California Energy Commission

CONSULTANT REPORT

2019 California Residential Appliance Saturation Study (RASS)

Volume I: Methodology

Prepared for**: California Energy Commission**
Prepared by: **DNV GL Energy Insights USA, Inc.**

July 2021 | CEC-200-2021-005-MTHLGY

California Energy Commission

Primary Authors:

Claire Palmgren, Project Manager
Bob Ramirez, Project Sponsor
Miriam Goldberg, Ph.D.
Craig Williamson

DNV GL Energy Insights USA, Inc.

155 Grand Avenue, Suite 600

Oakland, CA 94612

www.dnvgl.com/energy

Contract Number: 800-16-004

Prepared for:

California Energy Commission

Glen Sharp
Le-Huy Nguyen
Contract Managers

Le-Huy Nguyen
Project Manager

Amanda Poletti
Office Manager

Data Integration and Policy Office

Siva Gunda
Deputy Director

Energy Assessments Division

Drew Bohan
Executive Director

DISCLAIMER

This report was prepared as the result of work sponsored by the California Energy Commission. It does not necessarily represent the views of the CEC, its employees, or the State of California. The CEC, the State of California, its employees, contractors, and subcontractors make no warrant, express or implied, and assume no legal liability for the information in this report; nor does any party represent that the uses of this information will not infringe upon privately owned rights. This report has not been approved or disapproved by the California Energy Commission, nor has the California Energy Commission passed upon the accuracy or adequacy of the information in this report.

ACKNOWLEDGEMENTS

Los Angeles Department of Water and Power (LADWP)

- Joseph Lee, Rockeish McKenzie

Pacific Gas and Electric (PG&E)

- Brian Smith

Sacramento Municipal Utility District (SMUD)

- Thomas Adkins, Shannon McCann

San Diego Gas & Electric (SDG&E)

- Tim Vonder

Southern California Edison (SCE)

- Reginald Wilkins

Southern California Gas Company (SoCalGas)

- Christine Otis

APPRISE, Inc.

- David Carroll, Dan Bausch

Business Economic Analysis and Research (BEAR)

- Ken Parris

GC Green, Inc.

- Rick Halperin, Elizabeth Perez

Other DNV GL Energy Insights staff

- Jarred Metoyer, Paula Ham-Su, Robert Langton, Mikisha Doop, Tim Hennessy, Amber Watkins, Micaela Brown, Soofia Gilbert, Noel Stevens

ABSTRACT

From 2018 to 2020, the California Energy Commission funded and administered a Residential Appliance Saturation Study that serves as an update to the *2009 Residential Appliance Saturation Study*, with the following utilities participating: Pacific Gas and Electric Company, Southern California Edison, Sacramento Municipal Utility District, San Diego Gas & Electric Company, Southern California Gas Company, and Los Angeles Department of Water and Power. DNV GL Energy Insights (formerly KEMA, Inc.) was the prime consultant for this study as well as the 2009 and 2003 studies.

The research team implemented the study as online and mailed paper surveys. The surveys requested households to provide information on appliances, equipment, and general consumption patterns including electric vehicle charging and the presence of renewable energy technology such as a solar photovoltaic system. The research team completed data collection in early 2020 just before the full impact of COVID-19 events.

The study yielded energy consumption estimates for 28 electric and 9 natural gas residential end uses and appliance saturations for households. The team developed these consumption estimates using a conditional demand analysis, an approach that applied statistical methods to combine survey responses, household energy consumption data, and weather information to calculate average annual consumption estimates per appliance. The *2019 Residential Appliance Saturation Study* resulted in end-use saturations for 39,682 individually metered and 303 master-metered households. The team weighted survey and conditional demand analysis results to provide population-level estimates, representative of the participating utilities that allow comparison across utility service territories, forecasting climate zones, and other variables of interest including dwelling type, dwelling age group, and income.

Keywords: California Energy Commission, conditional demand analysis, CDA, unit energy consumption, UEC, residential, energy survey, online survey, appliance, saturations, degree-day normalization, AMI data, hourly load shapes, electric vehicles, EVs

Please use the following citation for this report:

DNV GL Energy Insights USA, Inc. 2020. *2019 California Residential Appliance Saturation Study*. California Energy Commission. Publication Number: CEC-200-2020-XXX.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: RASS Methodology Introduction

From 2018 to 2020, the California Energy Commission (CEC) funded and administered a Residential Appliance Saturation Study (RASS) that was implemented across the territories of the large investor-owned utilities (IOUs) and two of the largest municipal utilities. The 2019 study updated the *2009 RASS*. Participating utilities included Pacific Gas and Electric Company (PG&E), Southern California Edison Company (SCE), San Diego Gas & Electric Company (SDG&E), Southern California Gas Company (SoCalGas), Sacramento Municipal Utility District (SMUD), and Los Angeles Department of Water and Power (LADWP). DNV GL was the prime consultant.

The research team initiated the study the end of 2018 with the sampling plans and implementation beginning in the fall of 2019. The team collected data using online and direct mail approaches to a representative sample of California households. The survey asked households to provide information on appliances, equipment, and general usage patterns. The *2009 RASS* survey questionnaire was updated to reflect changes in available energy-consuming and generation-storage technologies in households including new questions about electric vehicles (EV), photovoltaic (PV) systems and battery storage systems. The team implemented a smaller and more focused nonresponse follow-up recruitment effort to a sample of nonrespondents after the initial contact methods were exhausted. The nonresponse effort consisted of telephone calls and in-person assistance with completing the survey. The team completed data collection in early 2020 just before the full impact of COVID-19 events.

The study yielded unit energy consumption (UEC) estimates for 28 electric and 9 natural gas residential end uses and appliance saturations for households. A UEC represents the amount of energy an appliance is estimated to use in a year. The team developed the UEC estimates using a conditional demand analysis (CDA), an approach that applied statistical methods to combine survey data, household energy consumption data, and weather information to calculate average annual consumption estimates per appliance.

The *2019 RASS* resulted in end-use saturations for 39,682 individually metered and 303 master-metered households. UEC estimates were provided for individually metered households only, while end-use saturations reflected individually and master-metered households. Survey and CDA results were weighted to provide population level estimates representative of the participating utilities that allow comparison across utility service territories, forecasting climate zones, and other variables of interest such as dwelling type, dwelling age group, and income.

By using a statewide survey instrument, the research team provided the CEC and other parties with a consistent set of questions and study results to use for statewide planning and cross-utility comparisons. The project required a joint effort among the study partners, as they collaborated on a research plan, program materials, and implementation strategy. Each utility provided the data necessary to create a unified sampling plan, as well as household-specific information for households that were selected for the sample. The research team provided anonymity to survey participants by assigning a generic identification code that represented the sampling stratification variables. Each participating utility was provided a key to the identification code that allowed the utilities to link survey respondents to a specific account.

Because the study was designed to support interests of a variety of users, the final report included a collection of research products:

- **The Executive Summary presents a summary of key findings.**
- **The Project Overview presents a detailed summary of key findings.**
- **Volume One** describes the study design and implementation methods, along with a detailed description of the data cleaning process, CDA method, and development of hourly electric and daily gas loads.
- **Volume Two** provides a brief description of the CDA along with tabulated results for end-use UECs and saturations and an overview of hourly electric and daily gas loads.
- All referenced **appendices** have been compiled into one document for convenience.
- The **RASS Website**, an updated version of the 2009 web tool, supports customized queries of the survey data including the ability to compare 2019 results to 2009 and 2003 results.

Volume One provides a comprehensive overview of the study design, implementation, and the methodological details of data preparation and analysis.

CHAPTER 2: Study Design and Implementation

Sampling Approach

Individually Metered Sample Design

The total population for the study consists of households in PG&E, SCE, SDG&E, SMUD, and LADWP electric service territories. The sample design was developed using data provided by the utilities.

The 2019 sample design and weighting processes were developed to maximize precision on key estimates of interest, ensure comparability in the estimates with those produced in the *2003 RASS* and *2009 RASS*, and support the survey strategy necessary to achieve the highest response rates possible. In addition to stratifying on the CEC's main variables of interest, DNV GL's design stratified on the presence or absence of email address in the utility records. This allowed the study team to conduct targeted outreach to households, minimizing data collection costs and increasing the number of responses. The sample design also included implicit stratification variables to allow for broader applicability of the results.

Individually Metered Sample Stratification

Explicit Stratification

DNV GL used a stratified random sample design. The total population of the five participating electric utilities was split into 127 sampling cells based on six explicit stratification variables. Target sample sizes are set for cells defined by combinations of these variables.

The RASS research plan suggested explicit sample stratification by the 2016 CEC forecasting climate zones (FCZ), electric consumption level, and presence of electric heat. The CEC forecasting climate zones are geographic areas defined by the CEC to assist energy forecasting and planning and sometimes are also called electricity demand forecast zones. The forecasting climate zones are nested within the electric utilities, so that stratification by FCZ also provides utility-level stratification. Electric utility is still considered a stratification variable, since estimates at the utility level are of key interest.

DNV GL's review of the data revealed that the "electric heat" was based on a grandfathered rate class that denotes all-electric residences (homes with no gas service) and was not a reliable indicator of current electric heating.

The presence of on-site solar electricity at a home means that the metered electric consumption is net, so that the actual total usage in the home is not known. Thus, the electric usage categories used for stratification are less meaningful for solar homes. Likewise, for virtual-net-metered homes, the tracked electric consumption is typically not the home's actual total consumption. Virtual net metering allows a multi-tenant building to allocate the direct benefits of onsite generation to the tenants, rather than all of the benefits going to the building owner. Presence of solar was identified from utility records. Virtual net metering was identifiable for PG&E customers but not for others. DNV GL defined a single stratification variable as "net metered" that included households identified in the utility record as having either onsite solar with net metering or virtual net metering.

DNV GL noted that the number of net-energy metering (NEM) customers identified in the sampling frame was roughly 75 percent of those that the participating utilities reported to the Energy Information Administration (EIA) for July 2018.[1]

To manage field costs, DNV GL budgeted to recruit homes with email addresses on record with the participating utilities at a somewhat higher proportion than their proportion in the population. To accomplish this differential sampling rate, it was necessary also to stratify by email status (email address available or not).

Based on these considerations, DNV GL used the following as explicit stratification variables:

- Electric Utility: PG&E, SCE, SDG&E, SMUD, or LADWP
- CEC forecasting climate zone: 15 zones, nested in the utilities
- Email address available: yes or no
- Net Metered: yes or no
- For Non-Net Metered,
  - Dwelling Unit type: single-family (SF) or multifamily (MF)
  - Usage level: The number of usage levels varied from 1 to 3 according to number of accounts in the population.
    - Groupings with fewer than 100,000 accounts apart from usage level stratification were not stratified by usage level.
    - Groupings with 100,000 to 300,000 accounts apart from usage level stratification were stratified into two usage levels (low and high), with the cut point at the 75th percentile of daily kWh.
    - Groupings with over 300,000 accounts apart from usage level stratification were stratified into 3 usage levels with the cut points at the 25th and 75th percentiles of daily kWh.

1 The counts of customers with solar reported to the EIA and found in the RASS sampling frame are shown below.

Participating Utilities	EIA Form 861 July, 2018	RASS Sampling Frame
Los Angeles Department of Water & Power	35,085	17,794
Pacific Gas and Electric Co.	378,803	291,779
Southern California Edison Co	272,902	244,559
San Diego Gas & Electric Co	134,822	64,504
Sacramento Municipal Utility District	22,584	11,118
TOTAL	844,196	629,754

A key stratification variable used in prior RASS was the CEC forecasting climate zone (FCZ). The FCZs were revised in 2016, and geographic boundaries were adjusted so the individual zones are not directly comparable to forecasting climate zones used in the prior RASS studies. As directed by the CEC, the 2016 CEC forecasting climate zones were used in the current study as an explicit stratification variable.

Implicit Stratification

Several additional variables were used for implicit stratification. The sample design did not set targets for combinations of implicit stratification variables with other explicit or implicit stratification variables. The sample was selected in such a way as to distribute the selections very close to proportionately across the implicit sampling dimensions and achieve approximate overall targets by these dimensions.

Implicit stratification variables included:

- **Title 24 (T24) building climate zone group:** DNV GL classified T24 building climate zones into three groups based on the related cooling and heating degree days.
- **Likely use of air conditioning (AC):** The research team determined the likely use of AC (high, medium, low) based on the ratio of summer to shoulder-month electric consumption. DNV GL believes that the extent of air conditioning use was a more important distinguishing characteristic than electric heating, which is less common and may be slight even when present.
- **California Alternative Rates for Energy (CARE) or Family Electric Rate Assistance Program (FERA) participant:** yes/no. This is an indicator of income level that is of direct interest.
- **Neighborhood low-income:** The team used American Community Survey (ACS) data to segment Census blocks into low, medium, or high proportions of low-income households.
- **Percentage of homes built before 1980:** ACS data were used to segment Census blocks by percentage of homes built before 1980 or after 1980. The use of the ACS data supported sampling a proportional mix of older and newer homes.
- **Percentage of owner-occupied dwellings:** ACS data were used to segment Census blocks by percentage of owner-occupied dwellings. The use of the ACS data helped sample a proportional mix of renters and homeowners.

Within each sampling cell, the sample was distributed roughly proportionately with respect to each of these implicit stratification variables.

Individually Metered Sample Allocation Process

The general design goal was to obtain accurate estimates by combinations of forecasting climate zone, dwelling type, and net metering status. Stratification by email status was for sampling efficiency, not to support separate estimates by this dimension.

As noted, for budget reasons DNV GL designed the sample to achieve a higher-than-proportional number of completed surveys from homes with email addresses available from utility records. The team implemented this by selecting homes with email addresses at higher rates than homes not having email addresses.

DNV GL found 41 percent of individually metered accounts had an email address on record, while 59 percent did not. Experience with online surveys in the last two years suggested that the initial planned response rates would be difficult to attain. Accordingly, DNV GL doubled the initial email send-out at minimal added cost. Because the United States Postal Service (USPS) mailed follow-ups added incremental costs for each follow-up household, DNV GL conducted mail follow-up with only half of the nonrespondents to the email push-to-web requests. **Table 1** indicates the planned recruitment and response rates.

Table 1: Response Assumptions for Sample Allocation

Recruitment Process Step	Email	Non-Email	Total
Initial send-out	352,953	119,393	472,346
Response rate without mailing	7%		
Response without mailing	24,707		
Mail follow-up to email	176,477		
Incremental response rate from mail follow-up	7%		
Incremental mail response	12,357		
Total completes	**37,064**	**14,458**	**51,522**
Completes/send-out	11%	12%	11%
Completes as % of total completes	72%	28%	100%

Source: 2019 California Residential Appliance Saturation Survey

With this approach, the anticipated ratio of completed surveys to recruitment would be lower for accounts with email addresses than for accounts without because the USPS mail follow-up would be conducted for only half of the nonrespondents with email addresses.

To determine the recruitment by sampling cell, DNV GL considered the differential response rates by strata from the *2009 RASS*. SMUD did not participate in the *2009 RASS*, so its response rates were mapped from corresponding LADWP strata. DNV GL rescaled these cell-wise response rates to the overall estimated completion rates for the email and non-email recruitments. DNV GL then applied these scaled completion rate assumptions to the 2019 sampling cells. **Table 2** presents the completion rate assumptions.

Table 2: 2019 Completion Rate Assumptions Based on *2009 RASS*

Utility	Dwelling Unit Type	Usage	Label	2009 Response Rate	2019 Completion Rate Assumption Email	2019 Completion Rate Assumption Non-email
LADWP	**ALL**	**ALL**	**LADWP_ALL_ALL**	**14.9%**	**8.2%**	**9.7%**
LADWP	ALL	H	LADWP_ALL_H	18.0%	9.9%	11.7%
LADWP	ALL	M	LADWP_ALL_M	12.3%	6.8%	8.0%
LADWP	ALL	L	LADWP_ALL_L	15.8%	8.7%	10.3%
PG&E	**ALL**	**ALL**	**PG&E_ALL_ALL**	**20.1%**	**11.0%**	**13.0%**
PG&E	MF	ALL	PG&E_MF_ALL	13.6%	7.5%	8.9%
PG&E	SF	H	PG&E_SF_H	22.1%	12.2%	14.4%
PG&E	SF	L	PG&E_SF_L	23.6%	13.0%	15.4%
SCE	**ALL**	**ALL**	**SCE_ALL_ALL**	**18.4%**	**10.1%**	**12.0%**
SCE	MF	ALL	SCE_MF_ALL	15.3%	8.4%	10.0%
SCE	SF	H	SCE_SF_H	19.8%	10.9%	12.9%
SCE	SF	L	SCE_SF_L	19.7%	10.8%	12.8%
SDG&E	**ALL**	**ALL**	**SDG&E_ALL_ALL**	**18.8%**	**10.3%**	**12.2%**
SDG&E	ALL	H	SDG&E_ALL_H	18.7%	10.3%	12.2%
SDG&E	ALL	M	SDG&E_ALL_M	20.4%	11.2%	13.3%
SDG&E	ALL	L	SDG&E_ALL_L	17.1%	9.4%	11.1%
SMUD	**ALL**	**ALL**			**8.2%**	**9.7%**
SMUD	ALL	H			9.9%	11.7%
SMUD	ALL	L			8.7%	10.3%

Source: 2019 California Residential Appliance Saturation Survey

Individually Metered Sample Frame and Recruitment

Table 3 illustrates the sample design for individually metered homes that DNV GL developed for the statewide RASS study.

The contents of **Table 3** are as follows:

- Columns A through G indicate the strata. For Column D (forecasting climate zone), a value of ”All FCZ” indicates records that were grouped together across forecasting climate zones. This step was not needed for other service territories where these dwellings were all in the same forecasting climate zone.
- Column H shows the count of dwelling units in the sampling cell.
- Column I displays this count as a percent of all units in the study population.
- Column J shows the target number of completes.

Column K contains the initial recruitment. This was determined based on assumed response rates along with the target number of completes. The initial recruitment is the initial number of emailed push-to-web invitations for those with email addresses in utility records, and the initial number of mailed push-to web invitation letters sent via USPS to those without email addresses.

Table 3: Sample Design, Individually Metered Homes — by Explicit Stratification Variables

A Utility	B Stratum No.	C Email	D Forecasting Climate Zone	E Net Metered	F Dwelling Type	G Usage	H Population	I Percent of Population	J Target Completed	K Recruitment
LADWP	001	No	16	No	MF	High	96,603	0.79%	155	1,323
LADWP	002	No	16	No	MF	Low	298,038	2.45%	464	4,518
LADWP	003	No	16	No	SF	High	42,971	0.35%	84	717
LADWP	004	No	16	No	SF	Low	130,613	1.07%	251	2,446
LADWP	006	No	16	Yes	SF	NA	6,903	0.06%	10	106
LADWP	007	No	17	No	MF	NA	134,674	1.11%	208	2,024
LADWP	008	No	17	No	SF	NA	111,655	0.92%	214	2,215
LADWP	009	No	17	Yes	SF	NA	7,851	0.06%	11	111
LADWP	010	Yes	16	No	MF	High	59,021	0.48%	306	3,104
LADWP	011	Yes	16	No	MF	Low	184,048	1.51%	920	10,597
LADWP	012	Yes	16	No	SF	NA	99,011	0.81%	531	6,494
LADWP	014	Yes	16	Yes	SF	NA	6,298	0.05%	26	312
LADWP	015	Yes	17	No	MF	NA	96,284	0.79%	460	5,302
LADWP	016	Yes	17	No	SF	NA	78,033	0.64%	405	4,953
LADWP	017	Yes	17	Yes	SF	NA	8,892	0.07%	33	398
LADWP	201	Yes	16	Yes	MF	NA	1,850	0.02%	8	33
PG&E	018	No	1	No	MF	High	77,150	0.63%	192	2,163
PG&E	019	No	1	No	MF	Low	80,601	0.66%	191	1,329
PG&E	020	No	1	No	MF	Middle	155,311	1.28%	384	2,579
PG&E	021	No	1	No	SF	High	127,080	1.04%	318	2,206
PG&E	022	No	1	No	SF	Low	129,956	1.07%	317	2,067
PG&E	023	No	1	No	SF	Middle	254,787	2.09%	636	4,279
PG&E	025	No	1	Yes	SF	NA	34,403	0.28%	74	497
PG&E	026	No	2	No	MF	NA	42,472	0.35%	104	1,175
PG&E	027	No	2	No	SF	High	42,529	0.35%	104	723
PG&E	028	No	2	No	SF	Low	128,705	1.06%	312	2,035
PG&E	030	No	2	Yes	SF	NA	9,234	0.08%	20	133
PG&E	031	No	3	No	MF	NA	8,852	0.07%	21	242

A Utility	B Stratum No.	C Email	D Forecasting Climate Zone	E Net Metered	F Dwelling Type	G Usage	H Population	I Percent of Population	J Target Completed	K Recruitment
PG&E	032	No	3	No	SF	NA	68,875	0.57%	165	1,112
PG&E	033	No	3	Yes	SF	NA	5,527	0.05%	12	80
PG&E	034	No	4	No	MF	NA	56,447	0.46%	137	1,542
PG&E	035	No	4	No	SF	High	75,137	0.62%	187	1,298
PG&E	036	No	4	No	SF	Low	230,423	1.89%	561	3,651
PG&E	038	No	4	Yes	SF	NA	30,591	0.25%	66	446
PG&E	039	No	5	No	MF	NA	53,460	0.44%	127	1,433
PG&E	040	No	5	No	SF	High	57,319	0.47%	142	985
PG&E	041	No	5	No	SF	Low	175,649	1.44%	426	2,771
PG&E	043	No	5	Yes	SF	NA	25,733	0.21%	55	367
PG&E	044	No	6	No	MF	NA	39,501	0.32%	97	1,098
PG&E	045	No	6	No	SF	High	34,717	0.29%	87	601
PG&E	046	No	6	No	SF	Low	105,822	0.87%	260	1,692
PG&E	048	No	6	Yes	SF	NA	11,854	0.10%	22	148
PG&E	049	Yes	1	No	MF	High	135,892	1.12%	887	11,856
PG&E	050	Yes	1	No	MF	Low	144,283	1.18%	887	7,294
PG&E	051	Yes	1	No	MF	Middle	275,899	2.27%	1,775	14,131
PG&E	052	Yes	1	No	SF	High	156,225	1.28%	1,036	8,519
PG&E	053	Yes	1	No	SF	Low	163,239	1.34%	1,034	7,973
PG&E	054	Yes	1	No	SF	Middle	314,251	2.58%	2,073	16,501
PG&E	056	Yes	1	Yes	SF	NA	74,657	0.61%	428	3,410
PG&E	057	Yes	2	No	MF	NA	52,990	0.44%	340	4,540
PG&E	058	Yes	2	No	SF	High	42,668	0.35%	272	2,241
PG&E	059	Yes	2	No	SF	Low	128,712	1.06%	818	6,304
PG&E	061	Yes	2	Yes	SF	NA	16,114	0.13%	91	722
PG&E	062	Yes	3	No	MF	NA	15,015	0.12%	95	1,269
PG&E	063	Yes	3	No	SF	NA	65,952	0.54%	416	3,311
PG&E	064	Yes	3	Yes	SF	NA	9,073	0.07%	53	420
PG&E	065	Yes	4	No	MF	NA	82,004	0.67%	516	6,900
PG&E	066	Yes	4	No	SF	High	78,777	0.65%	518	4,265

A Utility	B Stratum No.	C Email	D Forecasting Climate Zone	E Net Metered	F Dwelling Type	G Usage	H Population	I Percent of Population	J Target Completed	K Recruitment
PG&E	067	Yes	4	No	SF	Low	85,010	0.70%	518	3,996
PG&E	068	Yes	4	No	SF	Middle	158,097	1.30%	1,039	8,267
PG&E	070	Yes	4	Yes	SF	NA	55,303	0.45%	319	2,537
PG&E	071	Yes	5	No	MF	NA	73,306	0.60%	450	6,011
PG&E	072	Yes	5	No	SF	High	54,367	0.45%	355	2,921
PG&E	073	Yes	5	No	SF	Low	168,675	1.38%	1,066	8,218
PG&E	075	Yes	5	Yes	SF	NA	44,061	0.36%	247	1,966
PG&E	076	Yes	6	No	MF	NA	45,497	0.37%	293	3,924
PG&E	077	Yes	6	No	SF	High	33,193	0.27%	219	1,806
PG&E	078	Yes	6	No	SF	Low	102,549	0.84%	660	5,085
PG&E	080	Yes	6	Yes	SF	NA	16,252	0.13%	81	647
PG&E	202	Yes	All FCZ	Yes	MF	NA	13,171	0.11%	63	678
SCE	081	No	7	No	MF	High	136,896	1.12%	338	3,392
SCE	082	No	7	No	MF	Low	141,945	1.17%	338	3,384
SCE	083	No	7	No	MF	Middle	275,658	2.26%	678	6,795
SCE	084	No	7	No	SF	High	280,189	2.30%	698	5,425
SCE	085	No	7	No	SF	Low	291,533	2.39%	698	5,455
SCE	086	No	7	No	SF	Middle	562,856	4.62%	1,398	10,891
SCE	088	No	7	Yes	SF	NA	56,327	0.46%	137	1,141
SCE	089	No	8	No	MF	NA	52,986	0.44%	130	1,307
SCE	090	No	8	No	SF	High	41,379	0.34%	103	801
SCE	091	No	8	No	SF	Low	124,715	1.02%	309	2,418
SCE	093	No	8	Yes	SF	NA	9,773	0.08%	24	199
SCE	094	No	9	No	MF	NA	20,040	0.16%	47	474
SCE	095	No	9	No	SF	High	31,391	0.26%	77	599
SCE	096	No	9	No	SF	Low	98,431	0.81%	232	1,811
SCE	097	No	9	Yes	SF	NA	9,795	0.08%	24	198
SCE	098	No	10	No	MF	NA	80,741	0.66%	195	1,956
SCE	099	No	10	No	SF	High	77,738	0.64%	193	1,498
SCE	100	No	10	No	SF	Low	80,697	0.66%	193	1,507

A Utility	B Stratum No.	C Email	D Forecasting Climate Zone	E Net Metered	F Dwelling Type	G Usage	H Population	I Percent of Population	J Target Completed	K Recruitment
SCE	101	No	10	No	SF	Middle	155,642	1.28%	386	3,008
SCE	102	No	10	Yes	SF	NA	20,301	0.17%	50	421
SCE	103	No	11	No	MF	NA	48,537	0.40%	118	1,183
SCE	104	No	11	No	SF	High	68,724	0.56%	171	1,326
SCE	105	No	11	No	SF	Low	208,202	1.71%	513	4,007
SCE	107	No	11	Yes	SF	NA	29,723	0.24%	74	617
SCE	108	Yes	7	No	MF	High	59,668	0.49%	395	4,686
SCE	109	Yes	7	No	MF	Low	181,996	1.49%	1,186	14,074
SCE	110	Yes	7	No	SF	High	141,302	1.16%	941	8,652
SCE	111	Yes	7	No	SF	Low	145,530	1.19%	940	8,701
SCE	112	Yes	7	No	SF	Middle	284,545	2.34%	1,885	17,387
SCE	114	Yes	7	Yes	SF	NA	54,093	0.44%	350	3,460
SCE	115	Yes	8	No	MF	NA	19,985	0.16%	131	1,560
SCE	116	Yes	8	No	SF	NA	75,385	0.62%	501	4,957
SCE	117	Yes	8	Yes	SF	NA	8,207	0.07%	54	534
SCE	118	Yes	9	No	MF	NA	6,383	0.05%	40	475
SCE	119	Yes	9	No	SF	NA	55,551	0.46%	359	3,546
SCE	120	Yes	9	Yes	SF	NA	8,486	0.07%	55	542
SCE	121	Yes	10	No	MF	NA	33,503	0.28%	218	2,584
SCE	122	Yes	10	No	SF	High	42,732	0.35%	284	2,616
SCE	123	Yes	10	No	SF	Low	129,457	1.06%	853	7,896
SCE	124	Yes	10	Yes	SF	NA	20,658	0.17%	137	1,354
SCE	125	Yes	11	No	MF	NA	19,958	0.16%	130	1,538
SCE	126	Yes	11	No	SF	High	39,845	0.33%	265	2,441
SCE	127	Yes	11	No	SF	Low	120,672	0.99%	797	7,377
SCE	129	Yes	11	Yes	SF	NA	30,979	0.25%	206	2,032
SCE	203	Yes	All FCZ	Yes	MF	NA	2,033	0.02%	20	88
SDGE	130	No	12	No	MF	NA	70,323	0.58%	170	1,531
SDGE	131	No	12	No	SF	High	38,474	0.32%	94	773
SDGE	132	No	12	No	SF	Low	115,265	0.95%	283	2,551

A Utility	B Stratum No.	C Email	D Forecasting Climate Zone	E Net Metered	F Dwelling Type	G Usage	H Population	I Percent of Population	J Target Completed	K Recruitment
SDGE	134	No	12	Yes	SF	NA	10,726	0.09%	16	129
SDGE	135	Yes	12	No	MF	High	86,284	0.71%	548	5,325
SDGE	136	Yes	12	No	MF	Low	92,340	0.76%	548	5,852
SDGE	137	Yes	12	No	MF	Middle	174,854	1.44%	1,101	9,820
SDGE	138	Yes	12	No	SF	High	136,037	1.12%	855	8,310
SDGE	139	Yes	12	No	SF	Low	135,974	1.12%	854	9,120
SDGE	140	Yes	12	No	SF	Middle	267,135	2.19%	1,712	15,274
SDGE	142	Yes	12	Yes	SF	NA	105,026	0.86%	365	3,538
SDGE	204	Yes	12	Yes	MF	NA	6,732	0.06%	24	223
SMUD	143	No	13	No	MF	NA	45,491	0.37%	105	1,026
SMUD	144	No	13	No	SF	High	35,216	0.29%	86	732
SMUD	145	No	13	No	SF	Low	105,267	0.86%	257	2,500
SMUD	146	No	13	Yes	SF	NA	3,196	0.03%	6	57
SMUD	147	Yes	13	No	MF	NA	87,582	0.72%	466	5,371
SMUD	148	Yes	13	No	SF	High	57,574	0.47%	361	3,660
SMUD	149	Yes	13	No	SF	Low	176,923	1.45%	1,085	12,495
SMUD	150	Yes	13	Yes	MF	NA	157	0.00%	4	8
SMUD	151	Yes	13	Yes	SF	NA	13,558	0.11%	59	721
TOTAL							**12,179,433**		**51,522**	**472,346**

Source: 2019 California Residential Appliance Saturation Survey

Master-Metered Sample Design

Master-metered residential account refers to service being supplied through one meter to multiple dwelling units that do not have meters from the utility to the individual units. Master-metered rates have been closed to new installations and additions to existing meters since the mid-1980s and constitute a small portion of the annual energy use in California. For additional details on these customers, please view the individual utilities' definitions in the respective tariff sheets.[2]

The research team constructed the master-metered sample design by using the electric utility and the type of units the account served based on information provided by the utilities. Types of units serviced included master-metered accounts serving two to four units, multifamily complexes with 5 units to 20 units, multifamily complexes with more than 20 units, and mobile-home parks with 5 or more mobile homes. The team assigned target completed surveys for the sample based on the proportion of the population of units within that type (not accounts). The team divided the study population of master-metered accounts into 19 strata based on electric utility and type of units.

