Decarbonize This: Measuring the Building Electrification Baseline With a Comprehensive Meter Dataset

Brian F. Gerke, Mia Hermine[1] and Michael Deshay, Recurve
Nicholas Janusch and Ethan Cooper, California Energy Commission

ABSTRACT

Full electrification of space and water heating is critical to decarbonizing the economy. This transition, now underway and accelerating, will have profound impacts on customer energy costs, grid resiliency, and the cost of serving load. These impacts may be positive or negative depending on how electrification is executed. Ensuring positive outcomes demands an accurate understanding of present-day energy use patterns. For instance, the effects of deploying heat pumps depend on the prevalence not only of natural gas heating but also of air conditioning, since customers who lack AC will gain it, while customers with existing AC may see improved efficiency. High geographic resolution is also important for understanding distribution system impacts and for ensuring equitable customer outcomes. Today's energy-use characterizations rely primarily on survey results with low geographical resolution and little ground-truthing. In this study, we leverage a comprehensive warehouse of customer meter data compiled by the California Energy Commission, applying open-source methods from the OpenEEmeter to characterize residential energy consumption in detail. The result is an accurate, data-driven measurement of the prevalence, by ZIP code, of air conditioning, gas and electric space heating, and gas water heating. We also produce detailed distributions of space conditioning energy consumption, and we draw inferences about the prevalence of other fuels such as wood and propane. Putting meter data to work in this way on an ongoing basis will enable effective forecasting and tracking of electrification impacts and facilitate nimble course-correction on the way to affordable decarbonization.

Introduction

Rapid electrification of legacy fossil end uses to mitigate climate change is poised to spur dramatic near-term impacts for grid management and planning, with substantial load growth expected, along with drastic changes in the timing and seasonality of peak electricity demand (Mai et al. 2018, Kenney et al, 2021, Gerke et al., 2022). Electrifying natural gas space heating, for instance, will create significant new winter morning peaks in regions with high saturation of gas heating. Further, since most heating electrification is expected primarily to use heat pump technology, which also provides space cooling service, the conversion may significantly impact existing summer peak demand, by adding new space cooling load in some cases, while potentially reducing load through improved efficiency where it already exists.

Accurately forecasting the expected scale and timing of these changes is critical for cost-effective expansion of electricity generation capacity and transmission and distribution infrastructure. Planning for electrification-driven power system upgrades (or gas system retirements) requires a detailed understanding of the present-day geographic distribution of relevant gas and electric end-uses. For instance, a distribution circuit with a high saturation of gas space heating, and low AC saturation, may require significant capacity upgrades to

[1] Current affiliation: Moxion Power

accommodate new electric space heating and cooling load; whereas a circuit with high penetration of electric resistance space heating and AC may see overall reductions in peak demand after conversion to heat pump technology. Future generation capacity requirements will also be impacted by changes in the timing and seasonality of peak system loads, which will depend on the underlying geographic distribution of electrification impacts. A concrete example is the California Energy Commission's (CEC) load forecasting tool for electrification impacts, the Fuel Substitution Scenario Analysis Tool (FSSAT) (Sathe et al. 2020), which depends on a detailed geographical understanding of present-day AC saturation in California.

Historically, characterizing the mix of end uses in the customer population has depended mainly on customer surveys conducted at national (e.g., EIA 2022) or regional (e.g., Palmgren et al. 2021, NEEA 2019) scales. Surveys provide important information on the saturations and reported usage of typical energy end-uses in buildings, but they are ultimately limited by their sample sizes, response rates and respondent recollection. In particular, sampling variance makes it difficult to draw conclusions about end-use saturations on fine geographic scales. This limits the usefulness of these surveys for forecasting demand and planning grid upgrades, especially on the distribution-system level or in transmission-constrained pockets of the grid.

The emergence of advanced metering infrastructure (AMI) with daily, hourly, or even finer time resolution provides an alternative pathway to assessing energy-consumption patterns and end-use penetrations. Weather normalization and other modeling approaches applied to AMI data enable consumption from different end-use categories to be disaggregated from whole-building data, which can enable data-driven understanding of consumption patterns across the population (e.g., Recurve 2022). Barriers to unlocking these insights for policymakers and utilities include data accessibility, data scale, and computing power. Meter data is sensitive information requiring strong data security protocols, and applying modeling algorithms to hundreds of thousands or millions of customers rapidly outstrips the capabilities of consumer-grade computing infrastructure.

In this study, we estimated saturations of residential air conditioning, electric and gas space heating, and gas water heating in the service territory of Pacific Gas and Electric Company (PG&E), for a sample of some 3.6 million customers, based on a unique whole-population AMI dataset compiled by the CEC. We leveraged Recurve's secure, massively parallel computing platform to run the open-source OpenEEmeter software[2] on the dataset, which enabled disaggregation of temperature-dependent consumption for each customer. By analyzing the resulting population-level distributions, we identified threshold consumption criteria that appeared to indicate the presence of primary heating or cooling. We also developed criteria that allowed us to infer the presence of gas water heating. Applying these criteria across the population yielded a detailed and highly granular estimate of the penetration of these end uses within the population, providing a well-measured baseline against which the impacts of space and water heating electrification can be forecasted and measured.

It is important to note certain boundaries and limitations of this study. Perhaps most importantly, our criteria for identifying the presence of different end uses have not been validated against detailed site-level audits (although they compare favorably to existing survey results). Further, the present study is limited to residential customers of a single utility in Northern California. Although PG&E's service territory covers a very broad range of climates, the details of our end-use inference strategy may not be well tuned for use in other utility service

[2] https://lfenergy.org/projects/openeemeter/

territories or for non-residential customers. Further efforts in any of these areas would refine and improve the accuracy of the methods we demonstrate here. In the remainder of this paper, we first describe the CEC dataset and the approach to disaggregating temperature-dependent load. We then describe the distributions of disaggregated loads in the population and our approach to inferring primary heating and cooling fuels. Finally, we illustrate the power of these estimates for population-level analysis, presenting example population-level statistics on end-use penetration for residential customers by California building climate zone and by ZIP code.

