E dataset because the remainder addresses were in an unmatched climate zone or incorrect home types (multifamily or mobile homes). The consultant calculated distances between cities based on latitude-longitude. The clustering process involved the following steps.

1. Determine which CAHP houses have billing data by matching addresses from the CAHP project database to those included in the PG&E billing data.
2. Identify cities to be cluster centers amongst CAHP matched homes by manually reviewing the cities that CAHP homes are built in and ensuring that they are within a radius of 30 miles and in the same climate zone.
3. Filter the list of CAHP cluster center cities to remove clusters with less than 100 CAHP homes in its cluster. This step removes clusters with inadequate data for further statistical analysis. The consultant converted all PG&E addresses to latitude-longitude coordinates by uploading batches of 10,000 addresses to census bureau. For addresses that were not matched by the census bureau, the consultant used a paid service (csv2geo) to convert the remaining addresses.
4. Remove CAHP home addresses from the PG&E database to form non-CAHP database based on street address and co-ordinates information.
5. The addresses in the non-CAHP database thus formed were assigned to the CAHP cluster centers identified in step 3 based on the intercity distances being within a radius of 30 miles, while ensuring they are in the same climate zone. The non-CAHP homes that could not be clustered with any of the CAHP cluster centers were removed from the analysis.
6. Append conditioned floor area information to the clustered non-CAHP homes by purchasing data from a third-party entity called ‘Attom data’. As per budget allowance, consultant created a list of non-CAHP addresses based on even distribution across clusters (~10,000 non-CAHP addresses) to purchase information regarding their home size, number of bedrooms and number of bathrooms from Attom Data.

4.2.2 Data Cleaning

The consultant and subcontractor reviewed data provided by PG&E for completeness and developed a database to combine the billing data, home size data and CAHP program data. The subcontractor further filtered the data to ensure the analysis is not skewed due to inconsistency in parameters or incompleteness of data (Additional details in Appendix D Section 11.2).

- House size: Limited to houses with 1,000-5,500 sf of conditioned floor area
- Occupancy: Average power can be less than 110 W for no more than a month
- Completeness: At least one year with 355-375 days of data
- High consumption: Top 1% of consumption per sf is considered as an outlier
- Mixed fuel: Houses with no gas bill data were removed because lack of gas bills does not guarantee that a house is all electric

During the initial analysis for this study, the consultant and subcontractor identified that the data provided by PG&E was incomplete. This discovery occurred months after the data was provided by PG&E and as a result, the consultant and subcontractor developed a data request best practices memo (Appendix C – Data Request Approach and Process) to suggest process improvements for PG&E and third-party data requestors to minimize data integrity problems with future data requests.

4.3 Analysis Approach

The subcontractor used the following methods to address the study questions. All subcontractor analysis and summary data are provided in Appendix D – Subcontractor Unedited Analysis.

4.3.1 Differences in Energy Consumption and Bills

Visual Comparison

The first step in analyzing differences between CAHP and non-CAHP performance is a visual comparison using boxplots. They combine many summary statistics into a single graph: minimum and maximum, first and third quartiles, median, and outliers. Half of the data is contained in the box with a quarter above and a quarter below. In this analysis, a point is an outlier if it is more than 1.5 times the height of the box (interquartile range) below or above the first and third quartiles, respectively.

These graphs consider each cluster separately and use performance information normalized by floor area, but do not control for billing year or house vintage. Each house contributes one data point per complete year of billing data.

Regression

To formalize the findings from the visual comparison, single and multivariable regression models were run.

Linear regression models require that the dependent variable be normally distributed. However, the EUI and bills in this sample do not have a normal distribution, based on visual inspection and the Shapiro Wilk test; instead the distributions are right-skewed. One common approach to overcoming this is to use a log transformation on the dependent variable.[1] Therefore in these regression models the dependent variable is log (EUI) or log (normalized bills). Savings are then calculated as

$$\%\ savings = 1 - e^{regression\ coefficient}$$

For each cluster, five regression models were run

- Single independent variable
  - CAHP participation
  - Floor area (CFA)
  - Vintage[2]
  - Billing year
- Multivariable combining all four independent variables

In all models, vintage and billing year were coded as categorical variables.

4.3.2 Load Shapes

For each cluster, one sample weekday was chosen from 2017 in four categories – typical and extreme summer, typical and extreme winter – based on daily minimum, mean, and maximum outdoor temperatures for the cluster centers as reported by NOAA and Weather Underground. The typical summer day was chosen for having a mean temperature in the middle of the distribution for summer months, and the extreme summer day was chosen based on both mean and maximum. Winter typical and extreme days were chosen analogously.

Hourly load shapes were developed based on the average power of the four 15-minute increments making up that hour. The data was analyzed visually using boxplots to show the variability and compare CAHP vs non-CAHP. To compare typical and extreme days, only the median values were used.

Load shapes were only analyzed visually. These graphs consider each cluster separately, have data from the same days in each cluster, and use power normalized by floor area, but they do not consider house vintage.

4.3.3 Persistence of Performance

To track relative energy consumption of CAHP and non-CAHP houses over time, houses were segmented by cluster and vintage, using the first year of complete billing data as a proxy. Then within each segment of data, the EUI was plotted for each billing year to show the relative performance of CAHP and non-CAHP houses during the first, second, etc. years of occupancy.

4.3.4 Modeled vs. Measured Consumption

For CAHP houses, the results from the Title 24 energy model used to qualify for CAHP were compared to the billing data.

1 Lewis, K et al. 2018. Comparison of energy data for green-certified and non-certified buildings in the 2012 Commercial Buildings Energy Consumption Survey (CBECS). ACEEE Summer Study in Buildings. https://aceee.org/files/proceedings/2018/assets/attachments/0194_0286_000154.pdf

2 Throughout the analysis, the first complete year of billing data was used as a proxy for house vintage.

Visual Comparison

The comparison was first done visually with one scatter plot per cluster, with each house contributing one data point per complete year of billing data. This comparison was done for annual electricity consumption, gas consumption, and EUI. Because each house is only compared to itself, the energy information is not normalized by floor area for the electricity and gas.

The comparison does not control for vintage or billing year.

Regression

To formalize the findings from the visual comparisons, single variable regressions were run for each cluster with measured kWh, therms, or EUI as the dependent variable and modeled kWh, therms, or EUI as the independent variable.

The comparison does not control for vintage or billing year.

5. Results

The consultant provided data including billing data, usage data, home size, quantity of bedrooms and bathrooms for CAHP homes and non-CAHP homes. The subcontractor was responsible for analysis of the datasets to attempt to answer the following questions.

Study Question	Study Result
Are CAHP buildings using less energy than non-CAHP homes?	Addressed
What is the cost to operate difference between CAHP and non-CAHP homes?	Addressed
What are the differences between modeled energy consumption and measured energy consumption of CAHP homes?	Addressed
Do energy use reductions for high performance homes persist beyond the first year of occupation?	Partially addressed
How do daily load profiles of CAHP and non-CAHP houses compare on typical and extreme summer and winter weekdays?	Partially addressed
Are energy efficiency or home design features associated with lower consumption in CAHP homes?	Not addressed

The following sections address each of the study questions and include the statistics discussion to document any additional caveats or limitations to sample accuracy. The subcontractor provided the data analysis and reported findings without edit from the consultant for this report. Selected graphs and tables are shown with the analysis results (unedited and in subcontractor source formatting) in Appendix D – Subcontractor Unedited Analysis.

5.1 Are CAHP Buildings Using Less Energy than Non-CAHP Homes?

The subcontractor provided descriptive statistics about the sample group (CAHP homes) and control group (non-CAHP homes)—the important differences between the two groups are that CAHP homes tend to be newer than non-CAHP homes and tend to be larger or smaller depending on the cluster. Newer homes are built to stricter energy code requirements, and floor area has a large impact on EUI. The clusters are not equivalent populations; Fresno and Bakersfield have the highest population of non-CAHP homes, San Francisco and Santa Rosa have the lowest population of non-CAHP homes. Fresno and Gilroy have the highest population of CAHP homes.

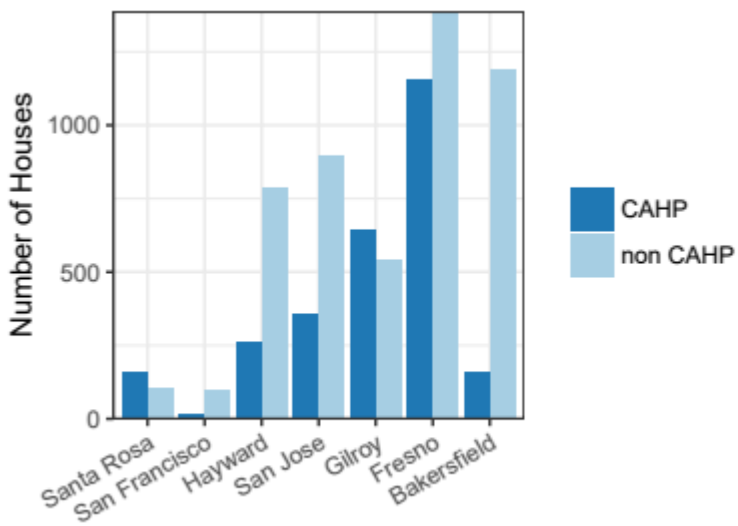

Figure 3. Number of houses in sample and control groups by cluster city

For each home in the sample and control groups, the subcontractor quantified and plotted the energy use intensity (EUI, kBtu/sf) converting electricity and natural gas data into equivalent energy units (kBtu). Figure 4 shows the EUI spread for sample and control groups (each point represents one year of data from one home). The median EUI for CAHP homes (sample) is substantially lower than non-CAHP homes (control) in Fresno and Bakersfield, slightly lower in Gilroy and Hayward and equivalent in San Jose and San Francisco. The median EUI for sample homes is slightly higher than control for Santa Rosa. The most significant finding is that variation within an analysis group is large compared to the variation between sample and control groups.

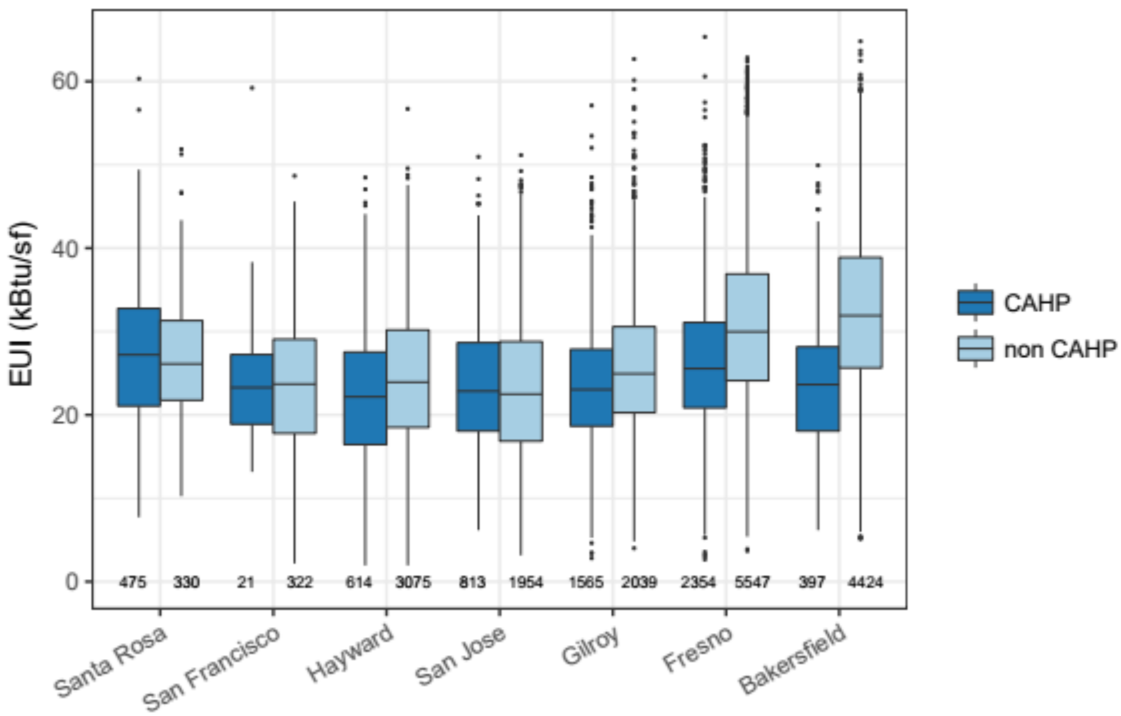

Figure 4. EUI for sample and control groups by cluster city[1]

The EUI differences are not dictated by climate zone: Gilroy and San Jose are both located in climate zone 4, but Gilroy's median CAHP EUI is lower than non-CAHP while San Jose's CAHP EUI is equivalent to non-CAHP; the same is true comparing Hayward and San Francisco. The variation in median EUI between clusters for non-CAHP homes is twice that of CAHP homes – this could be a result of CAHP influence.

	Max Median EUI	Min Median EUI
CAHP (sample)	27 (Santa Rosa)	22 (Hayward)
Non-CAHP (control)	32 (Bakersfield)	23 (San Jose)

Figure 5. Median EUI variation in sample and control groups

The proportion of gas use of total EUI in all groups (sample and control) is greater than 50% except for sample homes in Fresno and Bakersfield. The gas consumption proportion of sample homes is substantially lower in Fresno, Bakersfield and San Francisco. The study cannot identify the source of this finding without further data on the end use systems and appliances in the sample and control groups. This finding could imply that the CAHP changes affect heating and water heating more than other end uses, but then it is surprising that this difference is more present in the cooling-dominated climates than the milder ones. It could also be, for example, that CAHP homes are more likely to have electric appliances than non-CAHP. The analysis would need to separate homes with electric heating and electric water heating from homes with gas appliances to positively identify the cause of the gas proportion results shown in Figure 6.

1 One data point per house per year – the same house contributes multiple data points if we have more than one complete year of data for it. Analysis does not control for vintage but accounts for CFA by looking at EUI and weather by grouping within clusters.

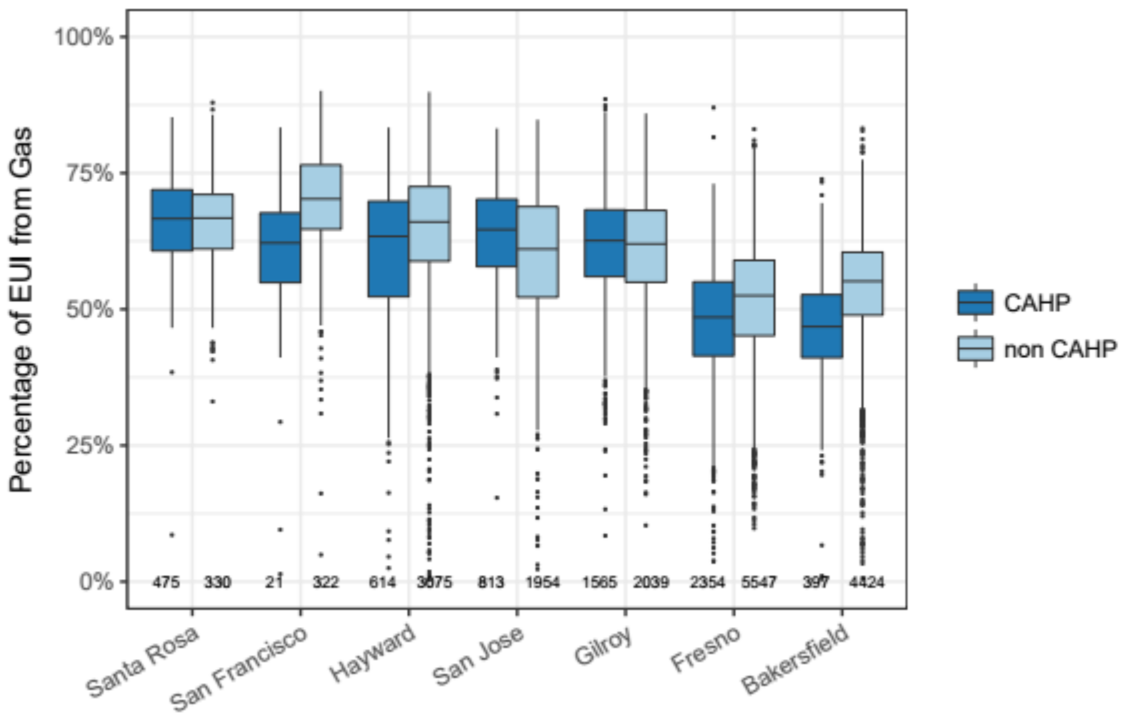

Figure 6. Proportion of gas consumption in total EUI by cluster city[1]

Figure 7 and Figure 8 show the annual electricity consumption and gas consumption (respectively) for sample homes and control homes across the cluster cities. These figures show San Francisco CAHP homes use higher electricity than non-CAHP homes and that CAHP homes in San Francisco, Fresno and Bakersfield use less gas. All other clusters show approximately equivalent consumption between CAHP and non-CAHP homes.

1 One data point per house per year – the same house contributes multiple data points if we have more than one complete year of data for it. Analysis does not control for vintage but accounts for CFA by looking at EUI and weather by grouping within clusters.

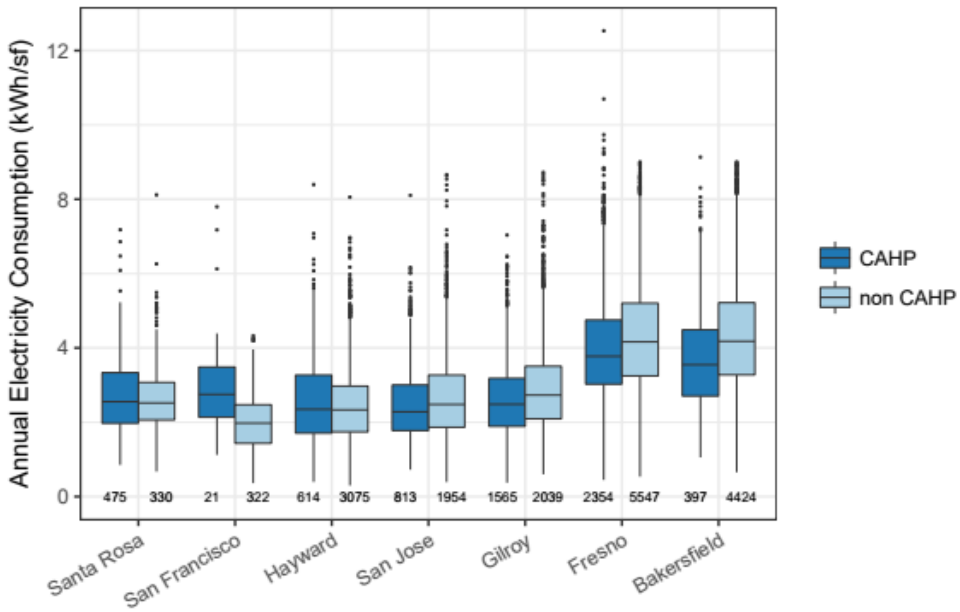

Figure 7. Electricity consumption comparison sample vs control by cluster city[1]

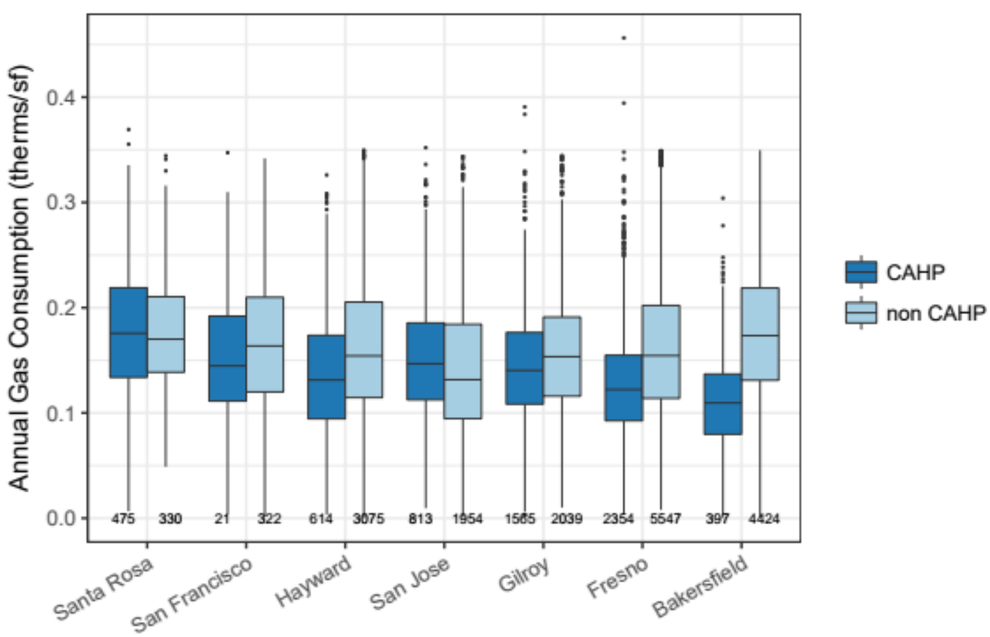

Figure 8. Gas consumption comparison sample vs control by cluster city[2]

1 One data point per house per year – the same house contributes multiple data points if we have more than one complete year of data for it. Analysis does not control for vintage but accounts for CFA by looking at EUI and weather by grouping within clusters.

2 One data point per house per year – the same house contributes multiple data points if we have more than one complete year of data for it. Analysis does not control for vintage but accounts for CFA by looking at EUI and weather by grouping within clusters.

5.2 What is the Difference in Cost to Operate between CAHP and Non-CAHP Homes?

The subcontractor found that CAHP homes and non-CAHP homes have significant differences in cost to operate in most clusters. CAHP participation is statistically significant predictor[1] of annual utility cost in five of the seven cluster cities when controlling for the year the house the built, the billing year, and CFA.

- San Francisco – 30% increase in normalized bills
- Hayward – 4.6% increase in normalized bills
- San Jose – 3.3% increase in normalized bills
- Fresno – 2.6% decrease in normalized bills
- Bakersfield – 11% decrease in normalized bills

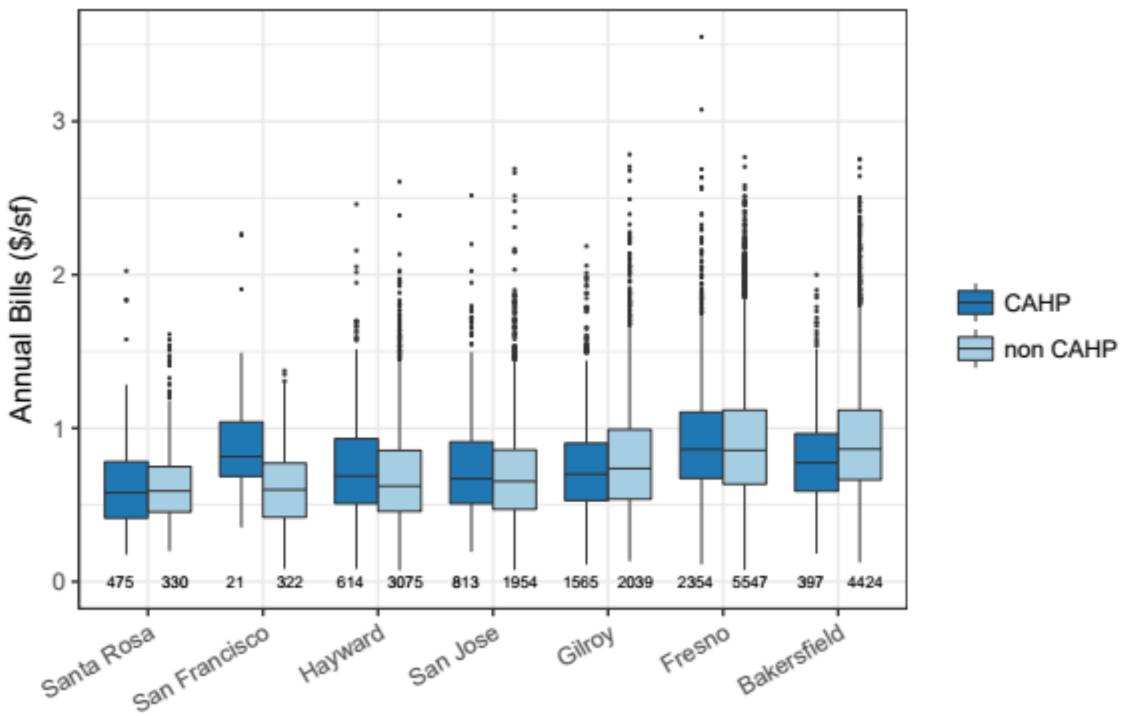

Figure 9. Annual utility (electricity and gas) cost for sample vs control by cluster city[2]

Across the population, electricity is more expensive than gas. This could explain the higher cost to operate a CAHP home in San Francisco, because a higher percentage of EUI comes from electricity in CAHP homes and the overall cost is higher. In Fresno and Bakersfield, CAHP homes tend to have lower EUI and (as shown in Figure 6) the majority of EUI savings comes from electricity, and cost to operate is equivalent in Fresno and only slightly lower in Bakersfield.