The team surveyed the master-metered accounts differently based on what type of units they served. Accounts serving two to four units were surveyed the same way the individually metered households were in that one survey was mailed to the account contact.

Master-metered accounts serving more than four units were surveyed using a phone interview with the account contact to obtain information about central systems. The team determined the number of phone calls per stratum by the number of target-completed surveys.

Mobile home accounts serving two to four units were classified and surveyed using the same process as other master-metered accounts serving two to four units. Mobile home accounts serving five or more units were assigned to their own stratum within each utility.

Master-Metered Sample Allocation Process

The sample of mastered-metered accounts serving two to four units was randomly pulled proportionately to the number of units served in each service territory. The team randomly selected the sample of the other types of master-metered accounts for the phone call survey of the account contact.

Table 4 presents the master-metered sample frame.

---

2 Tariff sheets can be found at
https://www.pge.com/tariffs/assets/pdf/tariffbook/ELEC_SCHEDS_EM%20(Sch).pdf
https://www.sce.com/residential/rates/multi-family
https://www.sdge.com/total-electric-rates.

Table 4: Sample Design, Master-Metered

Utility	Home Type	Number of Meters	Number of Units	Pro-portion of Total Units	CEC Target Phone Surveys	CEC Target Unit Mail Surveys	Survey Mailouts (Units)
PG&E	2-4 units	16,550	37,092	7.9%	-	130	1,500
PG&E	Multifamily 5-20 units	2,158	9,918	4.3%	50		
PG&E	Multifamily >20 units	1,042	62,949	13.4%	86		
PG&E	Mobile home >4 units	,425	97,946	20.9%	60		
PG&E	**Sub-Total**	**21,175**	**217,905**	**46.5%**			
SDGE	2-4 units	3,352	7,504	1.6%	-	26	425
SDGE	Multifamily 5-20 units	662	6,926	1.5%	-		
SDGE	Multifamily >20 units	188	10,109	2.2%	40		
SDGE	Mobile home >4 units	59	38,622	8.2%	-		
SDGE	**Sub-Total**	**4,661**	**63,161**	**13.5%**			
SCE	2-4 units	4971	11,739	2.5%	-	41	500
SCE	Multifamily 5-20 units	1,484	14,239	3.0%	105		
SCE	Multifamily >20 units	599	39,788	8.5%	65		
SCE	Mobile home >4 units	1,295	97,172	20.7%	60		
SCE	**Sub-Total**	**8,349**	**162,938**	**34.8%**			
LADWP	2-4 units	2,692	5,734	1.2%	-	20	200
LADWP	Multifamily 5-20 units	342	4,127	0.9%	23		
LADWP	Multifamily >20 units	184	6,900	1.5%	25		
LADWP	Mobile home >4 units	22	3,075	0.7%	1		
LADWP	**Sub-Total**	**3,240**	**19,836**	**4.2%**			
SMUD	1-4 units	6	14	0.0%	-	-	-
SMUD	Multifamily 5-20 units	1	7	0.0%	-	-	-
SMUD	Multifamily >20 units	-	-	0.0%	-	-	-
SMUD	Mobile home >4 units	81	4,786	1.0%	3		
SMUD	**Sub-Total**	**88**	**4,807**	**1.0%**			
TOTAL	2-4 units	27,571	62,083	13.2%	-	218	2,625
TOTAL	Multifamily 5-20 units	4,647	45,217	9.6%	173		
TOTAL	Multifamily >20 units	2,013	119,746	25.6%	216		
TOTAL	Mobile home >4 units	3,282	241,601	51.6%	124		
TOTAL	**Total**	**37,513**	**468,647**	**100.0%**	**513**	**218**	**2,625**

Source: 2019 California Residential Appliance Saturation Survey

Nonresponse Follow-Up Sample Design

The nonresponse effort of the study sought to help reduce nonresponse bias by obtaining responses from a portion of households that had not responded after the email solicitations (for those with email addresses) and the first paper survey mailing. The research team selected a subset of 6,000 individually metered households from those who were sent the first survey packet and had not responded as of January 3, 2020. For this sampling, a household was considered a nonrespondent if neither an online survey submittal nor paper survey receipt had been recorded. In some cases, a paper survey might have been in transit and not yet recorded.

The team divided nonresponders into two groups: clustered and unclustered. The more densely populated areas of the state were clustered for follow-up sampling by ZIP code groups to allow more efficient in-person data collection:

- In 2009, there were eight ZIP-3 groups in the unclustered group: 934, 935, 939, 949, 954, 955, 960 and 961. These were left unclustered in 2019.
- In addition, a group of 782 low-density ZIP codes in 42 ZIP-3 groups were added to the unclustered group. This grouping was done at the ZIP-5 level. The entire ZIP-3 group was not necessarily added to the unclustered group. This was done to increase the population in the unclustered ZIP codes to 10 percent of the total, with the purpose to not have net metered households dominate the unclustered sample. This process allowed net-metered households to be proportionally represented in the unclustered sample.
- Households in the remaining 724 ZIP-5s (belonging to 55 ZIP-3s) were clustered by ZIP-5 or contiguous ZIP-5 groups.

As mentioned above, net-metered households, regardless of ZIP code, were included in the unclustered pool for follow-up sampling. The team took this step for two reasons: the first was to reserve the spots in the higher-cost clustered follow-up for households that could be used in CDA estimation. The second reason is that since net metering is more prevalent in some neighborhoods than others, leaving them in the cluster sample could result in particularly high or low representation of net metering, based on which clusters were randomly selected for follow-up.

Clusters were created within the ZIP-3 areas designated for cluster sampling as follows.

- Any ZIP-5 within these areas that had less than 900 square miles and at least 54 nonresponding sample cases was its own cluster. The 900 square-mile limit was based on traversing an area no larger than 30 miles by 30 miles within a one-day work packet.
- Any ZIP-5 with fewer than 54 nonresponding sample cases was combined with an adjacent cluster.

The goals for collapsing clusters were to have a minimum of 54 nonrespondents per cluster, with no more collapsing than necessary, and, if possible, keep the total cluster area under 900 square miles. In a few cases this limit was exceeded, for example, in situations where the total area of a ZIP-5 was large, but almost all the population was in a small portion of the area, adjacent to another small ZIP-5. In such cases the area to be traversed for the combined cluster was still well within the limit. This process resulted in the creation of 869 clusters.

Sample Allocation Between Clustered and Unclustered Groups

The research team targeted 6,000 households for nonresponse follow-up. Ten percent of these (600) were allocated to the unclustered sample and 90 percent (5,400) to the clustered sample.

Sample Selection From Unclustered Nonrespondents

The 600 follow-up households from the unclustered nonrespondents were selected with a uniform sampling rate across all unclustered nonrespondents.

Sample Selection From Clustered Nonrespondents

The 5,400 clustered follow-up sample was selected in two stages.

1. From the 869 clusters, the team selected 270 with probability proportional to size. That is, clusters with more nonrespondents were sampled at a higher rate.
2. Within each selected cluster, the team selected 20 households with uniform sampling probability.

With this two-step approach, each household from the clustered nonrespondents had an equal chance of being selected for nonresponse follow-up.

Project Implementation

While the agreements with the utilities for data transfer were being negotiated and the sample frame was being developed and finalized, the study team collaborated on updates of the survey materials and planned the overall project implementation. This section details the results of that planning and implementation.

Materials Design and Pretest

The materials for the *2019 RASS* were based on the materials from the *2009 RASS*. All materials were reviewed by the CEC and the participating utilities for content and appearance. The cover letters were revised to refresh the content and incorporate the push to the web survey. The taglines on the outer envelopes were updated. The survey instrument was also revised to reflect additional areas of interest and updated technologies. The SMUD logo was added to the cover of the survey and the request for a participant's phone number was removed from the back page. The instructions were updated along with the addition of a reference to the web survey on the inside cover. Specific changes to the survey included the following, listed by survey section:

Home and Lifestyle:

- Type of building: Removed "exterior" from townhouse description
- Updated home vintage categories
    - Updated home vintage categories to incorporate more recent construction
    - Adjusted categories between 1975 and 2000 to align with the CEC's forecasting requirements
    - Created break at change in HVAC codes in 2006
    - Maintained category for most recent construction at about five years, as sample is drawn from only partial 2018 data, so expect few homes built in 2019
    - Expanded to 2019 to cover any homes that have been rebuilt after the wildfires
- Added "unknown" as a response option for wall and attic insulation
- Added triple-pane windows as a window option
- Update peak hours for operation of electrical appliances to reflect new peak hours of 4 p.m. to 9 p.m. to align with the California Public Utilities Commission Database for Energy Efficient Resources (DEER).

Electric Vehicles:

- Replaced previous questions about electric vehicles with new battery of questions:
  - Do you currently own or lease a plug-in battery electric vehicle or plug-in hybrid electric vehicle?
  - How many electric vehicles of each type does your household own or lease?
  - On an average day, how many total miles do you drive your electric vehicles?
  - How often do you charge your electric vehicle(s) at home, work or somewhere else?
  - Is your primary charger used at home a Level 1 (120V) or Level 2 (240V)?
  - When is/are the EV(s) normally charged using this primary charger?

Space Heating:

- Added "steam, hot water, fin tubes" to description of "hot water radiator"
- Updated types and descriptions of thermostats to include smart thermostats (thermostats that can be used with home automation systems and may sense when you are in the home, such as Nest, Ecobee, and so forth)

Space Cooling:

- Added "mini-split" or "ductless" AC to the categories of central AC
- Updated types and descriptions of thermostats to include smart thermostats
- Added "68°F – 69°F" category to better capture lower temperature settings

Water Heating:

- Updated showerhead flow rate to 1.8 gallons per minute

Laundry:

- Added question of age of clothes dryer
- Split question about number of dryer loads to correspond with loads at low, medium, or high temperature setting

Food Preparation:

- Added outdoor oven
- Added question of age of dishwasher

Refrigerators:

- Removed "frost-free or manual defrost?" as a characteristic of refrigerators

Swimming Pools:

- Added question about type of pool filtration pump (single-speed, variable speed drive, and so forth.)

Entertainment and Technology:

- Update TV and accessories question (K1):
  - Updated TV categories to include organic light-emitting display (OLED)/ light-emitting diode (LED) TVs and digital light processing (DLP)
  - Added a medium size category for LED and liquid crystal display (LCD) TVs
  - Removed converter box (digital to analog)
  - Adjusted categories of devices connected to TVs; combined less prevalent items that previously were listed separately
  - Added streaming media players, home theater system, and soundbar as separate accessories
- Adjusted size categories of combined hours of TV use to match updated size categories used in K1
- Deleted question about activities conducted on computers (K5)
- Adjusted list to update home office equipment technologies
  - Combined printer, scanner, multifunctional machine into a single category
  - Dropped answering machine
  - Dropped dial-up internet connection
  - Added tablet computer, e-reader
  - Added hubs, controllers, with examples
  - Added smart home devices (excluding lighting)
  - Added smart cell phone and other cell phone categories (replaced the general cell phone category)

Lighting:

- Replaced L1 with new question to indicate estimate of portion of light bulbs of different types (incandescent, compact fluorescent lamp (CFL), LED), instead of asking for count of lamps.
- Added "smart" (connected) light bulbs
- Added external LED fixtures
- Dropped 2009 survey question L5 about having replaced a CFL with an incandescent

Miscellaneous Appliances:

- Updated/revised list of appliances in M1
  - Reduced list by removing electric blanket, aquarium, trash compactor, security system, garage door opener, and lawn mower
  - Added rechargeable vacuum

- Added small cooking appliances
    - Added portable air purifier
    - Added "fountain" to pond or water garden pump line
- Replaced "vehicle" with "bicycle, skateboard" and removed references to car/vehicle in M6

Renewable Energy Technologies (New Section):

- Added question about currently installed renewable energy technologies such as solar, battery storage, wind generator, and fuel cells
- Added question about plans to install renewable energy technologies

Household Information:

- Added SMUD as electricity provider for a vacation home
- Added question about natural gas provider for vacation home
- Added "or Pacific Island" to "Asian" in question about primary language spoken in home
- Updated income categories to accommodate high-income categories
- Added "Prefer not to answer" to income question

Back Cover Page

- Removed request for phone number
- Added a list of frequently asked questions (FAQ)

The 2019 survey instrument is contained in Appendices Volume, Appendix A. Lists of variables from the 2019, 2009, and 2003 surveys are available in Appendices Volume, Appendix B.

The research team pretested the updated materials with 10 energy customers in San Diego and Oakland. Results of the pretest were shared with the study team, and final modifications were made to the materials. The results from the materials pretest are in Appendices Volume, Appendix C.

The direct mail solicitation package included the following items:

- An outgoing envelope (7.5 inches by 10.5 inches) with a window opening
- Standard first- or second-mailing cover letter in English and Spanish
- A 20-page scannable survey in English (6.75 inches by 9.75 inches)
- A business-reply envelope (7 inches by 10 inches)

The direct mail materials are included in the Appendices Volume, Appendix D.

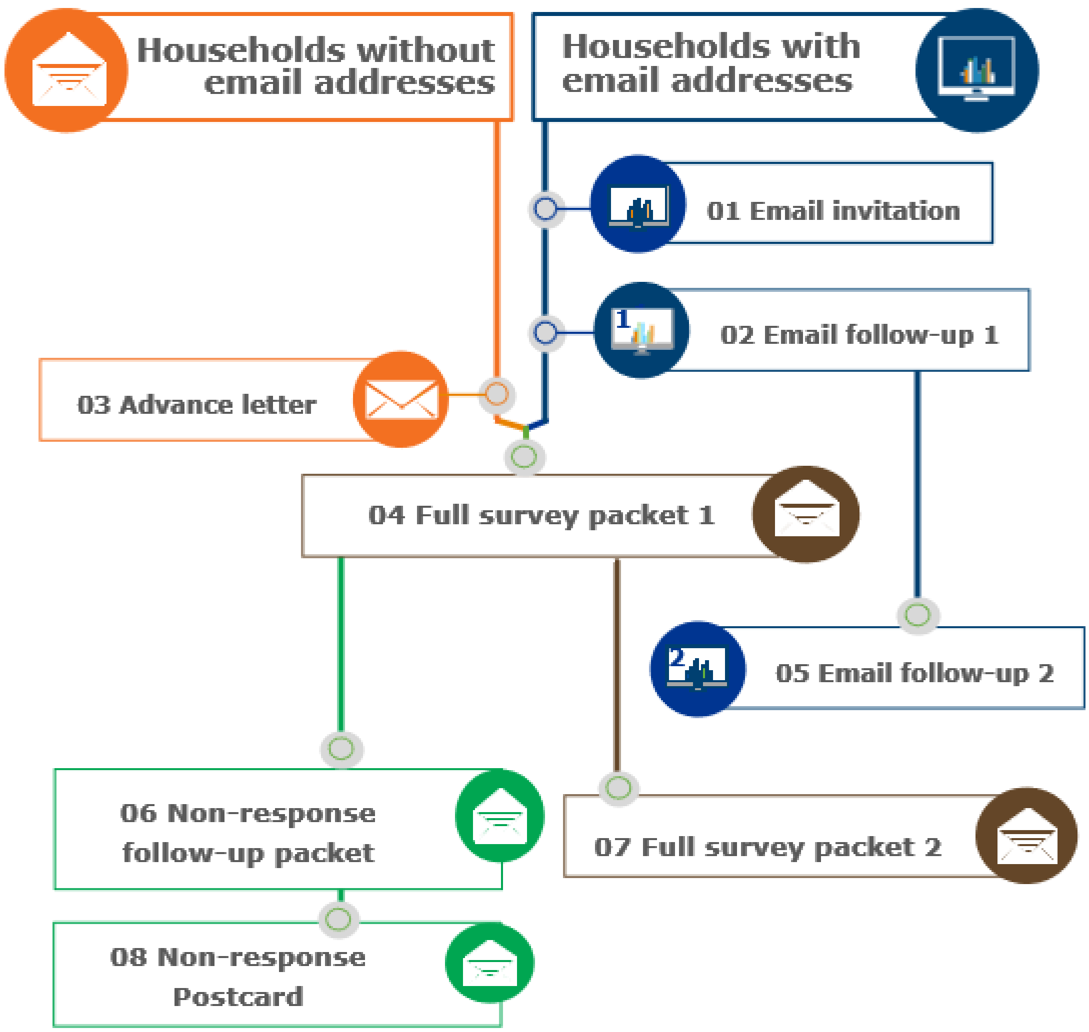
Web Survey

Once the survey content was finalized, the web version of the survey was developed in English and Spanish on the Form.com platform. The web survey followed the structure and format of the questions on the print survey as closely as the software permitted. On the introductory screen, participants elected to take the English version or the Spanish version of the survey. Screen shots of the survey are presented in the Appendices Volume; the English online version is in Appendix E, and the Spanish version is in the Appendix F.

Survey Implementation

As outlined in the Sample Design section, the sample contained households that had an email address on file with their electricity provider and households that did not. The survey implementation was different depending on whether the household had an email address on record. **Figure 1** presents the recruiting efforts for the survey implementation.

Figure 1: Survey Implementation



Source: 2019 California Residential Appliance Saturation Survey

The *2019 RASS* survey implementation began with an effort to contact the sample households and encourage them to complete the survey online. Recruitment for participation in the online survey was conducted in two ways, depending upon the availability of email addresses. Customers that had an email address on file with their electric provider were sent an invitation to start the survey via an embedded link. Customers without an email address received a direct mailing that requested they start the online survey by typing in an access code provided in the mailing. Once a survey was initiated, the participants could return to the survey to review their responses or complete the survey and submit.

The initial email invitations were sent out from an email address with a DNV GL domain (calhes.dnvgl.com) in staggered batches by utility to optimize the survey delivery experience for participants. Batches were sent starting on September 12, 2019, through October 8, 2019. Reminders to encourage survey completion were sent periodically to households who had started, but not completed the survey, and households who had not started the survey. The first reminders were sent about one month after the initial invitations were sent, followed by a second reminder sent about a month later. Reminder emails were sent from an email address with a CEC domain (energy.ca.gov) to provide authenticity to the request to participate.

Direct Mailings

Households that did not have a valid email address on file with the participating electric utilities were sent an advance letter that invited them to complete the online RASS survey. Customer names, mailing addresses, service addresses, and access codes were printed on the advance letters. The access codes had to be entered when prompted to gain access to the online survey. The advance letters were sent on November 13, 2019.

A direct mail package consisting of an outer window envelope, a cover letter, a survey and a business reply envelope followed the advance letter mailing. Customer names and mailing addresses were printed on the cover of the surveys in an area where they would show through the window of the outgoing envelope. A bar code, containing the access code, and the service address were also printed on the survey cover. Instructions on the inside cover directed respondents to complete the survey for the service address printed on the cover. A cover letter identified study sponsors, provided background information on the study, and encouraged readers to complete the survey online.

The direct mail packages were assembled, presorted, and mailed third-class from a direct mail service. The business reply envelopes included in the survey packets ensured the paper surveys would be delivered to the survey processing center. The bar codes on the surveys were scanned as they arrived, and a list was created of the surveys received.

Two groups of households were sent paper survey packets: all households that received an advance letter and 50 percent of the households that were sent an email invitation but did not start a survey online. The survey packet was sent to the advance letter group November 18, 2019. The second wave of survey packets was sent December 6, 2019, to the group who had been sent an email invite but did not start the survey online.

If households had not submitted a survey by late January and had not been selected for the nonresponse follow-up, a second survey packet was sent February 6, 2020. The materials for the second mailing were identical to the first mailing except the outer envelope had an alternate phrase on the cover.

Data collection protocols are contained in Appendices Volume, Appendix G.

Nonresponse Follow-Up Implementation

The *2019 RASS* included an intensive follow-up effort with the objective of reducing nonresponse bias by obtaining survey responses from a sample of selected households that did not respond to the initial RASS survey invitations. This intensive follow-up was designed and implemented to provide additional and compelling opportunities for selected households to complete the survey, with the goal of maximizing the sample representation of residential energy consumers using the available budget. This effort is referred to as the "nonresponse follow-up" for the remainder of this report.

Nonresponse Follow-Up Data Collection

As shown in **Figure 2**, two stages comprised the nonresponse follow-up effort. The first stage was the initial outreach. A priority mailing package was sent to the 6,000 sampled households with instructions on how to complete the survey, a $5 cash pre-incentive, a paper survey packet and return envelope, and a promise to receive an additional $15 reward card upon completion of the survey. The nonresponse priority mail materials are in Appendices Volume, Appendix H.

The priority mail packets were sent on January 10, 2020. A few days later, reminder postcards were sent out to everyone in the sample for the nonresponse follow-up. Following the mailings, a telephone attempt was made to remind them of the mailing and encourage their participation.

The next stage of the nonresponse effort entailed field staff making in-person visits to cluster areas and making additional telephone attempts to the unclustered sample. A sample of 5,400 households in 270 cluster areas was selected for in-person field visits. Each cluster included 20 eligible households in close geographic proximity. A sample of 600 households located in other areas not selected for the in-person visits received phone calls encouraging them to complete the surveys.

The target goal for the intensive outreach was to obtain a final completion rate of at least 20 percent (1,176 completed interviews).

The nonresponse follow-up effort was managed by APPRISE and conducted by APPRISE and GC Green under the supervision of DNV GL. The follow-up effort was implemented from mid-January 2020 to mid-March 2020.

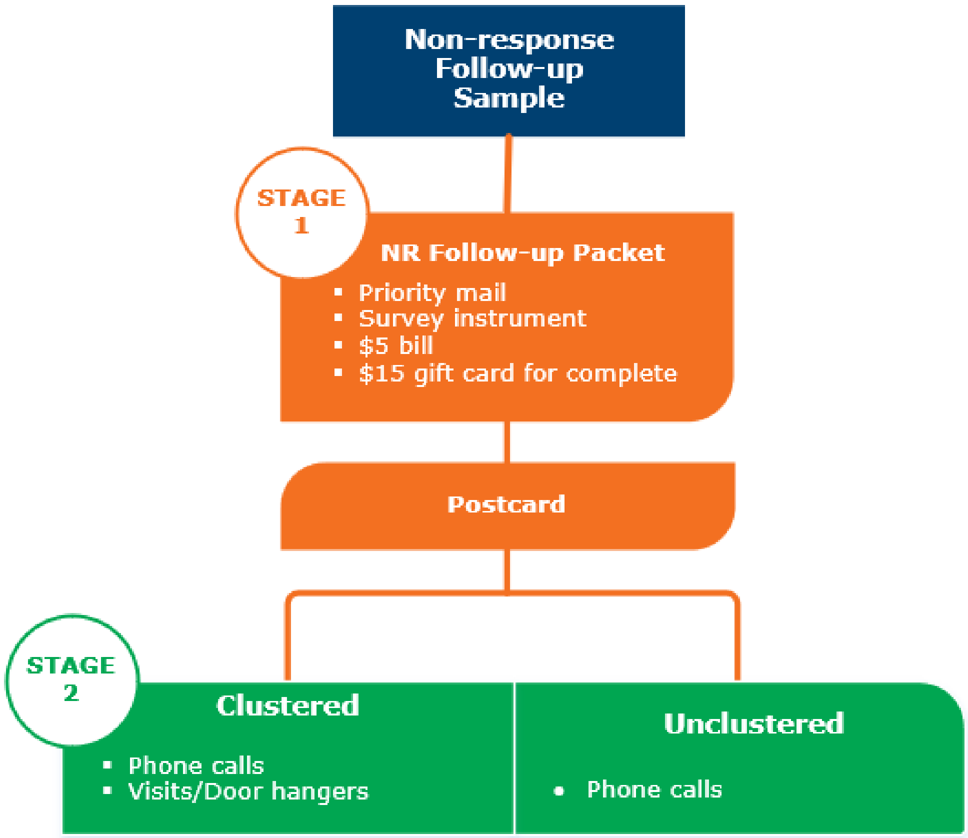
Mailings — The priority mailing package and postcard reminder were sent to all 6,000 sampled households in early January 2020.

Inbound Telephone Calls — APPRISE responded to inbound telephone calls from respondents that received the mailing materials.

Initial Outbound Telephone Calls — APPRISE staff began telephone calls to all sampled cases following the priority package mailing. Initial telephone calls to all cases with a valid phone number were completed in mid-February.

Field Visits — APPRISE and GC Green staff conducted in-person field visits to households in the clustered sample that had not responded to the mailings or initial telephone calls. One cluster area was reassigned to the unclustered sample based on the geographic distance from other cluster areas. Two cluster areas were excluded from field visits because the respective survey completion rates exceeded 50 percent. A total of 267 cluster areas were targeted for field visits. The field visits occurred from January 24, 2020, to March 8, 2020. Materials for the nonresponse field effort are in Appendices Volume, Appendix I.

Figure 2: Nonresponse Follow-Up Data Collection



Source: 2019 California Residential Appliance Saturation Survey

Nonresponse Follow-Up Protocols

Field staff members were assigned cluster areas and completed the visits using the following protocols:

- Review Cluster Assignment and Prepare Materials — Staff reviewed materials for a cluster area they were assigned and prepared materials for conducting the visits.
- Navigate to Targeted Address — Staff traveled to the clustered addresses using prepared assignment sheets and maps.
- Confirm Materials for Targeted Address — Staff confirmed the materials they need for the addresses they are visiting.
- Assess Safety and Access — Staff were trained to access safety at the location and whether the home is accessible to approach.
- Attempt Contact — Staff attempted contact at each located address where there was access to the home or building's main entrance.
- Provide Information/Leave Materials — Staff left materials via bags hung from door hardware if no one was home or answered the door. If staff spoke to an adult resident, he or she presented information about the survey and how to participate. Staff had materials available to provide to the respondent as needed.
- Document and Report Results — Staff recorded the outcome for each address on paper and reported the clustered results to office staff these via telephone.

- Additional Telephone Outreach — Staff made additional telephone attempts to households in the unclustered sample and households who were unable to be visited due to restricted access or safety concerns. Staff made up to four total telephone call attempts to households from mid-February to mid-March.

Protocols for the nonresponse follow-up are in Appendices Volume, Appendix J.

Nonresponse Follow-Up Survey Completes

The combination of the three main activities of the nonresponse follow-up resulted in an overall completion rate of 37 percent. More detailed outcomes include:

- Mailings — All 6,000 households were sent the mailing materials. One hundred ten priority mailing packages were returned as undeliverable, and 22 respondents mailed back their materials without participating.
- Telephone Calls — Staff completed calls to all households with valid telephone numbers. A total of 20 interviews were completed over the telephone, including 6 Spanish interviews.
- Field Visits — Field visits were completed for all targeted clusters. A total of 3,900 households were visited by field staff. Field staff obtained 78 completed surveys during the field visits.

Table 5 displays the intensive outreach results by utility and sample group. The overall survey completion rate was 37 percent for the nonresponse effort, exceeding the target goal by 17 percentage points.

- The completion rate for the clustered sample was 37 percent, ranging from 33 percent for LADWP households to 42 percent for SDGE households.
- The unclustered sample completion rate was 42 percent, ranging from 20 percent for SMUD households to 64 percent for SDGE households.

Table 6 presents the number and percentage of completed interviews by survey mode for each sample group. The most common mode was mail, with 61 percent of the clustered sample respondents and 65 percent of the unclustered sample respondents mailing a completed paper survey packet. About one-third of both the clustered and unclustered respondents completed the survey using the online web instrument. For the clustered sample, 4 percent of respondents completed the survey during an in-person visit by field staff. About 1 percent of respondents completed telephone interviews or completed the survey using multiple modes.