Methodology

Approach to Estimating Heating and Cooling Loads

To disaggregate estimated customer space heating and space cooling loads we applied the OpenEEmeter version 3.0[3] to each customer's gas and electricity consumption data. OpenEEmeter 3.0, an open-source project of the Linux Foundation Energy, is the software implementation of the CalTRACK 2.0 modeling framework,[4] which specifies an approach to performing weather-normalized modeling of customer energy consumption data. The framework was developed through an open, stakeholder-driven process with participants from across the demand-side management industry. OpenEEmeter consists of two separate models—a *daily* model and an *hourly* model—that are used to model data of different frequency. The daily model is a change-point temperature regression model that models the response of a customer's daily electricity or gas consumption to the average daily outdoor air temperature (at a nearby weather station) using piecewise linear regression in three different temperature regimes. Energy consumption is assumed to decrease with temperature below the heating balance-point temperature, increase with temperature above the cooling balance point, and be insensitive to temperature between the two balance points. This structure is illustrated in Figure 1. The balance points and the slopes of the heating and cooling response are parameters of the model that are optimized by the model fitting function. The algorithm also considers models with only a heating or cooling balance point, or neither, to account for buildings without heating or cooling from the fuel being analyzed. The hourly model fits energy consumption as a function of time of week and temperature, using piecewise linear regression for each hour of the week in several fixed temperature bins. It also performs a preliminary step to infer occupied and unoccupied hours and fits separate models for each occupancy state.

The daily model enables disaggregation of temperature-dependent usage from the customer's total energy consumption. As illustrated in Figure 1, the model identifies a flat base consumption between the two balance points, with consumption increasing as the temperature moves away from the balance points in either direction. Consumption above the base level can thus be identified as space heating load below the heating balance point and as space cooling load above the cooling balance point. Because of the structure of the hourly model, it is less straightforward to identify heating and cooling consumption in its outputs. Thus, in the rest of this study, we will rely on heating and cooling loads disaggregated using the daily model to identify customers with heating and cooling. We applied the daily model to the hourly electricity AMI data by aggregating the hourly data up to a daily level, and we applied it to the monthly gas data by dividing the monthly billing consumption by the number of days in each billing period to

[3] This version of OpenEEMeter, the latest at the time of analysis, has since been superseded by OpenEEMeter 4.0.
[4] https://docs.caltrack.org/en/latest/

yield approximate daily gas consumption, consistent with the published CalTRACK methods. The model fits resulted in estimated electric heating and cooling consumption, and estimated gas heating consumption, for each of the 3.6 million residential sites selected for analysis.

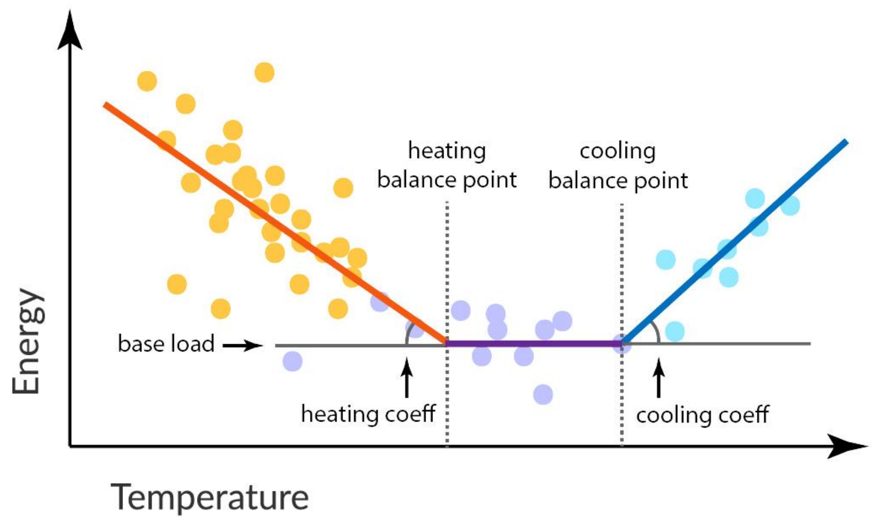

Figure 1. A schematic diagram of the piecewise change-point regression model used for daily temperature normalization modeling in OpenEEmeter 3.0.

For comparison purposes, we also computed each site's aggregate electricity and gas consumption over the full year and in the summer (June-September), winter (November-February) and shoulder (October and March-May) seasons. For customers with rooftop PV, we compute aggregate *net* electricity consumption (i.e., the net total of all imports and exports) and aggregate *delivered* energy consumption (i.e., the total of imports only). The delivered value is more practically useful as a basis for comparison, as it cannot be negative and is typically well above zero, providing a better (if imperfect) representation of the customer's behind-the-meter consumption. For customers without rooftop PV, net and delivered consumption are identical.

Dataset

The dataset used in this study consists of hourly electricity and monthly natural gas and consumption data for all residential customers in PG&E service territory, covering a period of one year for each customer, starting in October of 2020 (exact start dates may vary according to the specifics of customers' billing cycles). In addition, the dataset contains metadata describing certain characteristics of each customer and property, such as geographical information, presence of rooftop photovoltaic (PV) systems, gas and electric rate codes, and enrollment in low-income energy rate programs.

We selected a subset of the population, consisting of all residential customers who:

- Receive electricity distribution service from PG&E
- Do not fall within the service territory of a different gas utility
- Have sufficient energy consumption, defined as at least 500 kWh over the analysis year, to infer that the household was plausibly occupied for part of the year

- Have sufficient data for modeling and can be accurately modeled by OpenEEmeter[5] over the analysis year

This selection includes customers of California Community Choice Aggregation (CCA) electricity providers in PG&E service territory, since PG&E delivers electricity to these customers. It also includes PG&E all-electric customers (i.e., customers without gas service) in the specified territory. It *excludes* the following customer categories:

- PG&E gas customers who receive electricity from a different utility
- Certain PG&E customers who are in SoCalGas service territory
- Customer sites that were unoccupied or minimally occupied during the analysis year
- New construction and sites with customer move-out or move-in occurring during the analysis year, since these customers will not have sufficient data for modeling
- Customers who experienced lengthy power shutoffs during the analysis year
- Any other sites that have insufficient gas or electricity data in the analysis year to support modeling with OpenEEmeter

The resulting dataset includes 3.6 million households in PG&E service territory. With this selection, we proceeded to estimate the heating and cooling loads for each customer.

Approach to Estimating End-use Saturations

Figure 2 shows the distribution of estimated cooling consumption, as a percentage of each customer's delivered summer electricity consumption, for customers in each California Title 24 climate zone covered by the analyzed population.[6] As shown in Figure 2, each distribution includes some customers with very low (but nonzero) cooling consumption. This may indicate temperature-dependent consumption other than direct expansion air conditioning, such as use of fans or evaporative cooling, or it may simply reflect noise in the model outputs. Inspecting the distributions in Figure 2 suggests the presence of two distinct populations, with summer cooling fractions falling below and above 20%, respectively. Based on this observation, we assumed that customers were likely to use air conditioning as their primary source of space cooling if their cooling usage exceeded 20% of their delivered summer electricity usage.