1 These savings estimates are based on a regression model predicting log(CFA-normalized bills) using CFA, billing year, vintage, and CAHP participation as the independent variables. Weather is accounted for by using a separate regression model for each cluster.

2 One data point per house per year – the same house contributes multiple data points if we have more than one complete year of data for it. Analysis does not control for vintage but accounts for CFA by looking at EUI and weather by grouping within clusters.

5.3 What are the Differences in Modeled Energy Consumption and Measured Energy Consumption of CAHP Homes?

To answer this question, the subcontractor compared the modeled and measured consumption of CAHP homes by fuel type. The results show a much wider range of measured consumption than modeled consumption for both fuel types. For example, in Bakersfield the difference between the highest and lowest consumer was 14,000 kWh/yr. for the measured data and only 8,000 kWh/yr. for the modeled data. This difference is more extreme in other clusters. Across all clusters, the model underestimated electricity consumption in 94% of cases (Figure 10).

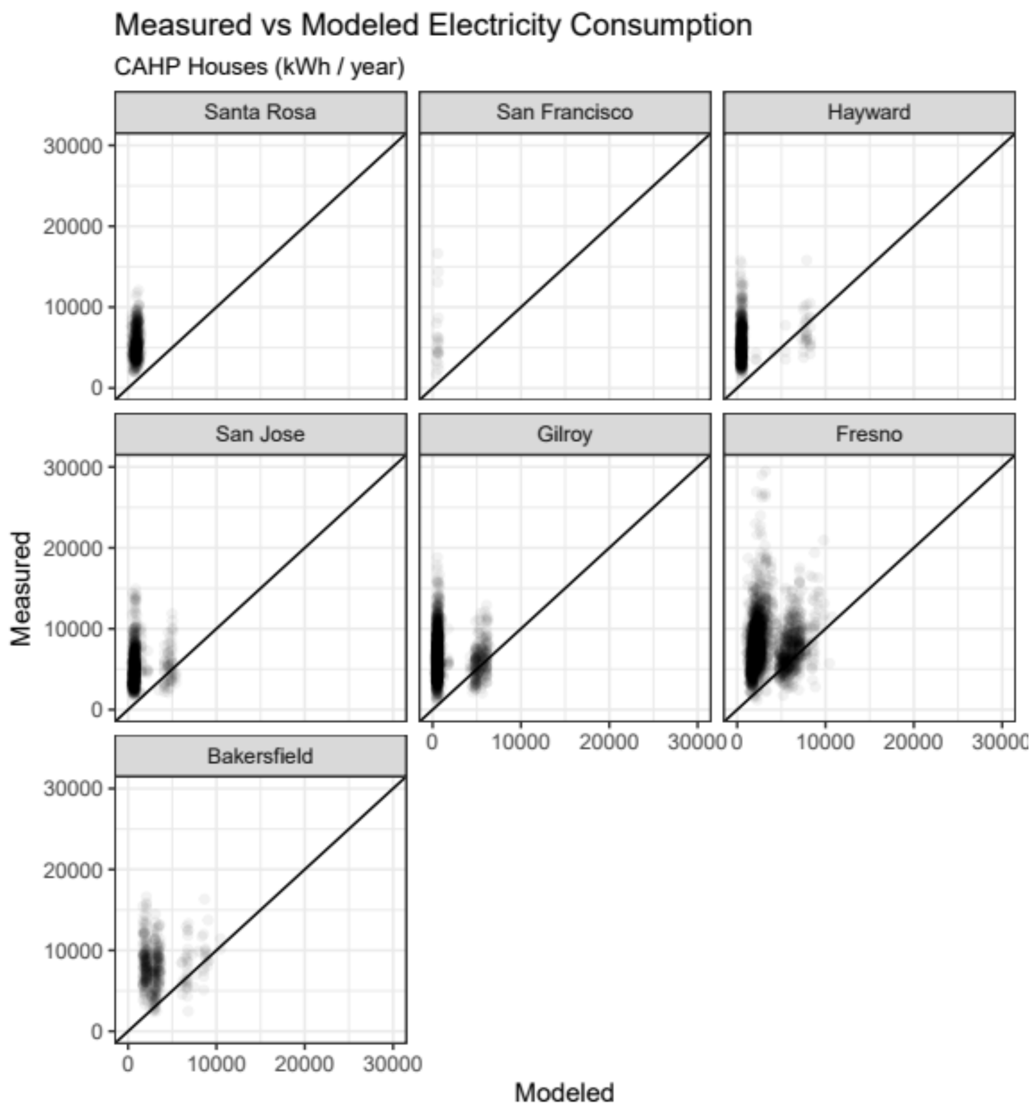

Figure 10. Measured vs modeled electricity consumption for CAHP homes[1]

1 This analysis shows one data point per house per year, without controlling for vintage or billing year. Comparing each house to itself controls for CFA.

Similarly, for gas consumption the variation in measured results is wider than the variation in modeled results. Unlike electricity consumption findings, the model overestimated gas consumption 83% of cases (Figure 11).

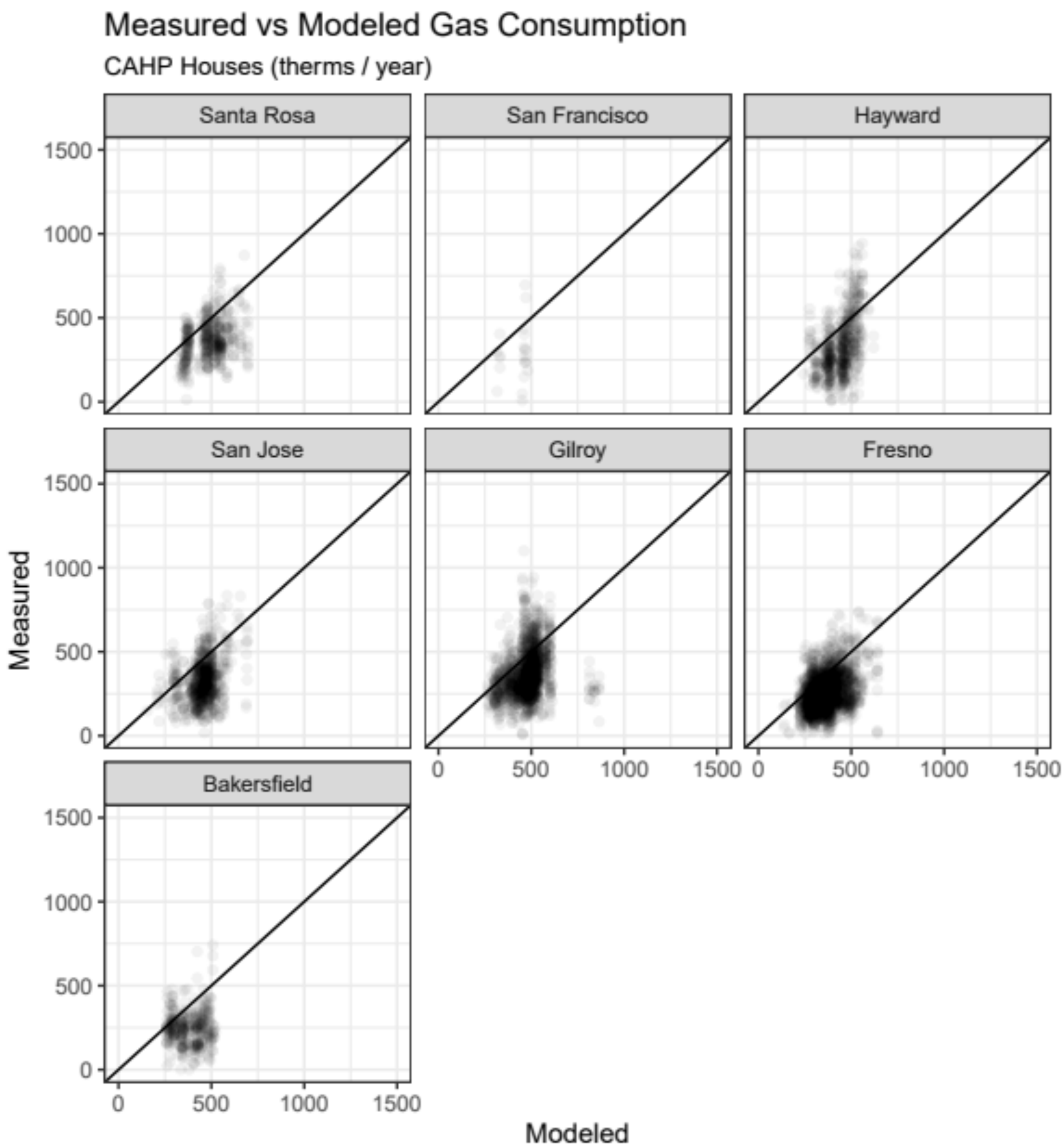

Figure 11. Measured vs modeled gas consumption for CAHP homes[1]

1 This analysis shows one data point per house per year, without controlling for vintage or billing year. Comparing each house to itself controls for CFA.

5.4 Load Profiles

With 15-minute interval data, the subcontractor was able to prepare load profiles[1] to observe the trends controlling for factors such as house size, climate region for four weekdays in 2017: typical summer, extreme summer, typical winter and extreme winter[2]. We aggregated all load profiles for a given cluster to produce the median load profiles. The data show CAHP homes have lower peak demand and lower afternoon ramp in cooling dominated clusters (Fresno and Bakersfield). The results also show that the difference between CAHP homes and non-CAHP homes are negligible for the other clusters, except on extreme summer days in Santa Rosa and Hayward where the median CAHP home has a higher peak demand than non-CAHP homes.

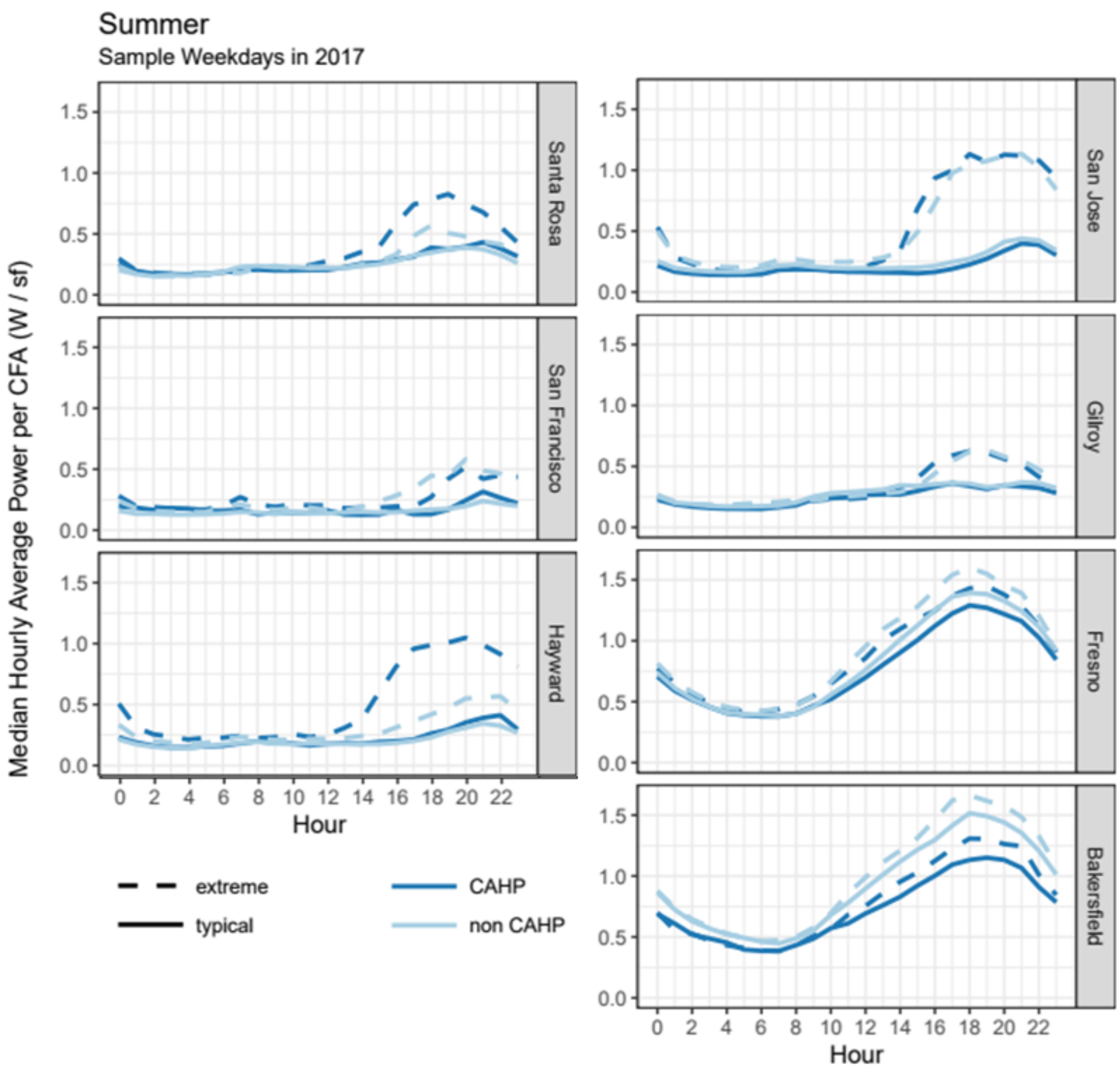

Figure 12. Median load profiles for summer by cluster city[3]

1 These graphs have one data point per house and do not control for vintage. Power is normalized by CFA to account for house size, and weather is controlled for by looking within clusters on a single day.

2 The actual days chosen to vary by cluster based on measured weather.

3 Each house contributed one data point to calculating the median, without controlling for vintage. Power is normalized by CFA to account for house size, and weather is controlled for by looking within clusters on a single day.

For the summer days (both typical and extreme), the larger variations in hourly average power per sf (W/sf) occur in the later afternoon and evening, and the difference between CAHP and non-CAHP homes is small. For Bakersfield, the CAHP homes have lower median power requirements throughout the day including peak afternoon periods in summer (Figure 13). The winter load profiles for CAHP and non-CAHP homes are more closely aligned indicating that the influence of CAHP leads to homes with lower peak demand during summer (cooling-dominated) periods and similar power demands of non-CAHP homes for winter (heating-dominated) periods. The load profiles for all clusters are available in Appendix D. The variation within analysis groups is significant and the graphs shown below are cropped because there are some extreme outliers.

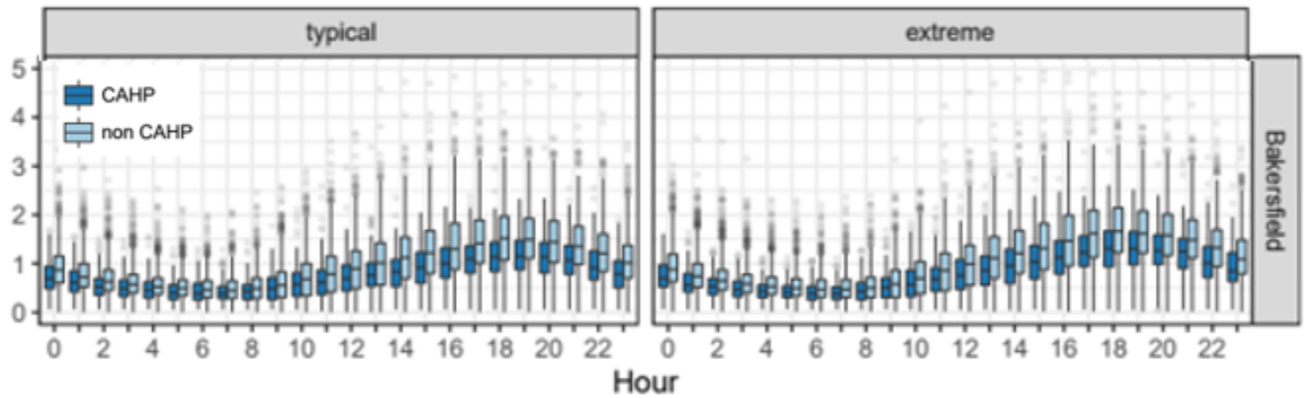

Figure 13. Summer Load Profile (W/sf): Bakersfield

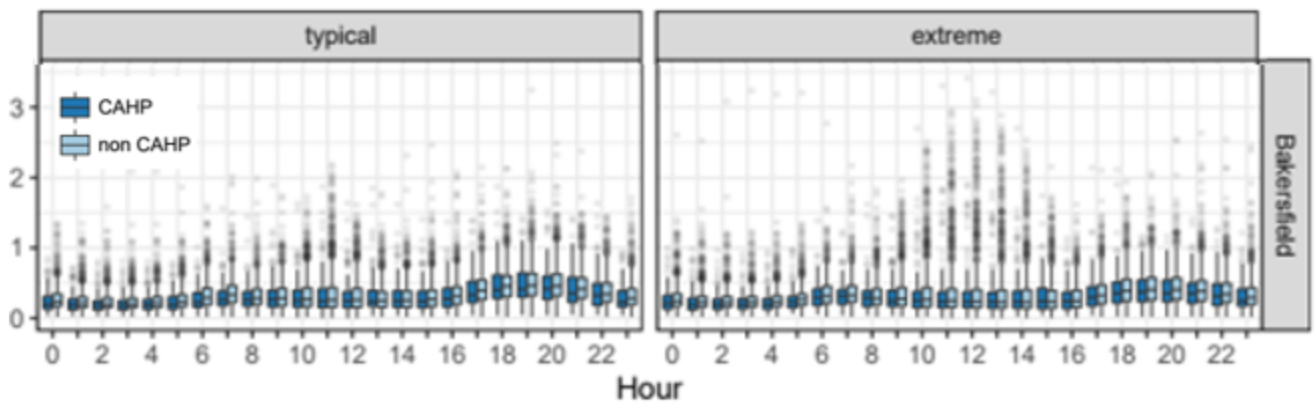

Figure 14. Winter Load Profile (W/sf): Bakersfield

We see similar results for the other cooling dominated cluster (Fresno). The Fresno results are consistent with Bakersfield results however the difference between CAHP and non-CAHP is less extreme in Fresno than in Bakersfield. Fresno and Bakersfield show the median CAHP home load profile is markedly lower than the non-CAHP homes. This influence is attributed to the program requirements to must meet or exceed the Title 24 requirements for envelope performance, efficient air conditioning systems, envelope tightness (including ducts), efficient lighting (which reduces indoor heat gain) and window solar heat gain coefficient (SHGC).

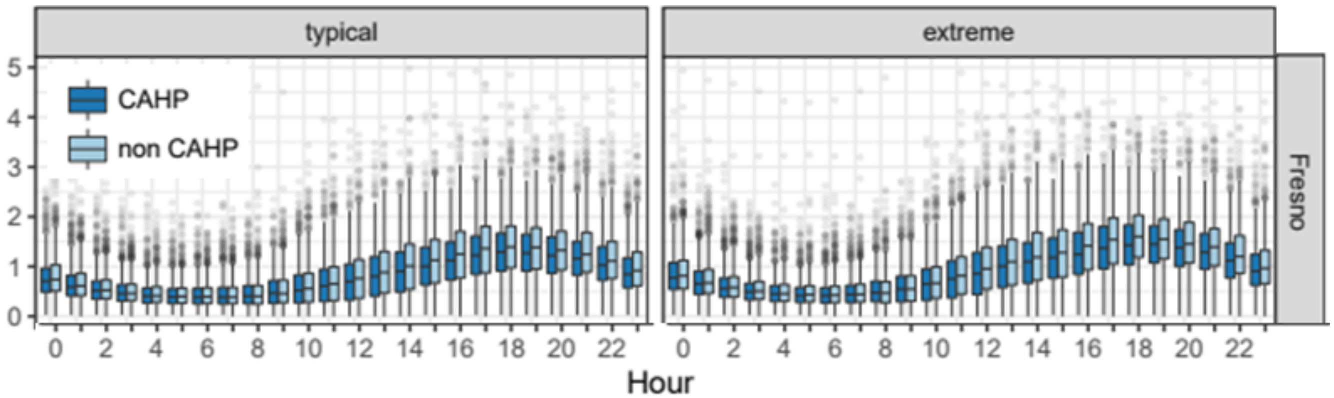

Figure 15. Summer Load Profile (W/sf): Fresno

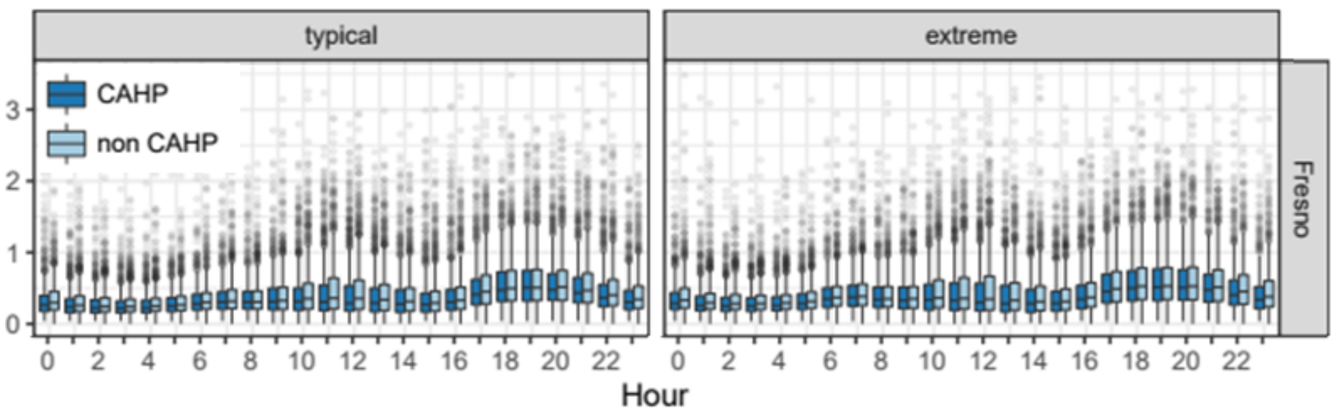

Figure 16. Winter Load Profile (W/sf): Fresno

CAHP peak demand is higher than non-CAHP in Santa Rosa and Hayward, which could indicate that CAHP houses are more likely to have cooling than non-CAHP houses (since AC isn't standard in those cluster's climate zones). We would need to document the end use systems and fuel type in the population groups to investigate the cause further.