Table 5: Nonresponse Follow-Up Completion Rates

Sample Type	Utility	Sample	Completed Surveys	Completion Rate
Clustered Sample	LADWP	660	215	33%
Clustered Sample	PG&E	1,723	596	35%
Clustered Sample	SCE	2,013	766	38%
Clustered Sample	SDGE	665	276	42%
Clustered Sample	SMUD	319	129	40%
Total Clustered Sample		**5,380**	**1,982**	**37%**
Unclustered Sample	LADWP	8	2	25%
Unclustered Sample	PG&E	373	156	42%
Unclustered Sample	SCE	201	79	39%
Unclustered Sample	SDGE	33	21	64%
Unclustered Sample	SMUD	5	1	20%
Total Unclustered Sample		**620**	**259**	**42%**
Total Non-Response Sample	LADWP	668	217	32%
Total Non-Response Sample	PG&E	2,096	752	36%
Total Non-Response Sample	SCE	2,214	845	38%
Total Non-Response Sample	SDGE	698	297	43%
Total Non-Response Sample	SMUD	324	130	40%
Total Nonresponse Sample		**6,000**	**2,241**	**37%**

Source: 2019 California Residential Appliance Saturation Survey

Table 6: Nonresponse Follow-Up Results by Survey Mode

Survey Mode	Clustered Sample Count	Clustered Sample Percent	Unclustered Count	Unclustered Sample Percent	Total Count	Total Percent
Web	652	33%	84	32%	736	33%
Mail	1,208	61%	168	65%	1,376	61%
Mail & Web	27	1%	4	2%	31	1%
Phone	17	1%	3	1%	20	1%
Field	76	4%	NA	-	76	3%
Field & Web	2	<1%	NA	-	2	<1%
Total	**1,982**	**100%**	**259**	**100%**	**2,241**	**100%**

Source: 2019 California Residential Appliance Saturation Survey

Table 7 shows the distribution of completion results for the 270 clusters. More than 55 clusters (20 percent) had completion rates that exceeded 50 percent, 154 (57 percent) had completion rates between 30 percent and 50 percent, and 52 (19 percent) had completion rates between 20 percent and 30 percent. Only 9 of the 270 clusters had completion rates below 20 percent.

Table 7: Nonresponse Follow-Up Completion Results for Clusters

Completion Rate	Count of Clusters	Percent of Clusters
0%-9%	1	<1%
10%-19%	8	3%
20%-29%	52	19%
30%-39%	87	32%
40%-49%	67	25%
50%-59%	41	15%
60%-69%	13	5%
70%-79%	1	<1%
80%-89%	0	0%
90%-100%	0	0%
Total	270	100%

Source: 2019 California Residential Appliance Saturation Survey

Survey Weights

Individually Metered Sample Weights

Sample Stage

The completed RASS individually metered sample includes respondents in the stages indicated in **Table 8**. Each of these has a different sampling process and therefore a different weight calculation. All individually metered strata included mailed ("paper survey") standard and nonresponse follow-up stages, though some had no completes in one or the other of these. Only email strata had email-only sample.

Paper responses had a lag between when the household completed it and when it was received. The response stages are defined based on when the response was recorded. The key cut-offs for assigning the response stage were the mailing of the first and second paper survey packets. The nonresponse follow-up subsample was selected from all households who received the first paper survey.

The sampling stages were as follows.

- Email only (E). Web survey response to email recruitment registered before the first paper survey packet mailing. Applies only to households in email strata.
- Main early (M). Not an email-only respondent. Web or paper survey response registered **before** the second paper survey mailing. For non-email strata, includes responses received before the second paper survey packet mailing, including for households that did have email addresses provided and were recruited by email prior to the first paper survey.
- Main late (L). Not in the nonresponse follow-up sample, and web or paper survey response received after the second paper survey package was sent.
- Nonresponse follow-up clustered (C). Response from a household in the nonresponse follow-up clustered subsample.
- Nonresponse follow-up unclustered (U). Response from a household in the nonresponse follow-up unclustered subsample.

While these are referred to as stages, the main late and nonresponse follow-up stages were fielded simultaneously. The clustered and unclustered nonresponse follow-up stages were geographically distinct, with the exception that households in the net-metered strata were included only in the unclustered nonresponse follow-up. Some strata had only clustered or unclustered, while others had both stages of nonresponse follow-up.

For weighting, with this structure,

- Within email strata, all households receiving the first paper survey (main early send-out) are treated as a nonresponse follow-up sample from the initial email survey.
- All households receiving the second paper survey (main late send-out) are treated as nonresponse follow-up from the first paper survey.
- The nonresponse follow-up subsample is a nonresponse follow-up sample to the first paper survey.

Because nonresponse follow-up recruitment occurred in parallel with ongoing standard recruitment efforts, the nonresponse follow-up respondents include households that would have responded to the main late effort (the second paper survey mailing) without the additional outreach efforts. These households cannot be identified and are likely to be a small share of the sample. The study team estimated the portion of the non-response follow-up respondents who would not have responded without the follow-up effort based on the difference in response rate between the nonresponse follow-up and main late stages.

General Weighting Process

The general weighting process used the following steps within each sampling stratum:

1. For each of the sample stages, calculated a set of weights that would be used if that stage alone were to represent the full stratum population. These "solo weights" are calculated separately for each sampling stratum, for each stage with completed responses in that stratum.
2. For each sampling stratum, calculated a "blending fraction" (proportion of the stratum) for each of the sample stages with completed responses in that stratum. The sum of the blending fractions is 1 within each stratum.
3. Calculated the combined weights for each stratum and sample stage as the product of the solo weight and the blending fraction. The result is the *base weight*.
4. Calibrated the base weights so that the totals align closely with population counts by key explicit and implicit stratification variables.

For each stratum and stage, the blending fraction is based on the fraction of the stratum population directly represented by that stage. For the email-only, main early, and main late stages, the fraction directly represented is the response rate to the send-out from that stage, multiplied by the portion of the population not represented by the prior stages. The blending fraction for each stage is the fraction represented, divided by the sum of the represented fractions across all stages.

For the nonresponse follow-up stages, the fraction represented is calculated as the difference in response rate between that stage and the main late stage, multiplied by the fraction not represented by the email-only or main early stages. This calculation is an approximation, with the assumption that the nonresponse follow-up sample represents only households that would not have responded to the main late recruitment without the follow-up. That is, the study team calculated the represented population as if the nonresponse follow-up sample excluded households that would have responded without the follow-up.

Table 8 provides the specific calculations. The solo weights, blending fractions, and final weights were calculated within each stratum, as indicated in **Table 9**.

Table 8: Sample Stages

Notation	Sample Stage	Description	Population Directly Represented by Responding Sample	Effective Sendout = (Denominator of Response Rate)	Fraction of Population Represented (Numerator of Blending Fraction)
E	Email only	For email strata only. Selected for initial emailing and email follow-up, but not for paper survey	Customers who would respond to an email survey request without paper follow-up	S_E = # in initial email sendout	$f_E = r_E = n_E/S_r$ defined as $r_E = 0$ for non-email strata
M	Main early respondents	Initial email sendout selected for paper mailings, plus non-email strata sample, who responded prior to non-response follow-up sample pull. Non-email stratum households who had email addresses provided and were included in initial email attempt are included with those that never had email addresses for purposes of weighting.	Customers who would respond by the time of non-response follow-up sample pull (but for email strata, who would not respond to email prior to first paper send-out)	S_M = # of 1st paper sendout	$f_M = r_M(1-r_E)$ $r_M = n_M/S_M$

Notation	Sample Stage	Description	Population Directly Represented by Responding Sample	Effective Sendout = (Denominator of Response Rate)	Fraction of Population Represented (Numerator of Blending Fraction)
L	Main late respondents	Eligible for non-response follow-up (before selecting clusters or unclustered non-response follow-up) but not selected for follow-up and did respond. = Initial email sendout selected for paper mailings, plus paper strata sample, who responded after the non-response follow-up pull and were not selected for the non-response follow-up sample	Customers who would respond only in the later time period, without non-response follow-up	S_L = # from Main Early with no response received by the time of the 2nd mailing, minus the non-response follow-up sendout $S_L = S_M - n_M - S_{Non\text{-}Response}$	$f_L = r_L(1-r_M)(1-r_E)$ $r_L = n_L/S_L$
O	Overlap (unidentifiable)	non-response follow-up respondents who would have responded in later period without non-response follow-up	Customers who would respond only in the later time period, without non-response follow-up		$f_O = r_L(1-r_M)(1-r_E)$
C	Non-response follow-up clustered excluding overlap (unidentifiable)		Customers in clustered areas who would respond only with non-response follow-up		$f_C = (r_C-r_L)(1-r_M)(1-r_E)(N_C/N)$
U	Non-response follow-up unclustered excluding overlap (unidentifiable)		Customers in unclustered areas who would respond only with non-response follow-up		$f_U = (r_U-r_L)(1-r_M)(1-r_E) \times (N_U/N)$
Total Sample			**Total represented = denominator of blending fraction**		$\mathbf{f_{SUM} = f_E + f_M + f_L + f_C + f_U}$

Source: 2019 California Residential Appliance Saturation Survey

Table 9: Weight Calculations for Each Individually Metered Sample Stage

Notation	Sample Stage	Solo Weight WO	Fraction Represented f	Blending Fraction b	Combined Weight W
E	Email only	$W0_E = N/n_E$	f_E	$b_E = f_E/f_{SUM}$	$W_E = b_E W0_E$
M	Main early respondents	$W0_M = N/n_M$	f_M	$b_M = f_M/f_{SUM}$	$W_M = b_M W0_M$
L	Main late respondents	$W0_L = N/n_L$	f_L	$b_L = f_L/f_{SUM}$	$W_L = b_L W0_L$
C	Non-response follow-up Clustered	$W0_C = N/n_C$	f_C	$b_C = f_C/f_{SUM}$	$W_C = b_C W0_U$
I_NC	Non-response follow-up unclustered	$W0_U = N/n_U$	f_U	$b_U = f_U/f_{SUM}$	$W_U = b_U W0_U$

Source: 2019 California Residential Appliance Saturation Survey

In these tables,

N = stratum population count

n_E, n_M, n_L, n_C, n_U = the completed sample counts for email-only, main early, main late, nonresponse follow-up clustered (n_C), and nonresponse follow-up unclustered (n_U) stages in the stratum, respectively.

S_E, S_M, S_L, S_C, S_U = the corresponding respective send-out counts, that is, the number of households who were solicited to complete a survey by that method.

r_E, r_M, r_L, r_C, r_U = the corresponding respective response rates. The response rate is the ratio of the number of completed responses from that stage to the send-out.

Within a stratum, the same population count N is used in the numerator of the solo weight for each of the stages. If a stratum has no completes for a stage, that stage is collapsed with an adjacent stage for weighting to ensure the entire population is represented in the weights.

Collapsing of Stages and Strata

Within a stratum, stages were collapsed for weighting for the following situations.

1. If a stratum had fewer than four responses totaled across all stages, the stratum was combined with another. These strata were collapsed by combining multifamily net-metered strata across forecasting climate zones within a utility. One exception was SMUD, for which there was a single multifamily net-metered response, with a population of 101 and only one response. This case was left as its own stratum. When strata were collapsed, they were combined stage by stage.
2. Within a (possibly collapsed) stratum, if a stage had positive send-out but zero or only one respondent, the 0- or 1-response stage was combined with an adjacent stage in the following priority
   a) Combine clustered and unclustered nonresponse follow-up stages if either has $n \leq 1$.
   b) Combine nonresponse follow-up with main late stages if nonresponse follow-up has $n \leq 1$.
   c) Combine email only with main early stages if EO has $n \leq 1$.

Confirmation of Weights

The base weights are constructed in a way that makes the sum of these weights match the frame counts for any combination of the explicit stratification variables. The study team confirmed those matches as a quality control check.

Weights Calibration: ("Raking")

After the base weights were created, the team calibrated these weights to implicit as well as explicit stratification variables using iterative proportional fitting, or "raking." (For a specific stratification variable, the raking process rescales all the weights so that the sum of sample weights matches the population count for each level of the variable. Rescaling is applied to align each stratification variable in succession. The process iterates until the weights stabilize.)

Raking the base weights produced the final weights used for the full sample. The following raking variables were used:

- Vintage (percentage of homes in the census tract that were built in 1970 or earlier)
- Forecasting climate zone
- Email Available yes/no
- Single or multifamily
- Net metered yes/no
- Likely AC
- Home ownership rate (percentage of homes in the Census tract that are occupied by owners)
- Low income (Census tract percentage bin)
- Electricity usage level
- Utility

Table 10 summarizes by strata the full sample weights for the individually metered respondents. The summary includes the minimum weight, maximum weight, and average for each stratum.

The load profile sample is the subset of the full sample for which interval data were available. This subsample represents the same population as the full sample and requires its own set of weights. The starting point for these weights was the full-sample base weight for each household in the load profile sample. The same raking procedure was applied to the load profile subsample base weights, with the same raking variables, as was used to produce the full sample final weights.

In addition to the full-sample and load-profile weights, the team developed a third set of weights solely for use in the calibration step of the conditional demand analysis described in Chapter 5. The CDA calibration weights were calculated by the same process as the load-profile weights. The starting point was the full-sample base weight for each household in the CDA estimation subsample. Since net-metered homes were not used in the CDA estimation, the team excluded net-metering strata from the CDA calibration weights.

Table 10: Individually Metered Full Sample Weights

Utility	Stratum	Email Available	Forecasting Climate Zone	Net Metered	Dwelling Type	Electricity Usage Level	Population	Standard Email and Mail Surveys	Non-Response Follow-Up Surveys	Minimum Weight	Average Weight	Maximum Weight
LADWP	001	No	16	No	MF	High	86,614	85	7	45.8	1,019.0	12,723.5
LADWP	002	No	16	No	MF	Low	258,961	349	27	33.0	742.0	14,948.1
LADWP	003	No	16	No	SF	High	47,790	74	7	25.0	645.8	7,218.6
LADWP	004	No	16	No	SF	Low	156,318	276	13	36.9	566.4	12,945.2
LADWP	006	No	16	Yes	SF	All	13,219	23	0	39.7	574.7	1,749.6
LADWP	007	No	17	No	MF	All	149,691	160	17	30.6	935.6	12,188.2
LADWP	008	No	17	No	SF	All	111,121	254	24	18.5	437.5	11,833.5
LADWP	009	No	17	Yes	SF	All	6,797	19	0	107.9	357.8	1,302.3
LADWP	010	Yes	16	No	MF	High	64,192	123	11	8.2	521.9	8,916.9
LADWP	011	Yes	16	No	MF	Low	197,422	442	30	10.7	446.7	11,720.4
LADWP	012	Yes	16	No	SF	All	92,932	377	22	5.0	246.5	6,416.2
LADWP	014	Yes	16	Yes	SF	All	5,218	34	1	11.6	153.5	511.4
LADWP	015	Yes	17	No	MF	All	111,005	240	31	3.3	462.5	6,153.1
LADWP	016	Yes	17	No	SF	All	54,903	303	28	3.8	181.2	5,979.7
LADWP	017	Yes	17	Yes	SF	All	3,957	35	0	21.7	113.1	275.0
LADWP	201	Yes	16	Yes	MF	All	2,609	2	0	1,000.8	1,304.7	1,608.6
PG&E	018	No	1	No	MF	High	82,094	193	8	93.3	425.4	8,653.2
PG&E	019	No	1	No	MF	Low	82,147	151	10	94.5	544.0	8,827.4
PG&E	020	No	1	No	MF	Medium	185,670	275	14	117.2	675.2	13,557.6
PG&E	021	No	1	No	SF	High	139,475	273	15	99.8	510.9	9,393.7
PG&E	022	No	1	No	SF	Low	149,304	362	12	78.3	412.4	13,384.3
PG&E	023	No	1	No	SF	Medium	223,537	604	27	48.9	370.1	10,836.7
PG&E	025	No	1	Yes	SF	All	42,353	104	0	252.5	407.2	522.4
PG&E	026	No	2	No	MF	All	49,146	137	5	91.9	358.7	7,307.8
PG&E	027	No	2	No	SF	High	49,109	107	2	308.1	459.0	634.6
PG&E	028	No	2	No	SF	Low	107,171	362	13	47.9	296.1	10,153.5
PG&E	030	No	2	Yes	SF	All	9,704	34	1	219.2	285.4	407.5

Utility	Stratum	Email Available	Forecasting Climate Zone	Net Metered	Dwelling Type	Electricity Usage Level	Population	Standard Email and Mail Surveys	Non-Response Follow-Up Surveys	Minimum Weight	Average Weight	Maximum Weight
PG&E	031	No	3	No	MF	All	8,273	24	2	250.7	344.7	514.2
PG&E	032	No	3	No	SF	All	70,263	172	8	40.3	408.5	14,938.6
PG&E	033	No	3	Yes	SF	All	4,580	15	0	235.0	305.4	402.4
PG&E	034	No	4	No	MF	All	56,335	149	14	43.9	378.1	4,437.1
PG&E	035	No	4	No	SF	High	71,194	160	13	62.3	445.0	5,492.4
PG&E	036	No	4	No	SF	Low	222,203	528	29	60.9	420.8	9,062.6
PG&E	038	No	4	Yes	SF	All	28,510	92	4	51.9	309.9	6,100.2
PG&E	039	No	5	No	MF	All	48,981	118	15	61.0	415.1	4,187.2
PG&E	040	No	5	No	SF	High	50,171	102	8	87.9	491.9	8,849.8
PG&E	041	No	5	No	SF	Low	176,546	305	20	112.2	578.8	8,683.1
PG&E	043	No	5	Yes	SF	All	24,356	61	0	250.4	399.3	506.9
PG&E	044	No	6	No	MF	All	40,152	119	12	55.2	337.4	5,777.9
PG&E	045	No	6	No	SF	High	17,572	90	3	32.0	195.2	5,154.9
PG&E	046	No	6	No	SF	Low	113,510	268	15	74.7	423.5	10,615.5
PG&E	048	No	6	Yes	SF	All	11,611	31	1	243.5	374.5	547.3
PG&E	049	Yes	1	No	MF	High	140,989	678	36	24.9	207.9	3,789.3
PG&E	050	Yes	1	No	MF	Low	122,110	443	26	19.9	275.6	6,988.1
PG&E	051	Yes	1	No	MF	Medium	261,384	799	54	16.6	327.1	7,692.7
PG&E	052	Yes	1	No	SF	High	155,719	626	24	27.4	248.8	6,585.9
PG&E	053	Yes	1	No	SF	Low	154,020	702	34	19.1	219.4	6,027.0
PG&E	054	Yes	1	No	SF	Medium	304,753	1410	46	11.9	216.1	11,727.3
PG&E	056	Yes	1	Yes	SF	All	78,627	507	14	23.8	155.1	4,799.6
PG&E	057	Yes	2	No	MF	All	57,772	294	26	23.3	196.5	3,139.8
PG&E	058	Yes	2	No	SF	High	43,096	174	8	17.5	247.7	6,793.8
PG&E	059	Yes	2	No	SF	Low	129,487	758	21	17.2	170.8	8,835.5
PG&E	061	Yes	2	Yes	SF	All	17,798	140	2	32.6	127.1	4,184.9
PG&E	062	Yes	3	No	MF	All	13,515	65	0	97.4	207.9	361.0
PG&E	063	Yes	3	No	SF	All	68,924	262	8	25.8	263.1	13,696.5
PG&E	064	Yes	3	Yes	SF	All	7,742	57	0	67.3	135.8	243.4

Utility	Stratum	Email Available	Forecasting Climate Zone	Net Metered	Dwelling Type	Electricity Usage Level	Population	Standard Email and Mail Surveys	Non-Response Follow-Up Surveys	Minimum Weight	Average Weight	Maximum Weight
PG&E	065	Yes	4	No	MF	All	73,854	279	20	23.4	264.7	7,850.7
PG&E	066	Yes	4	No	SF	High	73,407	278	12	34.4	264.1	6,650.9
PG&E	067	Yes	4	No	SF	Low	94,684	384	21	23.6	246.6	6,018.7
PG&E	068	Yes	4	No	SF	Medium	182,999	739	25	29.8	247.6	7,780.6
PG&E	070	Yes	4	Yes	SF	All	49,428	328	9	26.5	150.7	4,092.3
PG&E	071	Yes	5	No	MF	All	71,666	225	32	22.3	318.5	3,078.5
PG&E	072	Yes	5	No	SF	High	57,419	167	10	50.6	343.8	4,984.8
PG&E	073	Yes	5	No	SF	Low	182,025	547	35	38.2	332.8	6,031.6
PG&E	075	Yes	5	Yes	SF	All	42,435	183	3	41.6	231.9	10,064.4
PG&E	076	Yes	6	No	MF	All	45,490	237	15	34.1	191.9	2,719.3
PG&E	077	Yes	6	No	SF	High	34,804	166	9	25.2	209.7	4,477.8
PG&E	078	Yes	6	No	SF	Low	110,431	500	25	24.0	220.9	5,086.7
PG&E	080	Yes	6	Yes	SF	All	16,550	119	2	40.1	139.1	3,883.5
PG&E	202	Yes	All PG&E	Yes	MF	All	12,282	64	1	107.2	191.9	340.1
SCE	081	No	7	No	MF	High	151,611	266	27	64.1	570.0	6,006.4
SCE	082	No	7	No	MF	Low	129,739	267	29	57.2	485.9	7,327.1
SCE	083	No	7	No	MF	Medium	307,191	564	65	58.8	544.7	7,800.7
SCE	084	No	7	No	SF	High	248,545	506	29	60.6	491.2	11,347.0
SCE	085	No	7	No	SF	Low	306,523	640	41	49.5	478.9	10,180.2
SCE	086	No	7	No	SF	Medium	488,569	1312	103	23.6	372.4	9,134.2
SCE	088	No	7	Yes	SF	All	69,513	196	11	64.9	354.7	5,862.8
SCE	089	No	8	No	MF	All	48,735	111	7	54.6	439.1	7,675.4
SCE	090	No	8	No	SF	High	45,783	102	6	65.4	448.9	7,309.9
SCE	091	No	8	No	SF	Low	129,866	349	17	68.7	372.1	8,559.1
SCE	093	No	8	Yes	SF	All	7,949	54	2	50.6	147.2	1,980.0
SCE	094	No	9	No	MF	All	19,768	42	2	289.3	470.7	630.4
SCE	095	No	9	No	SF	High	34,452	59	6	62.5	583.9	5,385.6
SCE	096	No	9	No	SF	Low	97,978	242	8	61.3	404.9	14,245.8
SCE	097	No	9	Yes	SF	All	8,336	39	1	143.3	213.8	269.2

Utility	Stratum	Email Available	Forecasting Climate Zone	Net Metered	Dwelling Type	Electricity Usage Level	Population	Standard Email and Mail Surveys	Non-Response Follow-Up Surveys	Minimum Weight	Average Weight	Maximum Weight
SCE	098	No	10	No	MF	All	84,332	131	6	112.3	643.8	10,763.2
SCE	099	No	10	No	SF	High	82,661	107	15	66.8	772.5	6,111.9
SCE	100	No	10	No	SF	Low	89,412	171	10	65.7	522.9	9,401.2
SCE	101	No	10	No	SF	Medium	162,159	300	17	72.8	540.5	12,514.1
SCE	102	No	10	Yes	SF	All	21,122	67	2	61.3	315.2	8,097.6
SCE	103	No	11	No	MF	All	53,104	68	3	101.2	780.9	13,989.0
SCE	104	No	11	No	SF	High	83,336	131	5	106.7	636.2	10,706.1
SCE	105	No	11	No	SF	Low	198,039	470	25	31.4	421.4	8,335.6
SCE	107	No	11	Yes	SF	All	35,488	93	7	47.3	381.6	7,413.7
SCE	108	Yes	7	No	MF	High	54,377	239	17	17.5	227.5	3,224.8
SCE	109	Yes	7	No	MF	Low	159,989	824	59	17.5	194.2	3,693.6
SCE	110	Yes	7	No	SF	High	140,202	540	37	24.6	259.6	4,236.5
SCE	111	Yes	7	No	SF	Low	180,736	770	59	20.1	234.7	3,273.6
SCE	112	Yes	7	No	SF	Medium	320,763	1283	72	16.9	250.0	5,327.2
SCE	114	Yes	7	Yes	SF	All	54,754	423	8	19.1	129.4	4,707.5
SCE	115	Yes	8	No	MF	All	16,817	104	5	14.9	161.7	3,764.8
SCE	116	Yes	8	No	SF	All	79,021	467	12	26.5	169.2	5,780.3
SCE	117	Yes	8	Yes	SF	All	4,396	92	3	7.0	47.8	1,245.4
SCE	118	Yes	9	No	MF	All	5,222	30	1	106.4	174.1	255.4
SCE	119	Yes	9	No	SF	All	58,598	290	22	10.2	202.1	2,856.4
SCE	120	Yes	9	Yes	SF	All	5,720	62	1	41.5	92.3	153.3
SCE	121	Yes	10	No	MF	All	28,402	118	7	24.7	240.7	4,174.2
SCE	122	Yes	10	No	SF	High	39,785	134	11	27.4	296.9	3,865.8
SCE	123	Yes	10	No	SF	Low	118,500	562	35	12.7	210.9	4,045.8
SCE	124	Yes	10	Yes	SF	All	15,094	134	3	19.7	112.6	3,545.4
SCE	125	Yes	11	No	MF	All	19,067	89	8	14.6	214.2	2,474.0
SCE	126	Yes	11	No	SF	High	39,111	139	4	47.7	281.4	6,654.7
SCE	127	Yes	11	No	SF	Low	112,392	574	28	26.0	195.8	3,892.0
SCE	129	Yes	11	Yes	SF	All	26,023	210	9	13.5	123.9	4,477.1

Utility	Stratum	Email Available	Forecasting Climate Zone	Net Metered	Dwelling Type	Electricity Usage Level	Population	Standard Email and Mail Surveys	Non-Response Follow-Up Surveys	Minimum Weight	Average Weight	Maximum Weight
SCE	203	Yes	All SCE	Yes	MF	All	1,981	10	0	119.9	198.1	441.0
SDGE	130	No	12	No	MF	All	70,318	196	10	85.6	358.8	6,064.3
SDGE	131	No	12	No	SF	High	37,382	102	4	128.8	366.5	5,047.5
SDGE	132	No	12	No	SF	Low	116,015	470	23	60.4	246.8	4,154.9
SDGE	134	No	12	Yes	SF	All	11,464	29	0	294.3	395.3	630.1
SDGE	135	Yes	12	No	MF	High	87,844	284	15	24.3	309.3	5,656.3
SDGE	136	Yes	12	No	MF	Low	87,006	323	21	31.2	269.4	4,012.4
SDGE	137	Yes	12	No	MF	Medium	179,234	548	45	25.0	327.1	4,060.6
SDGE	138	Yes	12	No	SF	High	135,571	641	44	19.2	211.5	3,305.9
SDGE	139	Yes	12	No	SF	Low	140,558	780	40	21.6	180.2	3,263.2
SDGE	140	Yes	12	No	SF	Medium	262,752	1343	82	16.9	195.6	3,960.0
SDGE	142	Yes	12	Yes	SF	All	104,889	446	15	31.9	235.2	6,240.5
SDGE	204	Yes	12	Yes	MF	All	6,137	10	0	460.3	613.7	908.1
SMUD	143	No	13	No	MF	All	43,825	96	13	48.1	456.5	5,165.8
SMUD	144	No	13	No	SF	High	35,880	95	2	104.8	377.7	9,403.3
SMUD	145	No	13	No	SF	Low	106,486	395	22	46.7	269.6	4,504.2
SMUD	146	No	13	Yes	SF	All	2,975	12	0	169.8	247.9	476.7
SMUD	147	Yes	13	No	MF	All	89,256	256	19	30.4	348.7	4,987.2
SMUD	148	Yes	13	No	SF	High	56,910	275	14	24.4	206.9	3,347.3
SMUD	149	Yes	13	No	SF	Low	175,707	1123	61	19.2	156.5	4,978.1
SMUD	150	Yes	13	Yes	MF	All	83	1	0	82.7	82.7	82.7
SMUD	151	Yes	13	Yes	SF	All	13,845	113	1	44.2	122.5	318.5
TOTAL							12,179,421	39,682	2,263			

Source: 2019 California Residential Appliance Saturation Survey

Master-Metered Sample Weights

The research team developed basic weights for the master-metered completed surveys as the ratio of the stratum population divided by the number of completed surveys. The population counts were from the initial population data as provided by the participating utilities.

Table 11 presents the number of completed surveys per stratum along with the sample weight for the master-metered sample.

Table 11: Master-Metered Weights 2-4 Units

Utility	Population (Units)	Completes	Weight
LADWP	2,308	21	109.9
PG&E	16,550	179	92.5
SDGE	3,352	51	65.7
SCE	4,971	54	92.1
Total	**27,181**	**305**	

Source: 2019 California Residential Appliance Saturation Survey

Comparison of Results Across Sampling and Study Groups

Nonresponse Follow-Up Comparison

A nonresponse follow-up effort can effectively reach segments of the population that do not respond to the initial mailings. **Table 12** compares the households that completed their surveys without nonresponse follow-up to the households responding with nonresponse follow-up. The nonresponse households had similar major equipment and energy usage in their households to the initial mail responders. Key differences of nonresponse follow-up households were that they are:

- Less likely to own their home.
- Likely to have fewer seniors in the household.
- More likely to have a head of household that is Hispanic.

Web Survey Comparison

The 2019 survey used a push-to-web strategy as the initial recruitment option, with follow-up paper surveys mailed to households who did not respond to initial emailed requests or to an advance letter for those without email addresses available. Online surveys can reach respondents not inclined to complete a paper and pencil survey and have lower recruitment costs.