A drawback of this assumption is that some customers with very large overall consumption (e.g., customers with electric vehicles, pools, and spas) might have significant AC usage that nevertheless represents less than 20% of their summertime consumption. Such customers would be miscategorized by a simple 20% threshold. To mitigate this effect, we also set a threshold in absolute annual cooling consumption of 500 kWh, above which customers were inferred to use air conditioning as a primary source of space cooling. Figure 3 is a scatter plot of summer cooling percentage vs. total annual cooling consumption. There is a strong correlation between the two parameters, but with significant scatter. A simple threshold in either dimension would exclude customers who appear to have significant cooling according to the

[5] Specifically, OpenEEmeter 3.0 requires 7446 hours of data for the hourly model, and 328 days of data for the daily model, and we required that the model fitted to the data have a coefficient of variation of the root mean squared error (CVRMSE) no greater than 1.0.
[6] Not shown in Figure 2 are customers for whom the best model included no cooling balance point, whose estimated cooling load is exactly zero. Such customers make up about 18% of the analyzed population.

other metric but have scattered above or below the main trendline. Applying both thresholds, as shown in Figure 3, allows us to capture a more complete selection.

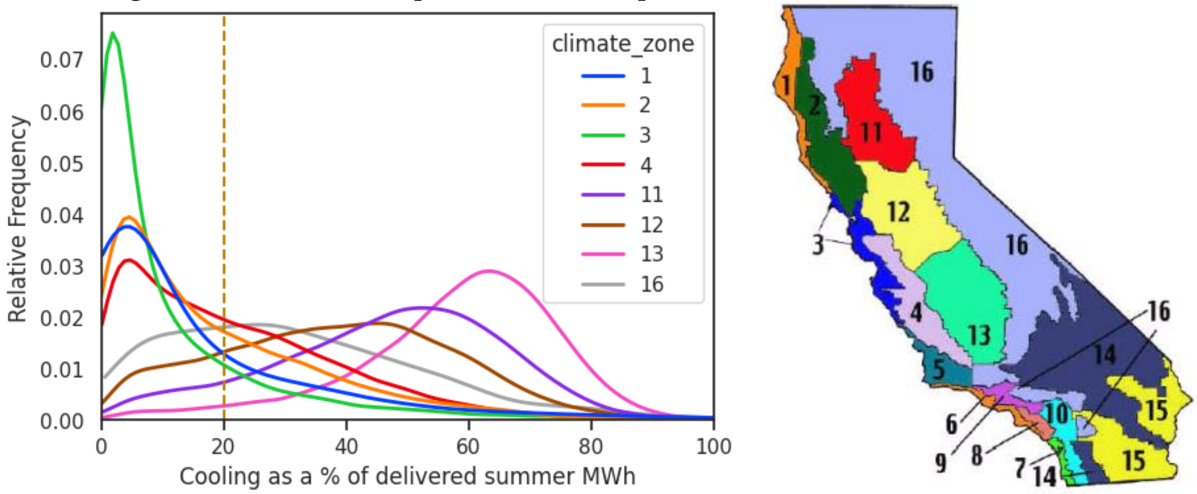

Figure 2. *Left:* Distributions of customer cooling consumption as a percentage of each customer's total summer electricity consumption, by California Title 24 climate zone. *Right:* A map of the Title 24 climate zones. The population analyzed in this study has customers in climate zones 1, 2, 3, 4, 11,12, 13, and 16.

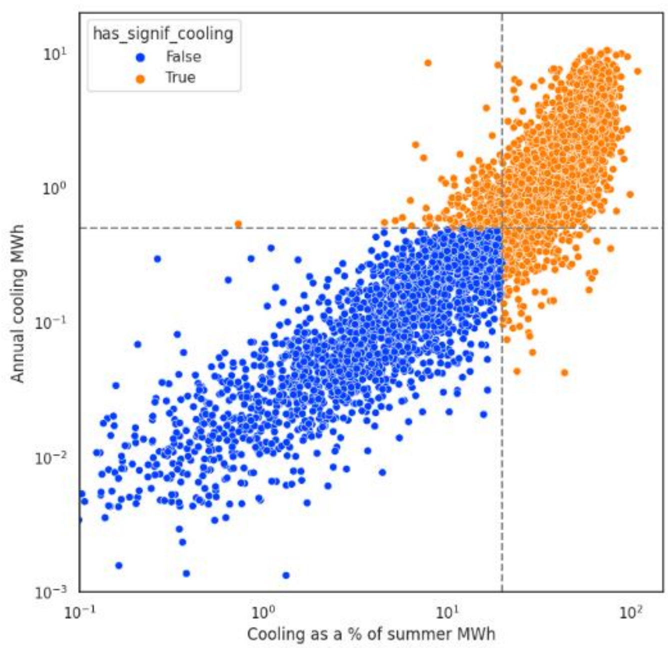

Figure 3. Scatter plot showing customers' cooling load as a fraction of their summer delivered electricity, versus absolute annual cooling consumption, on a log-log scale. The dashed lines denote the thresholds used to identify significant air conditioning usage in this study, and the colors indicate the selection.

We took a similar approach to inferring space heating saturations. For both electricity and gas consumption, we inspected the distribution of heating consumption (as estimated using OpenEEmeter) as a percentage of total delivered wintertime usage, and we selected thresholds that appeared likely to indicate the use of each fuel as a primary heating source. We additionally defined thresholds in absolute heating consumption to account for customers with unusually large non-temperature-dependent wintertime consumption. Table 1 summarizes the various thresholds used to identify primary cooling and heating end uses in this study.