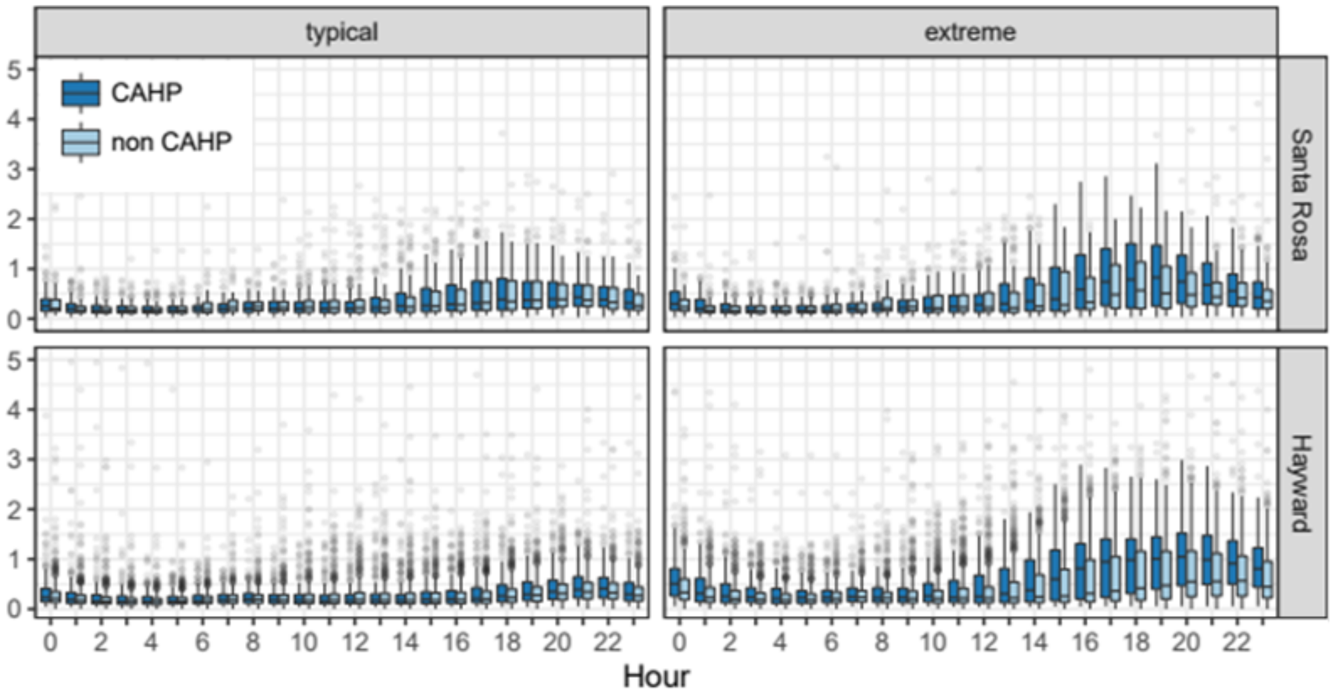

Figure 17. Summer Load Profile (W/sf): Santa Rosa and Hayward

During extreme summer days outside of Fresno and Bakersfield, the median power draw is basically flat until early to mid-afternoon and the evening peak is earlier than typical summer day (around 6-7pm vs 8-9 pm), lasts longer (4-5 hours vs 1-2 hours), and begins ramping sooner (around 3 pm). Overall the differences between the typical and extreme days are much larger than differences between CAHP and non-CAHP.

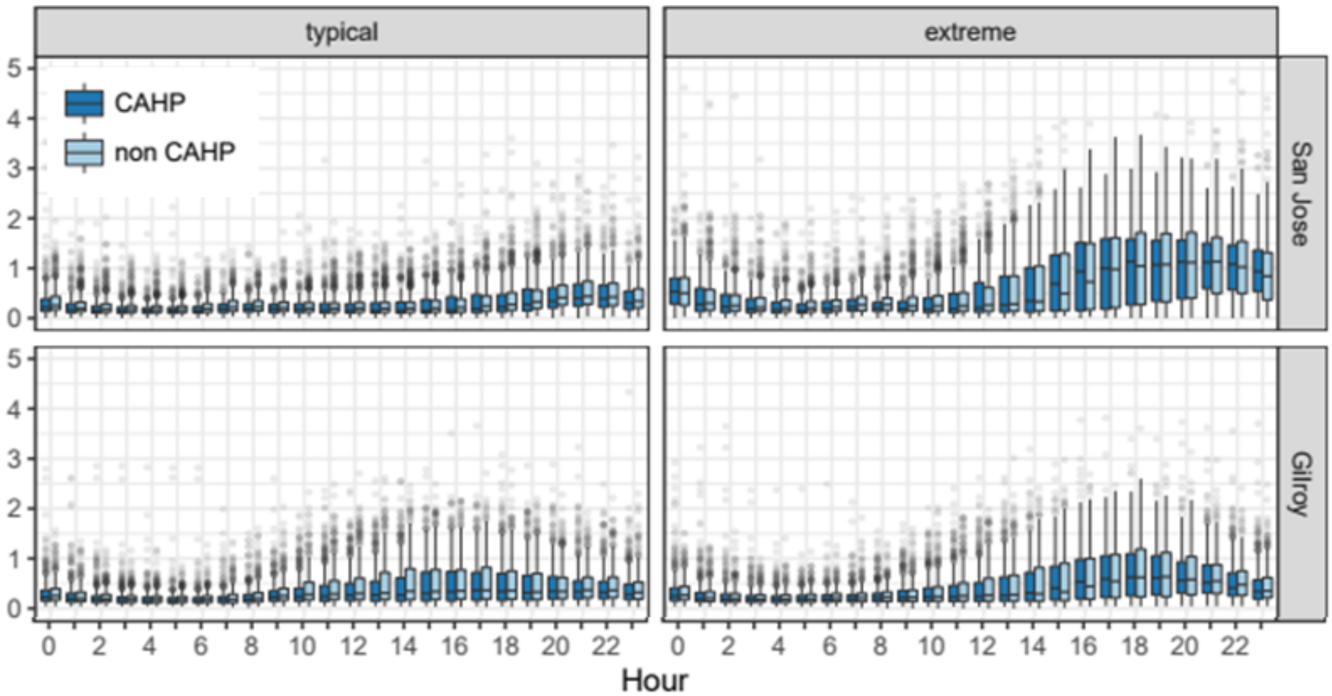

Figure 18. Summer Load Profile (W/sf): San Jose and Gilroy

The hourly load profiles for winter are shown in Appendix D. The variation between CAHP and non-CAHP homes is small and the difference between CAHP and non-CAHP homes is tiny compared to variation within each study group. The variation in hourly average power per sf is more constant throughout the day than for the summer days, although it is still slightly higher in the evening in general.

The median hourly load profiles in winter (Figure 19) show an early morning peak which is not present in the summer median profiles. The maximum winter peak is much lower than summer across all clusters. Typical and extreme winter days show consistent median power except in San Francisco and Bakersfield.

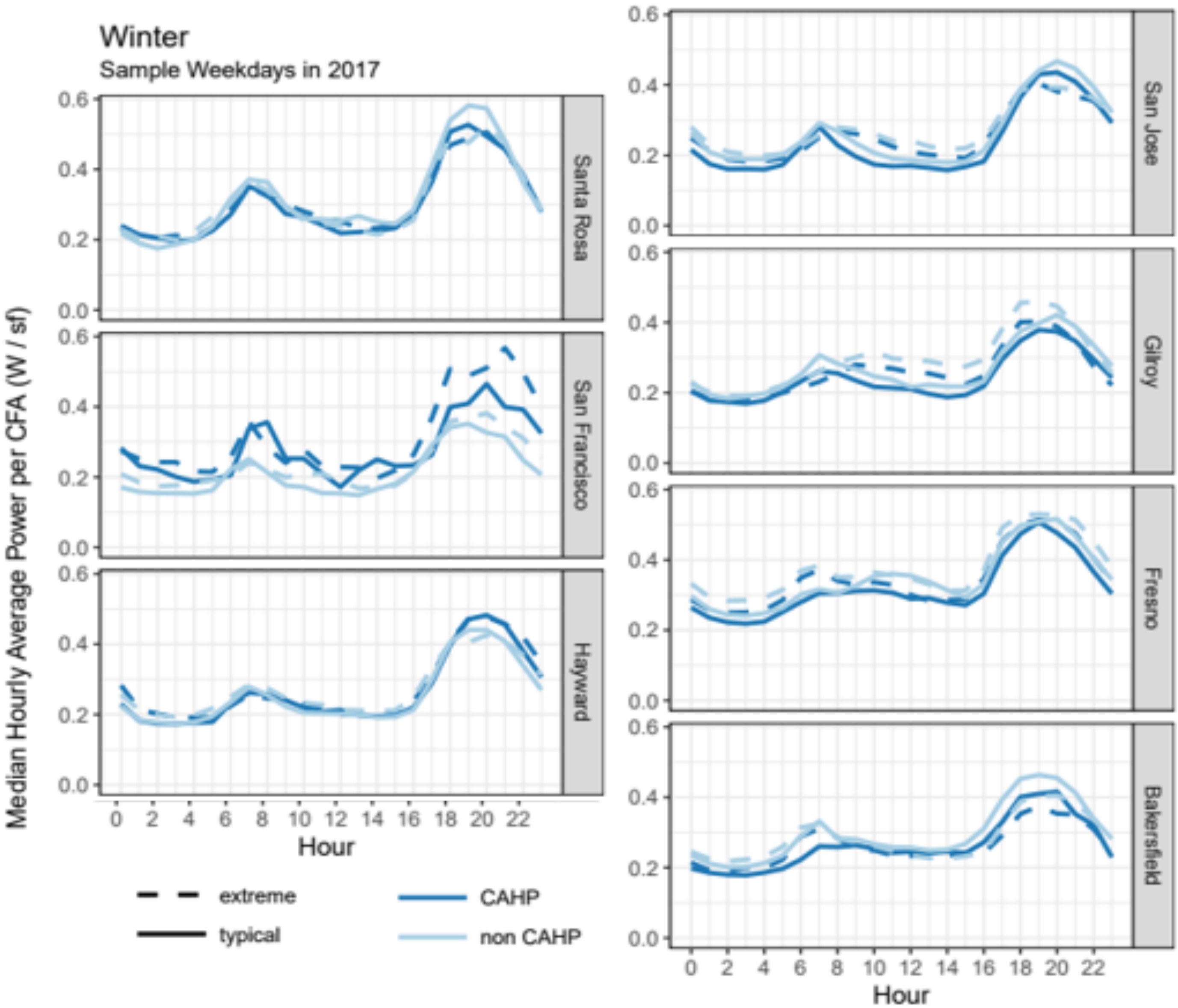

Figure 19. Median load profiles for winter by cluster city[1]

1 Each house contributed one data point to calculating the median, without controlling for vintage. Power is normalized by CFA to account for house size, and weather is controlled for by looking within clusters on a single day.

5.5 Study Limitations

The study results were limited due to the following factors:

- The study team was not be able to control for the demographics of those who purchase CAHP homes such as income level, age of occupants, family size, or cultural background.
- The consultant and subcontractor could not disaggregate consumption data for consumption and generation data. This data source limitation truncated the population of this study to contain only homes with no solar PV systems integrated.
- The challenge of obtaining measure-level data for all homes limited the study. Future studies could inform program design if the measure-level analysis could correlate specific energy measures to the energy savings demonstrated by billing data.
- This study only covers homes in PG&E territory, which does not contain all 16 California climate zones.
- The scope of this study only analyzed billing data and did not include resident surveys, therefore analysis to identify whether resident knowledge of the performance of the home has an impact on energy consumption was outside the scope of this research.
- The study team did not perform analysis to conduct gas end use appliance accounting. Any systematic difference between which equipment and appliances are gas in CAHP and non-CAHP houses will greatly impact our results.
- The study team omitted several projects from Fresno and Bakersfield clusters because we could not positively determine if homes with only electric service from PG&E were all-electric homes, or if the homes received gas from Southern California Gas (SCG). We recommend obtaining (SCG) billing and usage data to combine with PG&E billing and usage data to get more accurate picture of Fresno and Bakersfield results.

6. Conclusions

The purpose of the CAHP program is to motivate the building industry to adopt building practices which result in better performing buildings. Electricity and natural gas savings were the metric of success of the program during the 2008 and 2013 code cycles (the period of analysis), but as the Title 24 code moves to net-zero the energy savings opportunities vanish. CAHP can influence the load profile of energy use favorably, thereby delivering grid benefits of reduced peak demand and more consistent power draw throughout the day. The data show CAHP influence has produced more favorable load profiles especially in cooling dominated climates. The program should be considered a viable approach to flattening the duck curve and reducing electricity consumption during the hours when electricity generation is most costly (and most GHG-intensive). The avoided costs of these operational changes to the grid are significant.

CAHP participation is a statistically significant predictor of EUI in 3/7 clusters (San Jose 4.1% increase, Fresno 4.6% decrease and Bakersfield 15% decrease). Home size (sf) is the variable that explains most of the variability in EUI. CAHP is a statistically significant predictor of normalized bills ($/sf) in 5/7 clusters:

- San Francisco – 30% increase in normalized bills
- Hayward – 4.6% increase in normalized bills
- San Jose – 3.3% increase in normalized bills
- Fresno – 2.6% decrease in normalized bills
- Bakersfield – 11% decrease in normalized bills

Cost increase or decrease compared to non-CAHP (control group) homes is correlated to the proportion of electricity consumption. San Francisco for example shows a higher proportion of electricity consumption and higher normalized bills (similar trend is true for Hayward and San Jose clusters). Time-of-use utility rates (where the consumer pays a higher price during certain hours) may impact the cost to operate CAHP and non-CAHP homes.

Differences (or lack thereof) between CAHP and non-CAHP are relatively stable across years within clusters and vintages (see Appendix D Section 11.7). The data show that CAHP home performance is more persistent than non-CAHP home performance.

Energy modeling predicts EUI better than gas consumption and much better than electricity consumption. Actual consumption variations were much a greater range than simulated predicted consumption range (i.e. the outliers were much wider apart in actual consumption data than in predicted consumption data).

Load profiles show that in summer, the difference between extreme and typical days is much greater than the difference between CAHP and non-CAHP homes. For winter periods, the load shape is largely consistent between extreme and typical days. For the typical summer and winter day, CAHP homes have lower peak power in Fresno and Bakersfield and higher peak power in San Francisco. For extreme summer days, CAHP homes have higher peak power in Santa Rosa and Hayward.

7. RECOMMENDATIONS

This study attempted to correlate specific efficiency measures to measurable energy savings among the CAHP participating homes, but due to limited sample sizes the study team was not able to assess the impact of individual measures. The consultant has energy simulation models of all CAHP projects, but we could not parse individual measures out of each simulation file within the budget of this study. More recent projects (using energy simulation tool CBECC-Res) have data output files that can be parsed more easily, but the sample of projects in that group of projects was not large enough to include in this study. To answer this question, the consultant would need to compare homes features from energy simulation output files with homes features verified by third party Home Energy Rating System (HERS) raters. This study would provide more granular feedback about the specific home features correlate to energy consumption and load profile.

The data show that CAHP home load profiles have significantly reduced afternoon demand, the consultant would recommend further study of this result. If CAHP homes and non-CAHP homes were compared using only the hottest 15 days of the year we could quantify the impact that CAHP program influence has on the most expensive electricity used in the year. The study team recommends further research on this result because the load shape performance of CAHP homes could be a significant approach to reducing cost and GHG-emissions from electricity generation during peak days and hours.).

The consultant does recommend a future study be funded to evaluate impact of resident demographics on energy consumption in CAHP and non-CAHP homes. We would like to address the question: which occupant behaviors correlate with energy consumption patterns in CAHP and non-CAHP homes? This study could not explain the reasons why EUI, or utility cost variations occurred, but we hypothesize that occupant demographics differences between San Francisco and other clusters may be one significant factor driving the result that San Francisco shows the highest utility spend difference between CAHP and non-CAHP homes of any cluster. We suspect that EUI results may also be influenced by demographics—if certain occupants are less cost constrained, perhaps they use energy more frequently and have a higher plug load density. The demographics study could help CAHP predict use patterns and improve performance within certain customer segments with targeted program treatment.

The data show that the energy models do not accurately predict consumption. Predicting energy consumption in new construction is challenging because the occupant's behavior (and normal operating conditions) are unknown. Currently there is no routine feedback for the energy model software to learn from actual use patterns. We recommend further study to investigate whether energy simulation can more accurately predict the pattern of energy use in homes, or some controls mechanism to influence occupant behavior.

The study team did not perform analysis to conduct gas end use appliance accounting. Any systematic difference between which equipment and appliances are gas in CAHP and non-CAHP houses will greatly impact our results. We recommend future study to rerun the analysis presented in this project by clustering homes by fuel-type and by end use. We expect that the EUI and utility cost performance of homes is influenced by whether the water heater and the heating system uses electricity or natural gas. We would expect to find that the results would show that homes with natural gas water heating cost less money to operate throughout the year.

8. Appendix A – Data Request Memo

January 25, 2018

To: Eric Panlasigui (PG&E)

Cc: Anna LaRue, Margaret Pigman (Resource Refocus), Michelle Ortland, Sylvia Lau, Conrad Asper and Lucy Morris (PG&E)

From: Michael Maroney, Avani Goyal, Matthew Christie, Abhijeet Pande (The consultant)

Re: Billing data request for California Advanced Homes Program (CAHP) Billing Analysis Research

Overview of Data Request

This memo is the first data request for the PG&E California Advanced Homes Program (CAHP) Billing Analysis Research project. The consultant (Consultant) in collaboration with Resource Refocus (Subcontractor) will analyze the impact of CAHP participation on metered home energy use compared with energy use of homes that did not take part in CAHP by comparing actual billing data. The research analysis will be supported by a large existing dataset of CAHP home code compliance energy models of homes built in PG&E territory during the 2008 and 2013 code cycles, as well as metered billing data from PG&E. The billing data for CAHP and non-CAHP homes will be normalized on their climate zone and house sizes for a fair comparison.

The consultant requests PG&E to supply billing data information of single-family homes served by PG&E to support the CAHP billing analysis of CAHP and non-CAHP homes. This includes electricity and gas consumption information of all homes built on or after January 1, 2010 (as identified by the first account created on date) in the following climate zones: 2, 3, 4, 11, 12, and 13.

The database developed from this Data Request will be shared with Resource Refocus, the Subcontractor.

Request Details

PG&E will provide the following residential customer billing data:

- Billing data from the time that the first customer moved in, for current and past PG&E customers, which contains 15-minute (or hourly, if 15-minute data is not available) interval data of kWh consumption as well as the customer's rate plan
- If the house is PV integrated and net-metered, individual consumption and generation data is requested in addition to the net-metered data
- Daily (or monthly) therms consumption data as well as the customer's rate plan
- Utility bill data in dollars

For each data set, PG&E will filter the results based on the following:

- Include only single-family homes
- Filter out homes built before 2010

9. Appendix B – Data Request Memo Follow-up

September 13, 2018

To: Eric Panlasigui (PG&E)

Cc: Anna LaRue, Margaret Pigman (Resource Refocus), Josh Harmon, Brian Smith, Doreen Caruth, and Conrad Asper (PG&E)

From: Michael Maroney, Avani Goyal, Matthew Christie, Abhijeet Pande (The consultant)

Re: Billing data request for California Advanced Homes Program (CAHP) Billing Analysis Research

Overview of Data Request

This memo is a follow-up data request for the PG&E California Advanced Homes Program (CAHP) Billing Analysis Research project. The consultant (Consultant) in collaboration with Resource Refocus (Subcontractor) will analyze the impact of CAHP participation on metered home energy use compared with energy use of homes that did not take part in CAHP by comparing actual billing data. The research analysis will be supported by a large existing dataset of CAHP home code compliance energy models of homes built in PG&E territory during the 2008 and 2013 code cycles, as well as metered billing data from PG&E. The billing data for CAHP and non-CAHP homes will be normalized on their climate zone and house sizes for a fair comparison.

The consultant requests PG&E to supply billing data information of single-family homes served by PG&E to support the CAHP billing analysis of CAHP and non-CAHP homes. This includes electricity and gas consumption information of all homes built on or after January 1, 2010 (as identified by the first account created on date) in the following climate zones: 2, 3, 4, 11, 12, and 13.

The database developed from this Data Request will be shared with Resource Refocus, the Subcontractor.

Request Details

PG&E will provide the following residential customer billing data:

- Net energy meter agreement flag – to verify whether a customer has rooftop solar installed
- Net energy meter agreement date – to determine when the rooftop solar was installed and connected

For each data set, PG&E will filter the results based on the following:

- Include only single-family homes
- Filter out homes built before 2010

10. Appendix C – Data Request Approach and Process

May 31, 2019

To: Josh Harmon (PG&E)

From: Mike Maroney, Abhijeet Pande (The consultant)

CC: Margaret Pigman (Resource Refocus)

Data Request Approach and Process

The consultant requested the following data for all homes in PG&E service territory built after January 2010: kWh, kW, therms, electricity cost, natural gas cost and date account created on. The consultant requested the data using a wide filter and criteria selection to be able to match addresses from PG&E to address records that the consultant manages to implement the California Advanced Homes Program (CAHP). The consultant reviewed the initial data provided by PG&E and followed-up to request a file to be able to translate account ID into physical address. The consultant needed the physical address to match records between PG&E data and the consultant data.

After beginning analysis, the consultant identified that approximately 11,000 records could not be matched to the data provided by PG&E. The consultant discovered this data gap after Resource Refocus started the data analysis process and halted the analysis. The data request and analysis process should ideally happen in sequence. To ensure that future data requestors receive complete and accurate data, the consultant recommends the following improvements to this data request process:

- PG&E should provide a list of fields available for data request. For example, the data requests asked for CFA and income level of occupants. If PG&E could identify that these data are not available, we could have started working earlier on the plan for getting the additional data from another source.
  - PG&E should provide documentation of the various fields in the data. For example, PG&E should define how the “energy direction” field in the interval data should be interpreted.
- The data requestor should confirm with PG&E that the data requested can be pulled as expected. The data requestor should identify for PG&E the essential variables requested and the optional variables requested. This update would have identified that the address to account ID map would be necessary for this project.
  - If appropriate, PG&E should consider providing the data request script, or query to the data requestor to confirm that the data request intent is captured correctly and completely.
- Once data is provided by PG&E, the first step is that the data requestor should confirm that all the expected variables exist within the dataset provided, and confirm which units are used to capture the data. For example, the data requestor should confirm that electricity usage units are documented clearly.
  - PG&E should provide a contact person assigned to interact with each data request so that the requestor can ask follow-up data questions.
  - As part of each data request project, PG&E should ask the data requestor to document the questions asked and the answers. PG&E should integrate this information into an FAQ document to send with future data requests. One question we have: why do some accounts have greater than or fewer than 365 days in a given year?
- Once variables are confirmed to exist and the units are understood, the data requestor should confirm that the data request filters bound the data appropriately and does not constrain the dataset

requested. For example, if the data request is asking for records from all PG&E service territory, the data requestor should first confirm that the addresses are distributed appropriately, and if the range of data is unexpected, notify PG&E. This change would have identified that certain cities and regions of PG&E territory were omitted from the data provided for this project.

- If appropriate, PG&E should consider requesting the data requestor provide a disposition report that outlines the data variables included and the number of records of data supplied.

♦ Lastly, the data requestor should review that the data provided by PG&E is complete by comparing the number of records expected to the data provided. When a data requestor does not have insight into the expected number of records expected, PG&E should consider asking the data requestor to provide a minimum threshold of number of records to perform analysis.

- During analysis, PG&E should ask the data requestor to document data anomalies. For example, PG&E filtered out most of the multifamily units, but there were still multifamily records in the data. After the data requestor thoroughly reviews the data and develops the initial sample for analysis, we can confirm the actual sample size based on relevant records in the data. For example, we have removed all multifamily (apartment units) records and all records that have null values (indicating that the residence is unoccupied).

The data requestor should produce a data integrity report for all data requests documenting the screening and filtering analysis, the completeness of data and the data match rate once the data requestor develops the initial sample for the analysis.

11. Appendix D – Subcontractor Unedited Analysis

This appendix contains a relatively informal description of the analysis and a high-level view of the process employed.

11.1 Key Takeaways

11.1.1 Overall

- Differences between CAHP and non-CAHP (energy consumption, bills, load shapes) are small compared to the within-group variation
- The same or similar asset yields a wide range of actual performance in terms of electricity consumption (kWh), gas consumption (therms), EUI, and bills

11.1.2 Are CAHP buildings using less energy than non-CAHP homes? What is the cost to operate difference between CAHP and non-CAHP homes?

- Small decrease in EUI and bills in Fresno and Bakersfield for CAHP compared to non-CAHP
- CAHP participation is a statistically significant predictor of EUI in 3/7 clusters after controlling for CFA, billing year, and vintage
  - San Jose – 4.1% increase in EUI
  - Fresno – 4.6% decrease in EUI
  - Bakersfield – 15% decrease in EUI
  - CFA is the single variable that explains most of the variability in EUI
- CAHP participation is a statistically significant predictor of normalized bills ($/sf) in 5/7 clusters after controlling for CFA, billing year, and vintage
  - San Francisco – 30% increase in normalized bills
  - Hayward – 4.6% increase in normalized bills
  - San Jose – 3.3% increase in normalized bills
  - Fresno – 2.6% decrease in normalized bills
  - Bakersfield – 11% decrease in normalized bills
  - The direction of the difference is related to how much of the difference between CAHP and non-CAHP consumption is from electricity vs gas
  - The billing year is the single variable that explains most of the variability in normalized bills

11.1.3 Do energy use reductions for high performance homes persist beyond the first year of occupation?

- Differences (or lack thereof) between CAHP and non-CAHP houses as whole groups are relatively stable from year to year.