Table 12: Comparison by Surveying Method and Dwelling Type*

Survey Questions	Single Family Initial Mail	Single Family Non-Response	Multifamily (2-4 Units) Initial Mail	Multifamily (2-4 Units) Non-Response	Multifamily (5+ Units) Initial Mail	Multifamily (5+ Units Non-Response	Mobile Homes Initial Mail	Mobile Homes Non-Response
Completed Surveys	24,942	1,321	5,592	428	6,083	480	802	34
Weighted to Population	2,383,860	4,916,557	578,286	1,599,575	668,606	1,789,568	83,897	159,074
Average Electric Consumption	7,257	7,268	4,565	4,475	3,810	3,880	6,632	6,775
Average Gas Consumption	438	439	270	257	180	208	337	410
Average Dwelling Size	1,968	1,848	1,228	1,212	969	942	1,295	1,316
Average Dwelling Age	47	47	45	42	41	39	38	40
Average Number of People	2.8	3.2	2.4	2.8	2.2	2.5	2.2	2.4
Average Number of Seniors	0.51	0.35	0.39	0.21	0.36	0.21	0.64	0.38
Average Income	118,723	116,664	89,255	85,259	76,982	74,091	45,181	38,704
Owners	89%	82%	49%	39%	28%	16%	87%	82%
Central Cooling	59%	61%	40%	38%	33%	33%	54%	36%
Gas Space Heating	75%	74%	61%	53%	37%	35%	58%	52%
Primary Language English	82%	81%	76%	68%	73%	74%	85%	89%
Head of Household Latino	16%	26%	20%	34%	19%	28%	18%	30%
College Grad or Higher	59%	58%	54%	50%	53%	47%	27%	22%

* Individually Metered Accounts
Source: 2019 California Residential Appliance Saturation Survey

Table 13 compares respondents by whether they completed a paper survey or submitted a survey through the website. The results suggest that households completing a survey online were more likely to have a higher average annual income, more likely to have fewer seniors in the household, and less likely to have gas space heating.

Table 13: Comparison by Survey Mode

Survey Questions	Single-Family Mail Survey	Single-Family Online Survey	Multifamily (2-4 Units) Mail Survey	Multifamily (2-4 Units) Online Survey	Multifamily (5+ Units) Mail Survey	Multifamily (5+ Units) Online Survey	Mobile Homes Mail Survey	Mobile Homes Online Survey
Completed Surveys	14,018	12,245	3,083	2,937	3,506	3,057	588	248
Weighted to Population	4,717,578	2,582,839	1,333,228	844,632	1,488,161	970,012	194,176	48,795
Average Electric Consumption	7,269	7,256	4,480	4,521	3,821	3,934	6,400	8,955
Average Gas Consumption	445	425	271	242	221	169	387	374
Average Dwelling Size	1,891	1,881	1,362	1,264	1,097	1,109	965	925
Average Dwelling Age	48	46	45	39	43	42	40	38
Average Number of People	3.1	3.1	2.6	2.8	2.7	2.8	2.3	2.5
Average Number of Seniors	0.46	0.30	0.32	0.16	0.34	0.12	0.49	0.40
Average Income	110,586	129,666	85,014	88,380	69,982	82,389	38,456	50,830
Owners	86%	82%	45%	36%	22%	14%	83%	88%
Central Cooling	62%	57%	40%	37%	39%	24%	39%	54%
Gas Space Heating	79%	66%	62%	45%	42%	26%	51%	67%
Primary Language English	83%	78%	72%	66%	78%	67%	87%	92%
Head of Household Hispanic	23%	22%	29%	31%	26%	25%	21%	44%
College Grad or Higher	56%	63%	49%	54%	45%	54%	24%	20%

Source: 2019 California Residential Appliance Saturation Survey

Energy Consumption Comparison

The team compared household energy consumption between RASS survey respondents and the target population. The team obtained the energy consumption data from the original population files, from which the sample frame was developed. The energy consumption of the survey respondents was then compared to the average energy consumption of the population by stratum. **Table 14** compares the average energy consumption for respondents to that of the target population. The average energy consumption of respondents in the higher energy consumption strata was slightly lower than the population averages for their respective stratum.

The All stratum column includes households that were aggregated across strata because of low numbers in the more detailed strata. Because these strata reflect combinations of household types, the average for the respondents is more likely to vary more from the population compared to the more homogeneous strata.

Table 14: Comparison of Energy Consumption for RASS Respondents and Target Population

Utility	Study Characteristic	Single family Dwelling Low	Single family Dwelling Medium	Single family Dwelling High	All	Net Metered	Multi-family Dwelling Low	Multi-family Dwelling Medium	Multi-family Dwelling High	All	Net Metered	Utility Total
LADWP	Population Count	130,613		42,971	288,699	29,944	482,086		155,624	230,958	1,850	1,362,745
LADWP	Population kWh/Yr	4,726		18,157	9,232	6,696	2,454		7,933	4,677	4,731	5,702
LADWP	Sample Count	276		74	934	111	791		208	400	2	2,796
LADWP	Respondent kWh/Yr	5,345		16,420	8,476	4,625	2,593		6,557	4,356	3,038	5,332
LADWP	StdErr of Respondent kWh/Yr	303		1,160	766	465	114		427	273	1,077	236
PG&E	Population Count	1,418,740	727,135	702,012	134,827	332,802	224,884	431,210	213,042	469,544	13,171	4,667,367
PG&E	Population kWh/Yr	4,747	6,112	13,337	8,300	4,216	1,334	3,009	6,254	4,528	2,638	6,032
PG&E	Sample Count	4,716	2,753	2,143	434	1,671	594	1,074	871	1,647	64	15,967
PG&E	Respondent kWh/Yr	4,878	6,107	12,606	10,157	3,973	1,352	2,904	5,928	4,335	1,990	5,949
PG&E	StdErr of Respondent kWh/Yr	135	129	380	1,437	460	52	64	208	157	343	131
SCE	Population Count	1,199,237	1,003,043	723,300	130,936	248,342	323,941	275,658	196,564	282,133	2,033	4,385,187
SCE	Population kWh/Yr	4,676	6,142	13,391	7,869	8,683	2,397	3,353	6,853	4,746	6,249	6,622
SCE	Sample Count	3,778	2,895	1,718	757	1,370	1,091	564	505	693	10	13,381
SCE	Respondent kWh/Yr	4,633	6,139	12,261	7,849	8,805	2,348	3,328	7,250	4,936	6,251	6,446
SCE	StdErr of Respondent kWh/Yr	127	79	268	497	455	98	83	456	448	1,400	112
SDGE	Population Count	251,239	267,135	174,511		115,752	92,340	174,854	86,284	70,323	6,732	1,239,170
SDGE	Population kWh/Yr	3,160	5,398	11,205		3,039	1,623	3,296	6,291	3,296	2,473	4,891
SDGE	Sample Count	1,250	1,343	743		475	323	548	284	196	10	5,172
SDGE	Respondent kWh/Yr	3,037	5,395	10,166		2,762	1,502	3,372	5,860	3,104	2,958	4,650
SDGE	StdErr of Respondent kWh/Yr	110	82	304		830	74	80	206	368	831	140
SMUD	Population Count	282,190		92,790		16,754				133,073	157	524,964
SMUD	Population kWh/Yr	7,383		18,042		7,389				5,731	7,057	8,848
SMUD	Sample Count	1,518		370		125				352	1	2,366
SMUD	Respondent kWh/Yr	7,177		16,035		7,907				5,610	21,356	8,371
SMUD	StdErr of Respondent kWh/Yr	216		705		1,252				540		306

Source: 2019 California Residential Appliance Saturation Survey;

Precision of RASS Estimates

Individually Metered Sample Precision

Table 15 presents the precision of estimates for the individually metered sample by electric utility at the 90 percent confidence intervals. The three columns on the right of the **Table 15** provide the percentage points to be added to and subtracted from an estimate of 50 or 50, 20 or 80, and 10 or 90 percent, respectively, to obtain the 90 percent confidence bounds.

Table 15: Precision of Estimates for the Individually Metered Sample

Utility	Population	Total Completes	90% Confidence Bounds (+/-) 50/50%	90% Confidence Bounds (+/-) 20/80%	90% Confidence Bounds (+/-) 10/90%
LADWP	1,362,747	2,796	1.6%	1.2%	0.9%
PG&E	4,667,374	15,967	0.7%	0.5%	0.4%
SCE	4,385,163	13,381	0.7%	0.6%	0.4%
SDG&E	1,239,171	5,172	1.1%	0.9%	0.7%
SMUD	524,966	2,366	1.7%	1.4%	1.0%
Total	12,179,421	39,682	0.4%	0.3%	0.2%

Source: 2019 California Residential Appliance Saturation Survey

Master-Metered Sample Precision

Table 16 presents the precision of estimates for the master-metered sample by electric utility at the 90 percent confidence intervals. The three columns on the right of the **Table 16** provide the percentage points to be added to and subtracted from an estimate of 50 or 50, 20 or 80, and 10 or 90 percent, respectively, to obtain the 90 percent confidence bounds.

Table 16: Precision of Estimates for the Master-Metered Sample

Utility	Population	Total Completes	90% Confidence Bounds (+/-) 50/50%	90% Confidence Bounds (+/-) 20/80%	90% Confidence Bounds (+/-) 10/90%
LADWP	2,308	21	17.9%	14.4%	10.8%
PG&E	16,558	179	6.1%	4.9%	3.7%
SCE	4,881	53	11.3%	9.0%	6.8%
SDG&E	3,285	50	11.6%	9.3%	7.0%
Total	27,032	303	4.7%	3.8%	2.8%

Source: 2019 California Residential Appliance Saturation Survey

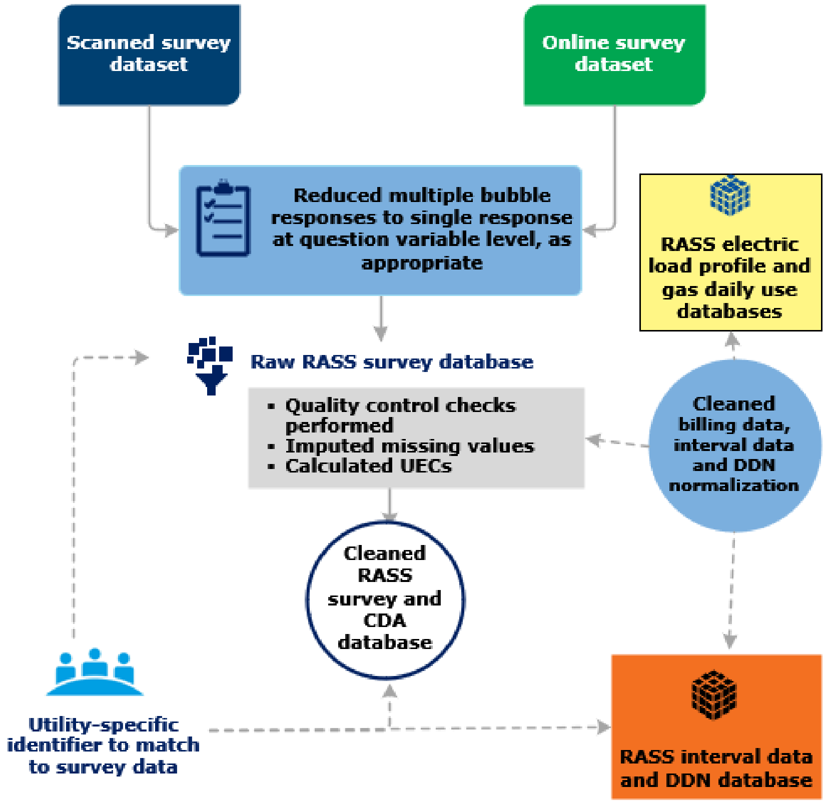
CHAPTER 3: Database Preparation

This section describes the data files submitted to the CEC and each of the participating electric and gas utilities upon the completion of the study. The categories of data files from the RASS included:

- Raw RASS survey database: This database contained RASS survey data that were subjected to minimal cleaning procedures, that is, limited changes to the household responses recorded from the online or paper survey.
- Cleaned RASS survey data and CDA database: This database contained cleaned RASS survey data, variables used in the CDA and household normalized annual consumption (NAC), and unit end consumption (UEC) estimates for end uses.
- Billing database: This contained monthly and bimonthly (LADWP) billing series data foreach of the participating utilities.
- Interval data and degree-day normalization (DDN) file set: This set of files contained interval data for each of the participating utilities (except LADWP) and estimated normalized annual consumption for each household. LADWP did not provide interval metered data because it is only in the initial stages of smart meter deployment.
- Electric load profiles and gas daily use file set: This set of tiles contained aggregated hourly electric and daily gas interval data.

The participating utilities received an additional database, which contained utility-specific information, allowing them to match the RASS survey and billing data to their specific customers. **Figure 3** provides an overview of how these databases were constructed.

Figure 3: Overview of RASS Database Preparation



Source: 2019 California Residential Appliance Saturation Survey

Database 1: Raw RASS Survey Database

Slightly more than half of the RASS surveys were completed as paper surveys and scanned electronically into a fixed-format text file. The team converted data from surveys completed online to the same fixed-format file structure as the scanned paper survey data file. Responses from the paper surveys and the online surveys were then combined into one dataset.

The survey had 1,413 potential responses to questions, each represented by a bubble on the paper survey that was recorded in a text file by the scanning program. The initial Statistical Analysis System (SAS) code created separate fields for each response bubble as the text files were read into SAS datasets.

The first data cleaning step entailed condensing each of the separate fields into a single variable by assigning a value based on the populated bubble. For questions in which a respondent marked one response, the variable was simply assigned the value of the single response. For cases where a respondent marked multiple responses, the study team developed a set of decision rules to select a single value to be assigned to each variable. The choice was typically programmed as either the minimum or maximum value of the multiple responses, depending on the specific variable. For example, for the variable indicating years of residence, the maximum value was chosen for respondents who had provided multiple answers. For some survey variables, the choice of the single value assigned from multiple values depended on responses to other questions within the survey, thereby providing logically consistent answers to each question. The variables contained in the Raw RASS Survey database are listed in the Appendices Volume, Appendix K.

Responses from the online survey were stored directly into a data file and exported as a text file. The SAS code read in the online survey data and performed manipulations as needed to align the format of the online data to the format of the paper survey data. The online survey data and the paper survey data were then stacked to create a single raw survey dataset for analyses.

Database 2: Cleaned RASS Survey and CDA Database

The study team conducted quality control checks and performed additional cleaning steps on the raw survey data to develop the cleaned RASS survey and CDA database. These steps resulted in the omission of surveys based on incomplete data, an inordinate number of multiple responses, or an excessive number of logical inconsistencies from the final dataset. The details of the survey cleaning processes are discussed in Chapter 4.

Algorithms were designed to fill and impute missing values for variables used in the CDA. In addition, responses indicating fuels used for space heating, water heating, and other appliances were cross-referenced with billing data to identify and correct fuel misreporting. Chapter 4 discusses the CDA data imputation and consumption cleaning processes. Chapter 5 covers the specifics of the CDA modeling.

Household and end-use UECs from the CDA and post-normalized annual household electric and gas consumption data were appended to the cleaned survey data. The variables contained in the cleaned RASS Survey and CDA database are listed in the Appendices Volume, Appendix L.

Nonresponse and Not Applicable Indicators

Some sets of questions in the RASS survey incorporated a skip pattern. For example, if a household did not pay for the energy use of a swimming pool, they were instructed to skip to the next survey section. The nonresponses for questions subject to the intentional skip pattern were assigned a value of 99 as being not applicable during the cleaning process. The simple respondent nonresponse was assigned a value of 97 as a missing value during the cleaning. Surveys that contained an excessive number of nonresponses were omitted from the cleaned survey dataset.

Logical Response Inconsistencies

Some survey questions were interrelated, to which the response to one question would presumably influence the response of another question. For example, if a household reported not having a gas line to the home, it would be logically inconsistent if the household reported having a gas range in the residence. Where possible, logically inconsistent responses were corrected using billing data or other survey information. In cases where a value could not be inferred, the response was assigned the missing value of 97, and a logical inconsistency flag was set. The number of logical inconsistency flags was counted. No surveys were omitted from the cleaned survey dataset from having too many logical inconsistencies.

Imputing Missing Values

Although missing survey values were recorded as 97s in the cleaned RASS survey database, retaining these missing values in the CDA would have resulted in a nonresponse bias. Therefore, an approach was developed to impute all the variables that were used in the CDA. Chapter 4 discusses the approach used to impute these variables.

Refining Fuel and System Types

Previous CDA studies conducted on the California residential population have shown that the misreporting of fuels used for space heating and water heating was common. Since space and water heating account for large shares of household energy consumption, the variables used in the CDA needed to accurately reflect the fuel type in the household for the results of the CDA to be accurate. Chapter 4 discusses the approach to fuel checking and imputing values.

Estimated UECs

The household and end-use UECs from the CDA were appended to the cleaned survey data. Normalized annual household electric and gas consumption variables were also added to the database.

Database 3: Billing Data Database

The study team conducted basic quality control checks on the electric and gas consumption data as they served as a validation of the interval data for the utilities that provided interval data. For LADWP, the billing data were used in the CDA, as they did not provide interval data. The variables contained in the Billing Data file set are listed in the Appendices Volume, Appendix M.

Database 4: Interval Data and Degree-Day Normalization Data File Set

The study team conducted quality control checks on the electric and gas interval data before performing the degree-day normalization (DDN). Since the weather-normalized annual usage was calculated independently for electric and gas consumption, the data were stored in two separate files. Chapter 4 discussed the DDN. The normalized annual consumption files contained the pre-normalized and normalized annual consumption. The DDN process used heating degree days and cooling degree days along with the degree day set points for each household, so these variables were also included in the files. The variables contained in the Interval Data and Normalization file set are listed in the Appendices Volume, Appendix N.

Database 5: Electric Load Profiles and Gas Daily Use File Set

These datasets contain estimates of average energy use at the hourly (electric) and daily (gas) levels that were developed using the interval data provided for the RASS for the 13 months from October 1, 2018, to November 30, 2019.

The study team developed the estimates using weights that were adjusted ("raked") separately for electric and gas. Using the weights for the entire set of survey respondents as a starting point, the weights were adjusted to account for the explicit and implicit stratification variables and a reduced sample size (the number of survey respondents for which the study team had interval data is lower than the number of survey respondents). Chapter 2 explains this process.

The electric load profiles dataset includes one record per day for individual strata, utility, and all utilities combined. The gas daily use dataset includes one record per day for the individual strata, the utility, and all utilities combined. The variables contained in the Electric Load Shapes and the Gas Daily Use files are listed in the Appendices Volume, Appendix O.

Database Formats

The volume of data generated by the RASS study demanded the use of software with the capacity to manipulate large datasets and the ability to support the analyses required by the study. The study team used the SAS software package from the SAS Institute to analyze the RASS data. All the survey data, billing data, interval data, and weather data were stored as SAS datasets and analyzed from within the SAS environment. The description of the SAS files and code is contained in the Appendices Volume, Appendix P.

The final databases were provided in two file formats: a SAS dataset format and a comma delimited (.csv) format. The .csv format facilitated importing the data into other software packages.

The study team also updated the web interface from the *2009 RASS* study that allows users to design their own queries to create reports directly from the RASS survey data. The website provides the ability to compare RASS results between the 2019 and 2009 or 2003 datasets and facilitates queries based on grouping by multiple variables.

Data Delivery

The RASS data were delivered via the DNV GL secure file transfer platform (Accellion) to the CEC and participating utilities. The files transferred are listed below.

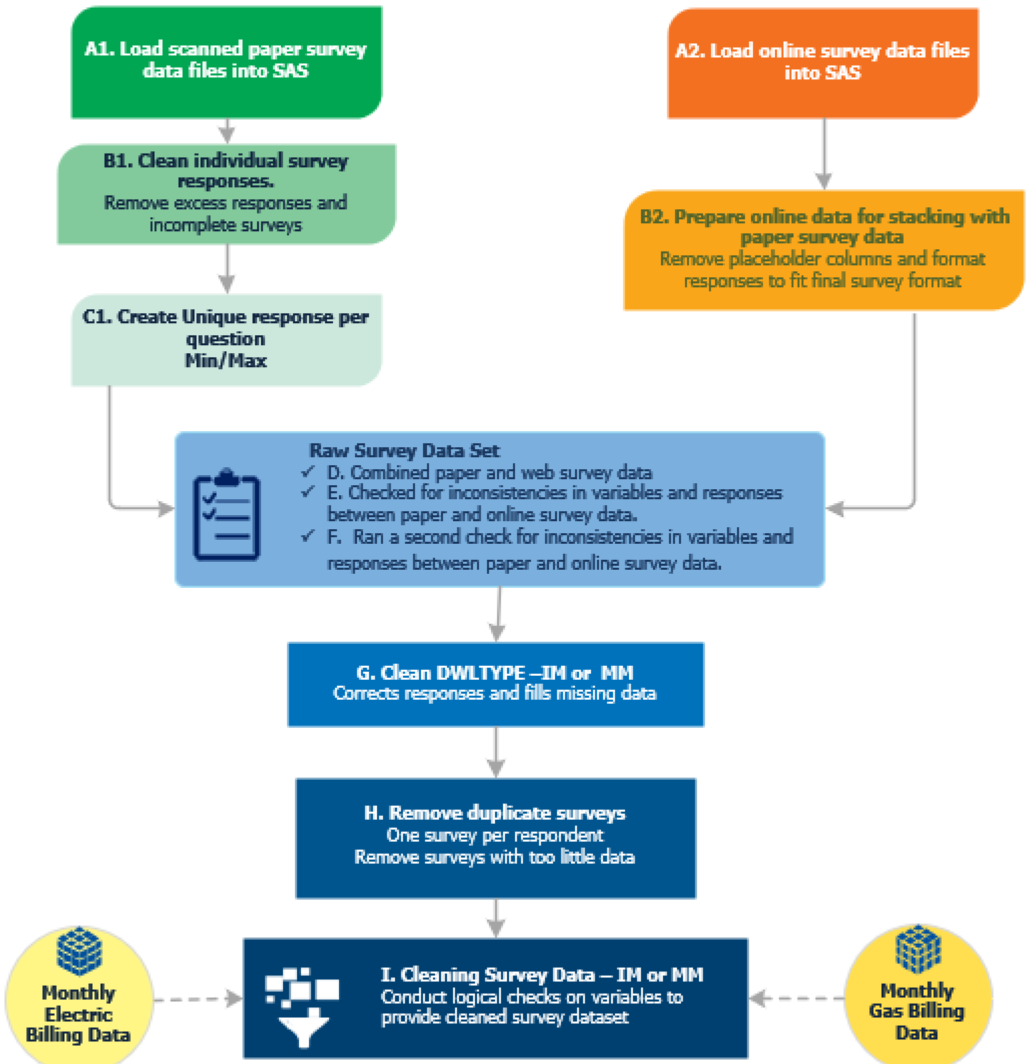
- Raw RASS survey database (.sas7bdat, .csv) and contents file (.xlsx)
- Cleaned RASS survey data and CDA database, including files with and without formats applied (.sas7bdat, .csv), list of formats (.xlsx) and commands to apply formats (SAS), and contents of data file (.xlsx)
- Billing database of monthly electric and gas billing data (.sas7bdat, .csv) and contents file (.xlsx)
- Interval data and Degree-Day Normalization (DDN) file set including electric hourly, daily and annual interval data (.sas7bdat, .csv), gas daily and annual data (.sas7bdat, .csv) and contents for all data files (.xlsx)
- Electric Load Profiles and Gas Daily Use file set including electric load shapes (.sas7bdat, .csv), gas daily use (.sas7bdat, .csv) and contents for each data file (.xlsx)

CHAPTER 4: Data Cleaning and Processing

Overview

The section outlines the processing steps applied to the survey data to ensure the data used to develop estimates were as accurate as possible. These steps included eliminating surveys that were determined to have excessive amounts of invalid data, cleaning RASS Survey variables, and creating new variables through the cleaning process and the combination of survey variables. **Figure 4** provides an overview of the general data cleaning process.

Figure 4: Overview of Data Cleaning Process



Source: 2019 California Residential Appliance Saturation Survey

Initial Survey Processing

Survey Processing

As surveys were received, the barcodes containing the access code were scanned to register the surveys received and queue them for processing. The study team scanned the paper survey forms using optical mark reading software to record the darkened circles on the forms. The responses were saved as text files, and images of the surveys were saved as PDF files. These text files were read into SAS data files such that each bubble response was reflected as a distinct variable.

Excess Responses

This step is illustrated in **Figure 4** as B1. The program *TooManyResponses.SAS* was run on the initial survey datasets. The program counted the number of questions with excess responses (in other words, more responses were given than the question requested). Surveys with 15 or more excess responses were flagged to be deleted. The number of questions with excess responses was recorded and passed along to the next program MIN_MAX.SAS and was contained in the output dataset: *Min_Max_output.sas7bdat.* The number of questions with excess responses was later used to assist in eliminating duplicate surveys. The process used to eliminate duplicate surveys is described in the section that describes cleaning individual survey questions.

Paper Survey Cleaning

As shown in C1 of **Figure 4**, the SAS program min_max.sas created a single SAS dataset that contained all scanned files and performed the following functions:

- Identified blank surveys — Surveys with responses missing for all questions were identified.
- Prioritization of multiple responses for each question (F of **Figure 4**) — The program selected a single response for questions required a single answer, but the respondent provided multiple answers. For most questions, a unique response was inferred based on a set of predefined criteria for each question that picked either the minimum or maximum response category for that question. For certain survey questions, however, the mean response was used in place of the minimum or maximum response category. This process resulted in the SAS data set Min_MAX_Output.sas7bdat.

Incomplete Surveys

This step is shown in B1 of **Figure 4**. The *Too_Many_Responses.SAS* code also contained a five-step process that identified incomplete surveys. This section outlines that process.

The first step in identifying incomplete surveys was to check a set of 15 variables and 5 composite indicator variables for missing values. The variables were selected to represent the beginning, middle, and end of the survey. The 15 variables listed in **Table 17** were selected as they were not subject to any skip patterns.

Table 17: Survey Variables Used to Identify Incomplete Surveys

Survey Section	Survey Variable	Description
A	DWLTYPE	Dwelling type
A	OWNRENT	Own or rent
A	YRS_RES	Years of residence
A	BUILTYR	Year home built
A	NUMROOM	Number of rooms
A	SQFT	Square footage
B	PAYHEAT	Pays for heat
C	PAYCOOL	Pays for cooling
D	PAYWH	Pays for hot water
E	LNDRYEQP	Laundry equipment in home
G	RFNUM	Number of refrigerators
I	SPTYP	Spa type
K	WORKHOME	Person works at home
M	WLWTRPMP	Well water pump
N	INCOME19	Household income

Source: 2019 California Residential Appliance Saturation Survey

The five composite indicator variables, shown in **Table 18**, were each based on a set of survey variables that represented either related survey questions or subcategories within a question. An example of the latter is the composite variable used to indicate missing values for both subcategories from the question on number of home computers, where the survey asks separately for the number of laptops and number of desktops.

Each composite variable considered the joint responses to the set of variables that alone could be missing, but the collection of the variables making up the composite indicator variable would constitute a missing response. If a respondent had missing values for each of the variables in the composite group, the composite indicator variable was coded as one. If at least one of the variables that made up the composite variable contained a nonmissing value, then the composite variable was coded as zero to indicate there was information for at least one variable in the group.

The team used similar logic to construct composite variables for the presence of natural gas service, usage of various cooking appliances, number of exterior lighting fixtures, and presence of miscellaneous appliances.

Table 18: Composite Indicator Variables Used to Identify Incomplete Surveys

Survey Section	Composite Variable	Survey Variables	Description
A	NGMISS	NGSERV	Natural gas service in area
A	NGMISS	NGLINE	Natural gas line to home
F	COOKMISS	WRNUSE	Weekly oven range stovetop use
F	COOKMISS	WOVUSE	Weekly oven use
F	COOKMISS	WMWUSE	Weekly microwave oven use
K	NPCSMISS	NDSKPCS	Number desktop PCs
K	NPCSMISS	NLASPPCS	Number laptop PCs
L	EXLIGHTMISS	EXINC	Number of exterior incandescent fixtures
L	EXLIGHTMISS	EXLED	Number of exterior LED fixtures
L	EXLIGHTMISS	EXCFL	Number of exterior CFL fixtures
L	EXLIGHTMISS	EXLOWWV	Number of exterior low voltage light systems
L	EXLIGHTMISS	EXHID	Number of exterior HID fixtures
M	M1MISS	CHRGRS	Number of plug-in chargers
M	M1MISS	RECHVAC	Number of rechargeable vacuums
M	M1MISS	SMACOOK	Number of small cooking appliances
M	M1MISS	FNPORT	Number of portable fans
M	M1MISS	FNCEIL	Number of ceiling fans
M	M1MISS	WNDATV	Number of wind turbine attic ventilators
M	M1MISS	FNATTIC	Number of attic fans
M	M1MISS	FNWHOLE	Number of whole-house fans
M	M1MISS	AIRCLEAN	Number of air cleaners
M	M1MISS	PORTPUR	Number of portable room air purifiers
M	M1MISS	HUMDEH	Number of humidifiers or dehumidifiers
M	M1MISS	WINCLR	Number of wine or beverage coolers
M	M1MISS	WHPURIFY	Number of water purification systems
M	M1MISS	DHWRPMP	Number of domestic hot water recirculating pumps
M	M1MISS	SAUNA	Number of saunas
M	M1MISS	POND	Number of pond or garden pumps
M	M1MISS	FIREPIT	Number of outdoor fire tables/fire pits

Source: 2019 California Residential Appliance Saturation Survey

The second and third steps in identifying incomplete surveys considered whether large portions of the survey were left blank. Responses to 18 questions (A1 — A18) from the Home and Lifestyle questions were checked to determine whether all questions were skipped. Any survey in which this entire section was left blank was flagged. Similarly, responses to the Laundry, Food Preparation, and Refrigerator sections (questions E1 — G2) were checked to determine if all responses were missing, and surveys in which all responses were missing were flagged.

The final step of this process was to select surveys for deletion if 10 or more of the variables from the 2 groups were missing or both sections reviewed in steps 3 and 4 were entirely missing. The number of questions with missing responses was also recorded in the *Min_Max_output.sas7bdat* dataset to assist in eliminating duplicate surveys from the same household once the cleaning of survey questions was complete.