While examining customer usage distributions, we also found a method to draw inferences regarding water heating usage. Figure 4 shows the distribution of estimated gas heating consumption across the entire analyzed population, as a percentage of each customer's annual gas consumption. A small but notable population exists around 100%[7] of total gas usage for space heating. Because water heating is also a large end use, these customers can safely be inferred *not* to have gas water heating. Therefore, we can conclude that most *other* customers in this population have gas water heating, so we set a maximum threshold of 95% in this parameter, below which customers are inferred to have gas water heating. A possible exception is customers who have very low annual gas consumption, inadequate for water heating. To exclude these customers, we set a minimum threshold at 50 Therms of annual gas consumption to infer the presence of gas water heating.[8] These thresholds are also summarized in Table 1. They are applied slightly differently than the space heating thresholds above, in that *both* thresholds must be met to infer the presence of gas water heating.[9]

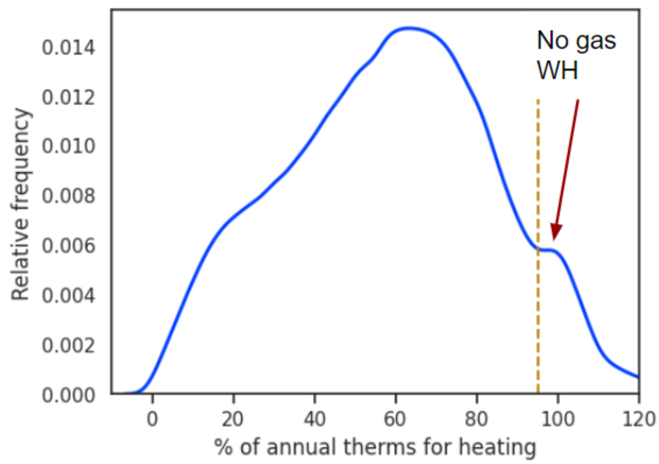

Figure 4. Distribution of gas space heating consumption as a percentage of each customer's total annual heating consumption. A notable population uses approximately 100% of their gas consumption for space heating, which implies that these customers do not have gas water heating.

With these thresholds in place, we can estimate the saturations of primary space heating and cooling end uses in different regions. First, we grouped customers by region, then we computed the fraction of customers who were inferred to have each end use. In the case of electric space heating, we found that customers with rooftop PV had estimated electric heating

---

[7] The distribution extends to above 100% in part because the curve has been smoothed but more substantively because the estimated heating consumption is the output of a regression model, which will naturally over or under-estimate consumption to some degree.

[8] This threshold is intended to represent a conservative bare minimum amount of gas consumption by a gas water heater. According to investigation of EnergyGuide data, an ultra-efficient condensing tankless water heater uses approximately 180 Therms annually, so this threshold would represent a seasonally occupied site with very efficient gas water heating.

[9] Because of the greater diversity of electrical end uses, there is no corresponding set of thresholds that can be used to infer the presence of electric water heating. The inverse of the gas water heating saturation can be used as a reasonable estimate (though it will be a slight overestimate since some customers use propane for water heating).

consumption patterns substantially different from the rest of the population. This was driven by a shortcoming of temperature regression for customers with PV: the correlation between temperature and solar irradiance can be incorrectly interpreted as heating load by the model. For this reason, we excluded customers with PV when estimating the saturation of electric space heating. Our conclusions regarding this end use rest on the assumption that PV customers do not have a materially different saturation of electric space heating from other customers.

Table 1. Thresholds used to infer primary heating and cooling end uses in this study.

End use	Percentage threshold	Threshold combination logic	Consumption threshold
Air conditioning	>20% of summer delivered electricity for cooling	OR	>500 kWh of cooling electricity consumption
Electric space heating	>40% of winter delivered electricity for heating	OR	>1000 kWh of heating electricity consumption
Gas space heating	>50% of winter delivered gas for heating	OR	>100 Therms of heating gas consumption
Gas water heating	<95% of annual delivered gas for heating	AND	>95 Therms of total annual gas consumption

Findings

Table 2 presents the population-level end use saturations estimated in this study using OpenEEmeter (denoted OEEM in the table), including fraction of households with nonzero heating or cooling consumption detected by the model, as well as the fractions that pass the significance thresholds in Table 1. For comparison, Table 2 also presents saturations reported for PG&E service territory in the CEC's Residential Appliance Saturation Study (RASS) (Palmgren et al. 2019), including estimates of both primary and auxiliary space heating and cooling loads.

Table 2. Saturations of heating and cooling end uses considered in this study, as detected by OpenEEmeter (OEEM) modeling and the significance thresholds described above, and as estimated in the RASS study for primary end-use fuels and primary plus auxiliary end-use fuels.

End use		OEEM detected	OEEM + significance	RASS* primary	RASS* primary + auxiliary
Space cooling		82%	51%	51%**	74%**
Electric space heating†		83%	22%	19%	28%
Gas space heating‡		97%	83%	77%	79%
Gas water heating‡		NA	90%	89%	NA

* RASS values as reported for PG&E service territory
** Here, RASS "primary" space cooling includes central AC; auxiliary space cooling includes room AC and evaporative cooling.

† OEEM-detected saturations for electric space heating are computed as a function of homes without rooftop PV.
‡ Gas space heating and water heating values are calculated as a percentage of homes with gas meters.

Notably, the detections of "significant" end uses by OEEM match very closely with the estimates of primary heating and cooling fuels from RASS. However, the OpenEEmeter detections of nonzero heating and cooling consumption are considerably higher than the RASS estimates of primary-plus-auxiliary end-uses. The discrepancy is especially large for electric space heating, where OpenEEmeter finds that 83% of customers have nonzero detected consumption, while RASS estimates that only 28% of customers use any form of electric heating. Although it is possible that some of the OpenEEmeter detections with very low consumption reflect noise in the modeling outputs, the size of the discrepancy suggests that RASS may also be missing a significant amount of auxiliary electric heating consumption. This may imply underreporting of portable electric space heater usage by survey respondents.

Results by California Building Climate Zone

OpenEEmeter allows us to detect heating and cooling end uses on the household level, so we can estimate saturations at any level of geographical granularity that is supported by the metadata available in the CEC database. We can also look at different combinations of end-uses. In this section, we characterize various end-use saturations by climate zone.

Figure 5 shows the saturations of detected and significant space cooling consumption in each climate zone. Not surprisingly, cooler coastal climate zones (e.g. 1 and 3) have very few customers with significant cooling load, whereas in hotter inland zones (e.g. 11 and 13) significant cooling is nearly universal. Notably, in nearly all climate zones, the majority of customers have nonzero cooling load, suggesting use of fans or some room air conditioning in the cooler climate zones. The exception is climate zone 1, where 70% of customers have no cooling load at all, consistent with the especially cool climate along California's north coast.