11.1.4 How do daily load profiles of CAHP and non-CAHP houses compare on typical and extreme summer and winter weekdays?

- Summer – difference in median load shape (magnitude, peak times) between extreme and typical days is much greater than the difference between CAHP and non-CAHP, except in Bakersfield where the differences are approximately equal.
- Winter – not much change in median load shape between typical and extreme days.
- Typical summer and winter days – very little difference in median load shape (magnitude, peak times) except for summer in Fresno and Bakersfield (CAHP has lower peak) and winter in San Francisco (CAHP has higher peak).
- Extreme summer days – in Santa Rosa and Hayward CAHP median power is substantially higher than non-CAHP on peak. Are CAHP houses in those clusters more likely than non-CAHP to have air conditioning? The climate is mild enough that many houses don't have cooling.

11.1.5 Does modeled performance predict measured performance of CAHP homes?

- Electricity – modeling does not predict measured electricity consumption well
  - Regression model accounted for 0.3-4% of the variation in measured consumption depending on the cluster
  - Actual consumption was higher than the prediction 94% of the time
- Gas – modeling predicts measured gas consumption slightly better than electricity consumption in 6/7 clusters
  - Regression model accounted for 0.1-23% of the variation in measured consumption depending on the cluster
  - Actual consumption was lower than the prediction 86% of the time
- EUI – modeling predicted EUI better than electricity or gas on its own in 6/7 clusters. The overprediction of gas consumption partially makes up for the underprediction of electricity consumption. Nevertheless, the model still underestimated EUI 60% of the time.
- The variation in predicted electricity and gas consumption as well as EUI was much smaller than the observed variation.

Notes

- One of the original research goals was to correlate energy use with energy efficiency features of CAHP homes. However, due to time and budget constraints this question was not addressed.
- All the billing data is from PG&E, but significant portions of the Fresno and Bakersfield clusters are in Southern California Gas territory. Given the complexity of the territory boundary, we cannot accurately distinguish between houses that only have electric bills because they are all electric and houses that only have electric bills because they are in SCG territory. In the other 5 clusters the sample size of all electric houses is too small for robust analysis, so the analysis in this report is limited to houses for which we have both electric and gas bills.
- A major limitation is that we don't know if there's a systematic difference between which equipment and appliances are gas in CAHP and non-CAHP houses. Gas has a large impact on EUI, so differences in gas equipment and appliances will impact our results.

11.2 Data Preparation

11.2.1 Clustering

Cluster centers were chosen by mapping the addresses of CAHP houses that we have data for and picking a major city in the areas with concentrated building.

Cities with CAHP houses were then assigned to a cluster center based on geographic proximity and climate zone. Clusters have a radius up to 30 miles.

After clusters were chosen based on CAHP houses, non-CAHP addresses were selected in each cluster.

	Unused – too few CAHP homes	CZ	Matched CAHP homes	Non-CAHP homes in sample
Redwood City	x	3	1	
Salinas	x	3	9	
Santa Cruz	x	3	20	
Paso Robles	x	4	27	
Napa	x	3	28	
San Francisco		3	99	126
Santa Rosa		2	207	254
Hayward		3	413	1,221
San Jose		4	684	1,387
Bakersfield		13	906	2,630
Gilroy		4	986	706
Fresno		13	2,537	2,400

Figure 20. CAHP cluster centers and sample sizes

Climate description

The "Guide to California Climate Zones"[1] lists heating and cooling degree days (base 65 °F) for the "reference city" in each of the 16 climate zones, as well as a few other major cities in that zone. Figure 21 shows the overall climate characteristics for the 7 cluster centers.

	CZ	HDD	CDD
Santa Rosa[2]	2	2,844	456
San Francisco	3	3,042	108
Hayward[3]	3	2,909	128
San Jose	4	2,335	574
Gilroy	4	2,278	913
Fresno	13	2,702	1,470
Bakersfield	13	2,430	995

Figure 21. Characteristics of climate zones

There's much more variation in CDD than HDD

- The difference between max and min CDD is almost twice the difference for HDD (764 vs 1362)
- Max CDD 13 times larger than min CDD; compared to 1.3 times larger for HDD

11.2.2 Available Data

- Monthly bills – electricity, gas, dollars
- Interval data – kWh every 15 min, daily therms
- non-CAHP – CFA, number of bedrooms
- CAHP – CFA, number of bedrooms, modeled energy consumption; measure-level efficiency information for a subset of houses

11.2.3 Data Cleaning

Dwelling type: Limited to single family houses, removing addresses with "apartment", "apt", "unit", or "#"; also "pump"

House size: Limited to houses with 1000 – 5500 sf CFA

---

1 https://www.pge.com/myhome/edusafety/workshopstraining/pec/toolbox/arch/climate/index.shtml

2 Napa

3 OAK

- Upper limit
  - Largest production CAHP homes (more than one home in the project) are 5,087 sf.
  - Largest non-CAHP home is listed as 208,659 sf.
  - Filter out homes larger than 5,500 sf – excludes 76 IDs
- Lower limit
  - Smallest is 120 sf
  - Filter out homes smaller than 1000 sf – excludes 70 IDs
- 14,393 unique addresses remaining

Occupancy: A house is considered "unoccupied" if average power for the month or billing period is less than 110 W. Cutoffs from 2 sources were considered

- Res-Intel[1]: 100 kWh/month
- Sheer et al[2]: 110 W average power for a month, or ~80 kWh/month
- We used the cutoff from Sheer et al because it is more conservative

Full year: A full year is defined as 355-375 days of data in the year, with no more than 1 month/30 days unoccupied during that time.

High consumption outliers: Top 1% of consumption per sf is considered an outlier.

- Electricity – top 1% of kWh/sf/yr. consumption is 9 kWh/sf/yr.
- Gas – top 1% of therms/sf/yr. consumption is 0.35 therms/sf/yr.

Gas bills: Limited to houses that have gas bills.

- Parts of the Fresno and Bakersfield clusters are served gas by Southern California Gas, so we don't have the bills for those houses. If a house doesn't have a gas bill, we don't know whether it's all electric or in SCG territory.
- Outside of the Fresno and Bakersfield clusters, we know that houses without gas bills are all electric, but there are so few of them (0-3 CAHP all electric houses per cluster) that we cannot do a statistically meaningful comparison.

1 Res-Intel. 2018. An Exploratory Comparative Assessment of the California Advanced Homes Program (CAHP). http://www.calmac.org/publications/SCE_CAHP_FINAL_Report.pdf

2 Sheer A, S Borgeson, K Rosendo. 2017. Customer Targeting for Residential Energy Efficiency Programs: Enhancing Electricity Savings at the Meter. https://pda.energydataweb.com/api/view/1945/Customer_Targeting_Final_Whitepaper_ResEE.pdf

Cluster	CAHP Mixed Fuel	non CAHP Mixed Fuel	Total Mixed Fuel	CAHP No gas bills	non CAHP No gas bills	Total No gas bills
Santa Rosa	160	103	263	2	46	48
San Francisco	19	101	120	0	3	3
Hayward	262	788	1,050	0	36	36
San Jose	360	898	1,258	1	14	15
Gilroy	646	544	1,190	3	16	19
Fresno	1,154	1,387	2,541	229	316	545
Bakersfield	163	1,188	1,351	221	1,018	1,239
Total	2,764	5,009	7,773	456	1,449	1,905

Figure 22. Number of CAHP and non-CAHP houses by cluster and billing fuels

Interval data – Treatment of missing values

- Electricity – for missing 15 min periods, the average consumption of the rest of the day is used
- Gas – for missing days, the average consumption of the rest of the month is used

11.3 Descriptive Statistics

We have two sources of energy consumption data: monthly bills and interval data (15 minutes for electricity, 1 day for gas). However, there are houses for which we have gas bills but no gas interval data. Therefore, the analysis in this report, including the descriptive statistics below, is based on the billing data except the analysis specifically on load shapes.

11.3.1 Overall sample size – houses

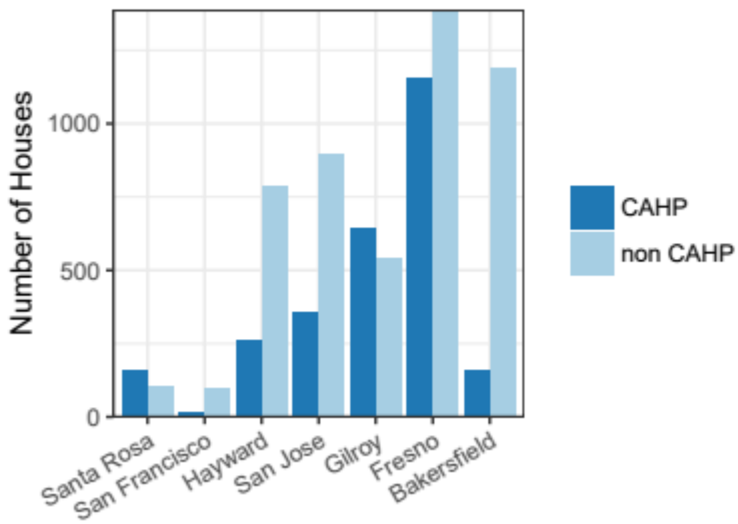

Figure 23. Number of houses by cluster and CAHP participation – graph

cluster	CAHP	non CAHP	Total
Santa Rosa	160	103	263
San Francisco	19	101	120
Hayward	262	788	1,050
San Jose	360	898	1,258
Gilroy	646	544	1,190
Fresno	1,154	1,387	2,541
Bakersfield	163	1,188	1,351
Total	2,764	5,009	7,773

Figure 24. Number of houses by cluster and CAHP participation – table

11.3.2 Floor area

- Gilroy, Fresno – similar distribution of CFA between CAHP and non-CAHP
- Santa Rosa, San Francisco – CAHP houses tend to be slightly smaller

- Hayward, San Jose, Gilroy, Bakersfield – CAHP houses tend to be slightly larger
- The difference is most pronounced in San Francisco; it is visible in the histograms. For the other clusters the difference is apparent in the table of summary statistics.

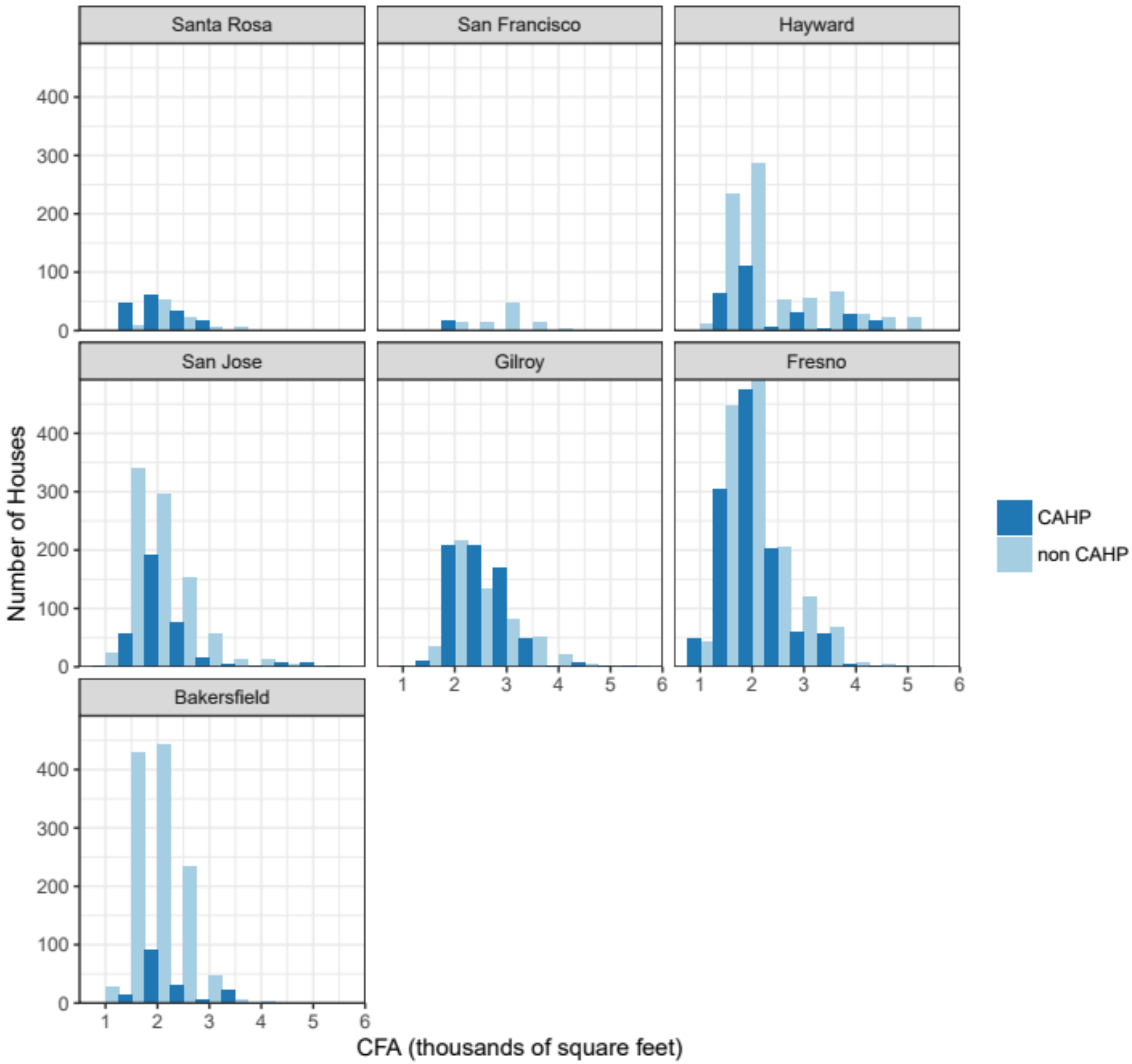

Figure 25. Number of houses by cluster, floor area, and CAHP participation

Cluster	Analysis Group	Min	Q1	Median	Q3	Max	Mean	N
Santa Rosa	CAHP	1,352	1,427	2,049	2,364	2,887	2,031	160
Santa Rosa	non CAHP	1,200	1,846	2,083	2,574	5,103	2,299	103
San Francisco	CAHP	1,547	1,824	2,001	2,001	2,133	1,924	19
San Francisco	non CAHP	1,783	2,418	2,910	3,120	4,990	2,889	101
Hayward	CAHP	1,467	1,800	2,044	3,220	4,674	2,488	263
Hayward	non CAHP	1,007	1,692	1,985	2,761	5,305	2,320	788
San Jose	CAHP	1,324	1,875	2,138	2,334	4,894	2,219	360
San Jose	non CAHP	1,097	1,534	1,892	2,284	4,848	1,996	898
Gilroy	CAHP	1,720	2,132	2,491	2,991	4,636	2,545	646
Gilroy	non CAHP	1,240	2,028	2,307	2,890	4,920	2,470	544
Fresno	CAHP	1,179	1,667	1,990	2,286	4,189	2,053	1,151
Fresno	non CAHP	1,108	1,630	1,900	2,387	5,407	2,069	1,387
Bakersfield	CAHP	1,694	1,990	2,198	2,482	3,969	2,331	163
Bakersfield	non CAHP	1,038	1,594	1,903	2,215	3,766	1,947	1,188

Figure 26. Floor area statistics by cluster and CAHP participation – range, quartiles, mean, N

11.3.3 Overall sample size – data points

The same house may have data for more than one year. The previous section describes the sample sizes and floor area characteristics of the houses themselves. This section counts the same house multiple times if we have multiple complete years of data for it.

Cluster	CAHP	non CAHP	All
Santa Rosa	475	330	805
San Francisco	21	322	343
Hayward	611	3,075	3,686
San Jose	810	1,954	2,764
Gilroy	1,565	2,039	3,604
Fresno	2,354	5,547	7,901
Bakersfield	397	4,424	4,821
Total	6,233	17,691	23,924

Figure 27. Number of data points by cluster and CAHP participation

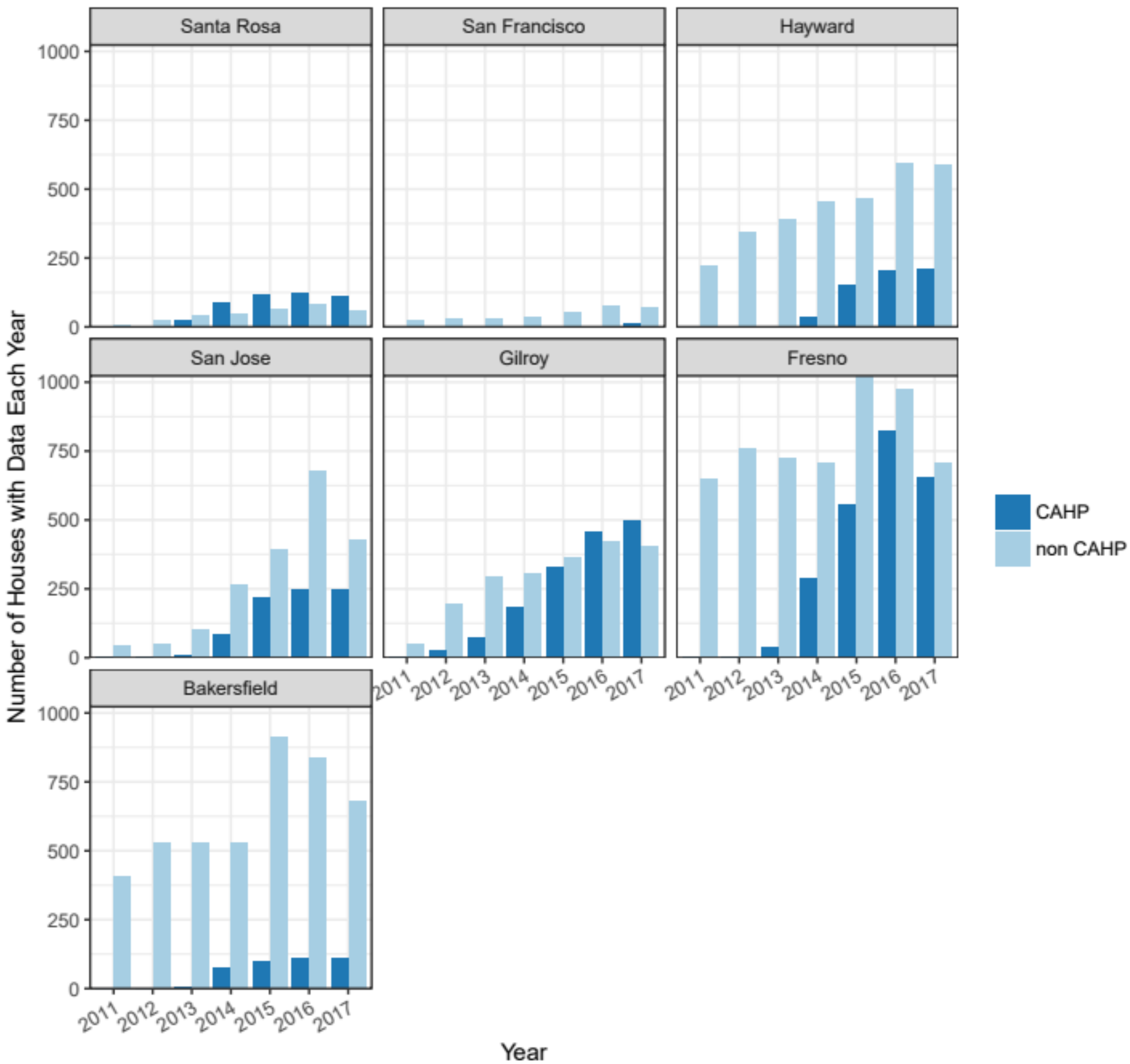

Figure 28. Number of houses with a full year of billing data in a given year by cluster and CAHP participation – graph

Cluster	Analysis Group	2011	2012	2013	2014	2015	2016	2017	All
Santa Rosa	CAHP	0	3	26	92	117	122	115	475
Santa Rosa	non CAHP	6	24	42	48	63	85	62	330
San Francisco	CAHP	0	0	0	0	1	4	16	21
San Francisco	non CAHP	26	30	30	36	54	76	70	322
Hayward	CAHP	0	0	1	35	153	208	214	611
Hayward	non CAHP	224	346	392	454	470	598	591	3,075
San Jose	CAHP	0	0	11	87	221	245	246	810
San Jose	non CAHP	42	52	99	266	393	675	427	1,954
Gilroy	CAHP	0	28	73	185	327	455	497	1,565
Gilroy	non CAHP	50	196	296	306	365	423	403	2,039
Fresno	CAHP	0	2	36	288	553	823	652	2,354
Fresno	non CAHP	648	758	727	707	1,023	976	708	5,547
Bakersfield	CAHP	0	0	3	74	97	113	110	397
Bakersfield	non CAHP	408	529	527	528	913	839	680	4,424
Total	CAHP	0	33	150	761	1,469	1,970	1,850	6,233
Total	non CAHP	1,404	1,935	2,113	2,345	3,281	3,672	2,941	17,691

Figure 29. Number of houses with a full year of billing data in a given year by cluster and CAHP participation – table

While at least 1,400 non-CAHP houses have billing data in every year, 2011-2017, none of the CAHP houses have data in 2011 and only 33 have data in 2012.

11.3.4 Vintage

As a proxy for vintage, we look at the first year we have billing data for a house. CAHP houses tend to be newer than non-CAHP houses

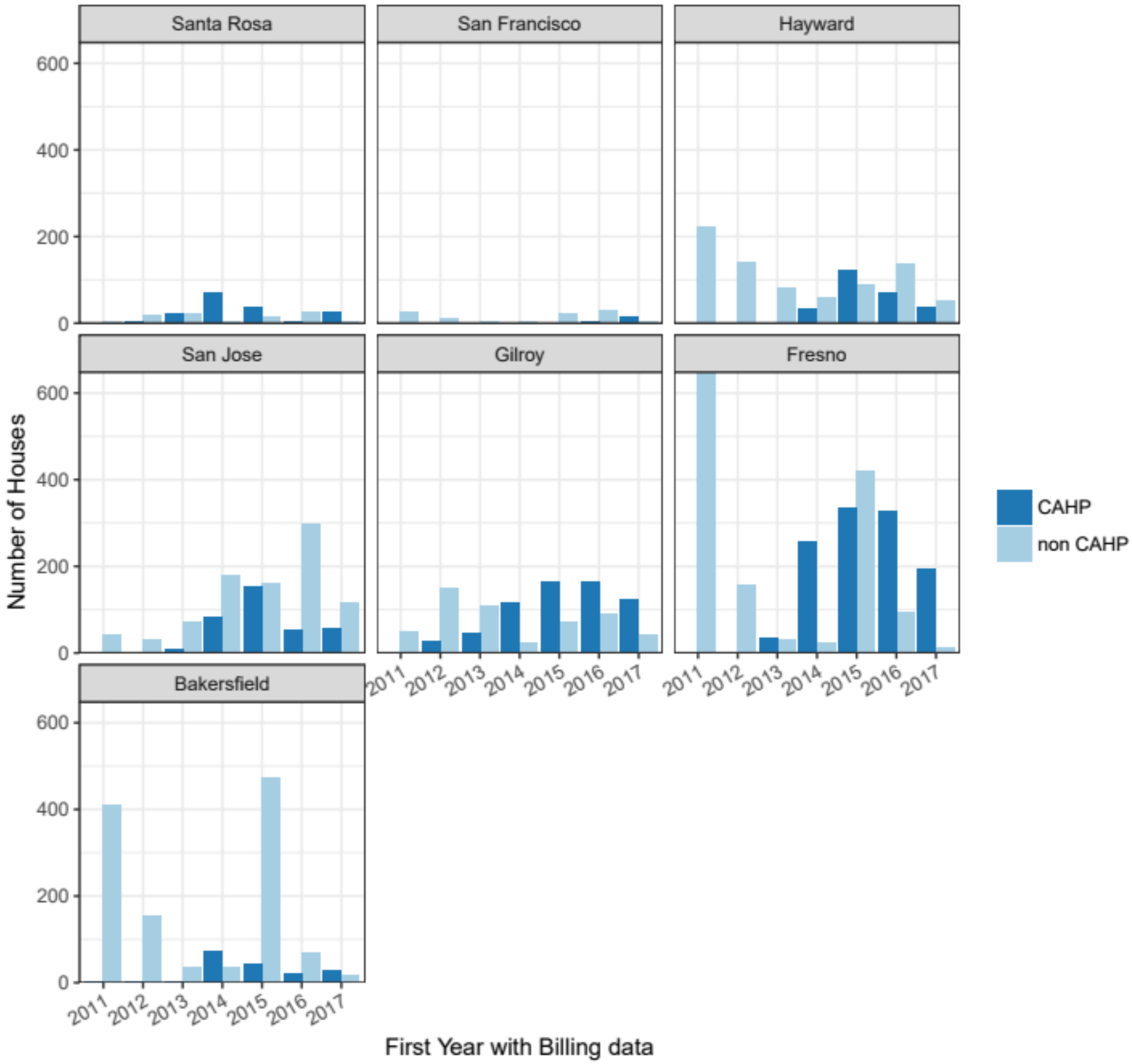

Figure 30. Number of houses by vintage[1], cluster, and CAHP participation – graph

1 The first year that we have billing data for a house is used as a proxy for vintage.

Cluster	Analysis Group	2011	2012	2013	2014	2015	2016	2017
Santa Rosa	CAHP	0	3	23	70	36	3	25
Santa Rosa	non CAHP	6	19	23	6	16	28	5
San Francisco	CAHP	0	0	0	0	1	3	15
San Francisco	non CAHP	26	12	4	3	21	29	6
Hayward	CAHP	0	0	1	34	121	70	36
Hayward	non CAHP	224	142	82	61	88	138	53
San Jose	CAHP	0	0	11	83	154	55	57
San Jose	non CAHP	42	30	71	178	161	299	117
Gilroy	CAHP	0	28	46	118	164	166	124
Gilroy	non CAHP	50	152	111	25	73	91	42
Fresno	CAHP	0	2	35	257	336	329	195
Fresno	non CAHP	648	157	30	24	419	96	13
Bakersfield	CAHP	0	0	3	71	41	21	27
Bakersfield	non CAHP	408	154	35	36	471	69	15
Total	CAHP	0	33	119	633	853	647	479
Total	non CAHP	1,404	666	356	333	1,249	750	251

Figure 31. Number of houses by vintage[1], cluster, and CAHP participation – table

1 The first year that we have billing data for a house is used as a proxy for vintage.

11.4 Energy Comparison

11.4.1 Visual Analysis

The boxplot below has one data point per house per complete year of data, so the same house may contribute more than 1 data point if it has multiple complete years of data.