Web Survey Processing

The study team downloaded survey responses as csv files from the web survey portal. These csv files were read into SAS datasets and mapped to the appropriate SAS variable. Differences between the web survey response collection and paper survey response collection required additional formatting (B2 of **Figure 4**).

Web Survey Cleaning

The nature of the web form used for collecting web-based survey data reduced the number of steps for formatting. For example, the web forms presented to the respondent control for question responses, such that the respondent cannot accidentally answer survey questions with too many responses. Given how the numeric responses were stored, additional postcoding steps were taken to map the web survey responses and paper-survey responses in an equivalent manner.

Several questions in the web format presented the respondent with an additional response when compared to the paper survey. For instance, Question J3 (How many hours per day do you operate your swimming pool filter?) presented four response options, whereas the paper survey presented three response options. Adjustments were applied to these questions to align the responses from the web survey with the responses in the paper survey.

Combined Survey Dataset

Once paper-survey-specific and web-survey-specific formatting and cleaning were complete, the datasets were stacked and programmatically checked for variable inconsistencies and response inconsistencies (D and E in **Figure 4**). If the team found inconsistencies, code updates were made in the previous steps to align the data, and the stacked survey data were rechecked for variable inconsistencies and response inconsistencies.

Once the paper and web survey data were successfully stacked, multiple surveys with the same unique identifier were identified (H of **Figure 4**). Duplicate surveys were programmatically checked for survey completeness as previously described for paper survey cleaning. The surveys with the most complete data for each unique identifier were carried through the cleaning process, while less complete surveys were removed.

Cleaning Individual Survey Questions

Figure 4 shows that the SAS program *Clean_Sample.sas* combined the unique survey responses (I of **Figure 4**) with monthly electric and gas billing data. This program was used to clean the survey questions, which consisted of the following steps, all of which are described in detail within this chapter:

- Refined Fuel System Types: The survey data set was combined with monthly electric and gas billing data to identify households in which fuel used for heating and water heat was misreported.
- Identified Year-Round Residents: As in the 2009 study, the current CDA restricts the analysis to year-round residents. Monthly electric data were used to identify partial year residents and remove them from the CDA and saturation estimates.
- Coded Nonresponse and Not Applicable Response: The cleaning process distinguished between nonresponses that resulted from the intentional skip pattern in the survey and

questions in which the response was left blank. The former was coded as 99, meaning not applicable, while truly missing responses were coded as 97.

- Determined Logical Response Inconsistencies: Many survey questions were interrelated, requiring responses to be logically consistent. For example, if a respondent indicated that he or she did not have natural gas service in the respective area, it would not be consistent for him or her to have a gas line to the home. Where possible, logically inconsistent responses were corrected using information contained in billing data or in other survey responses. In cases where a value could not be inferred, the response was set to 97 to reflect a missing value.
- The number of logical inconsistencies in each survey was counted using a cumulative flag that added 1 for each occurrence. The number of inconsistencies was used to identify surveys that contained too many errors to include in the CDA.
- Imputed Missing Values: Although missing survey values were recorded as 97 in the cleaned survey data set, retaining these as missing values in the CDA would result in nonresponse bias. Therefore, an approach was developed to impute missing values for all variables used in the CDA.

Invalid Surveys

Based on the criteria identified above, surveys were flagged if the number of excess, logically inconsistent, or missing responses exceeded the prescribed limit for the acceptable number of errors. The study team removed these surveys from the survey dataset used in the CDA and saturation tables.

Table 19 presents the number of surveys removed from the dataset according to the reasoning discussed above.

Table 19: Summary of Invalid Surveys

Reasons for Eliminating Survey	Number Eliminated
Too many multiple responses	7
Incomplete survey	319
Too many logical inconsistencies	2

Source: 2019 California Residential Appliance Saturation Survey

Survey-Specific Cleaning

This section describes the logic used to identify illogical responses and clean survey questions. The section is organized in the same order as the survey, presented in Appendices Volume, Appendix A, which is divided into the following sections:

- Your Home and Lifestyle
- Electric Vehicles
- Space Heating
- Space Cooling
- Water Heating
- Laundry

- Food Preparation
- Refrigerators
- Freezers
- Spas and Hot Tubs
- Pools
- Entertainment and Technology
- Lighting
- Miscellaneous Appliances
- Renewable Energy Technologies
- Household Information

Your Home and Lifestyle

The Your Home and Lifestyle section of the survey contained 21 questions, many of which are critical for other data cleanings and the CDA estimates. The process used to clean these variables is discussed below. Cleaning procedures used for some variables required cross references with other survey variables or billing data, or both. Cross references are clearly delineated below.

Type of Dwelling

The process used to clean the dwelling type variable (A1-DWLTYPE) is presented in detail because this variable is used extensively to estimate imputed values for other survey variables and serves as a key explanatory variable in the CDA process.

The original survey response values for the "type of building" (A1–DWLTYPE) included the following:

- 1 is a single-family detached house
- 2 is a townhouse
- 3 is a 2-4 unit apartment or condominium
- 4 is a 5+ unit apartment or condominium
- 5 is a mobile home
- 6 is other

Cleaning of the DWLTYPE variable addressed missing, inconsistent, and ambiguous responses. First, survey respondents that did not provide an answer to this question were coded 97, to reflect a missing value. Second, DWLTYPE was checked against several other survey questions to see if they contradicted each other. Third, attempts were made to match respondents who answered 6 (Other) to the DWLTYPE question to a less ambiguous response category.

Individually Metered Surveys

The variable RESIDENCE was created to reflect each household's corrected dwelling type. If there was no problem with the original DWLTYPE response, the original value for DWLTYPE was retained as RESIDENCE.

The process of creating the RESIDENCE variable used the following information:

- Survey responses to DWLTYPE, payment of heating, cooling, water heating, laundry systems, and square footage.
- Residence type code provided by each participating electric utility for the sample frame dataset.
- Household's service street address.

The dwell variable provided by the utility differentiated between single-family and multifamily homes. As such, the dwell variable was used as the RESTYPE in the cleaning algorithm. This resulted in two RESTYPE values: SF and MF.

The RESIDENCE variable for individually metered households was defined according to the rules outlined below.

- If DWLTYPE was equal to 2, 3, or 4 and the utility's RESTYPE code was SF, RESIDENCE equaled the individual's response for DWLTYPE. In this situation the survey response overrides the utility's RESTYPE code.
- If DWLTYPE equaled to 1 and the utility's RESTYPE was MF (indicating that it is a multifamily residence), the algorithm proceeded through the following checks:
  - Reviewed the service address. If address ended in a number 1—-4 or the letter A, B, C, or D, set RESIDENCE to 3.
  - Reviewed the service address. If the address ended in a number larger than 4 or a letter later than D, set RESIDENCE to 4.
  - Reviewed the service address. If the service address did not end in a letter or a number, checked if the respondent paid for a major system and if the survey response to square footage was less than 2,500.
  - If both checks were satisfied, then set RESIDENCE to 2.
  - If none of the above conditions was met, set RESIDENCE to 1.
- If DWLTYPE was equal to 6 and the utility's RESTYPE contained MF, the algorithm proceeded through the following checks:
  - Reviewed the service address. If the address ended in a number 1-4 or a letter A-D, set RESIDENCE to 3.
  - Reviewed the service address. If the address ended in a number larger than 4 or a letter later than D, set RESIDENCE to 4.
  - If the address did not end in a number or a letter and the survey response to square footage was greater than or equal to 2,500, set RESIDENCE to 1.
  - If the address did not end in a number or a letter and the survey response to square footage was less than 2,500, set RESIDENCE to 2.
- If DWLTYPE equaled 97 and RESTYPE contained SF, the algorithm proceeded through the following checks:
  - Reviewed the service address. If the address ended in a number 1-4 or letter A-D, set RESIDENCE to 3.

  - Reviewed the service address. If the address ended in a number larger than 4 or a letter later than D, set RESIDENCE to 4.
  - If the address did not end in a number or a letter, set RESIDENCE to 1.
- If DWLTYPE equaled 97 and RESTYPE contained MF, the algorithm proceeded through the following checks:
  - Reviewed the service address. If the address ended in a number 1-4 or a letter A-D, set RESIDENCE to 3.
  - Reviewed the service address. If the address ended in a number larger than 4 or a letter later than D, set RESIDENCE to 4.
  - If the address did not end in a number or a letter, set RESIDENCE to 2.

This process resulted in RESIDENCE values of 1, 2, 3, 4 or 5 for all households. Out of 39,682 individually metered survey respondents, 3,301 contained a RESIDENCE different from what was reported by the respondent in DWLTYPE. For these 3,301 households, 3,080 were changed from DWLTYPE=Other (6) or No Response (97 or missing). **Table 20** presents all such changes by DWLTYPE, reason for change, RESIDENCE and RESTYPE.

Table 20 shows that DWLTYPE was missing (No Response) for a total across all residence types of 2,617 survey respondents. From this total, 1,919 households were classified as single-family detached houses. Another 698 dwellings were reassigned to apartments, condominiums, townhouses, or duplexes from missing.

Master-Metered Surveys

The field RESIDENCE, the updated dwelling type of the master-metered respondents, was created similarly to the individually metered respondents. DWLTYPE recorded the responses of the master-metered survey participants, and RESTYPE was provided by the utilities. For this study, RESTYPE for the master-metered units contained only 2- to 4-unit duplex, triplex, or quadplex homes (RESTYPE = 1).

In some cases, DWLTYPE contained missing values or the value OTHER. Inconsistencies were found between DWLTYPE and RESTYPE in some other cases. For all these cases, the team assumed that the information provided by the utility, contained in field RESTYPE, was correct.

In particular, the following types of inconsistencies were found between DWLTYPE and RESTYPE:

- DWLTYPE=1 and RESTYPE=1
- DWLTYPE=3 or 4 and RESTYPE=1
- DWLTYPE=5 and RESTYPE=1

The cleaning code for master-metered respondents assumed that the values stored in RESTYPE were correct. Since DWLTYPE and RESTYPE could not be mapped perfectly, the following rule was adopted:

- If RESTYPE=1 then RESIDENCE=2.

Table 20: Dwelling Type Cleaning Results for Individually Metered Households

DWLTYPE	Reason for Change	RESIDENCE	RESTYPE Multifamily	RESTYPE Single-Family	RESTYPE Total
Single-Family Detached House	Landlord pays for at least one major system and SQFT =< 2,500	Townhouse, Duplex, or Rowhouse	33	0	33
Single-Family Detached House	Service address ends in a number or a letter	Apartment or Condominium, 2-4 Units	75	0	75
Single-Family Detached House	Service address ends in a number or a letter	Apartment or condominium, 5+ Units	113	0	113
Other	Address does NOT end in a number or a letter	Single-Family Detached House	0	239	239
Other	Landlord pays for at least one major system and SQFT =< 2,500	Townhouse, Duplex, or Rowhouse	78	0	78
Other	Service address ends in a number or a letter	Apartment or condominium, 2-4 Units	27	14	41
Other	Service address ends in a number or a letter	Apartment or condominium, 5+ Units	71	34	105
No Response	Address does NOT end in a number or a letter	Single-Family Detached House	0	1,919	1,919
No Response	Address does NOT end in a number or a letter	Townhouse, Duplex, or Rowhouse	151	0	151
No Response	Address does NOT end in a number or a letter	Apartment or condominium, 2-4 Units	110	24	134
No Response	Address does NOT end in a number or a letter	Apartment or condominium, 5+ Units	330	83	413
TOTALS			**988**	**2,313**	**3,301**

Source: 2019 California Residential Appliance Saturation Survey

Table 21 provides the counts of DWLTYPE by RESIDENCE for all surveys where DWLTYPE differed from RESIDENCE. There were 283 master-metered surveys where DWLTYPE differed from RESIDENCE. For example, there were 192 surveys that reported DWLTYPE="Single-Family" that were changed in RESIDENCE to townhouse, duplex, or row houses

Table 21: Dwelling Type Cleaning Results for Master-Metered Households

DWLTYPE	Townhouse, Duplex, or Rowhouse RESTYPE=1
Single-Family	192
Apartment or Condominium, 2-4 Units	24
Apartment or Condominium, 5+ Units	7
Mobile Home	2
Other	26
No Response	32
Total	**283**

Source: 2019 California Residential Appliance Saturation Survey

When no inconsistencies between DWLTYPE and RESTYPE were found, the survey response of DWLTYPE was carried over to the RESIDENCE variable. This process resulted in imputed values for the RESIDENCE variable for all valid surveys.

Cleaning Procedures for Questions A3 – A19

Table 22 summarizes the allocation of missing responses for nine of the major questions in the Home and Lifestyle section before and after the cleaning process. **Table 22** shows that some variables that were missing before cleaning were assigned values during the process, while others were set to "not applicable." In addition, some survey responses were found to be logically inconsistent with other responses or utility-provided information and changed to missing, not applicable, or reallocated to a new response. The procedures used to cross-reference and clean these variables are discussed below.

Table 22: Cleaning Results for Missing Home and Lifestyle Responses

Home & Lifestyle (A2 - A19)	Missing Post Cleaning	Not Applicable Post Cleaning	Re-Assigned Post Cleaning	Missing Post Cleaning	Not Applicable Post Cleaning	Re-Assigned Post Cleaning	Unchanged Post Cleaning
Own or rent home (OWNRENT)	1,392	0	0	0	0	0	38,593
How long at address (YRS_RES)	1,271	0	0	0	0	0	38,714
Seasonal Occupancy (SEASOCC)	16	0	1,397	0	0	0	38,572
Year home built (BUILTYR)	3,052	0	0	540	0	0	36,393
Number of bedrooms (NUMROOM)	1,353	0	0	1,065	0	37,567	0
How many square feet of living space (SQFT)	3,147	0	0	337	0	0	36,501
Attic/Ceiling Insulation (ACEILINS)	1,927	0	484	0	0	28	37,546
Ceiling Insulation (CEILINCH)	3,415	17,639	0	0	0	0	18,931
Remodeling (REMOD)	2,048	0	210	0	0	39	37,688
Is natural gas available (NGSERV)	1,416	0	1,826	0	0	1,093	35,650
Natural gas hookup in home (NGLINE)	1,359	3,608	2,907	3	1,257	1,063	29,788

Source: 2019 California Residential Appliance Saturation Survey

A3 – Years Respondent Lived in Home (YRS_RES)

Responses to Questions A3 and A6 indicated how long a respondent has lived in the home (YRS_RES) and the year that dwelling was built (BUILTYR). If the response to YRS_RES was greater than the age of the dwelling as indicated by BUILTYR, both variables were set to missing. In addition, the BUILTYR variable was cross-referenced with the age of the primary heating and water systems, HTSAGE and PRWHAGE respectively. If the age of the dwelling was less than the age of these systems, then BUILTYR and the respective system ages were set to missing.

A4/A5 – Seasonal Occupancy (SEASOCC) and (SEASJAN – SEASDEC)

As in the 2009 study, the CDA in this study was estimated using only dwellings occupied year-round. (See Chapter 5 for a detailed discussion of the CDA process.) Respondents were asked whether the home was occupied on a year-round or seasonal basis or serves as a vacation home (A4–SEASOCC). Cleaning this variable was a two-step process.

The first step of this process checked responses for SEASOCC against survey variables reporting the months a respondent indicated living in the residence (A5 — SEASJAN — SEASDEC). The following rules were used for this cross-reference:

- SEASOCC was set to 4 (vacation or rental home) for dwellings occupied for two or less months.
- SEASOCC was set to 2 (partial-year or seasonal residence) for dwellings occupied for 3–11 months.
- SEASOCC was set to 1 (year-round residence all months were left blank).

In the second step, responses for SEASOCC were checked against electric billing records to determine whether there was electric consumption for each month of the year. This check was performed on surveys where the response to SEASOCC was missing. If billing records had two or fewer consecutive months of missing or zero electric data, the missing response was updated to year-round occupancy.

A6 – Year Dwelling Was Built (BUILTYEAR)

Responses for A6 (BUILTYR) were cleaned using the same logic as A3 (YRS_RES). The variable BUILTYR was used to construct the variables AGEHOME, a continuous variable for the age of the dwelling, and NEWHOME, an indicator variable for new construction. The imputation of these variables is discussed further in the CDA data imputation section later in this chapter.

A7/A8 – Number of Bedrooms (NUMROOM) and Square Feet of Living Space (SQFT)

For a given dwelling type (DWLTYPE), the number of bedrooms (NUMROOM) was assumed to be constrained by the square footage (SQFT). The rules used to determine the logical consistency of these three variables are presented below. In cases in which the rules were violated, both the square footage and the number of bedrooms were set to missing.

- Single-Family Dwellings (DWLTYPE =1) — Less than 2,000 square feet with more than 8 rooms; or less than 250 square feet.
- Townhouses (DWLTYPE = 2) — Less than 2,000 square feet with more than 8 rooms; or less than 250 square feet with more than 1 room.
- Apartments (DWLTYPE = 3,4) — Less than 1,500 square feet with 4 or more rooms; or less than 250 square feet with more than 1 room.
- Mobile Homes (DWLTYPE = 5) — Less than 1,500 square feet with 4 or more rooms; more than 5 rooms; or less than 250 square feet with more than 1 room.

In addition, the SQFT variable was used to derive the continuous variable (SQFT_A). The SQFT_A variable is a continuous variable derived from the SQFT. These variables were used in the conditional demand analysis model, which required missing values to be imputed. The imputation of these variables is discussed further the CDA data imputation section later in this chapter.

A9 – Exterior Walls (EXTWLINS)

Responses to EXTWLINS were unchanged.

A10/A11 – Attic/Ceiling Insulation (ACEILINS and CEILINCH)

The variable ACEILINS, whether the attic or ceiling is insulated, was cross-referenced with CEILINCH, number of inches of insulation in the attic or ceiling. If the response to ACEILINS was either "no" or missing, but they provided the number of inches (that is, CEILIINCH was not missing), then the response for ACEILINS was changed to "yes."

A13/A14 – Remodeling (REMOD) and Type of Remodeling

Respondents were asked whether the home has been remodeled in the past 12 months (A13 – REMOD) and then asked to indicate the type of remodeling (A14). If a respondent skipped or answered "No" to A13 but indicated that a type of remodeling in A14, the response to A13 was changed to "yes."

A15 – Number of Occupants by Age Group (NR0-5, NR6-18, NR19_34, NR35-54, NR55-64, NR65-99)

The survey requested respondents to identify the number of individuals residing in the household according to six age groups. Although response categories included a "ZERO" option, it is common for respondents to simply skip age groups that do not apply to their household. Therefore, the following criteria were used to distinguish between skipped responses that are not relevant and those that did not respond to the set of questions:

- If a respondent skipped all questions pertaining to the number of residents by age group, then all values were set to 97 or missing.
- If at least one category was filled out, then the age groups that did not have a response were set to zero.
- The total number of residents was also set to missing if all age groups were missing or zero.

The following variables were created during the cleaning process to be used in the CDA and cross-tabulations of survey responses:

- Number of people living in the household (RESCNT)
- Number of people living in the household over 65 (SENIORS)
- Number of people living in the household under 19 (KIDS)
- Number of people living in the household 19–64 (ADULTS)

Missing values of the RESCNT variable were imputed for the CDA analysis, creating the new variable NUMI, which will be discussed in the CDA data imputation section later in this chapter.

A18 – Natural Gas Availability

Responses to natural gas service being available in their area (NGSERV) that were either missing or reported as "no" were changed to "yes" if the cleaning process used for question A19 (NGLINE) indicated that they had a natural gas line to the home.

A19 – Natural Gas Hookup in the Home (NGLINE)

Where possible, the presence of a natural gas line of the residence (A19 NGLINE) was checked against billing information sent by the three gas utilities based on the following:

- If the respondent indicated he or she did not have a gas line to the home but was found to have gas billing records, then the response was changed to yes.
- If the respondent reported gas service provided by one of the smaller utilities, responses to the number of natural and bottled gas appliances were used to confirm the presence of a natural gas line to the residence.
- The new variable NGLINE1 recorded the corrected response to NGLINE after verifying the survey response.

Additional Cleaning of Your-Home-and-Lifestyle Variables

A1 (subset) — Number of Stories (STORIES)

Respondents who live in single-family homes (A1-DWLTYPE= 1) were instructed to answer this question, while those living in all other dwelling types were instructed to skip it. If a respondent from one of the other dwelling types provided an answer to STORIES, the response was changed to 99 (Not Applicable).

A2 (subset) — Own or Rent Dwelling (OWNRENT)

Responses to the OWNRENT question are unchanged.

Electric Vehicles

This section covers the procedures used to clean and ensure consistency in responses in Section A – Electric Vehicles.

A20 — Currently Own/Lease an Electric Vehicle (*EVYN*)

If this question was missing or marked "No," the subsequent questions (A21 through A25) were checked for information. If the subsequent questions contained responses, the response to EVYN was updated to "Yes." If this question was marked "No" and the subsequent questions were left blank, then the subsequent questions were set to "Not Applicable." Following these steps, if A20 was still missing, the answers to this question and subsequent questions were set to "No Response."

Because the electric vehicle variables were used to construct engineering estimates of monthly kWh consumption used in the CDA, missing values were also imputed. The data imputation and the engineering estimates are described in the CDA section later in this chapter.

Space Heating

This section covers the procedures used to eliminate survey multiple responses and inconsistencies in responses in Section B – Space Heating of the survey. The cleaning process also revealed substantial fuel misreporting. Fuel misreporting is reviewed in the CDA variables section later in this chapter, which also discusses additional primary space heating system variables that were derived for the CDA model and data imputation.

B1 – Pay for Heat (*PAYHEAT*)

The question concerning how a household pays for heat (PAYHEAT) was critical to the process used to clean the remaining heating questions, and results are presented in **Table 23**. The following cross-references were used to evaluate logical inconsistencies and make corrections wherever possible.

Only households indicating that they pay for heat directly were asked to fill out the majority of the heating questions, while those who either indicated heat was included in the rent or that they do not have a heating system were asked to answer only questions concerning portable electric heaters. If information was provided for at least one heating system and PAYHEAT was either no or no – included in their rent, then a new variable PAYHEAT1 recorded the response as yes. The original PAYHEAT variable was preserved by the original pay for heat response.

For cases in which multiple responses were provided, the lowest numbered response was kept. This logic favored "yes — pay for heat" over "no — it is part of my rent/condo fee." Similarly, "no — included in rent" was chosen over "no — do not have heating system."

If a respondent indicated that he or she does pay for heat but did not list any heating systems, or PAYHEAT was missing and he or she did not list any heating systems, then system variables were set to missing (97). If a respondent did not pay for heat, then all heating system variables were set to not applicable (99).

Table 23: Heating Payment Question Cleaning

PAYHEAT	Total	% Total
Yes. — Pay for Heat	35,581	89.0%
No. — Included in Rent/Fee	780	2.0%
No. — Do not Have Heating	1,542	3.9%
Missing	2,082	5.2%
All	**39,985**	**100.0%**

Source: 2019 California Residential Appliance Saturation Survey

B2 — Type of Primary and Secondary Heating Systems

Natural Gas Heating Systems

If a respondent indicated having a primary or secondary natural gas heating system, the cleaned variable for a natural gas line to the home (NGLINE) was cross-referenced. As discussed above, this confirmed the heating system was consistent with the survey response for NGLINE as well as billing information. If a dwelling did not have gas service but indicated a natural gas system, the system response was set to missing (97). In addition, if a respondent reported having natural gas radiators, but either the radiators were not the primary heating system, or there was also a forced hot air system, then the radiators were not included as a heating system.

Electric Heating Systems

Survey responses for electric heating systems were checked to determine whether central heat pump heating and central forced air heating were indicated. If a respondent indicated having both types of systems, then the heat pump was selected as the primary heating system, and the central forced air system was set to missing. The survey allowed for up to six primary and six secondary electric heating systems. If a respondent reported having five or more electric space heaters, then all were set to missing.

Propane Heating Systems

If a respondent had natural gas in the home and indicated having propane heat, the propane systems were set to missing.

Other Heating Systems

The number of "other" space heating systems was restricted to two systems. If respondents indicated having more than two "other" systems, then all systems were set to missing.

Primary and Auxiliary Heating Fuels

If a respondent provided more than one primary heating system, then the first system selected was set as the primary system, and the subsequent responses were assigned to auxiliary heat. If only additional heating systems were provided, then the primary heat was set to the first additional heating system indicated.

The new variables primary heating fuel (PHTFUEL) and auxiliary heating fuel (AHTFUEL) were derived from the primary and additional heating system information. If the respondent indicated he or she had portable electric heaters and did not have natural gas auxiliary heat, then AHTFUEL was set to 2 for electric heat. The coding used for PHTFUEL and AHTFUEL is summarized in **Table 24**.

Table 24: Primary Heating Fuel Data Cleaning

PHTFUEL	Code	Total	% Total
Natural Gas	1	25,956	64.9%
Electric	2	7,183	18.0%
Bottled Gas	3	1,322	3.3%
Wood	4	724	1.8%
Solar	5	4	0.0%
Other	6	76	0.2%
Missing (Respondent failed to answer question)	97	2,857	7.2%
Not Applicable (Respondent does not pay for heat or does not have a heating system)	99	1,863	4.7%
All		**39,985**	**100.0%**

Source: 2019 California Residential Appliance Saturation Survey

B3 — Pilot Light for Primary and Secondary Natural Gas Heating System (*MAINPLT* and *SECPILT*)

The survey requested information on the use of pilot lights for primary and secondary natural gas heating systems, MAINPILT and SECPILT, respectively. If a respondent did not report having a natural gas system, then the MAINPILT and SECPILT were set to not applicable (that is, 99). For respondents who had either a primary or secondary natural gas system and no response was given for MAINPILT or SECPIL, respectively, then the value was set to missing (97).

B6 — Heat Temperature Setting

Responses to the heating temperature settings were cross-checked with information concerning how the household pays for heat, whether it has a heating system, and whether it skipped all temperature settings. Rules used to clean this section include the following:

- If respondents skipped PAYHEAT or indicated paying for heat but provided no indication regarding the type of heating system, then all temperature settings were set to missing (97).
- All temperature settings were set to Not Applicable (99) if respondents did not pay for heat, or if it is included in their rent, or if they do not have a thermostat.
- For respondents with a thermostat, temperature settings were evaluated to determine whether respondents answered a setting for at least one time of day. If they provided at least one setting, then all missing temperatures were set to off. If they did not provide any settings, then all were set to missing (97).

Space Cooling

This section covers the procedure used to eliminate survey multiple responses and inconsistencies in responses to survey Section B – Space Cooling. The space cooling section recorded information concerning central air conditioning and room air conditioning. Respondents were first asked how they pay for central air conditioning. Those who either did not pay for cooling or indicated it was included in their rent were asked to skip to the room air conditioning section.

C1 — Pay for Cooling

How a household pays for cooling (PAYCOOL) was evaluated similarly to the PAYHEAT variable and is summarized in **Table 25**. For households that did not indicate they pay for cooling but provided information on cooling systems, a new variable PAYCOOL1 recorded the response as "yes." For cases in which multiple responses were provided, the lowest numbered response was kept. If respondents indicated that they do pay for cooling but did not list any central cooling systems, or PAYCOOL was missing and they did not list any cooling systems, then system variables were set to missing (97). If a respondent did not pay for cooling, then all cooling system variables were set to not applicable (99).

Table 25: Cooling Payment Question Cleaning

PAYCOOL	Code	Total	% Total
Yes. — Pay for Cooling	1	22,911	57.3%
No. — Included in Rent/Fee	2	668	1.7%
No. — Do not Have Cooling	3	13,592	34.0%
Missing	97	2,814	7.0%
All	**All**	**39,985**	**100.0%**

Source: 2019 California Residential Appliance Saturation Survey

C2 — Central AC

The following checks were used to clean the central air conditioning section:

- Respondents that reported the central air conditioner, evaporative cooler, or heat pump was zoned but did not indicate the number of the respective systems were assigned one system of that type.
- If they have a central heat pump for heating, and they indicated having central air conditioning, then they were assigned a central heat pump.
- If the survey indicated the addition of a central air conditioning unit in the past 12 months and the household owns the dwelling, yet there are no central air conditioning units specified, then the number of central air conditioning units was set to one.

C3 — Age of Central Air Conditioner (*CLCNTAGE*)

The variable for age of central air conditioner was cross-referenced with the presence of a cooling system. If a system was reported, but no age was provided, then age was set to missing.

C4 — Central Air Conditioner Temperature Setting

Responses to the central air conditioner temperature settings were cross-referenced with information concerning how the household pays for cooling, whether they have a central air conditioning system, and whether they skipped all temperature settings. Rules used to clean this section include the following:

- If respondents skipped PAYCOOL or pay for heat but provided no indication of the type of cooling system, then all temperature settings were set to missing (97).
- All temperature settings were set to not applicable (99) if they did not pay for central cooling, or central cooling is included in their rent, or if they do not have a thermostat.
- For respondents with a thermostat, temperature settings were evaluated to determine whether respondents answered a setting for at least one time of day. If they provided at least one setting, then all missing temperatures were set to off. If they did not provide any settings, then all were set to missing (97).

C7 — Room AC

The first step in cleaning the room air condition questions was to ensure that information for the first air conditioner was populated before information for the second room air conditioner. If a respondent did not populate the first air conditioners information but provided responses for the second, this information was moved to the first unit. Similarly, if the information for the second air conditioner was left missing, but the third was populated, then this datum was moved to the second. This was done for both the type and age of each air conditioner.