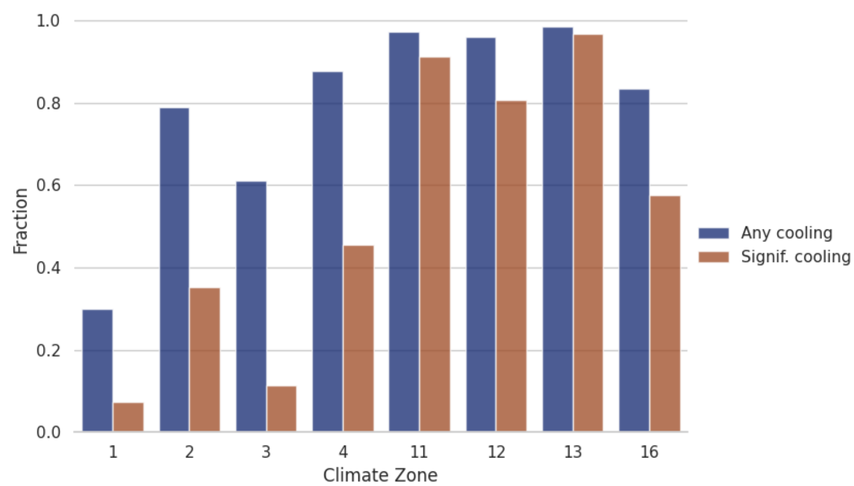

Figure 5. Space cooling saturations by California climate zone, as detected by OpenEEmeter modeling. Saturations are shown for customers who have any cooling load detected and for customers whose cooling load passes the significance thresholds defined for this study.

Figure 6 compares the fraction of customers with detected gas space and water heating to the fraction of customers with gas meters, by climate zone. There is considerable variation across the state in all three saturations, but notably the heating and water heating saturations are very close to the overall fraction of households with gas meters in each climate zone. The key finding from this figure is that, although access to gas service varies drastically by region, when customers have gas service, they almost universally use it for space and water heating.

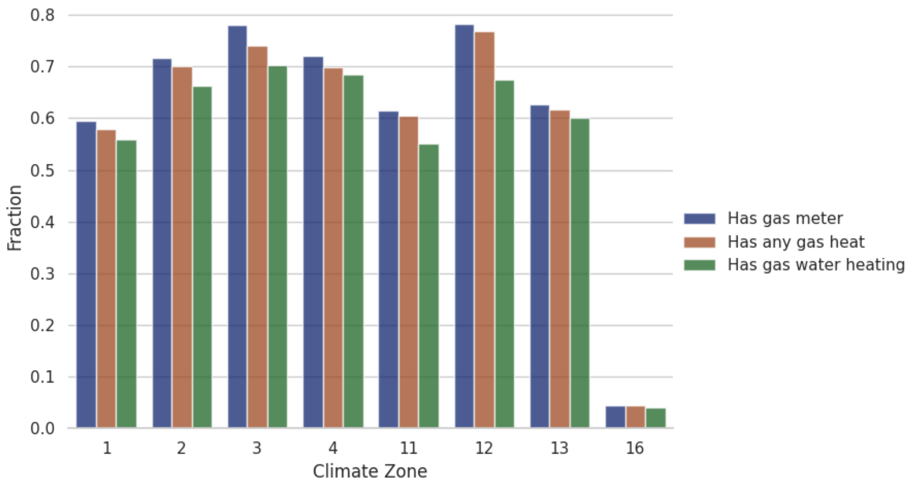

Figure 6. Estimated saturations of gas meters, gas heating, and gas water heating, by climate zone.

Figures 7 and 8 compare the fractions of households with detected gas and electric heat to the fractions with *significant* consumption from those end uses, and they also consider the fraction of customers without space cooling in each group. A few interesting patterns emerge. First, when households have detected gas heating, that usage is nearly always significant. For electric heating, the opposite is true: a large majority of households have at least some electric heating, but a small minority use significant amounts. This suggests that there is very little use of gas, but widespread use of electricity, for supplemental heating in homes.

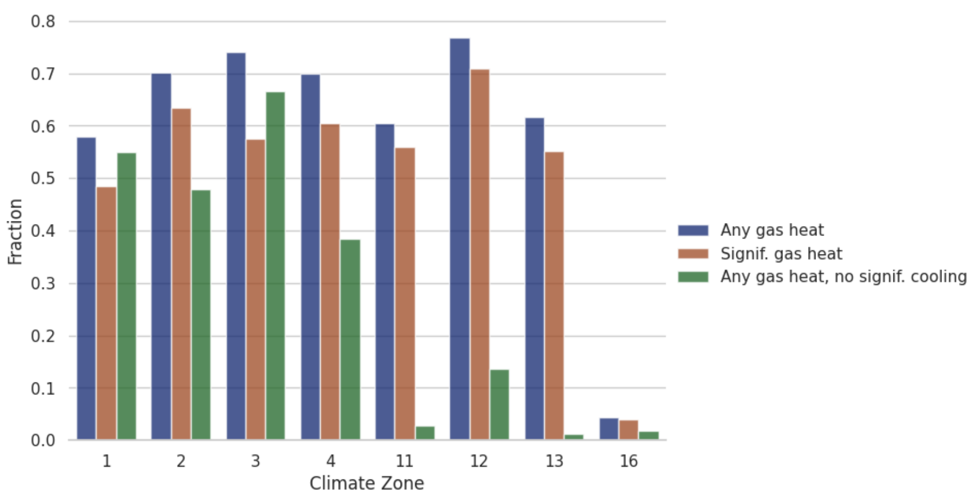

Figure 7. Fractions of households with detected gas heat, significant gas heat, and detected gas heat with no significant cooling consumption, by climate zone.

It is also interesting to look at the fractions of customers who do *not* have significant cooling load. Households with any gas heating and those with primary electric heating are potentially good targets for conversion to heat pumps, but those who do not currently have air conditioning will gain it after the switch, which may add significant summer peak load for the utility. Figures 7 and 8 indicate that customers in the hot climate zones 11,12, and 13 are unlikely to add significant new cooling when converting to heat pumps (in fact, they may reduce their cooling load via efficiency gains) but those in cooler climates have the potential to add significantly to summer peak demand.[10] Understanding this dynamic is critical to planning for generation and grid capacity expansion in the context of rapid electrification.