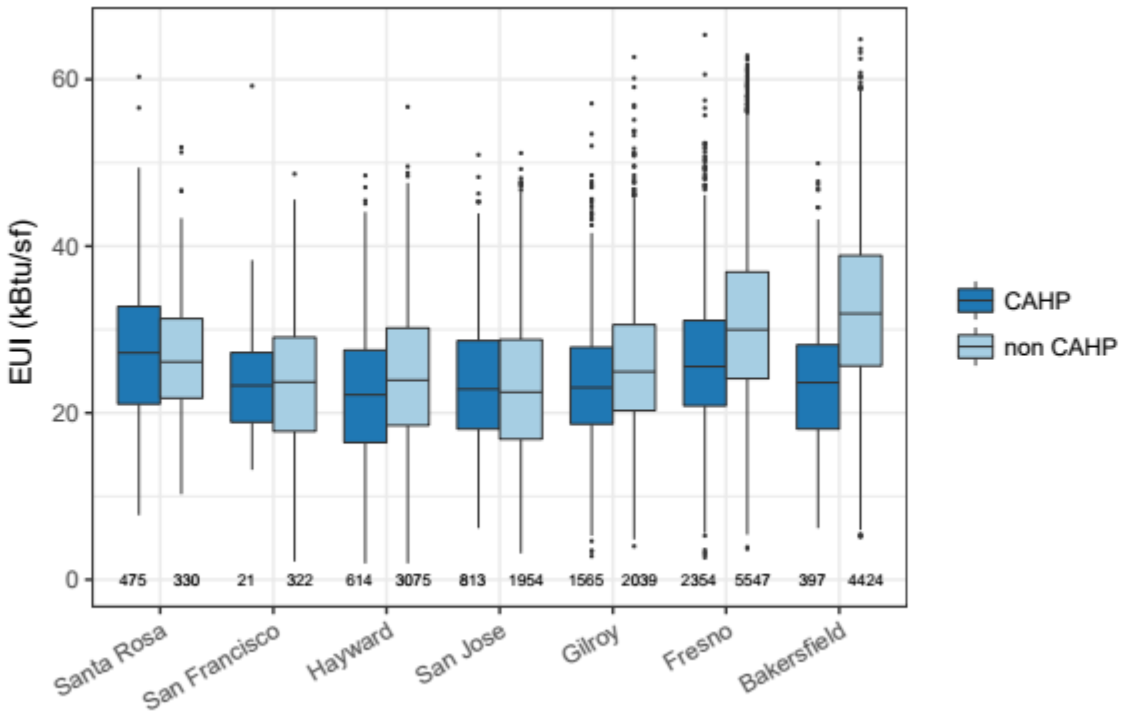

Figure 32. EUI of CAHP and non-CAHP houses by cluster

- As seen in the descriptive statistics, there is a large difference in sample sizes – only 21 CAHP houses in the San Francisco cluster compared to 5,547 non-CAHP in Fresno.
- There is a very large overlap in EUI between CAHP and non-CAHP house – the difference in medians is small compared to difference in interquartile range in all the clusters, except perhaps Bakersfield
- Q3 is on average 1.5 times as much as Q1.
- Medians
    - Substantially lower for CAHP in Fresno and Bakersfield
    - Slightly lower in Gilroy and Hayward
    - Same in San Jose and San Francisco
    - Marginally higher in Santa Rosa
    - These differences aren't dictated by climate zone – Gilroy and San Jose are both in CZ 4, but Gilroy's median CAHP EUI is lower than non-CAHP, while San Jose's is not. Similarly, for Hayward and San Francisco, which are both CZ 3
- The variation in median EUI between clusters for non-CAHP is almost twice that of CAHP

- Non-CAHP: max median EUI 32 (Bakersfield) vs min median EUI 23 (San Jose)
- CAHP: max median EUI 27 (Santa Rosa) vs min median EUI 22 (Hayward)

Cluster	Analysis Group	N	Mean	Std Dev	Min	Q1	Median	Q3	Max	Range
Santa Rosa	CAHP	475	27.3	8.0	7.7	21.1	27.2	32.8	60.3	52.6
Santa Rosa	non CAHP	330	26.6	7.3	10.3	21.8	26.1	31.3	51.9	41.6
San Francisco	CAHP	21	25.6	10.1	13.2	18.9	23.3	27.2	59.2	46.0
San Francisco	non CAHP	322	23.7	8.9	2.2	17.8	23.7	29.1	48.7	46.5
Hayward	CAHP	614	22.4	7.9	1.9	16.4	22.2	27.5	48.5	46.5
Hayward	non CAHP	3,075	24.5	8.6	1.9	18.5	23.9	30.2	56.7	54.8
San Jose	CAHP	813	23.6	7.6	6.2	18.1	22.8	28.7	51.0	44.8
San Jose	non CAHP	1,954	23.2	8.5	3.2	16.9	22.5	28.8	51.2	48.0
Gilroy	CAHP	1,565	23.5	7.1	2.8	18.7	23.0	27.9	57.1	54.3
Gilroy	non CAHP	2,039	25.6	8.0	4.0	20.3	24.9	30.6	62.7	58.7
Fresno	CAHP	2,354	26.3	7.8	2.7	20.9	25.6	31.1	65.3	62.7
Fresno	non CAHP	5,547	30.8	9.5	3.7	24.1	30.0	36.9	62.9	59.2
Bakersfield	CAHP	397	23.8	8.0	6.2	18.1	23.6	28.2	49.9	43.8
Bakersfield	non CAHP	4,424	32.4	9.5	5.1	25.6	31.9	38.9	64.8	59.7

Figure 33. Distribution of EUI by cluster and CAHP participation

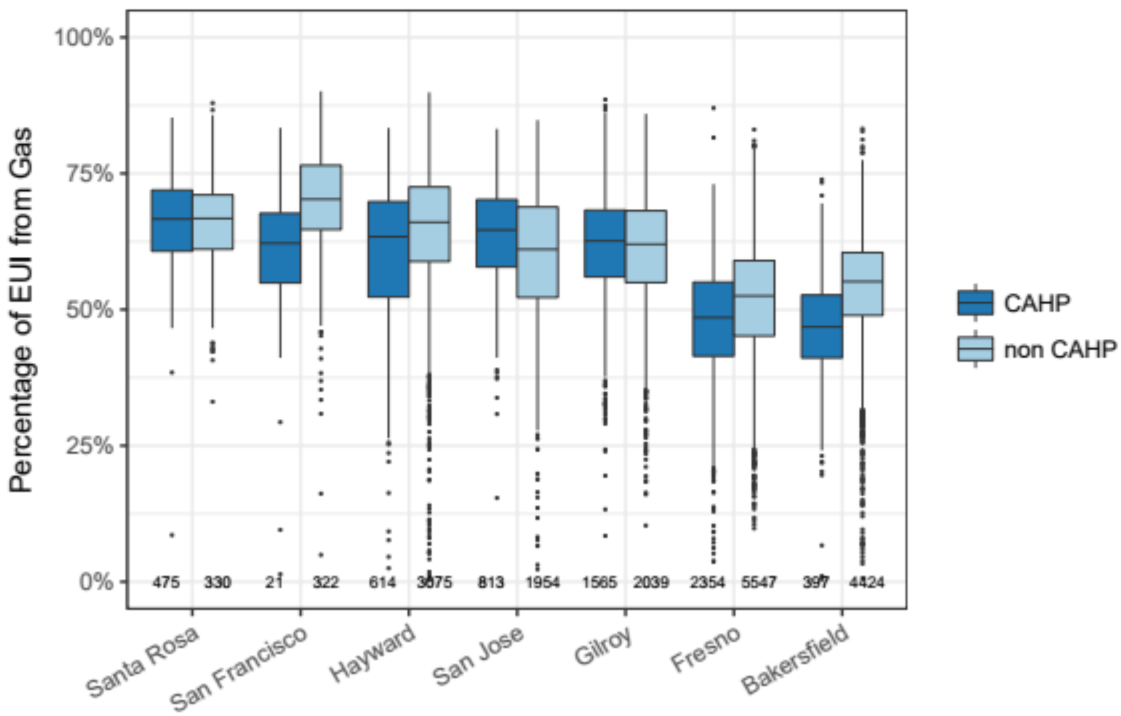

Figure 34. Percentage of EUI from gas in CAHP and non-CAHP houses by cluster and CAHP participation

- Gas consumption is a substantial portion of EUI in all clusters – the median is greater than 50% for all groups except for CAHP houses in Fresno and Bakersfield.
- Gas consumption is the smallest portion of EUI in Fresno and Bakersfield, the cooling-dominated climates
- Gas is a substantially lower portion of EUI for CAHP than non-CAHP in San Francisco, Fresno, and Bakersfield
- As seen in the descriptive statistics, there is a large difference in sample sizes – only 21 CAHP houses in the San Francisco cluster compared to 5,547 non-CAHP in Fresno.
- There is a very large overlap in EUI between CAHP and non-CAHP house – the difference in medians is small compared to difference in interquartile range in all the clusters, except perhaps Bakersfield
- Q3 is on average 1.5 times as much as Q1.
- Medians
  - Substantially lower for CAHP in Fresno and Bakersfield
  - Slightly lower in Gilroy and Hayward
  - Same in San Jose and San Francisco
  - Marginally higher in Santa Rosa
  - These differences aren't dictated by climate zone – Gilroy and San Jose are both in CZ 4, but Gilroy's median CAHP EUI is lower than non-CAHP, while San Jose's is not. Similarly, for Hayward and San Francisco, which are both CZ 3
- The variation in median EUI between clusters for non-CAHP is almost twice that of CAHP

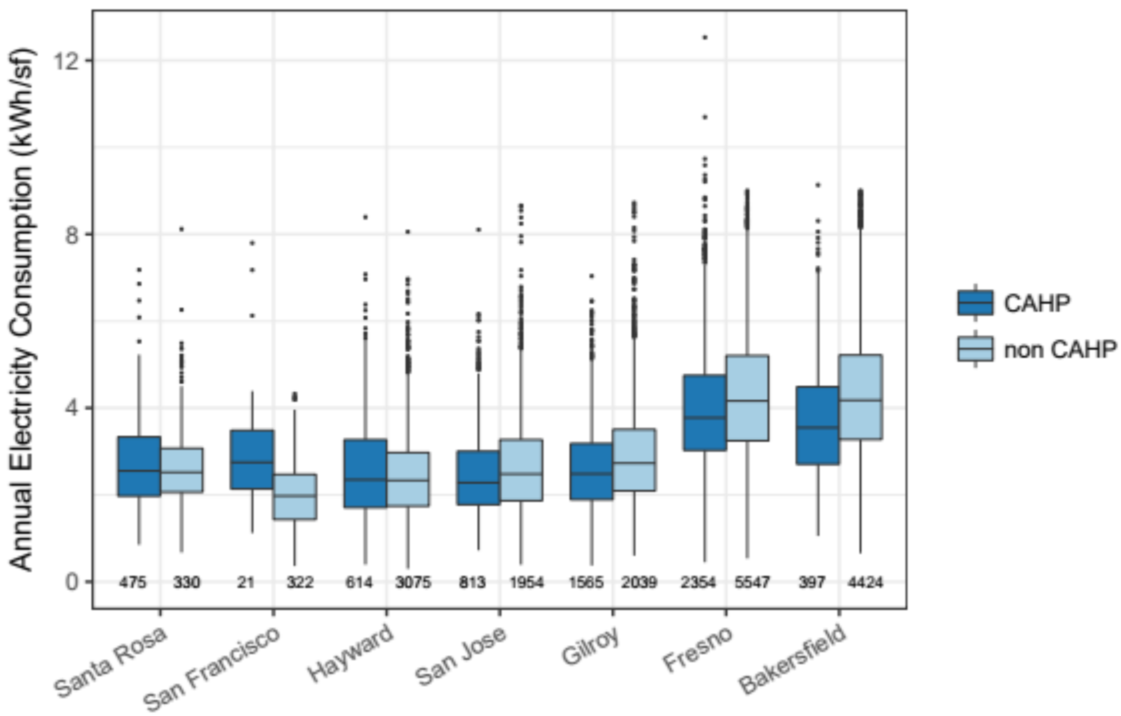

Figure 35. Annual electricity consumption per floor area of CAHP and non-CAHP houses by cluster

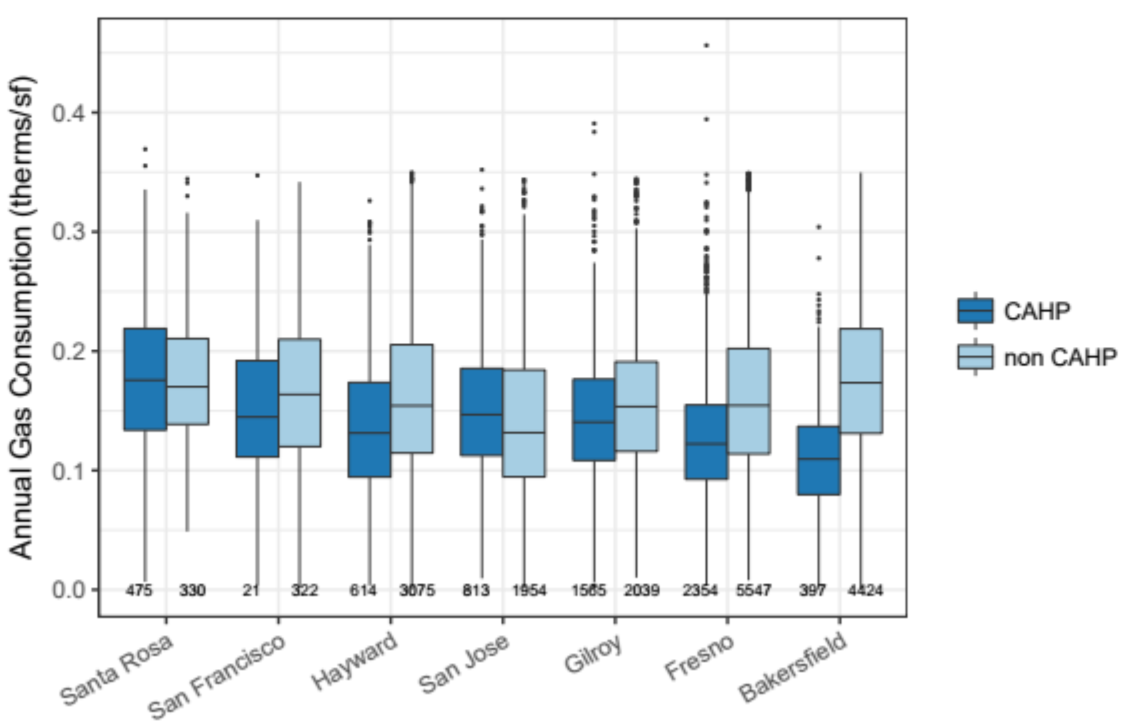

Figure 36. Annual gas consumption per floor area of CAHP and non-CAHP houses by cluster

- In general, the differences in consumption between CAHP and non-CAHP are bigger for gas than for electricity or EUI.

- This could mean that CAHP participation tends to affect gas end uses more than electric end uses, or there could be a systematic difference in how which appliances tend to be gas in the two kinds of houses.

11.4.2 Regression

Linear regression models require that the dependent variable be normally distributed. However, EUI in this sample does not have a normal distribution, based on visual inspection and the Shapiro Wilk test; instead the distribution is right-skewed. One common approach to overcoming this is to use a log transformation on the dependent variable.[1] Therefore in these regression models the dependent variable is log(EUI). Savings are then calculated as

$$\%\ savings = 1 - e^{regression\ coefficient}$$

Therefore, savings will be positive with negative regression coefficients.

For all the regression models described below, separate models were run for each cluster.

Single Variable Models

We considered four independent variables in separate single variable models:

- CAHP participation
- Floor area (CFA)
- Vintage (using the first year of billing data as a proxy)
- Billing year

For single variable models predicting log(EUI), CFA had the highest predictive power (accounting for 4.6-15% of the variation) in 6/7 clusters. In San Francisco, vintage (using the first year of billing data as a proxy) had the highest predictive power.

Outside of Fresno, CAHP participation is the independent variable with the least predictive power in these single variable models – floor area, vintage, and billing year all have higher R^2's.

1 Lewis, K et al. 2018. Comparison of energy data for green-certified and non-certified buildings in the 2012 Commercial Buildings Energy Consumption Survey (CBECS). ACEEE Summer Study in Buildings. https://aceee.org/files/proceedings/2018/assets/attachments/0194_0286_000154.pdf

Cluster	CAHP	CFA	Vintage	Billing Year
Santa Rosa	0.001	0.079	0.023	0.039
San Francisco	0.003	0.077	0.122	0.043
Hayward	0.006	0.082	0.020	0.023
San Jose	0.002	0.046	0.016	0.026
Gilroy	0.015	0.056	0.032	0.032
Fresno	0.045	0.100	0.081	0.055
Bakersfield	0.067	0.155	0.074	0.047

Figure 37. R^2 of single variable models predicting log(EUI) by cluster

Figure 38 shows the regression parameters for the single variable model with CFA predicting log(EUI). For every 500 sf of CFA added to the house, EUI savings increase by 6-15% depending on the cluster. This illustrates that as a metric, EUI privileges larger buildings. Some end uses (e.g. heating) scale by floor area, but others (e.g. laundry) scale by the number of occupants. So, the more space each person has, the lower the EUI will be, on average – even though the total consumption per person may be larger.

Cluster	Parameter	Estimate	Standard Error	EUI Percent Savings per Additional 500 sf	Significance
Santa Rosa	(Intercept)	3.61	0.04		***
Santa Rosa	CFA	0.00	0.00	8.3%	***
San Francisco	(Intercept)	3.73	0.12		***
San Francisco	CFA	0.00	0.00	11.3%	***
Hayward	(Intercept)	3.41	0.02		***
Hayward	CFA	0.00	0.00	6.4%	***
San Jose	(Intercept)	3.36	0.03		***
San Jose	CFA	0.00	0.00	6.6%	***
Gilroy	(Intercept)	3.50	0.02		***
Gilroy	CFA	0.00	0.00	6.8%	***
Fresno	(Intercept)	3.71	0.01		***
Fresno	CFA	0.00	0.00	9.5%	***
Bakersfield	(Intercept)	4.00	0.02		***
Bakersfield	CFA	0.00	0.00	15.5%	***

* Significant at p < .05

** Significant at p < .01

*** Significant at p < .001

Figure 38. Regression parameters for single variable models of CFA predicting log(EUI) by cluster

Multivariable Models

We saw from the descriptive statistics that the CAHP houses in the sample tend to be newer than the non-CAHP houses and smaller or larger depending on the cluster. To disentangle the effects of these variables and understand the impact of each variable above and beyond the other variables, we ran a full model with all 4 independent variables – floor area, vintage, billing year, and CAHP participation.

The multivariable models explain 9-21% of the variation in log(EUI) depending on cluster.

Cluster	R^2
Santa Rosa	0.158
San Francisco	0.189
Hayward	0.109
San Jose	0.087
Gilroy	0.110
Fresno	0.174
Bakersfield	0.207

Figure 39. R^2 of multivariable models predicting log(EUI)

Figure 40 lists the regression estimates from the multivariable models predicting log(EUI) in each cluster. Billing year and vintage are both categorical variables, with 2011 as the baseline. So "Year2012" compares log(EUI) in 2012 to 2011, and "first.year2015" compares the log(EUI) for houses with their first year of data in 2015 to houses with their first year of data in 2011. CAHP participation is also a categorical variable with non-participation as the baseline; "isCAHP" indicates the performance of CAHP houses compared to non-CAHP.