Next, the total number of room air conditioners was determined and used for the following checks:

- If at least one unit was present in the residence, then the variable NOROOMAC was set to zero. If no units were found, then NOROOMAC was set to one.
- If only one unit was found, then all information for the second and third units were set to not applicable (99).
- If the survey indicated the addition of a room air conditioning unit in the past 12 months (WWADD=1) but no room air conditioning units were specified, one was added and NOROOMAC was set to 0.

The type of room air conditioner was cross-referenced with the age of each unit. If the respondent filled in an age but left the type blank, then the unit was assumed to be a window/wall air conditioner. If type of room air conditioner was indicated, but age was left blank, then age was set to missing (97).

Water Heating

This section covers the procedure used to eliminate multiple and inconsistent responses to survey Section D — Water Heating. The cleaning process also revealed substantial fuel misreporting as reviewed in the CDA data imputation section later in this chapter.

D1 — Pay for Water Heat (*PAYWH*)

How a household pays for water heating is summarized in **Table 26**. The question concerning how a household pays for water heat (PAYWH) serves as the basis for cleaning the remainder of the section because households that indicated they do not pay for water heat, directly, were asked to skip this section. If information was provided for at least one water heater and PAYWH was either "no" or "no – included in their rent," then a new variable PAYWH1 recorded the response as "yes." The original PAYWH variable was preserved as the original pay for water heat response.

For cases in which multiple responses were provided, the lowest numbered response was kept. This logic favored "yes –pay for water heat" over "no — it is part of my rent/condo fee." Similarly, "no — included in rent" was chosen over "no — do not have water heating system."

If respondents indicated that they do pay for water heat but did not list any heating systems, or PAYWH was missing and they did not list any water heating systems, then system variables were set to missing (97). If a respondent did not pay for water heat, then all water heating system variables were set to missing (99).

Table 26: Water Heating Payment Question Cleaning

PAYWH	Code	Total	% Total
Yes.— Pay for Water Heating	1	15,144	37.9%
No.— Included in Rent/Fee	2	1,934	4.8%
No.— Do not Have Water Heating	3	270	0.7%
Missing	97	22,637	56.6%
All	**All**	**39,985**	**100.0%**

Source: 2019 California Residential Appliance Saturation Survey

D2 — Type of Primary and Secondary Water Heating Systems

The type of primary and secondary water heating system served as the basis for system fuel types and water heating UEC estimates. This section deals with the rules used to confirm the consistency of water heating systems with billing data and other survey variables. Specific rules used to clean natural gas, electric, propane, and other heaters are listed below.

Natural Gas Water Heating Systems

For respondents who indicated they have a primary or secondary natural gas water heater, the cleaned variable for a natural gas line to the home (NGLINE) was cross-referenced. As discussed above, this cross-referencing confirmed that the water heater was consistent with the survey response for NGLINE as well as billing information. If a household did not have gas service, but the survey response indicated a natural gas water heater, the system response was set to missing (97).

Propane Water Heaters

Households that had natural gas (NGLINE=1) were not allowed to have a propane water heater. For these households, all propane water-heating systems were set to missing.

Solar Water Heaters

Solar water heaters are allowed only for single-family dwellings. Responses for all other dwelling types that indicated solar were set to missing (97).

Other Water Heaters

Cases in which more than two “other” water heaters were indicated were set to missing.

Primary and Auxiliary Water Heating Fuels

If a respondent provided more than one primary heating system, then the first system selected was set as the primary system, and the subsequent responses were assigned to auxiliary heat. If only additional heating systems were provided, then the primary heat was set to the first additional heating system indicated. The maximum number of water heaters was set to four, such that surveys with five or more types of water heaters were considered erroneous. For these surveys, responses to all water heater type questions were set to missing (97).

The new variables primary heating fuel (PRWHFUEL) and auxiliary heating fuel (AWHTFUEL) were derived from the primary and additional heating system information. If the respondent indicated having portable electric heaters and did not have natural gas auxiliary heat, then AWHTFUEL was set to 2 for electric heat. Coding for PRWHFUEL is shown in **Table 27**.

D5 — Number of Showers and Baths Per Day

A limit was set on the number of showers/baths taken per day (SHWRDAY or BATHDAY) based on the cleaned number of residents (RESCNT). This limit was set at two showers or baths per day per person. Responses for households that exceeded this limit were set to missing (97).

Table 27: Water Heating Fuel Data Cleaning

PWHFUEL	Code	Total	% Total
Natural Gas	1	31,324	78.3%
Electric	2	2,513	6.3%
Bottled Gas	3	1,723	4.3%
Solar	4	13	0.0%
Other	5	35	0.1%
Missing	97	3,179	8.0%
N/A	99	1,198	3.0%
All		39,985	100.0%

Source: 2019 California Residential Appliance Saturation Survey

Laundry

This section covers the procedures used to clean Section E — Laundry Equipment. The cleaning process involved correcting for fuel misreporting on clothes dryers, as reviewed in the CDA data imputation section later in this chapter.

E1 — Presence of Laundry Equipment in Home (*LNDRYEQU*)

Question E1 (LNDRYEQU) asked whether laundry equipment was present in the home, not present, or was in a common area. Only respondents with laundry equipment in the home were instructed to fill out the remainder of the section. Therefore, responses were cross-referenced with responses to (E2) — Clothes Washer Type (CWTYP), (E3) — Clothes Washer Age (CWAGE), (E5) clothes dryer type (CDTYP), and (E6) clothes dryer age (CDAGE). If a respondent answered any of the questions pertaining to the type of laundry equipment in the home and answered, then LNDRYEQU was changed to "Yes."

E4 — Number of Clothes Washer Loads per Week

A limit was set on the number of loads washed per average week (sum of CWHWLD, CWWWLD, CWCWLD) based on the cleaned number of residents (RESCNT). This limit was set at five loads per week per person. Responses from households that exceeded this limit were set to missing (97).

E5 — Clothes Dryer Type

For respondents who indicated they have a primary natural gas dryer, the cleaned variable, a natural gas line to the home (NGLINE) was cross-referenced. As discussed above, this cross-referencing confirmed the clothes dryer was consistent with the survey response for NGLINE as well as billing information. If a household did not have gas service, but the survey response indicated a natural gas dryer, the system response was set to missing (97).

E7 — Number of Clothes Dryer Loads per Week

A limit was set on the number of loads dried per average week based on the cleaned number of residents (RESCNT). This limit was set at five loads per week per person. Responses from households that exceeded this limit were set to missing (97).

Food Preparation

For respondents who indicated they have a primary natural gas range or oven in question F1, the cleaned variable, a natural gas line to the home (NGLINE), was cross-referenced. As discussed above, this confirmed that the cooking equipment was consistent with the survey response for NGLINE as well as billing information. If a household did not have gas service, but the survey response indicated a natural gas range or oven, the system response was set to missing (97).

Refrigerators

Question G1 (RFNUM) asked respondents to indicate the number of refrigerators they own, while Question G2 contained a set of variables that recorded characteristics of up to three of those refrigerators. Refrigerators characteristics included the following:

- Door style (RF1STY, RF2STY, RF3STY)
- Cubic feet (RF1SZ, RF2SZ, RF3SZ)
- Age (RF1AGE, RF2AGE, RF3AGE)
- Other features (RF1OTH, RF2OTH, RF3OTH)

The number of refrigerators listed and refrigerator characteristics were cross-referenced to ensure they were logically consistent. If they were found to be inconsistent, the cleaning process attempted to impute the correct response given the available information. For example, if a respondent filled in the refrigerator characteristics, but the number of refrigerators (RFNUM) was missing or less than the number of refrigerators for which they provide characteristics, then the number of refrigerators was set to be consistent with the characteristics data. If NUMREF was missing and no characteristics were provided, then RFNUM was set to one, since nearly everyone has at least one refrigerator. Only those who specified that they had zero refrigerators were assumed not to have a refrigerator.

In addition to checking the total number of refrigerators, the characteristics of each refrigerator (one, two, and three) were checked against each other. For each set of refrigerator characteristics, if a respondent skipped the information for the lower number refrigerator (that is, Refrigerator 1) and populated the data for a higher refrigerator number (that is, Refrigerator 2), then the characteristics were assumed to apply to the lower number refrigerator. If the number of refrigerators, RFNUM, was larger than the set of refrigerator characteristics provided, the characteristics were set to "missing."

Table 28 summarizes the allocation of missing responses for the refrigeration section.

Table 28: Missing Refrigerator Number and Characteristics

Refrigerator Number	Refrigerators (G1 - G2)	Missing Post Cleaning	Not Applicable Post Cleaning	Missing Pre-Cleaning/ ReAssigned Post Cleaning	Missing Post Cleaning	Not Applicable Post Cleaning	Re-Assigned Post Cleaning	Unchanged Post Cleaning
N/A	How many refrigerators do you have plugged in? (RFNUM)	2,857	0	278	0	0	36,721	129
RF1	Door Style (RF1STY)	3,337	71	229	0	0	0	36,348
RF1	Size in Cubic Feet (RF1SZ)	5,116	71	288	0	0	0	34,510
RF1	Age (RF1AGE)	4,170	71	206	0	0	0	35,538
RF1	Other Features (RF1OTH)	3,258	71	22,502	0	0	0	14,154
RF2	Door Style (RF2STY)	3,481	25,985	17	175	74	2	10,251
RF2	Size in Cubic Feet (RF2SZ)	3,772	26,033	12	228	26	9	9,905
RF2	Age (RF2AGE)	3,592	26,025	25	155	34	7	10,147
RF2	Other Features (RF2ITH)	3,613	26,054	9,022	35	5	515	741
RF3	Door Style (RF3STY)	3,044	35,759	0	24	29	0	1,129
RF3	Size in Cubic Feet (RF3SZ)	3,035	35,769	0	41	19	0	1,121
RF3	Age (RF3AGE)	3,042	35,732	0	30	56	0	1,125
RF3	Other Features (RF3OTH)	3,067	35,787	1,063	1	1	42	24

Source: 2019 California Residential Appliance Saturation Survey

Because the refrigeration variables were used to construct engineering estimates of monthly kWh consumption used in the CDA, missing values were also imputed. The data imputation and the engineering estimates are described in the CDA section later in this chapter.

Freezers

Question F1 (FZNUM) asked respondents to indicate the number of freezers they own, while Question F2 contained a set of variables that recorded characteristics of up to two of those freezers. Freezers characteristics included the following:

- Door style — (FZ1STY, FZ2STY)
- Cubic feet — (FZ1SZ, FZ2SZ)
- Age — (FZ1AGE, FZ2AGE)

The number of freezers listed and freezer characteristics were cross-referenced to ensure they were logically consistent. If they were found to be inconsistent, the cleaning process attempted to impute the correct response given the available information. For example, if a respondent filled in the freezer characteristics, but the number of freezers (FZNUM) was missing or fewer than the number of freezers for which they provided characteristics, then the number of freezers was set to be consistent with the characteristics data. If FZNUM was missing and no characteristics were provided, then FZNUM was set to missing.

In addition to checking the total number of freezers, the characteristics of each freezer were checked against each other. For each set of freezer characteristics, if a respondent skipped the information for the lower number freezer, such as Freezer 1, and populated the data for a higher freezer number, such as Freezer 2, then the characteristics were assumed to apply to the lower number freezer. If the number of freezers, FZNUM, was larger than the set of freezer characteristics provided, the characteristics were set to "missing."

Table 29 Table 29 summarizes the allocation of missing responses for the freezer section.

Because the freezer variables were used to construct engineering estimates of monthly kWh consumption used in the CDA, missing values were also imputed. The data imputation and the engineering estimates are described in the CDA section later in this chapter.

Spas and Hot Tubs

Only respondents who indicated they pay for the use of a spa or hot tub were asked to complete the remainder of this section. If respondents indicated they have a spa or hot tub in a common area, or do not have a spa or hot tub, but filled in information provided by questions I2 — I7, the response to I1 (SPATYP) was changed to "Yes, I pay for its energy use."

For respondents who indicated they have a natural gas spa heater, the cleaned variable a natural gas line to the home (NGLINE) was cross-referenced. As discussed above, this cross-referencing confirmed whether the spa water heater was consistent with the survey response for NGLINE as well as billing information. If a household did not have gas service but indicated a natural gas spa heater, the system response was set to missing (97)

Respondents who lived in apartments were restricted from having a spa or hot tub.

Table 29: Missing Freezer Number and Characteristics

Freezer Number	Freezers (H1 - H2)	Missing Post Cleaning	Not Applicable Post Cleaning	Missing Pre-Cleaning/ Re-Assigned Post Cleaning	Missing Post Cleaning	Not Applicable Post Cleaning	Re-Assigned Post Cleaning	Unchanged Post Cleaning
N/A	How many freezers do you have plugged in? (FZNUM)	3,715	0	512	0	0	35,551	207
FZ1	Door Style (FZ1STY)	4,033	28,735	27	0	0	0	7,190
FZ1	Size in Cubic Feet (FZ1SZ)	4,350	28,735	32	0	0	0	6,868
FZ1	Age (FZ1AGE)	4,213	28,735	26	0	0	0	7,011
FZ2	Door Style (FZ2STY)	3,781	35,804	0	10	17	0	373
FZ2	Size in Cubic Feet (FZ2SZ)	3,798	35,807	0	18	14	0	348
FZ2	Age (FZ2AGE)	3,804	35,807	0	12	14	0	348

Source: 2019 California Residential Appliance Saturation Survey

Swimming Pools

Only respondents who indicated they pay for the energy use of a swimming pool were asked to complete the remainder of this section. If a respondent indicated they have a pool in a common area, or do not have a pool, but filled in information provided by questions J2 — J7, the response to I1 (PLTYP) was changed to "Yes, I pay for its energy use."

For respondents who indicated they have a natural gas pool heater, the cleaned variable a natural gas line to the home (NGLINE) was cross-referenced. As discussed above, this confirmed whether the pool heater was consistent with the survey response for NGLINE as well as billing information. If a household did not have gas service, but indicated a natural gas pool heater, the system response was set to missing (97)

Respondents who did not live in single-family dwellings were restricted from having a pool.

Entertainment and Technology

Responses to the entertainment and technology section were evaluated to determine whether respondents skipped appliances they do not have or skipped all questions. If respondents answered at least one technology question, then all missing values were set to zero. If they did not provide a response to any technologies, then all were set to missing (97,) resulting in the value of 3,261 for all questions in the first column of **Table 30**. Cleaning of these variables is summarized in **Table 30**.

Additional variables were constructed, and missing values were imputed for the CDA, as discussed in the CDA section later in this chapter.

Lighting

The lighting section consisted of a set of questions to gather information on interior and exterior lighting.

Cleaning of the interior lighting section differs from the *2009 RASS* because the section changed significantly since the 2009 study. Question L1 recorded the portion of compact fluorescent light bulbs (CFLs), light-emitting diode light bulbs (LEDs), and incandescent light bulbs inside the home. If a respondent did not provide an answer to the portion of CFLs, LEDs, or incandescent light bulbs, the variable was coded as missing (97). Since the survey responses were coded such that a value of one was equal to zero bulbs, and two was equal to one bulb, all non-missing responses were given a response value equal to one minus the survey response number. For example, if they provided a response value equal to one, they were coded as one minus one, or zero. This is because the first response was "zero." This logic applied to all light bulb types.

Question L2 recorded the number of interior lights used by time of day. Responses to L2 were cleaned using the same logic as L1. If a respondent did not provide an answer for a given time period, the variable was coded as missing (97).

Question L3 asked about interior lighting products, such as timers, sensors, and dimmers. If a respondent did not provide an answer for a given product, the variable was coded as missing (97). All valid responses were coded as the response value minus one.

The cleaning procedures used for exterior lights (L4) varied from those used for interior lighting. The process used to clean this section was consistent with the cleaning process used in the *2009 RASS* because a specific UEC was estimated for exterior lighting. The study team examined responses to all exterior lighting products to identify missing values. If the all values were skipped, then each value was coded as missing (97). If at least one value was provided for one of the products, then missing values were set to zero. The CDA model required additional lighting variables for exterior lighting, as discussed in the CDA section later in this chapter.

Table 30: Missing Entertainment and Technology Appliances

Entertainment and Technology Question (K1)	Missing Post Cleaning	Not Applicable Post Cleaning	Re-Assigned Post Cleaning	Missing Post Cleaning	Not Applicable Post Cleaning	Re-Assigned Post Cleaning	Un-changed Post Cleaning
Cathode-ray tube TVs (STDTV)	3,261	0	25,897	0	0	10,827	0
LCD TVs <36 inches (SMLLCDTV)	3,261	0	21,582	0	0	15,142	0
LCD TVs 36-59 inches (MEDLCDTV)	3,261	0	18,109	0	0	18,615	0
LCD TVs >59 inches (BIGLCDTV)	3,261	0	24,939	0	0	11,785	0
OLED or LED TVs <36 inches (SMLLEDTV)	3,261	0	18,901	0	0	17,823	0
OLED or LCD TVs 36-59 inches (MEDLEDTV)	3,261	0	16,296	0	0	20,428	0
OLED or LED TVs >59 inches (LRGLEDTV)	3,261	0	23,585	0	0	13,139	0
Plasma TVs (PLSMTV)	3,261	0	23,722	0	0	13,002	0
DLP or rear-projection TVs (DLPTV)	3,261	0	26,202	0	0	10,522	0
Cable or satellite TV set-top boxes or receivers, DVR (DVRBOX)	3,261	0	6,836	0	0	29,888	0
Stand-alone movie players (MOVPLA)	3,261	0	13,499	0	0	23,225	0
Gaming systems (GAMSYS)	3,261	0	15,356	0	0	21,368	0
Streaming media players (STREAM)	3,261	0	13,826	0	0	22,898	0
Home theater system connected to TV (HTSYS)	3,261	0	16,672	0	0	20,052	0
Soundbar connected to TV	3,261	0	15,519	0	0	21,205	0
Separate sound or stereo system connected to TV (AUDTOTV)	3,261	0	17,144	0	0	19,580	0
Stand-alone stereo, I-pod of MP3 docking station (SAMUSIC)	3,261	0	16,876	0	0	19,848	0

Source: 2019 California Residential Appliance Saturation Survey

Miscellaneous Appliances

Responses to the M1, number of miscellaneous appliances used, were evaluated to determine whether respondents skipped appliances they do not have or skipped all questions, summarized in **Table 31**. If a respondent answered at least one appliance question, then all missing values were set to zero. If he or she did not provide a response to any appliance questions, then all were set to missing (97), resulting in the value of 4,319 for all questions in the first column of the **Table 31**.

Additional variables were constructed, and missing values were imputed for the CDA, as discussed in the CDA data imputation section later in this chapter. **Table 31** presents the percentage of responses with missing values for the variables used to develop the appliance ownership indicator variables for the CDA.

Table 31: Missing Miscellaneous Appliances

Miscellaneous Appliances (M1)	Missing Post Cleaning	Not Applicable Post Cleaning	Re-Assigned Post Cleaning	Missing Post Cleaning	Not Applicable Post Cleaning	Re-Assigned Post Cleaning	Un-changed Post Cleaning
Chargers left plugged in all the time (CHRGRS)	4,319	0	1,647	0	0	34,019	0
Rechargeable vacuum cleaners (RECHVAC)	4,319	0	4,322	0	0	31,344	0
Small cooking appliances (SMACOOK)	4,319	0	3,967	0	0	31,699	0
Portable fan (FNPORT)	4,319	0	3,647	0	0	32,019	0
Ceiling fan (FNCEIL)	4,319	0	2,728	0	0	32,938	0
Wind-turbine attic ventilator (WNDATV)	4,319	0	5,277	0	0	30,389	0
Electric attic fan (FNATTIC)	4,319	0	4,952	0	0	30,714	0
Whole house fan (FNWHOLE)	4,319	0	5,041	0	0	30,625	0
Electric air cleaner (AIRCLEAN)	4,319	0	5,168	0	0	30,498	0
Portable room air purifier (PORTPUR)	4,319	0	4,861	0	0	30,805	0
Humidifier or dehumidifier (HUMDEH)	4,319	0	4,919	0	0	30,747	0
Wine or beverage cooler (WINCLR)	4,319	0	4,828	0	0	30,838	0
Water purification system (WHPURIFY)	4,319	0	4,863	0	0	30,803	0
Domestic hot water recirculation pump (DHWRPMP)	4,319	0	5,141	0	0	30,525	0
Sauna – electric (SAUNA)	4,319	0	5,318	0	0	30,348	0
Pond or water garden pump or fountain (POND)	4,319	0	5,008	0	0	30,658	0
Outdoor fire table/fire pit (FIREPIT)	4,319	0	5,268	0	0	30,398	0

Source: 2019 California Residential Appliance Saturation Survey

Renewable Energy Technologies

Questions M11 and M12, regarding Renewable Energy Technologies, were cleaned as follows. If all the answers in M11 and M12 were missing or all the answers were filled, then the responses were set to missing (97). If a renewable technology or plan to install a renewable technology was selected, along with "No renewable technology" or "No plans" (for M11 and M12, respectively), then all responses except the selected renewable technology or selected renewable technology plan was set to "Not Applicable." Lastly, for renewable technologies or plans to install renewable technologies that were not selected (left blank) by a respondent were set to "Not Applicable."

Household Information

The variables PTHME (N1), PTHMELOC, PTHMEUTL, and PTHMGUTL(N2) were cross-referenced for cleaning. If the respondent left vacation home (PTHME) blank but filled in data for location, electricity, or gas provider or a combination, the value was changed to "yes."

The team used the household income variable to create the variable AVGINC, which was used in the CDA analysis. **Table 32** provides a summary.

Table 32: Missing Household Information

Household Information (N1 - N7)	Missing Post Cleaning	Not Applicable Post Cleaning	Re-Assigned Post Cleaning	Missing Post Cleaning	Not Applicable Post Cleaning	Re-Assigned Post Cleaning	Unchanged Post Cleaning
Own vacation home (PTHME)	4,878	0	50	0	0	154	34,903
Location for vacation home (PTHMELOC)	5,084	32,905	0	0	77	0	1,919
Electric utility for vacation home (PTHMEUTL)	5,136	32,781	0	0	201	0	1,867
Natural gas utility for vacation home (PTHMGUTL)	5,237	32,407	0	0	575	0	1,766
Highest level of education (EDUC)	4,775	0	0	0	0	0	35,210
Primary spoken language (ETHNIC)	4,739	0	0	0	0	0	35,246
Number of occupants of home disabled (DISABLED)	4,751	0	0	0	0	0	35,234
Household total annual income (INCOME19)	5,478	0	0	0	0	0	34,507

Source: 2019 California Residential Appliance Saturation Survey

Consumption and Weather Data

This section discusses the development of the data that were stored in the RASS billing database. This section includes a description of the customer consumption data (interval and billing) provided by each utility, the methods used to clean the consumption data, the normalization routines employed to standardize the consumption amounts, and the merging of the consumption data with the survey data. The data provided by utility included consumption, read dates, status codes, and tariffs. Consumption data were provided as billing data and interval data from advanced metering infrastructure (AMI). Personal identifiable information, such as names, addresses, bill payment amounts, and payment information, was not used in the analysis or included in the datasets provided.

Billing Databases

The team requested billing data were from the three California IOUs (PG&E, SCE, and SDG&E) in addition to two California publicly owned utility (POUs) (LADWP and SMUD) for all sampled households. The section below describes the contents of the billing data from the utilities. In addition, the team also requested gas consumption data from SoCalGas for all electric respondents that could be matched to a gas account using an account matching process.

Pacific Gas and Electric Company

The billing data for PG&E respondents were provided in three datasets: one for each fuel and one with service address characteristics. The datasets covered the period June 2018 through January 2020 and contained information from 250,521 electric accounts and 174,589 gas accounts for a total of 134,496 gas and 181,868 electric premises. PG&E's billing data included the following information: premise number, account number, service agreement number, rate schedule, kWh and therm monthly consumption, kWh and therm consumption by tier, an indicator of whether each bill was estimated, an indicator for balanced payment plan, an indicator for participation in the California Alternative Rates for Energy or the Family Electric Rate Assistance Program (CARE/FERA) programs, and the number of days and start and the end date for the billing period. In addition, the electric billing data included indicators of NEM status and cost responsibility surcharge.

Southern California Edison

The billing data for SCE were provided in a single Microsoft Excel® binary workbook file. This dataset contained a total of 156,313 unique service account IDs within the SCE territory. SCE's billing data included the following set of information: a customer number, a premise number, a customer account number, a installed service number, a service account number, kWh monthly consumption, sum of all the months kWh consumption by premise, bill date, number of billing days, number of billing periods, rate code, bill code, an indicator of NEM status, an indicator of community choice aggregation status, an indicator of inclusion in the Family Electric Rate Assistance Program (FERA), an indicator if the customer was part of the California Alternative Rates for Energy (CARE) program, and an indicator if the customer was on a levelized payment plan. The billing data were provided in a wide format with each premise having a single line of data. The billing data covered the period June 2018 through November 2019.

San Diego Gas and Electric Company

The SDG&E electric and gas billing data were provided in two comma-separated datasets for each respective fuel. The SDG&E billing data contained data for 74,039 premises within SDG&E's territory. The data set had billing data for 62,389 households with electric consumption and 11,653 households with gas consumption. The gas and electric data included a premise identification number, a customer number, a service point identifier, an account number, kWh/therm consumption by tier, kWh/therm consumption, transaction type, procurement status, rate code, an indicator if a customer was on a levelized payment plan, an indicator if the customer has auto payment, an indicator if the customer was part of the California Alternative Rates for Energy (CARE) program, and the number of days in along with a start and an end date for the billing period. In addition, the electric billing data included an indicator of NEM and virtual NEM, an indicator of inclusion in the Family Electric Rate Assistance Program (FERA), and an indicator of inclusion of a demand response program. The billing data covered the period July 2018 through December 2019.

Sacramento Municipal Utility District

The SMUD billing data was provided in a comma-separated dataset. The dataset contained data on 26,569 unique premises. The SMUD billing data included a contract number, a premise identification number, a contract account number, kWh consumption, rate code, an indicator of NEM and virtual NEM, a low-income indicator, an indicator of the customer being on a payment plan, an indicator if the bill was an adjustment, the kWh of PV generation and the kWh returned to SMUD from PV generation, and the number of days in the billing period along with the start and end date of the period. The billing data covered December 2017 through November 2019.

Los Angeles Department of Water and Power

The LADWP billing data were provided in one comma-separated dataset. The dataset contained information on 44,652 premises within the LADWP territory recorded bimonthly. The LADWP data included a premise identification number; the account number; the service point number; kWh consumption by tier; kWh consumption by period; high, low, and base kWh consumption; rate code; the service agreement type; the account date; an indicator of whether a bill was estimated; an indicator of the type of bill; an indicator of NEM; and the number of billing days in the period in addition to the start and read dates for that billing period. The LADWP billing data covered June 2017 through December 2019.

Southern California Gas Company

The sample frame for the RASS study was developed from the residential electric population from the three IOUs and two POUs. As such, collecting natural gas billing data for respondents served by SoCalGas and PG&E Gas in the SMUD and SCE service territories involved a customer matching procedure between the RASS sample frame data and the SoCalGas and PG&E Gas residential populations. This procedure required each of the steps discussed below for SoCalGas and PG&E Gas.

Step 1 — Identify Gas Service ZIP Codes. The sample frame was sorted by ZIP code and merged with a file that contained the natural gas utility serving each ZIP code in California. The sample having SoCalGas as the gas utility was placed in one group and the sample having PG&E gas, but non-PG&E electric service was put in a second group for further analysis.

Step 2 — Disaggregate Customer Address. The service address variable in the sample frame, the SoCalGas residential population, and PG&E Gas not having PG&E electric service population were disaggregated into the following pieces:

- Street number and number fraction
- Street direction
- Street name
- Apartment or unit number
- ZIP code

The team developed code for each utility sample frame that created the six pieces of the address. These pieces, along with the customer name and account information, were matched against the gas population data for further analysis.

Step 3 — Customer Address Merging with Gas Population Files. The merging of sample addresses with gas population data to capture account number involved several phases. The team first merged the RASS sample and gas population files by ZIP code, street number, street number fraction, street direction, street name, and apartment/unit number to obtain the exact address matched cases in the first phase.

For the remaining unmatched sample, the second phase involved merging the files by ZIP code, street number, street number fraction, and street name followed by a case-by-case inspection to select matches. In Phase 2, accounts were located along with addresses that may have a missing street direction or different apartment/unit number designation (for example, D instead of 4). The customer name appearing in the RASS sample frame as well as the gas population were used in this phase to select the appropriate record.

For the remaining unmatched sample after Phases 1 and 2, the third phase involved merging the files by ZIP code, street number, street number fraction, street direction, and the first six characters of the street name followed by a case-by-case inspection to select matches using the same approach as was described in Phase 2. For the remaining unmatched sample after Phases 1 through 3, the next phase involved merging the files by ZIP code and customer last name followed by a case-by-case inspection to select matches that may have slightly different street name spellings between data sources. For the remaining unmatched sample after Phases 1 through 4, the final phase inspected the returned RASS nonmatched respondent addresses with gas population file addresses. For several family RASS nonmatched respondents served by SoCalGas, other units were in the SoCalGas population file but not the respondent unit.

Step 4 — Merge Gas Account Number to Sample Frame. The records with identified gas accounts using the five step 3 phases were merged into the RASS sample frame. For SoCalGas, 8,860 PG&E sample frame respondents, 36,931 LADWP sample frame respondents, 119,447 SCE sample frame respondents, and 4,789 SDG&E sample frame respondents were matched to SoCalGas accounts. For PG&E Gas not having PG&E electric service, 546 SCE sample frame respondents and 22,408 SMUD sample frame respondents were matched to PG&E Gas accounts.

Step 5 — Obtaining Gas Billing Data. The SoCalGas and PG&E Gas account numbers were provided to the appropriate utility so that monthly gas billing data could be extracted for the period of the Gas CDA history.