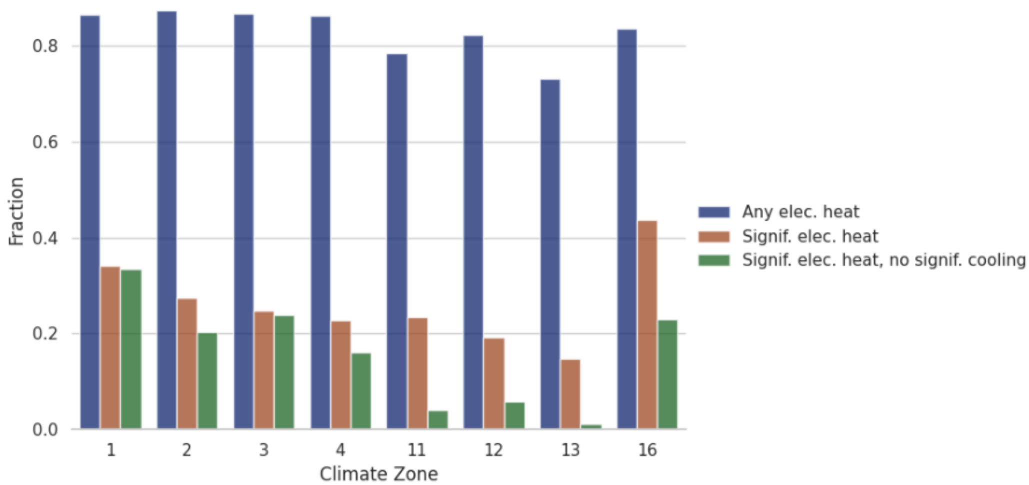

Figure 8. Fractions of households with detected and with significant electric heat, and with significant electric heat without cooling, by climate zone.

Finally, Figure 9 presents the fraction of households who have significant heating consumption from *both* gas and electricity. Such customers may be good candidates for a whole-building retrofit, incorporating weatherization alongside heat-pump conversion to improve overall comfort while reducing energy consumption. These customers represent a significant population—more than 10% of the total in coastal regions. Figure 9 also shows the fractions of customers without significant gas or electric heating and with no gas or electric heating detected at all. The latter category likely represents customers relying on delivered fuels (e.g., propane or wood) for all their heating, whereas the former category may represent customers using delivered fuels for their primary heating, as well as other customers who may be limiting their use of space heating due to a high energy-cost burden. Identifying these customer categories is important for ensuring equitable access and equitable outcomes for electrification programs. This figure demonstrates that analysis of meter data can be used indirectly to draw inferences about these often hard-to-reach populations.

[10] For preliminary measurements of this effect, see the paper by Kerrigan et al. in these proceedings.

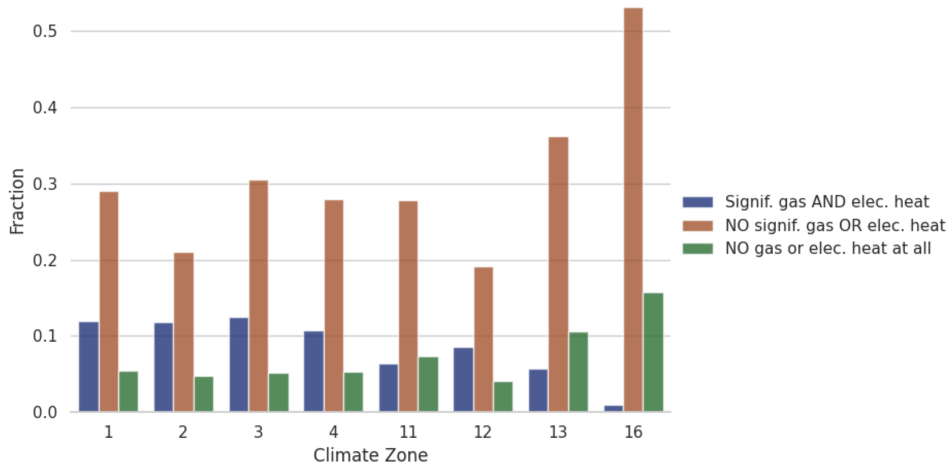

Figure 9. Fractions of customers with significant heating consumption from both fuels, with significant heating from neither fuel, and with no heating at all from either fuel.

Results by ZIP Code

It is also interesting to examine end-use saturations on a finer level of geographic granularity, for instance to identify areas that may be particularly well suited for community electrification, or to identify areas that may be in need of distribution-system upgrades to support electrification. In this section, we look at end-use saturations on a ZIP code level to gain a more fine-grained understanding of end-use variation. Figure 10 shows heat maps of the fractions of customers with significant air conditioning and gas heating consumption.

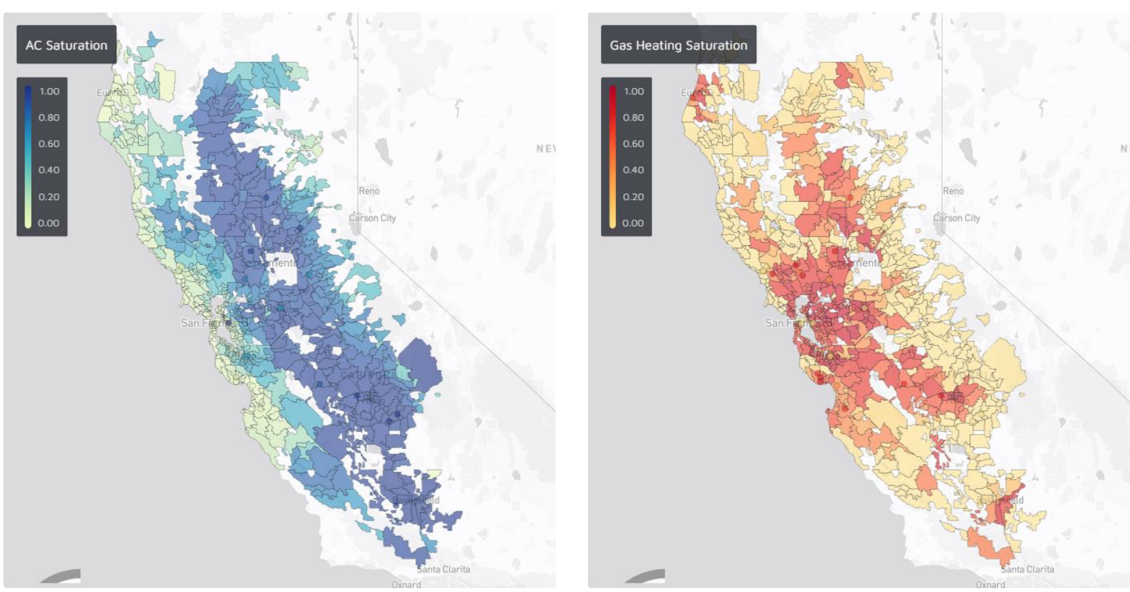

Figure 10. Heat maps of air conditioning (left) and gas heating (right) saturations, by ZIP code, across the analyzed territory.