- Billing year – at least one year is statistically significant only in Hayward, Fresno, and Bakersfield
- CAHP participation – statistically significant in San Jose, Fresno, and Bakersfield
  - San Jose – 4.1% increase in EUI
  - Fresno – 4.6% decrease in EUI
  - Bakersfield – 15% decrease in EUI
  - This is above and beyond the impact of the CAHP houses being substantially newer than the non-CAHP ones (and so are built to a stricter code)
- CFA – statistically significant in all clusters
  - The EUI percent savings values are very small because they're per square foot; they are much more impactful at larger intervals, e.g. per 500 sf as reported in Figure 38
- Vintage – at least one year is statistically significant in all clusters

Cluster	Parameter	Estimate	Standard Error	EUI Percent Savings	Significance
Santa Rosa	(Intercept)	3.724	0.125		***
Santa Rosa	Year2012	-0.023	0.137	2.3%	
Santa Rosa	Year2013	0.093	0.132	-9.8%	
Santa Rosa	Year2014	-0.089	0.131	8.5%	
Santa Rosa	Year2015	-0.057	0.131	5.6%	
Santa Rosa	Year2016	-0.019	0.131	1.9%	
Santa Rosa	Year2017	0.072	0.131	-7.5%	
Santa Rosa	isCAHP	0.022	0.019	-2.2%	
Santa Rosa	CFA	0.000	0.000	0.0%	***
Santa Rosa	first.year2012	-0.013	0.066	1.3%	
Santa Rosa	first.year2013	-0.090	0.066	8.6%	
Santa Rosa	first.year2014	-0.091	0.069	8.7%	
Santa Rosa	first.year2015	0.009	0.069	-0.9%	
Santa Rosa	first.year2016	0.033	0.075	-3.3%	
Santa Rosa	first.year2017	-0.311	0.087	26.8%	***
San Francisco	(Intercept)	3.773	0.153		***
San Francisco	Year2012	-0.007	0.118	0.7%	
San Francisco	Year2013	0.076	0.118	-7.9%	
San Francisco	Year2014	-0.135	0.114	12.6%	
San Francisco	Year2015	-0.135	0.110	12.7%	
San Francisco	Year2016	-0.074	0.109	7.1%	
San Francisco	Year2017	-0.058	0.111	5.7%	
San Francisco	isCAHP	-0.050	0.101	4.9%	
San Francisco	CFA	0.000	0.000	0.0%	***

Cluster	Parameter	Estimate	Standard Error	EUI Percent Savings	Significance
San Francisco	first.year2012	-0.235	0.072	20.9%	**
San Francisco	first.year2013	0.164	0.116	-17.8%	
San Francisco	first.year2014	-0.211	0.159	19.0%	
San Francisco	first.year2015	-0.295	0.078	25.5%	***
San Francisco	first.year2016	-0.027	0.085	2.7%	
San Francisco	first.year2017	-0.102	0.155	9.7%	
Hayward	(Intercept)	3.444	0.032		***
Hayward	Year2012	0.009	0.034	-0.9%	
Hayward	Year2013	0.048	0.034	-4.9%	
Hayward	Year2014	-0.095	0.033	9.0%	**
Hayward	Year2015	-0.097	0.033	9.3%	**
Hayward	Year2016	-0.049	0.032	4.7%	
Hayward	Year2017	0.008	0.032	-0.8%	
Hayward	isCAHP	0.002	0.015	-0.2%	
Hayward	CFA	0.000	0.000	0.0%	***
Hayward	first.year2012	0.045	0.019	-4.6%	*
Hayward	first.year2013	0.013	0.024	-1.3%	
Hayward	first.year2014	0.002	0.027	-0.2%	
Hayward	first.year2015	-0.079	0.026	7.6%	**
Hayward	first.year2016	-0.034	0.027	3.4%	
Hayward	first.year2017	-0.148	0.046	13.8%	**
San Jose	(Intercept)	3.593	0.064		***
San Jose	Year2012	-0.024	0.083	2.3%	
San Jose	Year2013	0.077	0.077	-8.0%	

Cluster	Parameter	Estimate	Standard Error	EUI Percent Savings	Significance
San Jose	Year2014	-0.071	0.071	6.8%	
San Jose	Year2015	-0.063	0.070	6.1%	
San Jose	Year2016	-0.047	0.070	4.6%	
San Jose	Year2017	0.078	0.070	-8.1%	
San Jose	isCAHP	0.040	0.012	-4.1%	***
San Jose	CFA	0.000	0.000	0.0%	***
San Jose	first.year2012	-0.096	0.050	9.2%	
San Jose	first.year2013	-0.204	0.044	18.5%	***
San Jose	first.year2014	-0.180	0.041	16.5%	***
San Jose	first.year2015	-0.177	0.042	16.2%	***
San Jose	first.year2016	-0.218	0.042	19.6%	***
San Jose	first.year2017	-0.216	0.050	19.4%	***
Gilroy	(Intercept)	3.628	0.052		***
Gilroy	Year2012	0.029	0.053	-2.9%	
Gilroy	Year2013	0.087	0.052	-9.1%	
Gilroy	Year2014	-0.073	0.051	7.0%	
Gilroy	Year2015	-0.057	0.051	5.5%	
Gilroy	Year2016	-0.041	0.051	4.0%	
Gilroy	Year2017	0.038	0.051	-3.8%	
Gilroy	isCAHP	-0.005	0.009	0.5%	
Gilroy	CFA	0.000	0.000	0.0%	***
Gilroy	first.year2012	-0.082	0.023	7.8%	***
Gilroy	first.year2013	-0.068	0.024	6.6%	**
Gilroy	first.year2014	-0.127	0.028	11.9%	***

Cluster	Parameter	Estimate	Standard Error	EUI Percent Savings	Significance
Gilroy	first.year2015	-0.129	0.027	12.1%	***
Gilroy	first.year2016	-0.190	0.028	17.3%	***
Gilroy	first.year2017	-0.249	0.036	22.0%	***
Fresno	(Intercept)	3.723	0.018		***
Fresno	Year2012	0.029	0.017	-3.0%	
Fresno	Year2013	0.036	0.017	-3.7%	*
Fresno	Year2014	-0.074	0.016	7.2%	***
Fresno	Year2015	-0.075	0.016	7.2%	***
Fresno	Year2016	-0.052	0.016	5.0%	**
Fresno	Year2017	0.018	0.016	-1.8%	
Fresno	isCAHP	-0.047	0.008	4.6%	***
Fresno	CFA	0.000	0.000	0.0%	***
Fresno	first.year2012	-0.034	0.013	3.3%	**
Fresno	first.year2013	0.068	0.021	-7.0%	**
Fresno	first.year2014	-0.099	0.017	9.4%	***
Fresno	first.year2015	-0.089	0.012	8.5%	***
Fresno	first.year2016	-0.097	0.018	9.2%	***
Fresno	first.year2017	-0.093	0.026	8.9%	***
Bakersfield	(Intercept)	3.828	0.028		***
Bakersfield	Year2012	0.033	0.020	-3.4%	
Bakersfield	Year2013	0.047	0.020	-4.8%	*
Bakersfield	Year2014	-0.046	0.020	4.5%	*
Bakersfield	Year2015	-0.063	0.019	6.1%	**
Bakersfield	Year2016	-0.032	0.020	3.2%	

Cluster	Parameter	Estimate	Standard Error	EUI Percent Savings	Significance
Bakersfield	Year2017	0.005	0.020	-0.5%	
Bakersfield	isCAHP	-0.157	0.015	14.5%	***
Bakersfield	CFA	0.000	0.000	0.0%	***
Bakersfield	first.year2012	-0.036	0.014	3.5%	**
Bakersfield	first.year2013	0.014	0.028	-1.4%	
Bakersfield	first.year2014	0.012	0.023	-1.2%	
Bakersfield	first.year2015	-0.057	0.014	5.6%	***
Bakersfield	first.year2016	0.009	0.028	-0.9%	
Bakersfield	first.year2017	-0.063	0.051	6.1%	

* Significant at p < .05

** Significant at p < .01

*** Significant at p < .001

Figure 40. Regression parameters of multivariable model predicting log(EUI)

11.5 Bill Comparison

11.5.1 Visual Analysis

The boxplot below has one data point per house per complete year of data, so the same house may contribute more than 1 data point if it has multiple complete years of data.

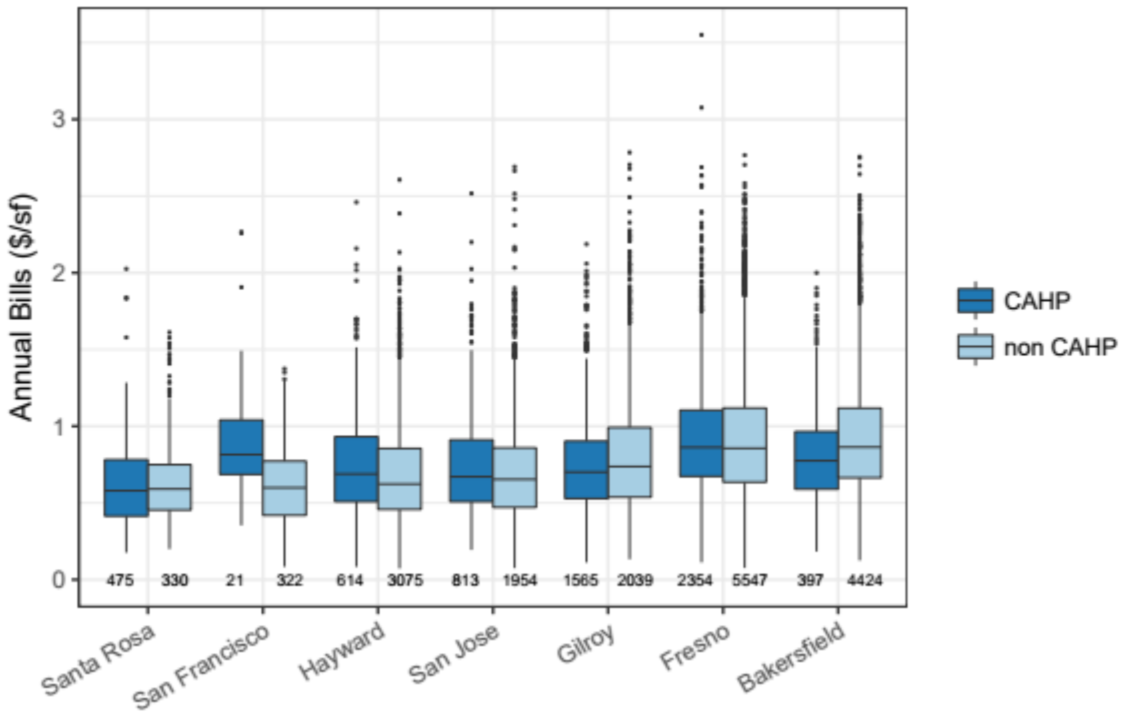

Figure 41. Annual bills per floor area of CAHP and non-CAHP houses by cluster

- Very little difference in annual bills between CAHP and non-CAHP except in San Francisco
- Huge differences in annual bills within each group – Q3 is 3 to 15 times as high as Q1.
- In general electricity is more expensive than gas; bills are affected by overall consumption but also the portion that is gas. Gains in efficiency (EUI) might not be seen in the bills if more of the savings are coming from electricity.
- Medians[1]
  - Santa Rosa – very little difference in EUI, % gas, or bills between CAHP and non-CAHP.
  - San Francisco – very little difference in EUI between CAHP and non-CAHP houses. But a higher percentage of EUI comes from electricity in CAHP houses, so the overall bills are higher.
  - Hayward – slightly lower EUI, slightly more of it comes from electricity, bills slightly higher
  - San Jose, Gilroy – very little difference in EUI between CAHP and non-CAHP

1 These descriptions are based on a comparison of Figure 32, Figure 34, Figure 35, and Figure 36 as well as Figure 41.

- Fresno – CAHP houses tend to have lower EUIs, but more of it comes from electricity – cancels out so that bills are basically the same
- Bakersfield – CAHP houses tend to have lower EUIs, but more of it comes from electricity –almost cancels out; bills are only slightly lower for CAHP houses

		N	Mean	Std Dev	Min	Q1	Median	Q3	Max	Range
Santa Rosa	CAHP	475	0.619	0.266	0.176	0.414	0.580	0.782	2.026	1.850
Santa Rosa	non CAHP	330	0.644	0.267	0.199	0.454	0.591	0.749	1.611	1.412
San Francisco	CAHP	21	1.005	0.553	0.354	0.686	0.815	1.040	2.266	1.913
San Francisco	non CAHP	322	0.615	0.268	0.083	0.419	0.599	0.772	1.374	1.291
Hayward	CAHP	614	0.742	0.337	0.083	0.509	0.689	0.932	2.460	2.377
Hayward	non CAHP	3,075	0.680	0.314	0.073	0.458	0.623	0.855	2.606	2.533
San Jose	CAHP	813	0.733	0.316	0.194	0.509	0.671	0.911	2.517	2.323
San Jose	non CAHP	1,954	0.705	0.332	0.075	0.471	0.654	0.859	2.691	2.615
Gilroy	CAHP	1,565	0.741	0.305	0.112	0.529	0.701	0.902	2.186	2.075
Gilroy	non CAHP	2,039	0.800	0.373	0.133	0.538	0.737	0.991	2.784	2.651
Fresno	CAHP	2,354	0.915	0.358	0.112	0.673	0.862	1.104	3.551	3.438
Fresno	non CAHP	5,547	0.911	0.377	0.076	0.635	0.856	1.118	2.767	2.690
Bakersfield	CAHP	397	0.811	0.324	0.183	0.590	0.775	0.965	2.000	1.817
Bakersfield	non CAHP	4,424	0.922	0.363	0.125	0.664	0.864	1.117	2.756	2.631

Figure 42. Distribution of annual bills per floor area of CAHP and non-CAHP houses by cluster

11.5.2 Regression

The regression methodology is analogous as that used for EUI (see Section ▯ for more details). In this case the dependent variable is the log of CFA-normalized bills (or log($/sf).

Single Variable Models

Once again, we considered four independent variables in separate single variable models:

- CAHP participation
- Floor area (CFA)
- Vintage (using the first year of billing data as a proxy
- Billing year

In the single variable models predicting log(normalized bills), billing year is the single variable with the highest predictive power (accounting for 4-7% of the variation) in 5/7 clusters. In San Francisco and Gilroy, the single variables with the highest predictive power are vintage and CFA, respectively. Since rates vary by year, it makes sense that billing year would be more predictive of bills than EUI.

Except for San Jose, the max R^2 for single variable models predicting log(EUI) are higher than the max R^2 for single variable models predicting log(normalized bills). In some cases, the difference is huge. For example, in Bakersfield 15% of variance in log(EUI) is explained by CFA, whereas only 3% of variance in log(normalized bills) is explained by year.

Cluster	CAHP	CFA	Vintage	Billing Year
Santa Rosa	0.003	0.010	0.023	0.052
San Francisco	0.048	0.090	0.108	0.024
Hayward	0.005	0.030	0.016	0.065
San Jose	0.003	0.036	0.022	0.055
Gilroy	0.004	0.017	0.012	0.012
Fresno	0.000	0.003	0.009	0.058
Bakersfield	0.008	0.004	0.004	0.036

Figure 43. R^2 of single-variable models predicting log(normalized bills)

Multivariable Models

The multivariable models have 4 independent variables – floor area, vintage, billing year, and CAHP participation.

The multivariable models explain 5-19% of the variation in log(normalized bills) depending on cluster. In all clusters the R^2 for the analogous model predicting log(EUI) is at least slightly higher.

Cluster	R^2
Santa Rosa	0.114
San Francisco	0.187
Hayward	0.104
San Jose	0.091
Gilroy	0.046
Fresno	0.073
Bakersfield	0.062

Figure 44. R^2 of multivariable models predicting log(normalized bills)

Figure 45 lists the regression estimates from the multivariable models predicting log(normalized bills) in each cluster.

- Billing year – at least one year is statistically significant except in Santa Rosa and San Francisco
  - It's surprising that it's not statistically significant in Santa Rosa since it's the single-variable model with the highest R^2
- CAHP participation – statistically significant in 5/7 clusters
  - San Francisco – 30% increase in normalized bills
  - Hayward – 4.6% increase in normalized bills
  - San Jose – 3.3% increase in normalized bills
  - Fresno – 2.6% decrease in normalized bills
  - Bakersfield – 11% decrease in normalized bills
- CFA – statistically significant in all clusters (like for EUI model)
  - The percent savings values are very small because they're per square foot; they are much more impactful at larger intervals
- Vintage – at least one year is statistically significant in all clusters

Cluster	Parameter	Estimate	Standard Error	Bills Percent Savings	Significance
Santa Rosa	(Intercept)	-0.295	0.175		
Santa Rosa	Year2012	0.050	0.191	-5.1%	
Santa Rosa	Year2013	0.187	0.185	-20.6%	
Santa Rosa	Year2014	0.163	0.183	-17.8%	
Santa Rosa	Year2015	-0.021	0.182	2.1%	
Santa Rosa	Year2016	0.113	0.183	-11.9%	
Santa Rosa	Year2017	0.315	0.183	-37.0%	
Santa Rosa	isCAHP	-0.027	0.026	2.7%	
Santa Rosa	CFA	0.000	0.000	0.0%	***
Santa Rosa	first.year2012	-0.056	0.092	5.4%	
Santa Rosa	first.year2013	-0.181	0.092	16.5%	*
Santa Rosa	first.year2014	-0.138	0.096	12.9%	
Santa Rosa	first.year2015	-0.045	0.097	4.4%	
Santa Rosa	first.year2016	-0.075	0.105	7.2%	
Santa Rosa	first.year2017	-0.553	0.121	42.5%	***
San Francisco	(Intercept)	0.167	0.172		
San Francisco	Year2012	0.026	0.133	-2.7%	
San Francisco	Year2013	0.077	0.133	-8.0%	
San Francisco	Year2014	0.013	0.128	-1.3%	
San Francisco	Year2015	0.063	0.123	-6.5%	
San Francisco	Year2016	0.198	0.122	-21.9%	

Cluster	Parameter	Estimate	Standard Error	Bills Percent Savings	Significance
San Francisco	Year2017	0.155	0.124	-16.8%	
San Francisco	isCAHP	0.269	0.113	-30.9%	*
San Francisco	CFA	0.000	0.000	0.0%	***
San Francisco	first.year2012	-0.260	0.081	22.9%	**
San Francisco	first.year2013	0.119	0.130	-12.6%	
San Francisco	first.year2014	-0.411	0.178	33.7%	*
San Francisco	first.year2015	-0.343	0.088	29.0%	***
San Francisco	first.year2016	-0.088	0.095	8.4%	
San Francisco	first.year2017	-0.312	0.173	26.8%	
Hayward	(Intercept)	-0.427	0.037		***
Hayward	Year2012	0.045	0.040	-4.6%	
Hayward	Year2013	0.102	0.039	-10.7%	**
Hayward	Year2014	0.083	0.038	-8.6%	*
Hayward	Year2015	0.164	0.038	-17.8%	***
Hayward	Year2016	0.284	0.037	-32.8%	***
Hayward	Year2017	0.395	0.038	-48.4%	***
Hayward	isCAHP	0.045	0.018	-4.6%	*
Hayward	CFA	0.000	0.000	0.0%	***
Hayward	first.year2012	0.040	0.022	-4.1%	

Cluster	Parameter	Estimate	Standard Error	Bills Percent Savings	Significance
Hayward	first.year2013	0.000	0.028	0.0%	
Hayward	first.year2014	-0.045	0.032	4.4%	
Hayward	first.year2015	-0.072	0.030	6.9%	*
Hayward	first.year2016	-0.011	0.031	1.1%	
Hayward	first.year2017	-0.151	0.054	14.0%	**
San Jose	(Intercept)	-0.269	0.075		***
San Jose	Year2012	-0.030	0.097	2.9%	
San Jose	Year2013	0.089	0.090	-9.3%	
San Jose	Year2014	0.104	0.084	-11.0%	
San Jose	Year2015	0.200	0.082	-22.1%	*
San Jose	Year2016	0.280	0.082	-32.4%	***
San Jose	Year2017	0.397	0.083	-48.8%	***
San Jose	isCAHP	0.033	0.014	-3.3%	*
San Jose	CFA	0.000	0.000	0.0%	***
San Jose	first.year2012	-0.045	0.059	4.4%	
San Jose	first.year2013	-0.163	0.051	15.1%	**
San Jose	first.year2014	-0.145	0.048	13.5%	**
San Jose	first.year2015	-0.119	0.049	11.2%	*
San Jose	first.year2016	-0.160	0.049	14.8%	**
San Jose	first.year2017	-0.186	0.059	17.0%	**
Gilroy	(Intercept)	-0.181	0.071		*
Gilroy	Year2012	0.051	0.073	-5.2%	
Gilroy	Year2013	0.091	0.071	-9.6%	
Gilroy	Year2014	0.072	0.070	-7.4%	

Cluster	Parameter	Estimate	Standard Error	Bills Percent Savings	Significance
Gilroy	Year2015	0.174	0.070	-19.0%	*
Gilroy	Year2016	0.236	0.070	-26.6%	***
Gilroy	Year2017	0.250	0.070	-28.5%	***
Gilroy	isCAHP	-0.016	0.013	1.6%	
Gilroy	CFA	0.000	0.000	0.0%	***
Gilroy	first.year2012	-0.089	0.031	8.5%	**
Gilroy	first.year2013	-0.080	0.033	7.7%	*
Gilroy	first.year2014	-0.185	0.038	16.9%	***
Gilroy	first.year2015	-0.132	0.037	12.4%	***
Gilroy	first.year2016	-0.215	0.039	19.3%	***
Gilroy	first.year2017	-0.262	0.049	23.1%	***
Fresno	(Intercept)	-0.307	0.023		***
Fresno	Year2012	0.102	0.021	-10.7%	***
Fresno	Year2013	0.126	0.021	-13.4%	***
Fresno	Year2014	0.154	0.021	-16.6%	***
Fresno	Year2015	0.226	0.020	-25.3%	***
Fresno	Year2016	0.311	0.020	-36.5%	***
Fresno	Year2017	0.417	0.021	-51.8%	***
Fresno	isCAHP	-0.026	0.011	2.6%	*
Fresno	CFA	0.000	0.000	0.0%	***
Fresno	first.year2012	-0.021	0.017	2.1%	
Fresno	first.year2013	0.088	0.028	-9.2%	**
Fresno	first.year2014	-0.065	0.022	6.3%	**
Fresno	first.year2015	-0.068	0.016	6.6%	***

Cluster	Parameter	Estimate	Standard Error	Bills Percent Savings	Significance
Fresno	first.year2016	-0.079	0.023	7.6%	***
Fresno	first.year2017	-0.061	0.034	5.9%	
Bakersfield	(Intercept)	-0.301	0.036		***
Bakersfield	Year2012	0.096	0.026	-10.0%	***
Bakersfield	Year2013	0.115	0.026	-12.2%	***
Bakersfield	Year2014	0.141	0.026	-15.1%	***
Bakersfield	Year2015	0.214	0.025	-23.8%	***
Bakersfield	Year2016	0.265	0.025	-30.4%	***
Bakersfield	Year2017	0.346	0.026	-41.3%	***
Bakersfield	isCAHP	-0.122	0.019	11.5%	***
Bakersfield	CFA	0.000	0.000	0.0%	***
Bakersfield	first.year2012	-0.018	0.017	1.8%	
Bakersfield	first.year2013	0.072	0.036	-7.4%	*
Bakersfield	first.year2014	0.001	0.029	-0.1%	
Bakersfield	first.year2015	-0.070	0.017	6.8%	***
Bakersfield	first.year2016	0.020	0.037	-2.0%	
Bakersfield	first.year2017	-0.045	0.065	4.4%	

* Significant at p < .05

** Significant at p < .01

*** Significant at p < .001

Figure 45. Regression parameters of multivariable model predicting log(normalized bills)

11.6 Load Shapes

For each cluster, we chose one sample weekday from 2017 in four categories – typical and extreme summer, typical and extreme winter – based on daily min, mean, and max outdoor temperatures for the cluster centers as reported by NOAA and Weather Underground. We looked at a couple of summer months and choose a day with a mean temperature in the middle of the distribution and a day with high mean and maximum temperatures. Similarly, for winter months.

The analysis is based on average hourly power per floor area.