All respondents in the study sample had electric billing account information but not all had gas service. The coincidence for the two services is shown in **Table 33**.

Table 33: Comparison of Gas and Electric Utility Providers for Survey Respondents

Gas Provider	LADWP (Electric)	PG&E (Electric)	SCE (Electric)	SDG&E (Electric)	SMUD (Electric)	All
PG&E	0	11,725	73	0	2,048	13,846
SoCalGas	2,347	779	10,187	0	0	13,313
SDG&E	0	0	0	5,257	0	5,257
All	**2,347**	**12,504**	**10,260**	**5,257**	**2,048**	**32,416**

Source: 2019 California Residential Appliance Saturation Survey

AMI Databases

The IOUs and SMUD provided hourly electric interval data, and PG&E, SoCalGas, and SD&GE provided daily gas consumption data for the survey respondents. The interval data were used in the degree-day normalization.

Pacific Gas and Electric Company

PG&E provided hourly electric AMI data as well as daily gas interval data in two text datasets per fuel for premises that completed surveys. The electric dataset included data for 15,749 accounts, and the gas data provided daily consumption data for 12,681 accounts. These datasets included the premise identification number, account number, service agreement identifier, usage date, and kWh/therm consumption. The electric data also included the hour during which consumption occurred for a given usage date, service point identifier, direction of energy flow, and an indicator as to whether the interval consumption was estimated. The electric and gas data covered August 2018 through November 2019.

Southern California Edison

SCE provided electric AMI data in 32 comma-separated files. In total SCE submitted data for 9,785 accounts. The datasets included an account number, a premise number, the installed service number, the meter channel, device identifier, interval start and end date along with the time of the read, length of the interval, kWh delivered and received, an indicator of NEM, and an indicator if the interval consumption was estimated. The SCE data covered August 2018 through November 2019.

San Diego Gas and Electric Company

SDG&E submitted electric and gas AMI contained in two submission of CSV files per fuel type. The daily gas data contained data for 3,028 premises, and the hourly electric data included consumption data for 4,265 premises. The gas and electric data included a customer identification number, a premise number, a service point identifier, the meter identifier, usage date, therm/kWh consumption for the interval, and a flag indicating if an interval is estimated. The electric data were submitted as a wide dataset with a column for every hour and included the meter channel identifier. The gas and electric data included data for August 2018 through November 2019.

Sacramento Municipal Utility District

SMUD provided hourly electric data in two comma-separated files. These files contained data for 2,150 accounts. The data sets included the following fields: the contract number, premise identifier, meter identifier, usage date and time, kWh consumption, an NEM indicator, and the meter channel identifier. In addition, the data contained quality flags to indicate if the hour was estimated, edited, missing, empty, passed/failed, occurred during a power outage, was deleted, or had an associated warning. The AMI data that SMUD provided covered August 2018 through November 2019.

Los Angeles Department of Water and Power

As mentioned, LADWP did not provide AMI data because it was only in the initial stages of smart meter deployment and evaluation.

Southern California Gas Company

SoCalGas provided daily gas data in a single SAS dataset. This file contained data for 11,613 gas accounts. The datasets contained the following fields: meter identifier, rate code, cubic feet of gas daily usage, therms of daily gas usage, usage date, gas network node identifier, and facility identifier. The interval data covered August 2018 through November 2019.

Data Cleaning and Preparation for Analysis (Weather Normalization)

DDN provided a way to generate a household's consumption for a standardized year's weather. This normalization accomplished two things: First, it converted consumption series that span varying numbers of days to a one-year period. Second, it provided annual consumption for long-run normal weather conditions. This normalization allowed comparisons across forecasting climate zones despite any unusual weather events that might have occurred in certain zones.

The normalization modeled electric or gas consumption individually for each household. Each household-level electricity model was a linear function of heating degree-days and cooling degree-days, with respect to heating reference temperature estimated specific to the location of the household. Each household level gas model was a linear function of heating degree-days, with respect to heating reference temperature estimated specific to the household location.

Preparing the data for the analysis required the following tasks:

Separate Electric and Gas

For the three utilities that supply natural gas (PG&E, SDG&E and SoCalGas), the study team identified the interval daily data series for electric and gas services. The weather-normalized annual consumptions were calculated separately for the two fuels.

Identify Potentially Problematic Billing Periods

The next step was to identify any series that might be problematic.

Validation, editing, and estimation (VEE) of each household's data was performed. The VEE was performed consistent with commonly acceptable methods so not to discard or capriciously exclude data from the analysis.

The VEE included:

- Checking for consistency between the AMI and billed data.
- Eliminating reads that feature long (greater than 40 days) or short (less than 20 days) billing cycles.
- Eliminated estimated billed or interval data.
- Eliminating duplicate records.
- Combining multiple unique records for a date.
- Eliminating spikes in the interval data.
- Requiring that a customer had a minimum of 180 days of data, including at least 30 summer days and 30 winter days

Prepare Data for Analysis

Data files from the different utilities were formatted so the files could be combined into a single electric and a single gas file. This formatting required standardizing variable names across the four electric and the three gas utility files. The two combined files were then sorted and transposed so that each record represented a single household, with all the periods in chronological order. Each period is described by three key variables: read date, number of days of service in that period, and kWh or therms used during that period. Each record in the combined analysis data sets was then matched to the respective correct weather series and corresponding survey data.

Before the normalization procedure, the team reviewed the AMI and billing data for anomalies. The goal of this process was to identify all data that could be used in the analyses. To be included in the analyses, each site was required to have interval and billing data (except LADWP). In addition, the following checks were made:

- Cycles with greater than 40 days or less than 20 days were eliminated (except LADWP).
- For non-net-metering customers, the interval data for a billing cycle usage were required to be within 10 percent.
- To calculate the electric daily use, the day needed at least 20 hourly intervals.
- Duplicate records were deleted.
- Multiple nonunique records (for example, reverse metering metering) were summed.
- Spikes in the interval data were eliminated.
- Electric daily usages with zero consumption were eliminated.
- Minimum required days of data for a site: 180 days total, 30 days winter (November-March), and 30 days of summer (April-October). For gas, the summer minimum data criteria were not used.
- Data were limited to July 1, 2018, through December 31, 2019.

Weather Normalization Procedure

To ensure that the analysis was comparable across utilities and climate zones, it was necessary to analyze the respondent energy consumption while controlling for the local weather, also called normalization. Techniques used for normalizing consumption include using calendarization, selecting a specific period for analysis that minimizes extreme weather, and modeling the relationship between weather and energy consumption. The team used modeling to normalize consumption in the current study.

The normalization of billing consumption data and the AMI daily interval data, for each fuel, followed the same procedure. The normalization process used was the degree-day normalization (DDN) similar to the Princeton Scorekeeping Model (PRISM™) technique. This method consists of two parts:

- Each household's energy consumption is modeled as a function of outdoor temperature over the study period using *actual year* weather data.
- Each household's fitted model is used to calculate energy consumption for a year of standardized temperatures using *normal year* weather data.

The results of the process provided normalized annual consumption (NAC) estimates for each household. These NACs reflected the households' estimated energy consumption for a typical (normal) year.

Temperature Data

The weatherization analysis was based on daily average dry-bulb temperatures. The temperature data used for weather normalization came from the most recent weather datasets created for the CEC and PG&E by White Box Technologies.[3] The normal year weather data for these datasets include data through 2017 that better captures the impacts of climate change. There are two datasets:

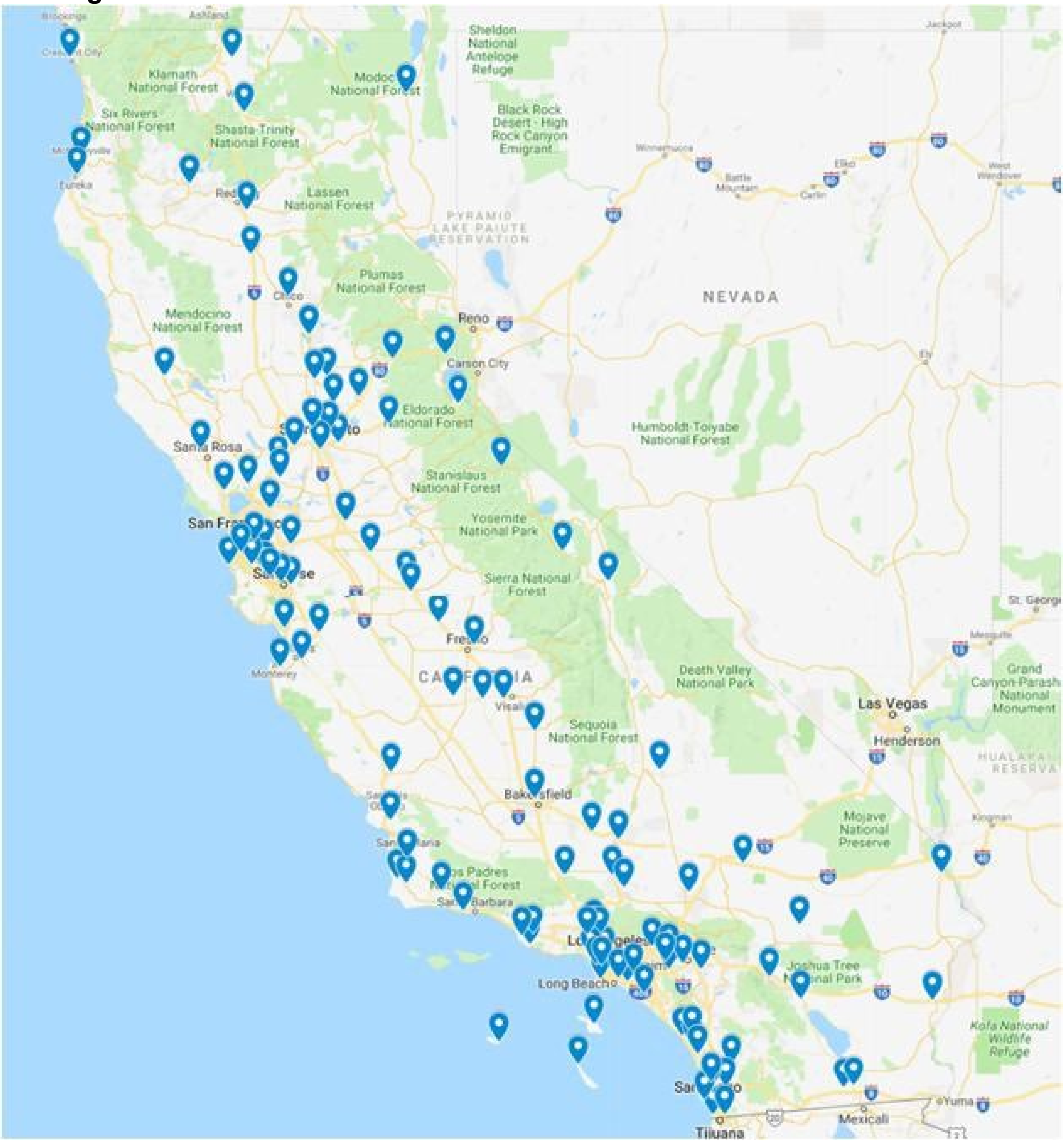
- The CALEE2018 weather dataset includes 117 locations, the normal year data cover a 12-year span (2006–2017), and the normal file was developed to represent a typical meteorological year for that city. The dataset also includes actual year weather data for 2016–2019.[4]
- The CZ2022 dataset contains only normal year weather data covering 20 years (1998–2017) for 97 weather station locations[5] and will be used for the T24 2022 Standards that will take effect January 1, 2023. The smaller set of 16 building climate zone weather files used for T24 compliance analysis is a subset of the full CZ2022 dataset. Weather stations for both datasets are shown in **Figure 5**, and **Table 34** illustrates the smaller building climate zone representative weather stations. Building climate zones were not used for the analysis, but every complete household survey was mapped to a T24 building climate zone using the service address information. The number of weather stations used for the CDA analysis within each T24 building climate zone is also noted in **Table 34**

3 "Update of California Weather Files for Use in Utility Energy Efficiency Programs and Building Energy Standard Compliance Calculations," White Box Technologies, March 6, 2010. CALMAC ID PGE0450.00. A webinar summarizing the changes can also be found here: https://pda.energydataweb.com/api/downloads/2280/Weather%20webinar%20CALEE2018%207-12-2019.pptx.

4 Actual year weather files were developed as part of the CALLEE2018 file development and are posted on CALMAC.org along with the CALEE2018 version of normal year weather data for each specific city: http://www.calmac.org/weather.asp.

5 CZ2022 weather files in EnergyPlus weather data format: http://weather.whiteboxtechnologies.com/custom-np-link-download?F1=5445946735*CZ2018_STATEWIDE_STYP20_97LOCS_EPW.zip.

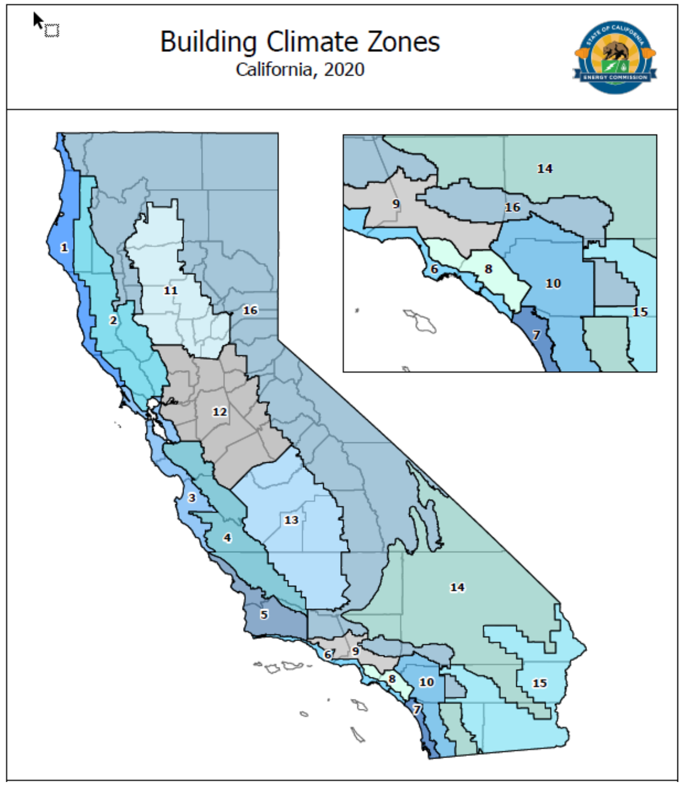
Figure 5: All Weather Stations for CALEE2018 and CZ2022 Weather Data



Source: Update of California Weather Files for Use in Utility Energy Efficiency Programs and Building Energy Standard Compliance Calculations

Table 34: Title 24 Building Climate Zones and Representative Weather Stations

T24 Building Climate Zone	T24 Building Climate Zone Representative City	Number Weather Stations used in the CDA
Zone 1	Arcata	2
Zone 2	Santa Rosa	3
Zone 3	Oakland	6
Zone 4	San Jose-Reid	2
Zone 5	Santa Maria	2
Zone 6	Torrance	8
Zone 7	San Diego Lindbergh	5
Zone 8	Fullerton	5
Zone 9	Burbank-Glendale	2
Zone 10	Riverside	5
Zone 11	Red Bluff	6
Zone 12	Sacramento	10
Zone 13	Fresno	5
Zone 14	Palmdale	4
Zone 15	Palm Springs-Intl	5
Zone 16	Blue Canyon	6



Source: CEC Cartography Unit

To ensure a robust weather dataset for normalization and the CDA analysis, the study team reviewed the weather data and selected a subset of weather stations that were in the electric service territories included in RASS and had good quality data for **both** the actual year and normal year. The actual weather was assessed for the study period (October 1, 2018, to September 30, 2019). This process yielded 76 weather stations with robust actual year and normalized (CZ2022) weather data. **Table 35** provides the complete list of weather stations used for weather normalization and the corresponding T24 building climate zones.

Table 35: Weather Stations Used for RASS Weather Normalization

#	WMO Station Number	Weather Station Name	T24 CZ	#	WMO Station Number	Weather Station Name	T24 CZ
1	725958	ALTURAS	16	39	723927	OXNARD-AP	6
2	725945	ARCATA-AP	1	40	723820	PALMDALE-AP	14
3	723840	BAKERSFIELD-MEADOWS-FLD	13	41	722868	PALM-SPRINGS-IAP	15
4	724837	BEALE-AFB	11	42	747187	PALM-SPRINGS-THERMAL-AP	15
5	724800	BISHOP-AP	16	43	723965	PASO-ROBLES-MUNI-AP	4
6	725845	BLUE-CANYON-AP	11	44	725910	RED-BLUFF-MUNI-AP	11
7	747188	BLYTHE-RIVERSIDE-CO-AP	15	45	725920	REDDING-MUNI-AP	11
8	722904	BROWN-FLD-MUNI	7	46	722860	RIVERSIDE-MARCH-AFB	10
9	724950	BUCHANAN-FLD-AP	12	47	722869	RIVERSIDE-MUNI	10
10	722880	BURBANK-GLNDLE-PASAD-AP	9	48	724830	SACRAMENTO-EXECUTIVE-AP	12
11	723926	CAMARILLO-AP	6	49	724833	SACRAMENTO-MATHER-FL	12
12	722927	CARLSBAD-MCCLELLAN	7	50	724839	SACRAMENTO-METRO-AP	12
13	722899	CHINO-AP	10	51	725930	SALINAS-MUNI-AP	3
14	723815	DAGGETT-BARSTOW-DAGGETT-AP	14	52	724938	SAN-CARLOS-AP	3
15	723810	EDWARDS-AFB	14	53	723830	SANDBERG	16
16	725940	EUREKA	1	54	722907	SAN-DIEGO-GILLESPIE	10
17	723890	FRESNO-YOSEMITE-IAP	13	55	722900	SAN-DIEGO-LINDBERGH-FIELD	7
18	722976	FULLERTON-MUNI-AP	8	56	722903	SAN-DIEGO-MONTGOMER	7
19	723898	HANFORD-MUNI-AP	13	57	722906	SAN-DIEGO-NORTH-ISLAND-NAS	7
20	747185	IMPERIAL	15	58	724940	SAN-FRANCISCO-IAP	3
21	722956	JACK-NORTHROP-FLD-H	8	59	724945	SAN-JOSE-IAP	4
22	725847	LAKE-TAHOE	16	60	722897	SAN-LUIS-CO-RGNL-AP	5
23	723816	LANCASTER-GEN-WM-FOX-FIELD	14	61	722977	SANTA-ANA-JOHN-WAYNE-AP	6
24	724927	LIVERMORE-MUNI-AP	12	62	723925	SANTA-BARBARA-MUNI-AP	6
25	722970	LONG-BEACH-DAUGHERTY-FLD	8	63	722920	SANTA-CATALINA-CATALINA-AP	6
26	722975	LOS-ALAMITOS-AAF	8	64	723940	SANTA-MARIA-PUBLIC-AP	5
27	722874	LOS-ANGELES-DOWNTOWN-USC	8	65	722885	SANTA-MONICA-MUNI	6
28	722950	LOS-ANGELES-IAP	6	66	724957	SANTA-ROSA(AWOS)	2
29	745046	MADERA-MUNI	13	67	724920	STOCKTON-METRO-AP	12
30	724815	MERCED-MUNI-MACREADY	12	68	722955	TORRANCE-MUNI-AP	6
31	724926	MODESTO-CITY-CO-AP	12	69	745160	TRAVIS-FLD-AFB	12
32	724915	MONTEREY-PENINSULA	3	70	725846	TRUCKEE-TAHOE	16
33	725957	MOUNT-SHASTA	16	71	725905	UKIAH-MUNI-AP	2
34	724955	NAPA-CO	2	72	724828	VACAVILLE-NUT-TREE	12
35	723805	NEEDLES-AP	15	73	722886	VAN-NUYS-AP	9
36	724930	OAKLAND-METRO-AP	3	74	723896	VISALIA-MUNI(AWOS)	13
37	747040	ONTARIO-IAP	10	75	745058	WATSONVILLE	3
38	745048	OROVILLE	11	76	724838	YUBA-CO	11

Source: 2019 California Residential Appliance Saturation Survey

Degree-Day Normalization Method

The normalization was based on AMI daily consumption data for all utilities except LADWP. For LADWP, the normalization was based on bimonthly billed consumption because AMI data were not available.

The DDN method modeled consumption (daily or monthly billed) as a function of monthly heating degree days and cooling degree days (HDD and CDD, respectively). The HDDs and CDDs for each household reflected the sum of daily degree-days for the interval of the consumption (daily or monthly). Heating degree-days for a specific day was the difference between the heating degree-day base τ_1 and the daily average temperature, if the daily average was below the base, and 0 if the daily average was above the base. Similarly, cooling degree-days for the day was the difference between the daily average temperature and the cooling degree-day base τ_2, if the daily average was above the base, and 0 if the daily average was below the base. The base or reference temperatures τ_1 and τ_2 were specific to each household, based on the model fit.

This relationship is shown in Equation 1. For each unique billing series, the coefficients β0, β_1, and β_2, and the parameters τ_1 and τ_2 were estimated to best fit the relationship between outdoor temperature and monthly energy consumption.

Equation 1: The DDN Heating and Cooling Model

$$U_i = \beta_0 + \beta_1 * \mathrm{HDD}_i(\tau_1, \mathrm{T_{ext}}) + \beta_2 * \mathrm{CDD}_i(\tau_2, \mathrm{T_{ext}}) + e_i$$

$$\mathrm{HDD}_i(\tau_1, \mathrm{T_{ext}}) = \sum_{d=firstdayi}^{lastdayi} \max(\tau_1 - \mathrm{T_{ext\,d}}, 0)$$

$$\mathrm{CDD}_i(\tau_2, \mathrm{T_{ext}}) = \sum_{d=firstdayi}^{lastdayi} \max(T_{\mathrm{ext\,d}} - \tau_2, 0)$$

Where:

- U_i = Electric usage during billing cycle i.
- T_{ext} = Series of external temperatures for each day of the study period
- Text d = External temperatures on day d
- HDD_i = Sum of heating degree days based on reference temperature τ_1 during consumption interval i.
- CDD_i = Sum of cooling degree days based on reference temperature τ_2 during consumption interval i.
- β_0 = Estimate of the average daily base load (temperature-invariant component of usage)
- β_1 = Increase in usage for each incremental increase in heating degree days
- β_2 = Increase in usage for each incremental increase in cooling degree days
- τ_1 = The heating "set-point"; the outside temperature at which the household's heating-related usage begins
- τ_2 = The cooling "set-point": the outside temperature at which the household's cooling-related usage begins

e_i = Residual Error

Roughly speaking, the product of β_1 times HDD corresponded to heating load, and the product of β_2 times CDD corresponded to cooling load. Non-heating and non-cooling use also varied over the year in ways that were correlated to some extent with heating and cooling degree-days. These estimated terms also included positive and negative seasonal effects associated with other uses. Heating and cooling consumptions were not assumed to be given by these terms, rather they were estimated via the cross-sectional CDA analysis applied to the total normalized annual consumption. Heating and cooling coefficients were used as indicators of the presence of heating and cooling, as described in the CDA section later in this chapter.

For some households, one or both degree-day terms showed little relationship to monthly consumption. The analysis determined for a household whether to include either the heating or cooling term, both terms, or neither term. The heating or cooling term was dropped from the model if the related coefficient was negative.

For electricity, the analysis tested for inclusion of both heating and cooling terms. Based on the diagnostics, the best-fit model for a household included both heating and cooling terms, only a heating term, only a cooling term, or neither of the terms. Inclusion of neither term meant that only a base term β_0 was estimated. For gas, the analysis assumed no gas cooling.[6] The gas model for a household therefore included only heating or only a base term.

For each set of reference temperatures, the normal-year HDD and CDD were calculated at all weather stations. The appropriate normal-year HDD and CDD series for each household were applied to the household's estimated coefficients from the DDN model to provide the predicted NAC, which formed the basis for the subsequent end-use analysis. The next section discusses the DDN models for the survey respondents, while the following section discusses the normalized consumption predicted for those households.

The DDN analysis used the full set of 76 weather stations as described but results in this section have been aggregated and summarized at the T24 building climate zone level for ease of review and comparison to the previous RASS studies.

Electric and Gas DDN Models

The distribution of households with electric and gas DDN models by T24 building climate zone and utility is shown in **Table 36** and **Table 37**. Only households with insufficient or unrealistic billing series did not have DDN models.

6 The CDD term is empirically small, and there is no gas cooling in single-family homes. Where there is no strong cool-weather-related trend, the CDD regression coefficient will reflect any deviations from the best-fit equation, essentially fitting to noise.

Table 36: Number of Households with Electric DDN Models by T24 CZ and Utility

T24 Building Climate Zone Representative City	Number of CDA Weather Stations	PG&E	SCE	SDG&E	SMUD	LADWP	Total
1 (Arcata)	2	173	-	-	-	-	173
2 (Santa Rosa)	3	923	-	-	-	-	923
3 (Oakland)	6	3,041	-	-	-	-	3,041
4 (San Jose-Reid)	2	1,357	-	-	-	-	1,357
5 (Santa Maria)	2	307	-	-	-	-	307
6 (Torrance)	8	-	1,742	228	-	337	2,307
7 (San Diego Lindbergh)	5	-	-	2,130	-	-	2,130
8 (Fullerton)	5	-	1,952	58	-	285	2,295
9 (Burbank-Glendale)	2	-	1,630	-	-	1,495	3,125
10 (Riverside)	5	-	1,254	769	-	-	2,023
11 (Red Bluff)	6	717	-	-	5	-	722
12 (Sacramento)	10	2,056	-	-	1,257	-	3,313
13 (Fresno)	5	844	245	-	-	-	1,089
14 (Palmdale)	4	-	524	20	-	-	544
15 (Palm Springs-Intl)	5	-	193	6	-	-	99
16 (Blue Canyon)	6	106	186	-	-	54	346
All California	**76**	**9,524**	**7,726**	**3,211**	**1,262**	**2,171**	**23,894**

Source: 2019 California Residential Appliance Saturation Survey

Table 37: Number of Households with Gas DDN Models by T24 CZ and Utility

T24 Building Climate Zone Representative City	Number of CDA Weather Stations	PG&E	SCG	SDG&E	LADWP	Total
1 (Arcata)	2	126	-	-	-	126
2 (Santa Rosa)	3	843	-	-	-	843
3 (Oakland)	6	3,032	-	-	-	3,032
4 (San Jose-Reid)	2	1,344	70	-	-	1,414
5 (Santa Maria)	2	-	313	-	-	313
6 (Torrance)	8	-	1,870	-	337	2,207
7 (San Diego Lindbergh)	5	-	-	2,028	-	2,028
8 (Fullerton)	5	-	2,269	-	285	2,554
9 (Burbank-Glendale)	2	-	3,113	-	1,495	4,608
10 (Riverside)	5	-	1,634	735	-	2,369
11 (Red Bluff)	6	565	-	-	-	565
12 (Sacramento)	10	3,630	-	-	-	3,630
13 (Fresno)	5	736	537	-	-	1,273
14 (Palmdale)	4	52	272	-	-	324
15 (Palm Springs-Intl)	5	-	261	2	-	263
16 (Blue Canyon)	6	10	164	-	54	228
All California	**76**	**10,338**	**10,503**	**2,765**	**2,171**	**25,777**

Source: 2019 California Residential Appliance Saturation Survey

Table 38 shows how the type of best-fit model for electric varied by T24 building climate zone. Two thirds (67 percent) of the households have a reaction to both HDD and CDD (that is, heating and cooling loads). Nearly a quarter (23 percent) had a reaction only to CDD, and 29 percent had a reaction to HDD.

Table 38: Best-Fit DDN Electric Model Type by T24 Building Climate Zone

T24 Building Climate Zone Representative City	Number of CDA Weather Stations	Heating and Base Load #Resp	Heating and Base Load %Resp	Cooling and Base Load #Resp	Cooling and Base Load %Resp	Heating, Cooling, and Base #Resp	Heating, Cooling, and Base %Resp	Total
1 (Arcata)	2	136	79%	12	7%	25	14%	173
2 (Santa Rosa)	3	184	20%	154	17%	585	63%	923
3 (Oakland)	6	1,267	42%	378	12%	1,396	46%	3,041
4 (San Jose-Reid)	2	190	14%	228	17%	939	69%	1,357
5 (Santa Maria)	2	134	44%	34	11%	139	45%	307
6 (Torrance)	8	436	19%	475	21%	1,396	61%	2,307
7 (San Diego Lindbergh)	5	180	8%	488	23%	1,462	69%	2,130
8 (Fullerton)	5	166	7%	579	25%	1,550	68%	2,295
9 (Burbank-Glendale)	2	179	6%	1,010	32%	1,936	62%	3,125
10 (Riverside)	5	37	2%	486	24%	1,500	74%	2,023
11 (Red Bluff)	6	27	4%	128	18%	567	79%	722
12 (Sacramento)	10	102	3%	594	18%	2,617	79%	3,313
13 (Fresno)	5	10	1%	283	26%	796	73%	1,089
14 (Palmdale)	4	9	2%	132	24%	403	74%	544
15 (Palm Springs-Intl)	5	1	1%	55	28%	143	72%	199
16 (Blue Canyon)	6	75	22%	73	21%	198	57%	346
All California	**76**	**3,133**	**29%**	**5,109**	**23%**	**15,652**	**67%**	**23,894**

Source: 2019 California Residential Appliance Saturation Survey

The model calculated increasing consumption with colder weather below the HDD setpoint and increasing consumption with hotter weather above the CDD setpoint. Some variation across the T24 building climate zones was found, as shown in **Table 39**. There was generally good agreement between the HDD setpoint for households with and without a CDD setpoint; similarly, CDD setpoints agreed between households with and without an HDD setpoint. For California as a whole, the average setpoint for cold weather (HDD) was 62° Fahrenheit (F), while the average hot weather set-point was 68°F.[7] This variation in set points demonstrated the value of choosing the best setpoints for each household rather than using 65°F for both HDD and CDD.