Dramatically different patterns are evident. As expected based on local climate, AC saturation is low to nonexistent at the coast, nearly universal in the Central valley, and low in the

Sierra foothills in the eastern part of the state. Gas heating has a markedly different distribution, with a high concentration in the densely populated regions around the Bay Area, Eureka, Fresno, and Bakersfield, with sharply lower penetration in rural areas. Examination of the two maps in tandem can allow identification of areas with high saturations of both gas heating and air conditioning, which may be particularly promising for community-scale electrification.

Figures 11, 12, and 13 illustrate the dramatic variation in household characteristics and end-use saturations that can occur on the ZIP-code level. Each figure shows summary statistics[11] for a single ZIP code, with numbers of particular interest highlighted in red. The first example (Figure 11) in Rocklin, a suburb of Sacramento at the base of the Sierra foothills, has very high penetration of rooftop PV, at 21%, and nearly universal penetration of air conditioning and gas service. Only 4% of customers in this ZIP code have gas heating without significant space cooling load, suggesting that electrifying space heating in this area could have significant emissions and grid benefits by jointly eliminating gas consumption and reducing peak electricity demand via improved cooling efficiency.

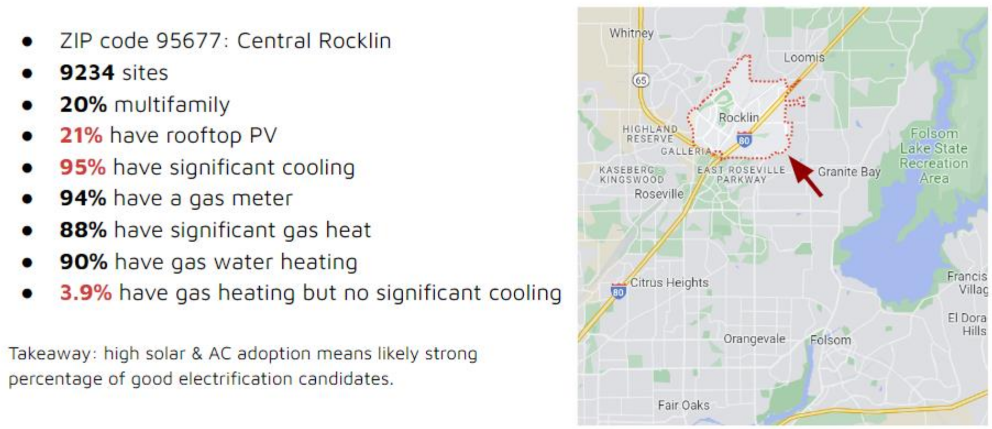

Figure 11. Estimated population characteristics for ZIP code 95677 in central Rocklin, northeast of Sacramento.

The second example (Figure 12) in coastal, semi-urban Berkeley, has a higher multifamily fraction and a sharply lower penetration of rooftop PV. Only 7% of customers have space cooling, while 90% have gas service and 64% use it significantly for heating. Electrifying space heating in this area would introduce a substantial number of new air conditioners into the building stock (albeit in a temperate climate); this will have implications for peak electricity load that are important to forecast and plan for.

The third example (Figure 13), a rural area northwest of Bakersfield in the San Joaquin Valley, has moderate penetration of rooftop PV, nearly universal cooling, and no gas service at all. Only a small fraction of customers have significant electric heating usage, suggesting widespread use of propane or wood for space heating. Electrifying space heating in this area would have significant positive impacts for emissions and the grid by potentially eliminating the use of high-emission heating fuels and reducing peak electricity loads. This example illustrates the value in taking a comprehensive approach to characterizing the population in an area for electrification targeting. A strategy focused on looking for customers with high gas heating usage, for example, would not have identified this area for attention.

[11] Multifamily fractions presented in the summaries were estimated by parsing customer address information to detect keywords and characters that indicate multifamily housing, such as *Apt.*, *Unit*, *#*, etc.

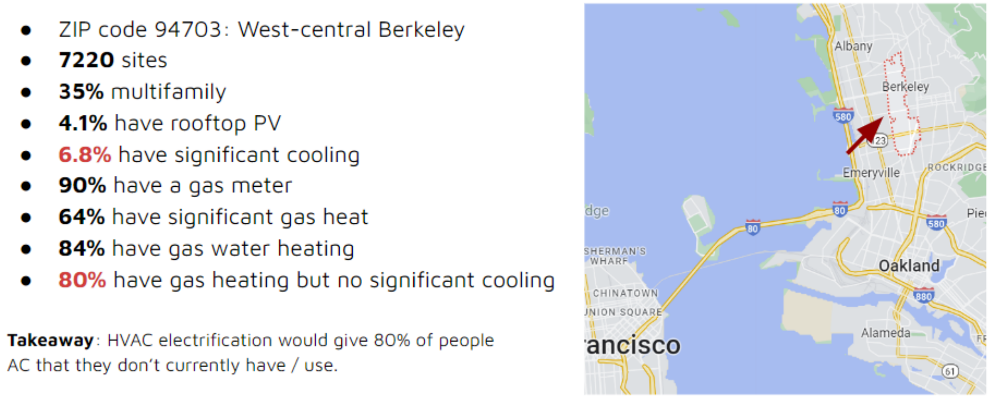

Figure 12. Estimated population characteristics for ZIP code 94703 in Berkeley.

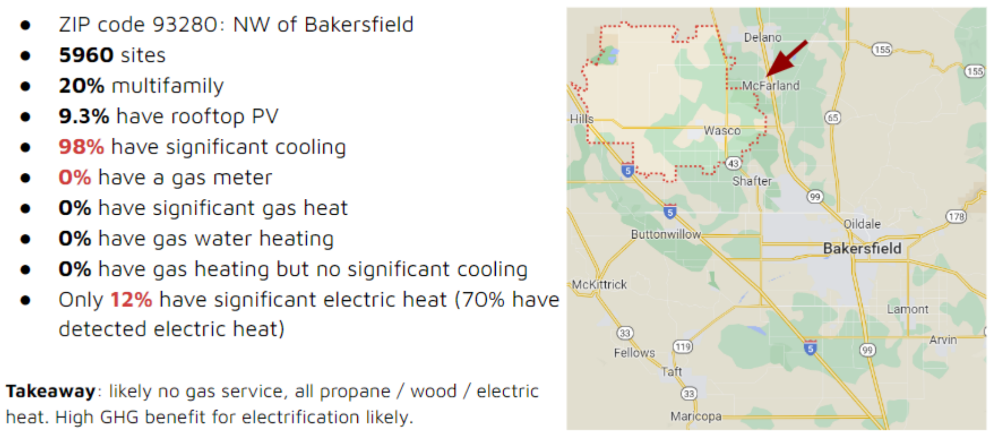

Figure 13. Estimated population characteristics for ZIP code 93280, a rural area northwest of Bakersfield.