11.6.1 Summer

Cluster	Extreme or Typical	Date	Max OAT	Ave OAT	Min OAT	CAHP	non CAHP
Santa Rosa	extreme	7/7/17	107	78	49	160	101
Santa Rosa	typical	7/27/17	90	71	51	160	101
San Francisco	extreme	9/1/17	106	88	69	19	100
San Francisco	typical	7/6/17	67	60	53	19	101
Hayward	extreme	9/1/17	103	86	69	261	784
Hayward	typical	7/12/17	77	68	58	259	786
San Jose	extreme	9/1/17	108	89	69	359	895
San Jose	typical	8/16/17	81	71	61	359	894
Gilroy	extreme	7/27/17	96	76	55	645	544
Gilroy	typical	7/4/17	89	72	55	645	544
Fresno	extreme	7/17/17	107	91	75	1,152	1,381
Fresno	typical	7/13/17	101	86	71	1,152	1,381
Bakersfield	extreme	7/17/17	106	93	79	162	1,183
Bakersfield	typical	7/27/17	103	90	77	162	1,181

Figure 46. Characteristics of the sample summer weekdays[1]

1 OAT is outdoor air temperature in °F. Source: NOAA and Weather Underground

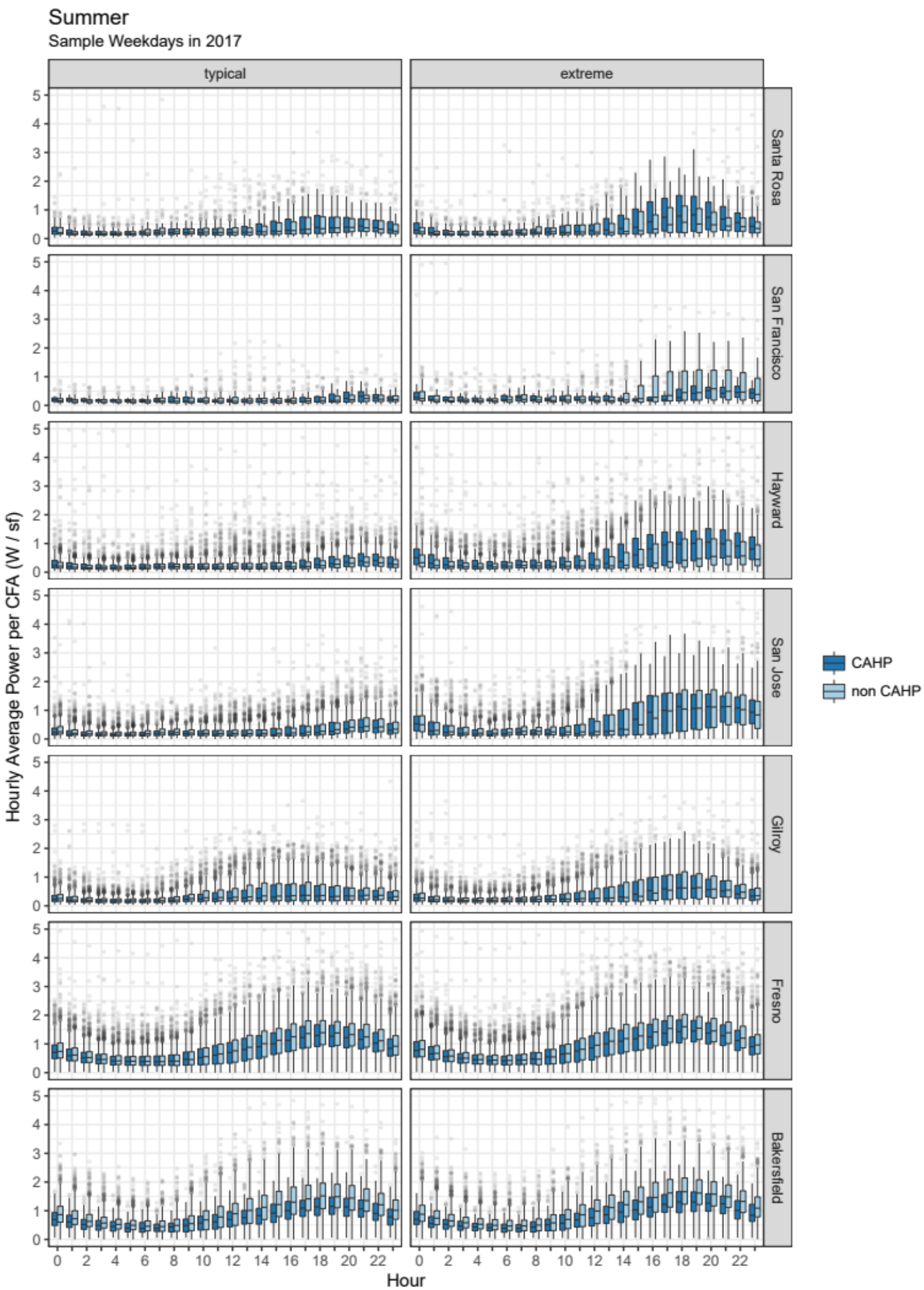

Figure 47. Load profiles for CAHP and non-CAHP houses on a typical and extreme summer day

- Graph is cropped and excludes 169 outliers (out of 371,544 data points). The maximum value is 18 W/sf.
- There is larger variation in hourly average power per sf in the afternoon and evening than during the morning and middle of the day.
- The difference between CAHP and non-CAHP is tiny compared to the within-group variation

Figure 48 shows the median load shapes for CAHP and non-CAHP houses for extreme and typical summer days.

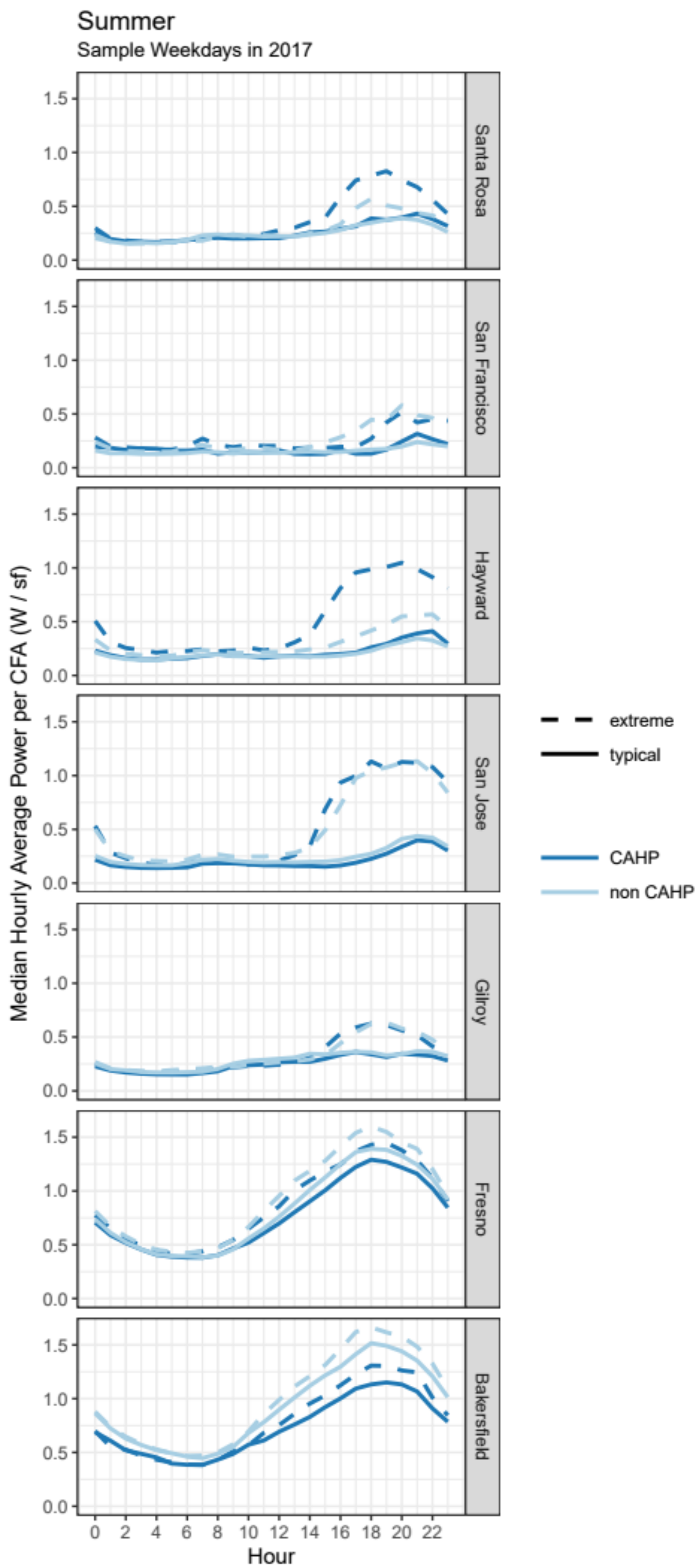

Figure 48. Median average hourly power per floor area on a typical and extreme summer day

- Medians outside of CZ 13 (Santa Rosa, San Francisco, Hayward, San Jose, and Gilroy)

  - Typical summer day
    - No difference between CAHP and non-CAHP
    - Median power is basically flat, except for a tiny peak around 8-9 pm
    - Except for Gilroy, where there was no evening peak
  - Extreme summer day
    - CAHP peak power is higher than non-CAHP in Santa Rosa and Hayward; no difference in other clusters. Perhaps CAHP houses are more likely to have cooling than non-CAHP houses (since AC isn't standard in those climates)
    - Median power is basically flat until early to midafternoon when it starts increasing towards the peak
    - Evening peak is earlier than typical summer day (around 6-7pm vs 8-9 pm), lasts longer (4-5 hours vs 1-2 hours), and begins ramping sooner (around 3 pm)
  - Differences between the typical and extreme days are much larger than differences between CAHP and non-CAHP
- Medians in CZ 13 (Fresno and Bakersfield)
  - Load shape is a stretched out "S" – minimum around 6 am, maximum around 6 pm
  - Timing of the peaks and ramping didn't change for typical vs extreme summer days by more than an hour
  - Bakersfield – more difference between CAHP and non-CAHP than between the typical and extreme days
  - Fresno – CAHP median power on the extreme day is about the same as the non-CAHP median power on the typical day

11.6.2 Winter

Cluster	Extreme or Typical	Date	Max OAT	Ave OAT	Min OAT	CAHP	non CAHP
Santa Rosa	extreme	12/13/17	62	45	27	160	101
Santa Rosa	typical	12/18/17	62	47	32	160	101
San Francisco	extreme	01/06/17	50	45	39	19	101
San Francisco	typical	02/01/17	56	53	50	19	101
Hayward	extreme	01/06/17	52	44	35	260	785
Hayward	typical	01/23/17	54	50	45	260	788
San Jose	extreme	12/22/17	58	46	34	359	895
San Jose	typical	12/06/17	64	51	38	359	895
Gilroy	extreme	01/16/17	55	44	32	645	544
Gilroy	typical	01/13/17	55	47	39	645	544
Fresno	extreme	12/22/17	54	42	29	1,152	1,380
Fresno	typical	12/27/17	61	48	35	1,152	1,382
Bakersfield	extreme	01/27/17	57	46	34	163	1,182
Bakersfield	typical	01/23/17	53	47	41	163	1,181

Figure 49. Characteristics of the sample winter weekdays[1]

1 OAT is outdoor air temperature in °F. Source: NOAA and Weather Underground

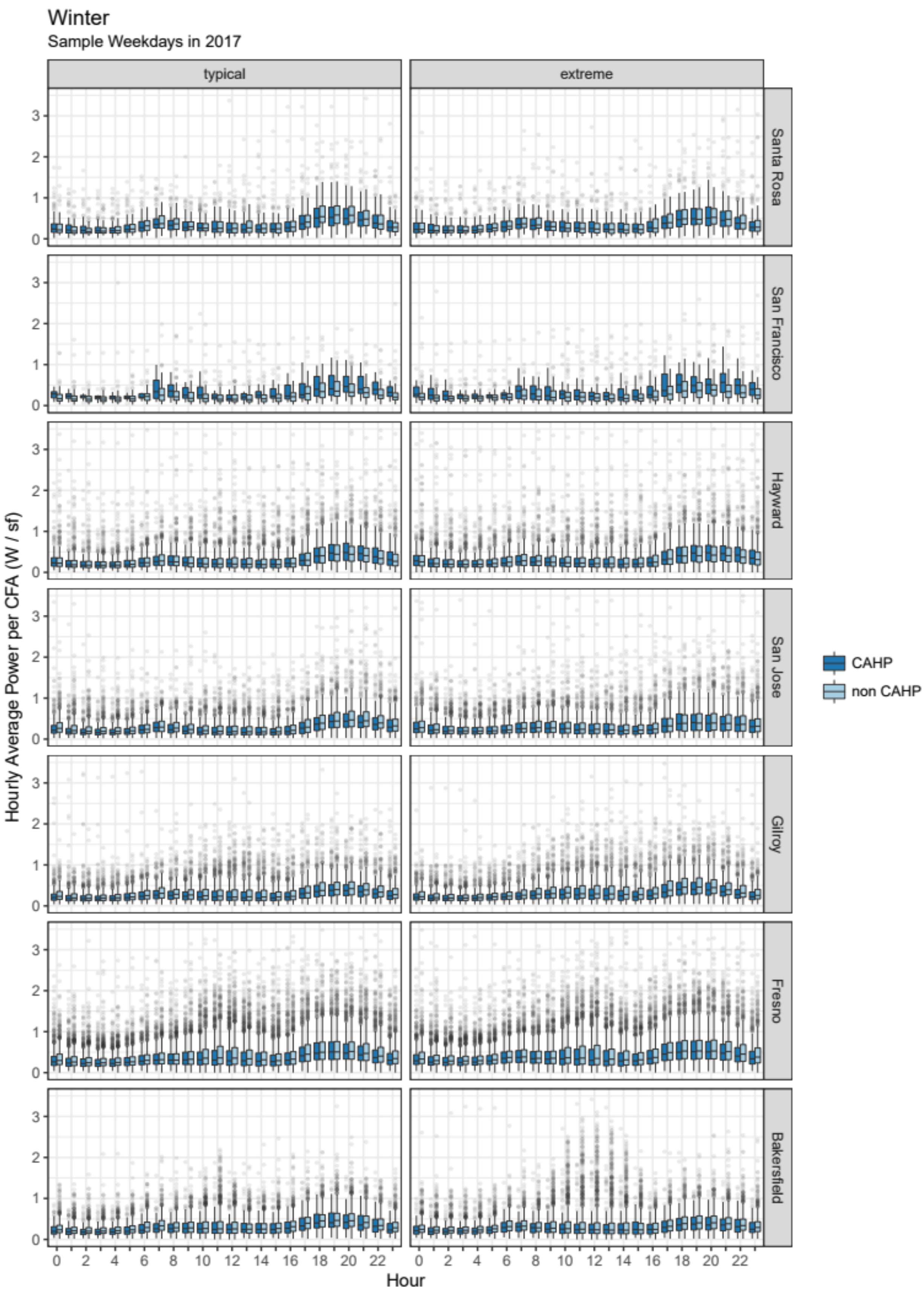

Figure 50. Load profiles for CAHP and non-CAHP houses on a typical and extreme winter day

- Graph is cropped and excludes 303 outliers (out of 371,664 data points). The maximum value is 25 W/sf.
- Note that the scale is different than for the summer load shapes (3.5 W/sf max instead of 5).
- The variation in hourly average power per sf is more constant throughout the day than for the summer days, although it is still slightly higher in the evening in general.
- The difference between CAHP and non-CAHP is tiny compared to the within-group variation

Figure 51 shows the median load shapes for CAHP and non-CAHP houses for extreme and typical winter days.

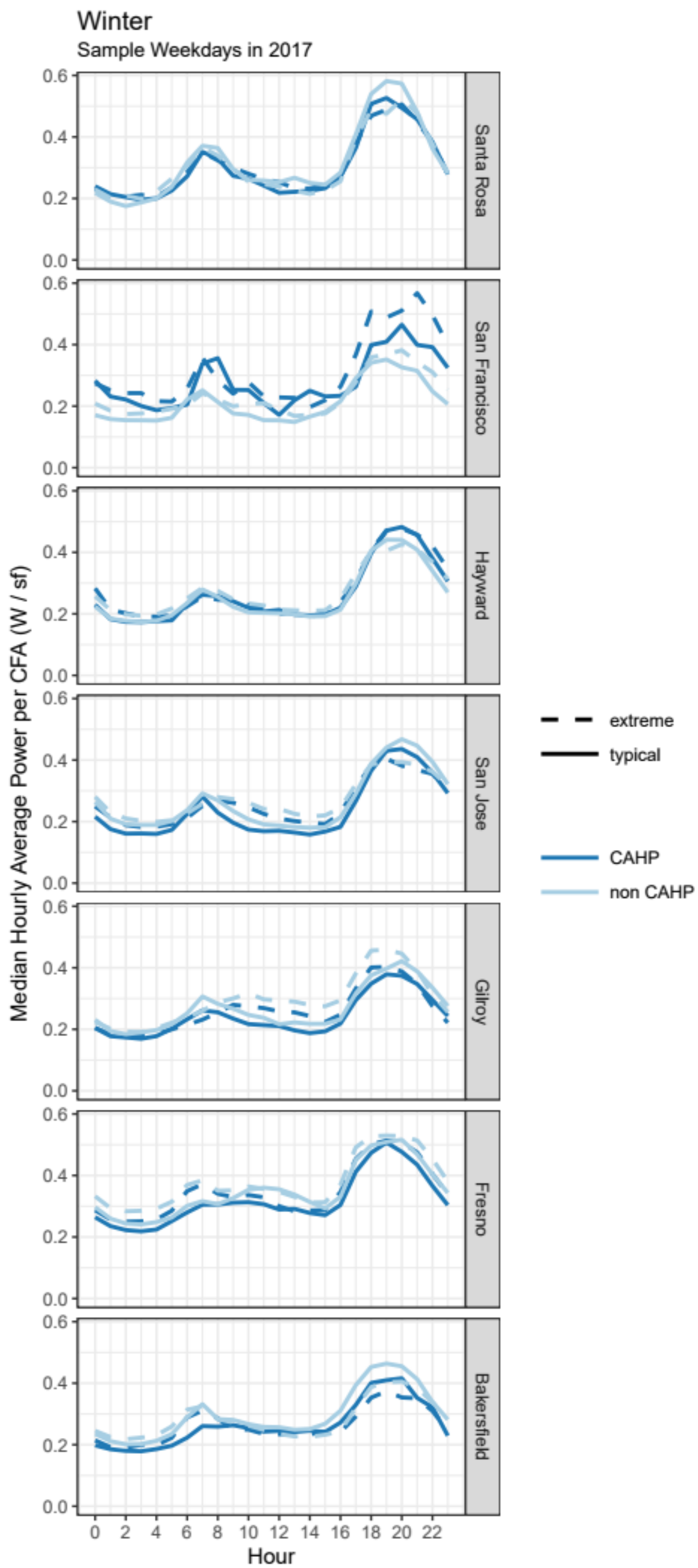

Figure 51. Median average hourly power per floor area on a typical and extreme winter day

- Note that the scale is smaller than for summer days
- Unlike in the summer, the median load shape in all clusters is the same – small morning peak and higher evening peak
  - The morning mini-peak is around 6-8 am. It tends to be later in the warmer CZs than cooler ones.
  - The evening peak is around 7-8 pm
- Very little variation in median load shape between CAHP and non-CAHP or typical and extreme winter days
  - The median CAHP power is higher than non-CAHP in San Francisco for both typical and extreme winter days

11.7 Persistence of Energy Savings

How many houses have more than one year of EUI data (from bills)?

CAHP houses tend to have 1-4 years of data, while a substantial number of non-CAHP houses have up to 7 years of data.

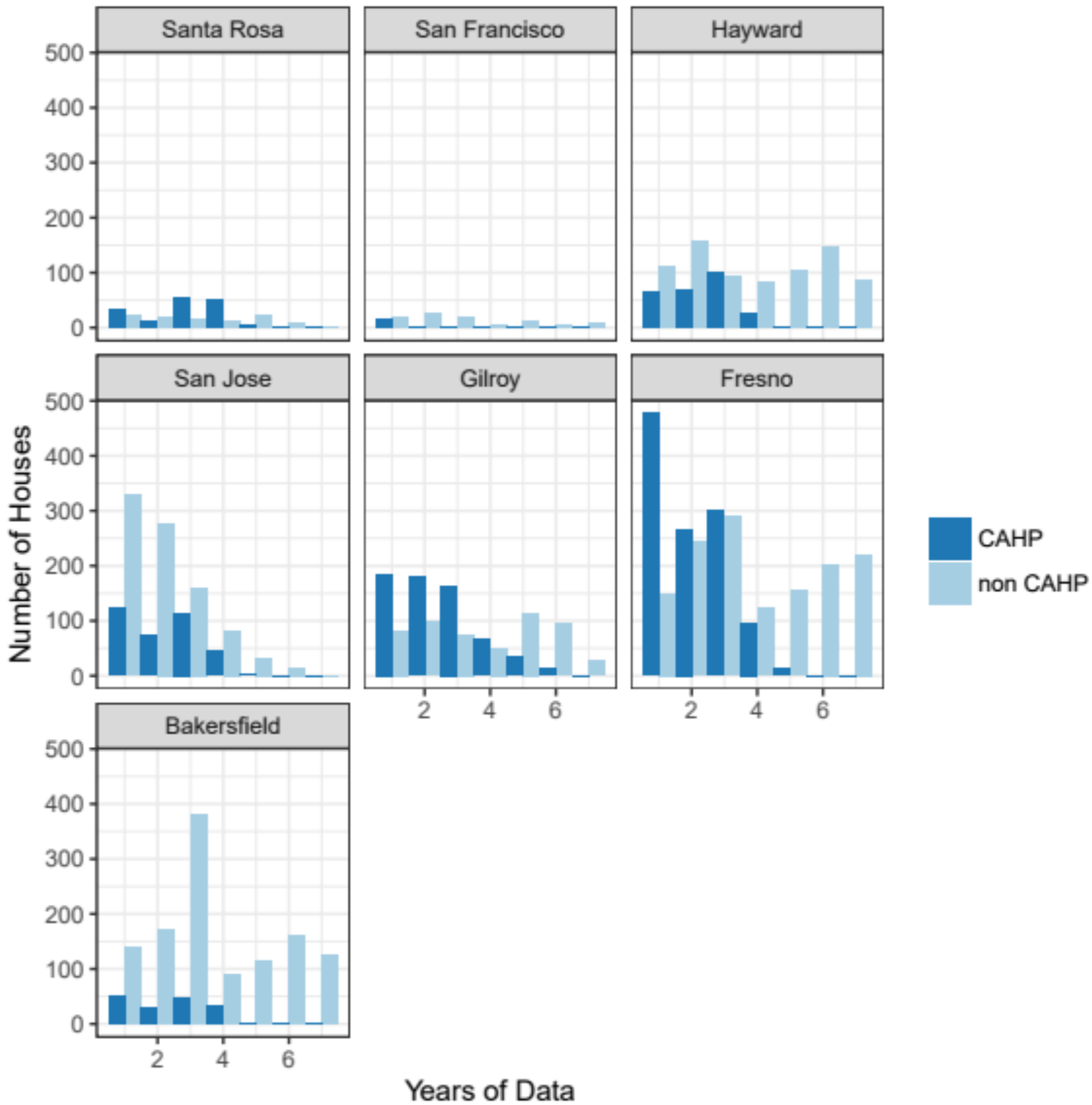

Figure 52. Number of houses with more than one year of data

We compared houses to others that have the same first year of billing data to see whether the relative position of the CAHP and non-CAHP houses is the same. In Figure 53, the x axis is the billing year, and the rows are the first year of data we have for a house. Only groups with at least 20 houses are shown (not all of them have data for each year after the first).

Differences (or lack thereof) between CAHP and non-CAHP are relatively stable across years within clusters and vintages. For example, for houses in Bakersfield for which was have the first year of complete bills in 2015, the median CAHP EUI is below the first quartile of non-CAHP EUI for all 3 years with data. For houses in Fresno with a first year of bills from 2015, there is no difference in EUI between CAHP and non-CAHP houses for any of the 3 years of data.

The performance of houses of different vintages in Fresno is very striking. For houses with the first year of complete bills in 2013, the median CAHP EUI is substantially lower than median non-CAHP EUI. However, houses with the first year of bills in 2015 and 2016 do not show this difference between CAHP and non-CAHP houses. The sample sizes are very different – 21-34 in the earlier vintage compared to a few hundred in the later vintages. Still, the change may be due to the new version of Title 24 that came into effect in mid-2014. However, this difference is not seen in Gilroy, the only other cluster with large enough sample sizes in the earlier vintage.

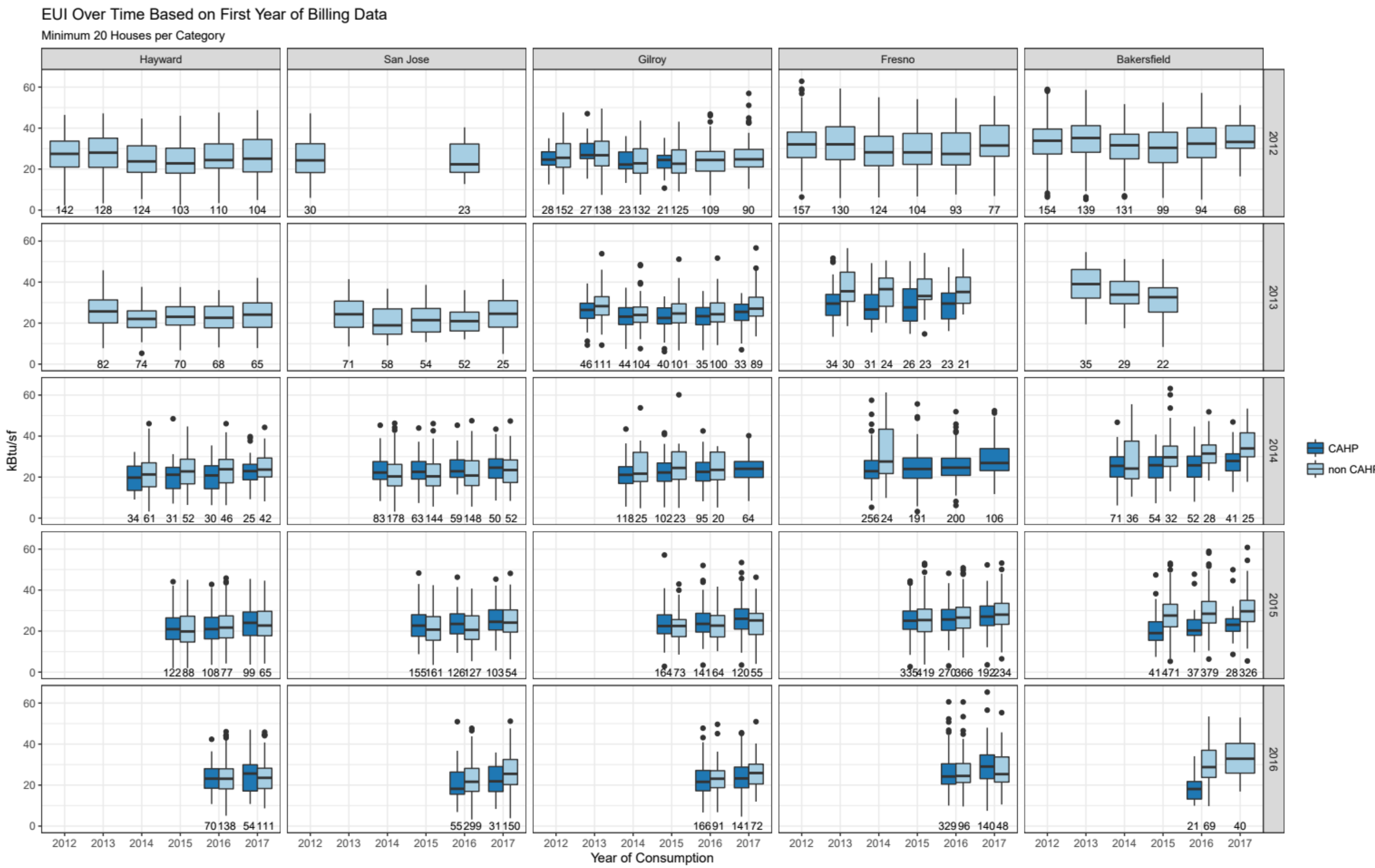

Figure 53. EUI over time based on first year of billing data

How consistent is the relative performance of individual houses? We looked at percentile ranking of each house each year (by cluster, CAHP participation, first year of data) to see how that changes over time. This approach considers year-to-year variability in weather because it's only comparing within the same year, same estimated vintage. However, the sample size is small enough in some groups that small differences in consumption correspond to large differences in percentile ranking.

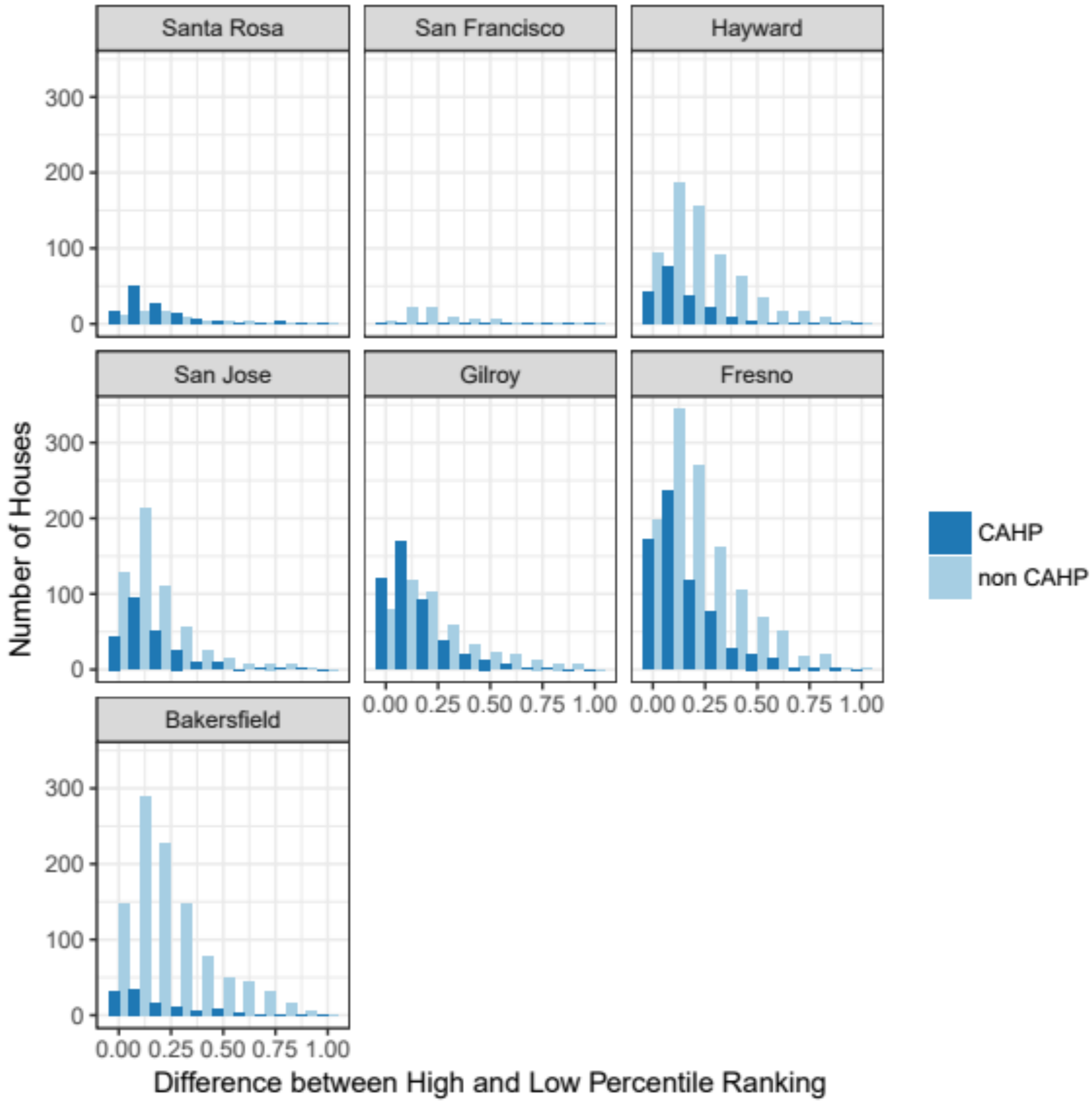

Figure 54. Difference between high and low EUI percentile rankings of houses with more than one year of data

- How much did relative performance vary over time? This is the difference between the highest and lowest percentile ranking that a house got when compared to others in its cluster with the same first year of data, during the same year. If the high consumers were consistently high consumers and the low consumers consistently low consumers, we'd expect the difference to be close to zero. If a house's relative performance in one year were uncorrelated with its relative performance in a different year, then we'd expect the difference to be all over the map – a flat distribution.
- For all the clusters, the peak is left skewed, which indicates some correlation. The most common difference between high and low percentiles was between 10 and 20.

- On average, about a quarter of houses in each cluster changed percentile ranking by more than 25 points, which means that they changed quartiles
- In some clusters, about 10% of houses went from being in the top half of performers to the bottom half
- Relatively stable comparison of CAHP and non-CAHP houses as groups but a lot of variation in the performance of individual houses

Cluster	Analysis Group	N	Percent of houses with a percentile change of more than 25
Santa Rosa	CAHP	127	25%
Santa Rosa	non CAHP	73	37%
San Francisco	non CAHP	76	38%
Hayward	CAHP	193	19%
Hayward	non CAHP	674	35%
San Jose	CAHP	237	20%
San Jose	non CAHP	567	21%
Gilroy	CAHP	458	17%
Gilroy	non CAHP	458	35%
Fresno	CAHP	672	22%
Fresno	non CAHP	1234	34%
Bakersfield	CAHP	113	27%
Bakersfield	non CAHP	1044	36%

Figure 55. Percentage of houses with a change in EUI percentile ranking of more than 25 points

Houses with percentile changes of more than 50 were in the top 50% of consumers in one year and the bottom 50% during a different year.

Cluster	Analysis Group	N	Percent of houses with a percentile change of more than 50
Santa Rosa	CAHP	127	6%
Santa Rosa	non CAHP	73	15%
San Francisco	non CAHP	76	12%
Hayward	CAHP	193	2%
Hayward	non CAHP	674	9%
San Jose	CAHP	237	4%
San Jose	non CAHP	567	5%
Gilroy	CAHP	458	3%
Gilroy	non CAHP	458	12%
Fresno	CAHP	672	5%
Fresno	non CAHP	1234	10%
Bakersfield	CAHP	113	5%
Bakersfield	non CAHP	1044	12%

Figure 56. Percentage of houses with a change in EUI percentile ranking of more than 50 points

11.8 Measured vs. Modeled Energy Consumption of CAHP Houses

For CAHP houses, we compared the predicted electricity and gas consumption from the Title 24 models to the actual measured consumption.

11.8.1 Electricity

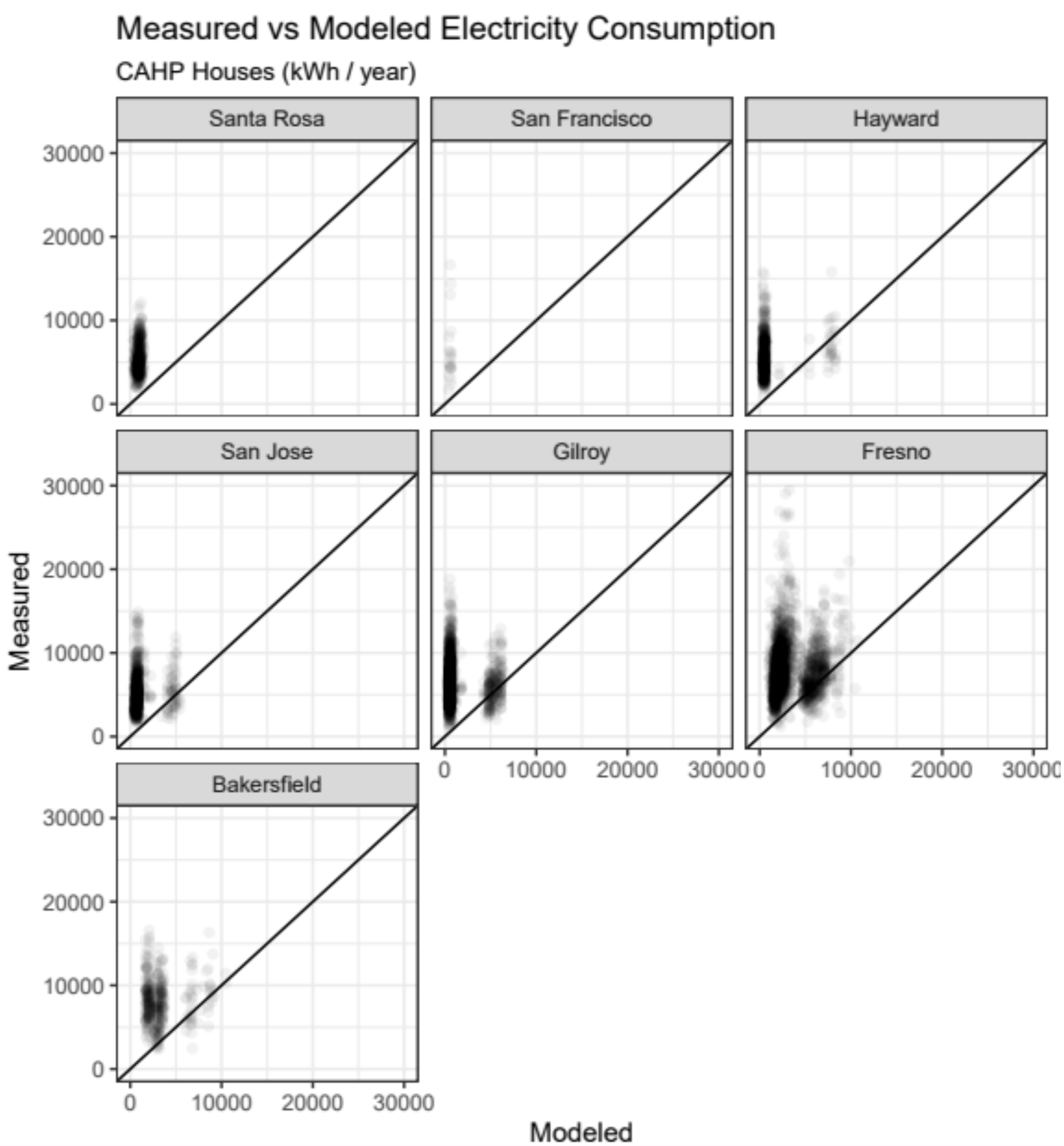

Figure 57. Measured vs. modeled electricity consumption of CAHP houses

- Figure 57 plots measured electricity consumption compared to what was predicted by the Title 24 model. If it were a perfect prediction, the data would all fall on the diagonal line.
- There is a much wider range of measured consumption than modeled.
  - For example, in Bakersfield, the difference between the highest and lowest consumer was 14,000 kWh/yr. for the measured data and only 8,000 kWh/yr. for the modeled data. The difference is even more extreme in other clusters.

- This indicates that the same or similar assets will yield a wide range of actual performance.

- The model underestimated electricity consumption 94% of the time

In addition to the visual analysis we ran single variable regressions with measured kWh, therms, or EUI as the dependent variable and modeled kWh, therms, or EUI as the independent variable. If the modeled values were a perfect predictor of actual consumption, we'd expect

- intercept = 0
- estimate (slope) = 1
- $R^2 = 1$

As shown in Figure 58, the actual regression parameters are rather different. This confirms the visual analysis that modeled electricity consumption is not a good predictor for measured electricity consumption. At best, modeled consumption is accounting for 4% of variation in measured consumption.

	Estimate (slope)		Intercept		R^2
Santa Rosa	1.7	***	3923	***	0.043
San Francisco	2.9		4580		0.006
Hayward	0.2	***	5629	***	0.016
San Jose	0.1		5219	***	0.003
Gilroy	-0.1	***	6792	***	0.006
Fresno	0.1	***	7630	***	0.006
Bakersfield	0.1		7756	***	0.004

Figure 58. Regression parameters for modeled electricity consumption predicting measured electricity consumption

11.8.2 Natural Gas

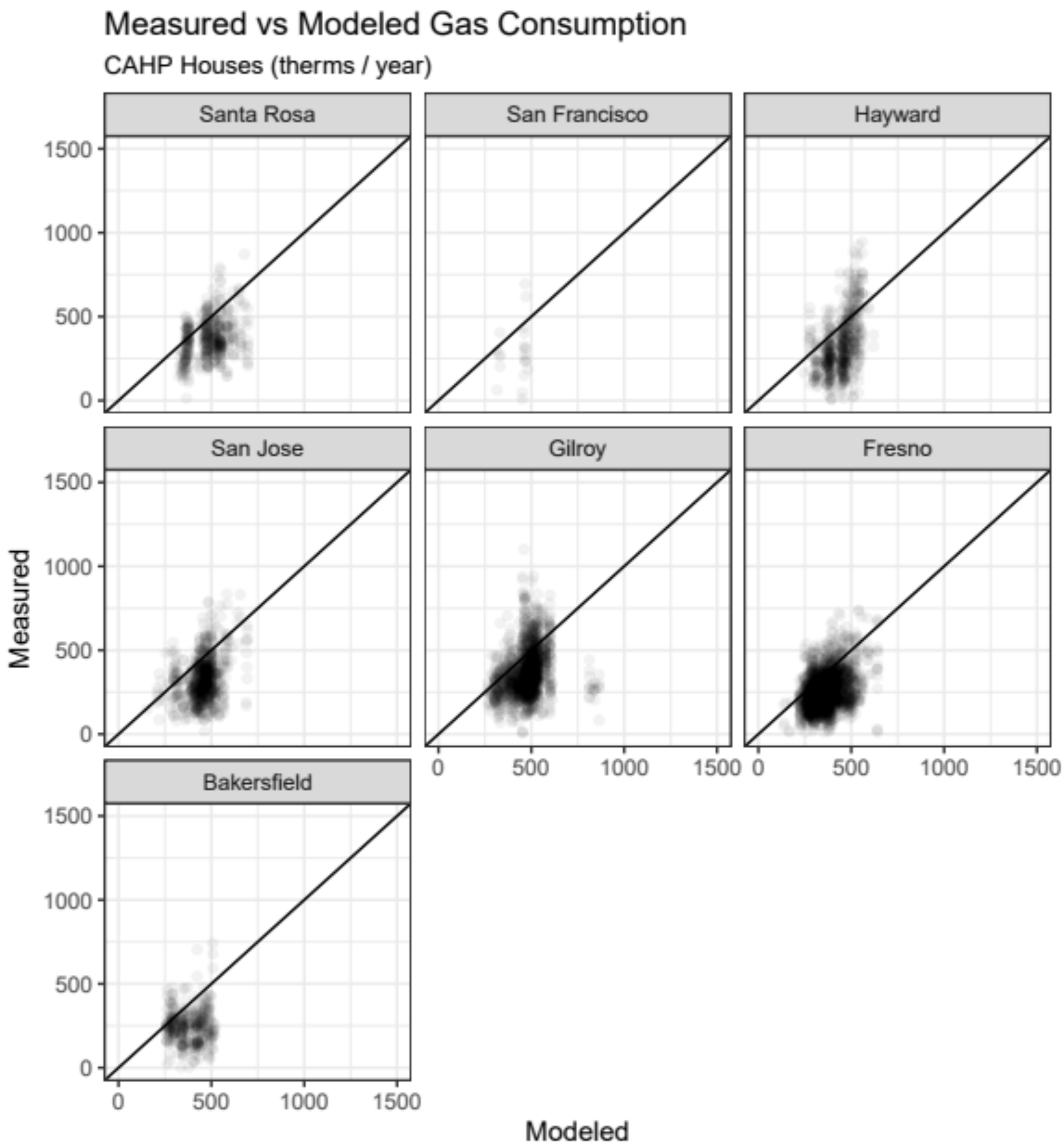

Figure 59. Measured vs. modeled gas consumption of CAHP houses

- Figure 59Figure 57 plots measured gas consumption compared to what was predicted by the Title 24 model. If it were a perfect prediction, the data would all fall on the diagonal line.
- Like electricity, wider range of measured consumption than modeled. Assets that were predicted to perform very similarly actually yielded a wide range of actual performance.
- The model overestimated gas consumption 83% of the time

The single variable regression model of modeled gas consumption predicting measured gas consumption is much better than the analogous one for electricity – it explains up to 13% of the variation in the measured values.

Cluster	Estimate (slope)		Intercept		R^2
Santa Rosa	0.4	***	172	***	0.092
San Francisco	0.5		91		0.031
Hayward	1.1	***	-166	***	0.225
San Jose	0.5	***	82	***	0.097
Gilroy	0.4	***	168	***	0.065
Fresno	0.5	***	76	***	0.130
Bakersfield	0.1		226	***	0.001

Figure 60. Regression parameters for modeled gas consumption predicting measured gas consumption

11.8.3 EUI

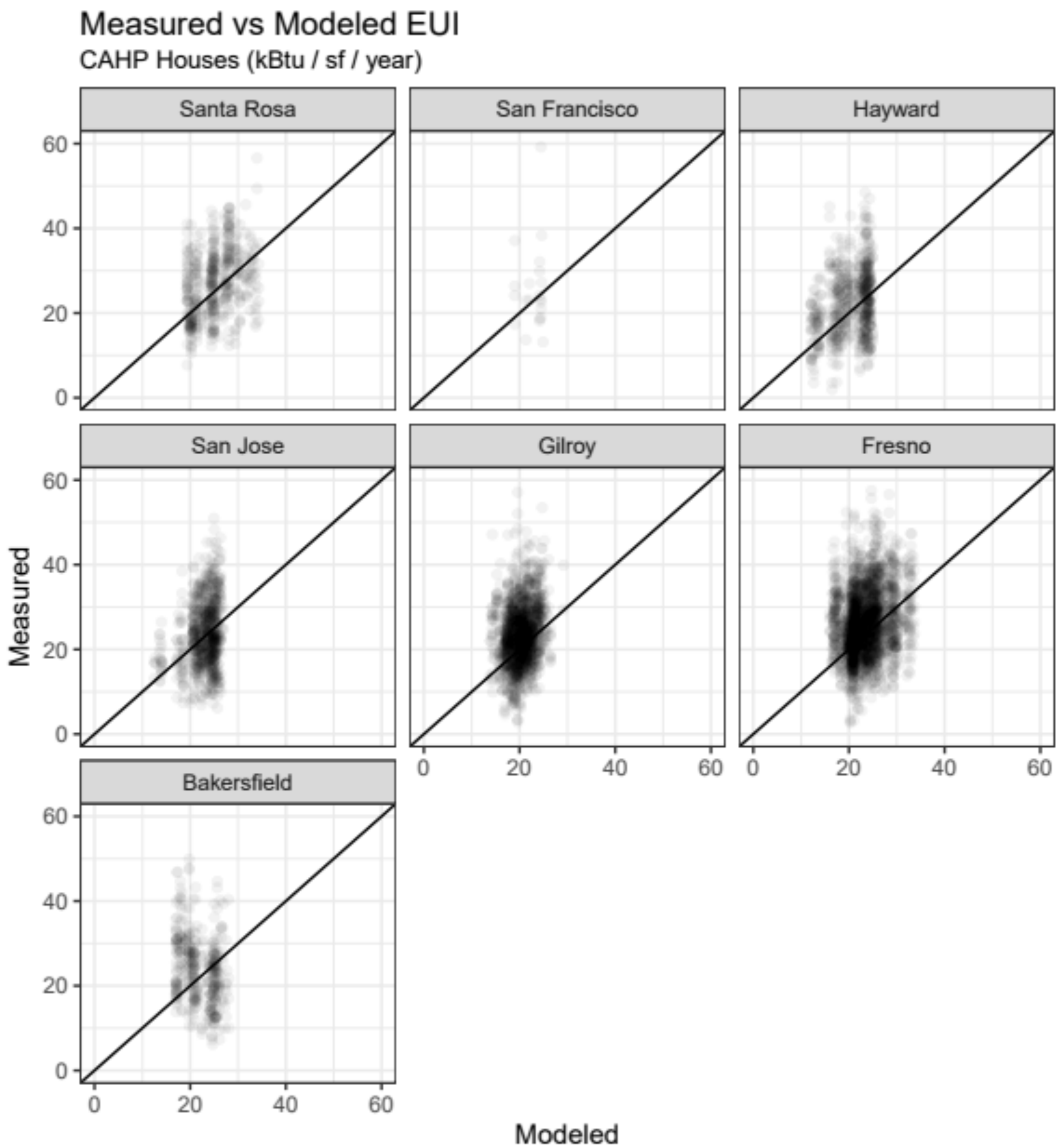

Figure 61. Measured vs. modeled EUI of CAHP houses

- Like electricity and gas, wider range of measured consumption than modeled.
- The model underestimated EUI 60% of the time. Underestimating electricity consumption and overestimating gas consumption means that the EUI predictions were slightly better.

Cluster	Estimate (slope)		Intercept		R^2
Santa Rosa	0.5	***	28724	***	0.099
San Francisco	0.8		12673		0.075
Hayward	1.0	***	3763		0.230
San Jose	1.2	***	-7426		0.152
Gilroy	0.6	***	28417	***	0.049
Fresno	0.8	***	15266	***	0.176
Bakersfield	0.0		49709	***	0.001

Figure 62. Regression parameters for modeled EUI predicting measured EUI

11.8.4 Modeled Metrics

Except for Bakersfield, Title 24 compliance software predicted EUI the best, slightly better than gas, and much better than electricity. Houses tended to use more electricity than predicted, but less gas than predicted.

Modeled EUI accounts for 0-18% of variation in measured EUI depending on cluster.

Cluster	Electricity	Gas	EUI	Electricity	Gas	EUI
	R^2	R^2	R^2	Houses with measured consumption greater than modeled (%)		
Santa Rosa	0.043	0.092	0.099	100%	17%	56%
San Francisco	0.006	0.031	0.075	100%	14%	52%
Hayward	0.016	0.225	0.23	96%	21%	59%
San Jose	0.003	0.097	0.152	96%	15%	52%
Gilroy	0.006	0.065	0.049	94%	20%	67%
Fresno	0.006	0.13	0.176	92%	15%	63%
Bakersfield	0.004	0.001	0.001	93%	11%	57%

Figure 63. Summary of regression models comparing modeled and measured electricity consumption, gas consumption, and EUI

11.9 Limitations

Except for a subset of CAHP houses, which we do not know which end uses use gas or whether it is systematically different between CAHP and non-CAHP houses.