These three examples are a small representation of the diversity of housing characteristics across the state. They illustrate the importance of having a detailed and fine-grained understanding of the population to support designing demand-side policies and programs that are most effective for the grid and the climate. Combining meter data with customer characteristics and with OpenEEmeter analytics can unlock these crucial insights for utilities and policymakers.

Conclusion

Weather-normalized analysis of customer meter data has had widespread applications in measuring the impacts of demand-side interventions and in targeting customers for program participation. In this study, we examined how such analysis, applied to the meter data for a whole population, can support planning and forecasting for policies and programs. Starting with a comprehensive dataset of electricity and gas meter data for households in PG&E service territory, collected by the CEC, we used the OpenEEmeter software, running on Recurve's large-scale analytics platform, to estimate disaggregated heating and cooling loads for each customer. We then defined thresholds in heating and cooling consumption that could be used to infer the use of gas or electricity as a significant source of heating or cooling, and we computed saturations of air conditioning, electric and gas space heating, and gas water heating at varying levels of geographic granularity. The results provide a detailed baseline understanding of the present-day population pre-electrification, with dramatic variation evident in end-use penetration and customer characteristics, both regionally by climate zone and sub-regionally by ZIP code. The analysis complements and improves upon traditional survey-based approaches by addressing the entire population of customers, rather than being limited to a statistical sample, and by being based directly on metered energy consumption, instead of self-reported information.

Granular understanding of energy consumption patterns across the population will be important in preparing for the impacts of widespread electrification. Analyses of the type we carried out here can provide important insights for utilities, grid planners, program designers, and regulators. The CEC has incorporated the results of this study in their modeling assumptions to obtain more accurate cost and energy impacts for their building electrification forecasting work. Additional applications include, among others: developing baselines for forecasting and tracking system-level demand or population-level emissions, understanding and predicting sources of strain on distribution infrastructure, designing programs to more effectively relieve grid stresses, identifying promising targets for community-level demand-side interventions, planning for appropriate siting and sizing of infrastructure upgrades, and designing outreach to ensure equitable program outcomes.

Essential ingredients for unlocking such insights are access to comprehensive, population-wide meter datasets, coupled with computing infrastructure that can perform analytics on the appropriate scale, which may range up to tens or hundreds of terabytes of data. Modern, parallel computing platforms, such as Recurve's, are up to the task and can enable a wide range of novel insights into energy consumption patterns, measured directly from customers' energy usage data. Such analyses can also be repeated on a regular cadence, more frequently than survey timelines allow, which will allow the utility industry to track progress and to respond nimbly to the and drastic changes that will result from rapid electrification.

Acknowledgments

We thank Ingrid Neumann, Joe Glass, Brett Kerrigan, and Adam Scheer for helpful feedback on this paper. This work was supported by the California Energy Commission (Agreement # 800-20-001). This document was prepared as a result of work sponsored by the California Energy Commission. It does not necessarily represent the views of the Energy Commission, its employees, or the State of California. The Energy Commission, the State of California, its employees, contractors, and subcontractors make no warranty, express or implied, and assume no legal liability for the information in this document; nor does any party represent that the use of this information will not infringe upon privately owned rights. This report has not been approved or disapproved by the Energy Commission nor has the Energy Commission passed upon the accuracy of the information in this report.

References

EIA. 2022. "2020 Residential Energy Consumption Survey: Household Characteristics Technical Documentation Summary." Washington, DC: EIA. https://www.eia.gov/consumption/residential/data/2020/pdf/2020%20RECS_Methodology%20Report.pdf

Gerke, Brian F., Sarah J. Smith, Samanvitha Murthy, Aditya Khandekar, Cong Zhang, Sunhee

Baik, Shreya Agarwal, Jingjing Liu, Richard E. Brown, Mary Ann Piette, and Peter Alstone. 2022. "Overview of Phase 4 of the California Demand Response Potential Study." In *Proceedings of the 2022 ACEEE Summer Study*. Pacific Grove, CA: ACEEE.

Kenney, Michael, Nicholas Janusch, Ingrid Neumann, and Mike Jaske. 2021. "California Building Decarbonization Assessment." Sacramento, CA: CEC. CEC Publication Number CEC-400-2021-006-CMF. https://efiling.energy.ca.gov/GetDocument.aspx?tn=239311

Mai, Trieu, Paige Jadun, Jeffrey Logan, Colin McMillan, Matteo Muratori, Daniel Steinberg, Laura Vimmerstedt, Ryan Jones, Benjamin Haley, and Brent Nelson. 2018. "Electrification Futures Study: Scenarios of Electric Technology Adoption and Power Consumption for the United States." Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A20-71500. https://www.nrel.gov/docs/fy18osti/71500.pdf.

NEEA. 2019. "Residential Building Stock Assessment II." Portland, OR: NEEA. https://neea.org/img/uploads/Residential-Building-Stock-Assessment-II-Single-Family-Homes-Report-2016-2017.pdf

Palmgren, Claire, Miriam Goldberg, Bob Ramirez, and Craig Williamson, "2019 California Residential Appliance Saturation Study (RASS)." Sacramento, CA: CEC. https://www.energy.ca.gov/publications/2021/2019-california-residential-appliance-saturation-study-rass

Recurve. 2022. "Utilizing Smart Meter Data to Improve Program Cost-Effectiveness and Customer Outcomes Executive Summary." Chicago: ComEd. https://innovate.comed.com/wp-content/uploads/2023/05/ComEd-Emerging-Tech-Recurve-NMEC-Customer-Targeting.pdf

Sathe, Amul, Karen Maoz, John Aquino, Abhijeet Pande, and Floyd Keneipp. 2020. Fuel Substitution Reporting Tools. California Energy Commission. Publication Number: CEC-200-2020-001. https://efiling.energy.ca.gov/GetDocument.aspx?tn=233241&DocumentContentId=65725