7 All temperatures are reported in degrees Fahrenheit.

Table 39: Average Degree-Day Reference Temperature (°F) by Electric Best-Fit Model Type and T24 Building Climate Zone

T24 Building Climate Zone Representative City	Number of CDA Weather Stations	Heating and Base Load (N=3,113) HDD Ref.	Cooling and Base Load (N=5,109) CDD Ref.	Heating, Cooling, and Base Load (N=15,652) HDD Ref.	Heating, Cooling, and Base Load (N=15,652) CDD Ref.
1 (Arcata)	2	60.7°	66.0°	59.3°	65.9°
2 (Santa Rosa)	3	61.4°	65.1°	58.8°	67.°
3 (Oakland)	6	62.°	65.8°	60.2°	67.6°
4 (San Jose-Reid)	2	62.4°	65.3°	59.8°	67.3°
5 (Santa Maria)	2	61.°	65.7°	60.2°	66.8°
6 (Torrance)	8	63.4°	66.8°	62.1°	68.2°
7 (San Diego Lindbergh)	5	65.1°	66.1°	61.7°	68.3°
8 (Fullerton)	5	64.°	66.8°	62.2°	68.6°
9 (Burbank-Glendale)	2	63.1°	68.4°	62.3°	69.4°
10 (Riverside)	5	64.2°	66.4°	60.3°	68.2°
11 (Red Bluff)	6	66.2°	66.7°	59.7°	68.6°
12 (Sacramento)	10	64.6°	65.8°	59.8°	68.2°
13 (Fresno)	5	67.7°	67.5°	60.4°	69.8°
14 (Palmdale)	4	67.6°	68.2°	61.1°	71.2°
15 (Palm Springs-Intl)	5	62.°	72.2°	65.1°	74.9°
16 (Blue Canyon)	6	64.5°	67.3°	60.7°	69.8°
All California	**76**	**63.7°**	**66.9°**	**60.9°**	**68.7°**

Source: 2019 California Residential Appliance Saturation Survey

Gas DDN Models

As shown in **Table 40**, all of the best fit DDN models for gas were heating plus base load models. This was expected given that most residential gas was used for household heating and water heating. 100 percent of households with gas had a cold-weather-dependent term (heating and base load model).

Heating reference temperatures were on average 62.8°F, with some variation by T24 building climate zone.

Table 40: Average Degree-Day Reference Temperature (°F) by Gas Best-Fit Model Type and T24 Building Climate Zone

T24 Building Climate Zone Representative City	Number of CDA Weather Stations	Heating and Base Load #Resp	Heating and Base Load HDD Ref Temp °F
1 (Arcata)	2	126	58.0°
2 (Santa Rosa)	3	843	60.2°
3 (Oakland)	6	3,032	60.8°
4 (San Jose-Reid)	2	1,414	61.3°
5 (Santa Maria)	2	313	60.3°
6 (Torrance)	8	1,870	63.3°
7 (San Diego Lindbergh)	5	2,028	64.2°
8 (Fullerton)	5	2,269	64.3°
9 (Burbank-Glendale)	2	3,113	64.5°
10 (Riverside)	5	2,369	62.7°
11 (Red Bluff)	6	565	62.9°
12 (Sacramento)	10	3,630	61.9°
13 (Fresno)	5	1,273	63.3°
14 (Palmdale)	4	324	62.2°
15 (Palm Springs-Intl)	5	263	67.4°
16 (Blue Canyon)	6	174	65.1°
All California	**76**	**23,606**	**62.8°**

Source: 2019 California Residential Appliance Saturation Survey

NAC Results

The NAC estimates derived from the electric and gas DDN models were generally close to the actual annualized consumption. Electric DDN models estimated NAC about 1 percent greater than the actual consumption for the sample. Gas DDN models predicted about 0.2 percent less gas consumption than was actually billed. These differences reflect that the normal temperatures were consistent with the actual test year (October 1, 2018, to September 30, 2019) temperatures.

CDA Variables and Data Imputation Process

This section addresses additional treatment of survey responses required for the conditional demand analysis (CDA) to produce unbiased unit energy consumption (UEC) estimates. The CDA was restricted to individually metered accounts; therefore, the following discussion applies only to individually metered survey responses.

The following processes are covered in this section.

- Creation of binary variables indicating the presence of a specific end use of the fuel being modeled
- Creation of continuous variables to reflect intensity of system use
- Data imputation processes for missing values

In this section, variables that received similar processing are grouped together. The first group includes four variables: square footage, age of dwelling, number of residents, and household income. These four variables were critical to all the UEC estimates; therefore, the study team gave special attention to the process used to impute missing values. Next, the space heating and water heating variables are discussed, with special attention given to the process used to identify fuel misreporting. The creation of indicator and continuous variables is also discussed. Finally, the refrigerators and freezers are discussed with attention to initial engineering estimates for energy use of these appliances.

Advanced Variable Imputation Process

The survey variables for square footage, household income, age of the dwelling, and number of residents were critical to developing UEC estimates. Because of the importance to the analysis, for these four variables it was essential to minimize bias that may result from imputing missing values. The team imputed missing values using a regression-based approach as described in the following steps:

1. The team created a binary variable for each of the four variables that took the value of 1 if the respondent answered the question and 0 if not.
2. The indicator variables served as dependent variables in a logistic regression used to estimate the likelihood of response to the specific question. Other survey responses served as the independent variables to the logistic regression.
3. The probability of response to each question was based on the logistic regression and used to estimate an inverse Mills' ratio for each respondent for each of the four questions.
4. The team estimated a linear regression model to provide a predicted response value for those who did not answer the question. The inverse Mills’ ratio was included in the linear regression as an explanatory variable, controlling for nonresponse bias.

If a survey was missing any of the explanatory variables needed to apply the regression-based approach, the team imputed the missing values using a conditional means process, which involved calculating the mean value by dwelling type.

Square Footage and Surface Area

The survey collected data on square footage in the SQFT variable for a series of size ranges. To use the size ranges in the CDA, the size range categories had to be converted to a continuous series, which was recorded in the SQFT_A variable. This variable typically used the midpoint of each size range to provide an estimate of the square footage of the dwelling. For responses in three of the size ranges, the following sizes were assumed:

- Dwellings in the smallest size group, less than 250 square feet, were assumed to be 200 square feet.
- Dwellings that were between 4,001–5,000 square feet were assumed to be 4,700 square feet.
- Dwellings greater than 5,000 square feet were assumed to be 6,000 square feet.

If SQFT was missing, the team imputed the value using the means value by dwelling type.

Once the continuous square foot series was assigned to each respondent, these estimates were converted to a new variable, AREA, which provided an estimate of the surface area of the dwelling. The surface area was calculated using the following equations from the *2003 RASS* presented below.

For single-story, single-family dwellings and mobile homes:

$$surface\ area = 5.9985 * SQFT\_A^{0.85288}$$

For multistory, single-family dwellings

$$surface\ area = 13.9694 * SQFT\_A^{0.73955}$$

For multifamily dwellings

$$surface\ area = 0.5955 * SQFT\_A^{1.10344}$$

The team imputed missing values for the SQFT_A variable using the regression-based approach outlined above.

Household Income

The survey collected data on household income in the INCOME variable for a series of income ranges. The team converted the INCOME variable into a continuous variable AVGINC, which was the midpoint of each of the income ranges, except for the highest income group of $150,000 or more. Responses in this income range were set to $175,000. If INCOME was missing, AVGINC was imputed using the mean value by dwelling type.

Dwelling Age

The survey collected data on the year a dwelling was built in the BUILTYR variable, to which respondents selected from a series of age ranges. The BUILTYR variable was converted into a continuous BUILT variable, which was the midpoint of each of the age ranges, except for the oldest and most recent age ranges. Responses in these categories were assigned values of 1935 and 2007, respectively.

During the data imputation, a new variable HOMEAGE was created that contained the value of BUILT or an imputed value using the regression-based approach discussed above. If no regression value could be derived, then HOMEAGE was imputed with the average age by dwelling type. In addition, the binary variable NEWHOME was set equal to 1 for all dwellings built after 2000 and zero otherwise.

Number of Household Residents

The survey collected data on the number of residents by age group in variables for each age group. These responses were summed to create a count of the total number of people in the household, RESCNT. RESCNT was required for the CDA, so the team imputed missing values using the regression-based approach to create a new variable (NUMI) set equal to RESCNT or the imputed value when RESCNT was missing. The CDA required a log transformation of the NUMI variable, so a new variable was created (NHH) and set equal to (log (NUMI + 1)).

Number of Electric Vehicles of Each Type (NUMBEV, NUMHEV)

If the respondent indicated that he or she owns or leases an EV in A20 but had missing data in this question, the missing values were replaced. If only one of the answers was missing, the number for that type of vehicle was set to 0. If both were missing, and the responses to A22, the number of miles driven question, was also missing, then EVYN was reset to "No." If both were missing, but the respondent did respond to A22, then the number of each was set to the proportion of that vehicle type in California, which is 0.56 for battery-electric vehicles (BEV) and 0.44 for plug-in hybrid electric vehicles (HEV).[8] This approach assumes that there is only one vehicle, but the combined number will reflect the appropriate proportions.

Electric Vehicle Miles Driven on an Average Weekday/Weekend Day (EVMDMF, EVMDSS)

If these responses were missing, they were filled in with the average for all other EV drivers who answered the question.

Electric Vehicle Charging Frequency at Home, Work, and Somewhere Else (EVTMCHH, EVTMCHW, EVTMCHO)

If all three responses were missing, they were left missing, but the engineering estimate assumed that all charging was done at home. If one or two responses were missing, then they were set to 0 charges per week. These values were converted to charges per week, then converted to a proportion of charging at home.

Correcting for Fuel Misreporting

Previous CDA studies conducted on the California residential population have shown that misreporting of fuels used for heating and water heating was common, particularly in multifamily units and areas with very low consumption. The variables used to model each observation must reflect an accurate profile of each observation for the statistical technique used for this CDA to provide accurate results. Since space and water heat account for such a large share of an individual's energy consumption, it is critical to identify cases in which a respondent inaccurately misreported whether he or she has electric or gas space and water heat.

After the fuel switching validation was completed, the study team derived binary and continuous variables from survey responses to identify the presence of electric and gas systems and the degree to which systems were used. These variables are also covered below.

---

8 Source: Auto Alliance data for California, at https://autoalliance.org/energy-environment/advanced-technology-vehicle-sales-dashboard/.

Space Heating

The space heating section gathered information for primary and secondary heating systems, as well as temperature settings that indicate intensity of use. Survey responses concerning an individual's primary and secondary space heating systems were cross-referenced with the following information to determine whether survey responses were accurate.

Primary Heat

Natural Gas Line

As discussed above, the variable NGLINE1 recorded whether a home was found to have natural gas by cross-referencing survey and utility information. For respondents who reported having primary electric heat, the presence of a gas line provides evidence that they may actually have gas heat.

Significant Coefficient on HDD From the Electric DDN

If a household's electric consumption was responsive to an increase in heating degree days, then the electric DDN model for that household was likely to have a significant coefficient on HDD. While those without a heating term may have electric heat that they do not use, the team considered the presence of a heating coefficient in the electric DDN model an indication of fuel misreporting.

The specific rules used to determine fuel misreporting are outlined below. The rules apply only to households for which the variable indicating that they pay for heat (PAYHEAT2) was "yes." The fuel misreporting rules did not apply to households that did not pay for heat because survey respondents asked to fill out the heating section only if they paid for heat. The variable PHTFUEL2 recorded the imputed primary heat fuel.

The following rules were applied to households that reported having primary electric heat and were identified as having gas in the home:

- If there was not a significant heating parameter in the electric DDN model, then PHTFUEL2 = 1 (gas)
- If the utility does not identify them as having electric heat, then PHTFUEL2 = 1 (gas).
- If there was a significant heating term in the electric DDN model and the utility indicated it has electric heat, then PHTFUEL2 = 2 (electric). A flag was constructed for these respondents to determine whether their CDA parameter estimates differed from those of other respondents with electric heat. Therefore, the only way a household with a gas line was allowed to have electric heat was if both the DDN model and the electric utility showed evidence of electric heat.

The following rules were applied to people who reported having primary electric heat who were identified as not having a gas line to the home based on the cleaning process outlined in earlier in this chapter.

- If there was not a significant heating parameter in the electric DDN model, then PHTFUEL2 = 6 (other).
- If there was a significant heating parameter in the electric DDN model, then PHTFUEL2 = 2 (electric). A flag was constructed for cases in which the utility did not provide

indication of electric heat to determine whether these CDA parameter estimates differed from those of other respondents with electric heat.

Table 41 shows the distribution of respondents according to reported PHTFUEL and cleaned PHTFUEL2.

Auxiliary Heat

Respondents were only allowed to have alternative electric heat if one of the following was true:

- The alternative electric heating was baseboard heating.
- The alternative electric heating was a heat pump with primary wood stove or fireplace.
- The alternative electric heating was a portable heater.
- For these cases, the imputed alternative heat variable (AHFUEL2) was set to 2 (electric), while for all other cases it was set to 6 (other).

Table 42 shows the distribution of respondents according to *AHTFUEL* and *AHTFUEL2*.

Space Heating Binary Variables

The following indicator variables were also derived from survey responses to reflect the presence or absence of each respective space heating technology:

Electric Heat

- DEHEAT — If the household pays for heat and the primary heating system is conventional electric heat, DEHEAT was set equal to one, zero otherwise.
- DEHP — If the household pays for heat and the primary heating system is a heat pump, DEHP was set equal to one, zero otherwise.
- NONELEBK — If the household has a primary electric heater and a non-electric backup, NONELEBK was set equal to one, zero otherwise.
- DEAUXHT — If the household has an additional electric heater, DEAUXHT was set equal to one, zero otherwise.
- ROOM — If the household has electric heat and the primary heater is a resistance heater, a through the wall heat pump, or a portable heater, ROOM was set equal to one, zero otherwise.

Table 41: Primary Space Heating Fuel Cleaning

PHTFUEL	PHTFUEL2 -Natural Gas	PHTFUEL2 -Electric	PHTFUEL2 -Bottled Gas	PHTFUEL2 -Wood	PHTFUEL2 -Solar	PHTFUEL2 -Other	PHTFUEL2 -Missing	PHTFUEL2 -N/A	Total	% Total
Natural Gas	25,739	-	-	-	-	-	-	-	25,739	64.9%
Electric	-	7,130	-	-	-	-	-	-	7,130	18.0%
Bottled Gas	-	-	1,302	-	-	-	-	-	1,302	3.3%
Wood	-	-	-	707	-	-	-	-	707	1.8%
Other	-	-	-	-	-	76	-	-	76	0.2%
Missing	18	20	-	2	4	-	2,841	17	2,902	7.3%
N/A	-	-	-	-	-	-	-	1,826	1,826	4.6%
All	**25,757**	**7,150**	**1,302**	**709**	**4**	**76**	**2,841**	**1,843**	**39,682**	**100%**

Source: 2019 California Residential Appliance Saturation Survey

Table 42: Auxiliary Space Heating Fuel Cleaning

AHTFUEL	AHTFUEL2 Natural Gas	AHTFUEL2 Electric	AHTFUEL2 Bottled Gas	AHTFUEL2 Wood	AHTFUEL2 Solar	AHTFUEL2 Other	AHTFUEL2 Missing	AHTFUEL2 N/A	Total	% Total
Natural Gas	4,365	-	-	-	-	-	-	-	4,365	11.0%
Electric	-	3,407	-	-	-	10,472	-	-	13,879	35.0%
Bottled Gas	-	-	229	-	-	-	-	-	229	0.6%
Wood	-	-	-	1,365	-	-	-	-	1,365	3.4%
Solar	-	-	-	-	188	-	-	-	188	0.5%
Other	-	-	-	-	-	28	-	-	28	0.1%
Missing	-	-	-	-	-	-	1,822	-	1,822	4.6%
N/A	-	-	-	-	-	-	-	17,806	17,806	44.9%
All	**4,365**	**3,407**	**229**	**1,365**	**188**	**10,500**	**1,822**	**17,806**	**39,682**	**100.0%**

Source: 2019 California Residential Appliance Saturation Survey

Gas Heat

- DGHEAT — If the household pays for heat and the primary heating fuel is natural gas, DGHEAT was set equal to one, zero otherwise.
- NONGBU — If the household has a primary gas heater and a non-gas backup, NONGBU was set equal to one, zero otherwise.
- DNGAUXHT – If the household has an additional natural gas heater, DNGAUXHT was set equal to one, zero otherwise.
- GROOM – If the household has gas heat and the primary heater is a floor or wall furnace, GROOM was set equal to one, zero otherwise.
- SETBK – An indicator variable accounting for people who lower the thermostat setting at night. SETBK was set equal to one for respondents whose nighttime heater setting (HNITESET) was lower than the average setting (HTTSET), otherwise SETBK was set equal to zero.

Space Heating Continuous Variables

- HTTSET – The average daily thermostat temperature was set equal to the weighted average of each household's thermostat temperature for each period during the heating season. Missing values for this variable were imputed with the mean value by dwelling type.

Primary Water Heat

Survey responses concerning a household's primary water heater were cross-referenced with billing information to determine whether survey responses were accurate. The following steps were used to evaluate whether respondents that indicated they had an electric hot water heater actually had a gas hot water heater.

- Summer months were identified as the three warmest months of the year by climate zones.
- Any respondent that indicated having an electric hot water tank and had natural gas in the home were identified.
- The average monthly gas consumption over the three warmest months of the year for those households was calculated.
- Households with more than 10 therms per month over the summer were flagged by setting GWH_FLAG =1.
- When GWH_FLAG =1, the new variable PRWHFUEL2 was set to 1 (gas), and for all other households, PRWHFUEL2 was set = PRWHFUEL.

Table 43 shows a comparison of PRWHFUEL and PRWHFUEL2.

Table 43: Primary Water Heating Fuel Cleaning

PRWHFUEL	PRWHFUEL2 Natural Gas	PRWHFUEL2 Electric	PRWHFUEL2 Bottled Gas	PRWHFUEL2 Solar	PRWHFUEL2 Other	PRWHFUEL2 Missing	PRWHFUEL2 N/A	Total	% Total
Natural Gas	26,202	-	-	-	-	-	-	26,202	66.0
Electric	1,189	2,481	-	-	-	-	-	3,670	9.2
Bottled Gas	56	-	1,698	-	-	-	-	1,754	4.4
Solar	13	-	-	12	-	-	-	25	0.1
Other	5	-	-	-	34	-	-	39	0.1
Missing	2,649	-	-	-	-	3,155	-	5,804	14.6
N/A	992	-	-	-	-	-	1,196	2,188	5.5
All	**31,106**	**2,481**	**1,698**	**12**	**34**	**3,155**	**1,196**	**39,682**	**100.0**

Source: 2019 California Residential Appliance Saturation Survey

Water Heating Binary Variables

The following indicator variables were also derived from survey responses to reflect the presence or absence of electric or natural gas water heating:

The CDA contains a gas, an electric, and a solar water heating fuel indicator variable.

Electric Water Heating

- DEWH – Set equal to one for respondents with an electric hot water heater, zero otherwise.
- DWHSOLAR – Set equal to one for respondents with a solar hot water heater with an electric backup, zero otherwise.
- ADDWHEL – Set equal to one for respondents with more than one electric water heater, zero otherwise.

Gas Water Heating

- DGWH – Set equal to one for respondents with a gas hot water heater, zero otherwise.
- DGWHSOLAR – Set equal to one for respondents with a solar hot water heater with a natural gas backup, zero otherwise.

Water Heating Continuous Variables

The primary drivers of energy consumption for water heaters are clothes washers, dishwashers, and showers or baths. The following continuous variables were constructed to account for hot water usage due to these appliances:

- CWASHU – Clothes washer usage constructed from the number of loads per day by water temperature.
- DWASHU – Dishwasher usage constructed from number of loads per day.
- WHTSHWRS – Total number of baths and showers taken per day.

For respondents that did not answer the usage questions, DWASHU, CWASHU, or WHTSHRS was imputed using the mean value by dwelling type.

Binary and Continuous CDA Variables

This section reports on several survey variables for which continuous and binary variables were constructed. The binary variables reported on the presence of each respective appliance, while the continuous variables provided an indication of the amount or intensity of appliance use.

Central Air Conditioning

- DCAC – Set equal to one to indicate the presence of a central air conditioner or minisplit system including both heat pumps used for cooling and conventional central AC, zero otherwise.
- TSETC – Continuous variable for the weighted average of the thermostat temperature for each period during the cooling season. If the household had central air conditioning and did not report the temperature, the mean value by dwelling type was assigned.

Room Air Conditioning

- DRAC – Set equal to one to indicate the presence of room air conditioning, zero otherwise.
- RACCNT – Count of the number of room air conditioners.
- TSETUSE – Continuous variable for the weighted average of the room air conditioner use. If the household had a room air conditioner and did not report the temperature, the mean value by dwelling type was assigned.

Electric Vehicles

The variables related to electric vehicle charging were set up as follows.

- DEVCHG – Set equal to one to indicate the presence charging of an electric vehicle at the home.
- EVTYPEOFEVFACTOR – the sum of 1 multiplied by the number of BEV and 0.55 multiplied by the number of PEV. This sum reflects that PEV use an average of 55 percent of the household's energy from electricity.
- EVMILES – the miles driven per year, calculated based on EVMDMF and EVMDSS.
- PROPCHGHOME – Proportion of charging done at home in a typical week, based on the answers to EVTMCHH, EVTMCHW, and EVTMCHO.

Clothes Dryers

CDA variables for clothes dryers included the following indicator variables reporting the presence of electric or gas dryers and two continuous variables reporting the number of loads per week.

- GDRY — Set equal to one to indicate the presence of a gas clothes dryer that was not in a common area, zero otherwise.
- EDRY — Set equal to one to indicate the presence of an electric clothes dryer that was not in a common area, zero otherwise.
- GDRYU — The weekly usage of the gas dryer. If the survey response to DRYLDS was missing, and the household had a gas dryer, GDRYU was imputed using the mean value by dwelling type.
- EDRYU — The weekly usage of the electric dryer. If the survey response to DRYLDS was missing, and the household had an electric dryer, EDRYU was imputed using the mean value by dwelling type.

Outdoor Lighting

CDA variables for outdoor lighting included the following indicator and continuous variables.

- DOLT — Set equal to one to indicate the presence of exterior lighting, zero otherwise.
- OLTFIX — Total number of exterior fixtures.
- OPROPHID — Continuous variable for the proportion of exterior lighting fixtures that were HID lights.

- OPROPSEN — Continuous variable for the proportion of exterior lighting fixtures that were on sensors.
- OPROPTIM — Continuous variable for the proportion of exterior lighting fixtures that were on timers.

Televisions

CDA variables for televisions included the following indicator and continuous variables.

- DTV — Set equal to one to indicate the presence of either standard, small LCD, large LCD or plasma TV, zero otherwise.
- TVHRS — The sum of the total number of hours watching small and large screen TVs per day.
- TVKW — Variable that accounts for electricity use based upon number of hours of usage. Standard and small LCD TVs were assumed to use 0.1 kW per hour, and large screen LCD and plasma TVs were assumed to use 0.25 kW per hour. If the household had multiple types of TVs, the usage numbers were multiplied by the proportion of TVs of each type.

If the household had one or more televisions and usage information was missing, the mean value by dwelling type was assigned.

Personal Computers and Home Offices

The current RASS collected data on the number desktop and laptop PCs as well as the number of hours each was used. These variables were converted to the same variables used in the 2003 study to maintain continuity. For each:

- DPC — Set equal to one to indicate the presence of either a desktop or laptop personal computer, zero otherwise.
- PCHRS — Continuous variable for the sum of desktop and laptop PC hours.
- PCNUM — The sum of the number of hours of usage for desktop and laptop PCs.
- DHMOFF — Set equal to one to indicate that someone in the household operated a business or worked from home, zero otherwise.
- HMOFFHRS — Continuous variable for the numbers of hours a week someone works out of the home.

Where applicable, for any respondent who did not provide a response to any of the three continuous variables, PCHRS, PCNUM, and HMOFFHRS, the value was imputed with the mean value by dwelling type.

Swimming Pools

CDA variables for pools included the following indicator and continuous variables.

- DPLPMP — Set equal to one to indicate the presence of a pool if the respondent indicated that he or she pays for its energy use. Only single-family households were allowed to have pools. All other pools listed in the survey were assumed to be pools in common areas and were disallowed in the CDA.5F

- PLFILT — The number of hours per day used to filter the pool. This variable differs between summer months (May–October) and winter months (November–April).
- PLSIZE — The pool size variable was set to 18,000 gallons for small pools, 30,000 for medium-sized pools, and 42,000 for large pools.
- EPLHT — Set equal to one to indicate that the pool was heated with electric heat, zero otherwise.
- DGPLHT — Set equal to one to indicate that the pool was heated with natural gas, zero otherwise.
- GPLHTFREQ — The gas CDA also analyzed the effect of the frequency of pool heating. This variable was allowed to differ between summer and winter months.
- PLCOV — Set equal to one to indicate the use of a pool cover. A pool cover may reduce the heating needs because of an increase in pool temperature or it may indicate a pool that is used more frequently, leading to an increase in heating needs.

Spas and Hot Tubs

The CDA required the following indicator variables regarding the presence of a spa or hot tub and the fuel type.

- DSPA — Set equal to one if the respondent lived in a single-family house, townhouse, or mobile home, had a spa or hot tub, and paid for its energy use, was set equal to one, zero otherwise.
- DEHTSPA — Set equal to one for spas heated with electricity or solar with electric backup, zero otherwise.
- SPASOLAR — Set equal to one for spas heated by solar with electric backup, zero otherwise.
- DGHTSPA —S et equal to one for spas heated with natural gas or solar with natural gas backup, zero otherwise.
- SPAGSOLAR — Set equal to one for spas heated by solar with natural gas backup, zero otherwise.

The CDA also required the following continuous variables regarding spa filter and heat usage by fuel type and spa size.

- SPCOV — Set equal to one if the spa had an insulated cover, zero otherwise.
- SPAFREQ — The frequency of spa filtering.
- SPAEHTFREQ — The frequency of electric heating was allowed to differ between summer and winter months.
- SPAGHTFREQ — The frequency of natural gas heating was allowed to differ between summer and winter months.
- SPASIZE — Continuous variable based on the number of people the spa holds. The number of people was set to 2 for small spas, 5 for medium spas, and 8 for large spas.

Fans

The CDA incorporated three types of fans: forced air fans, attic fans, and ceiling fans.

- DFFAN — Set equal to 1 if the primary heating fuel was natural gas or bottled gas and the heater is a central heater, zero otherwise.
- DATTFAN — Set equal to 1 if the household has an attic or a whole-house fan, zero otherwise.
- DCEILF — Set equal to 1 if the household has at least one ceiling fan, zero otherwise.

Seasonal Home Indicator

The CDA accounted for differences in energy consumption between year-round and seasonal homes using the following variable:

- SEASONAL — Set equal to 1 for anyone that reported the residence was not their year-round home and lived there less than 12 months of the year, zero otherwise.

Double-Pane Windows and Dwellings in Colder Zones

- DPWIN — Set equal to 1 for respondents who indicated their WINDTYPE was all or mostly double paned or triple pane, or a mixture of triple, double, and single, zero otherwise.
- T24 Building Climate Zone — Set equal to 1 if the residence was in Climate Zone 1 or 16 and zero otherwise. T24 has building requirements that apply to new homes in CEUS Weather Zones 1, 161 and 162, which may offset the colder climates in these zones.

Kitchen Appliances

The following indicator variables were defined for kitchen appliances:

- DERGOV —Set equal to 1 for households with either an electric range or oven, zero otherwise.
- DGRGOV — Set equal to 1 for households with either a natural gas range or oven, zero otherwise.
- DMWV — Set equal to 1 for households that indicated they had a microwave oven, zero otherwise.

Laundry

- DCW — Set equal to 1 to indicate the presence of laundry equipment in the home and either a top-loading or a front-loading washer, zero otherwise.

Energy Consumption for Refrigerators and Freezers

Engineering estimates used in the CDA model accounted for differences in energy consumption of refrigerators and freezers with differing characteristics. This step was necessary because these appliances had roughly 100 percent saturation, eliminating differences among households with and without refrigerators and freezers. Without such differences, the statistical model used to estimate the UECs would be unable to identify the energy consumption of the appliances. Using predefined engineering estimates for refrigerators and freezers with different characteristics, the model was better able to detect variation in consumption among households with different refrigerators and freezers.

Due to the significant age of the previous estimates and changes to refrigerator-freezer standards since 2009, the engineering estimates for the current study were revised. Annual energy use estimates were generated using the U.S. Department of Energy's (DOE) ENERGY STAR® refrigerator calculator, which includes five refrigerator-freezer door configurations, five size-volume ranges, and six age ranges (linked to changes in appliance standards). Energy use values for all combinations were generated and then either mapped directly to the RASS survey combinations or adjusted to better align with the survey questions. For example, where the survey age or size range did not exactly match with the calculator ranges, then the study team used an average of two calculator-derived values. An adjustment to annual energy use was also made for secondary units because they are typically located in an unconditioned space.

CHAPTER 5: Data Analysis Method

The CDA used to derive electric and natural gas UEC estimates employed a statistically adjusted engineering (SAE) analysis modeling technique. The SAE model implemented was similar to that used in the *2009 RASS*,[9] in which engineering estimates were based on the *2003 RASS* equations.

This chapter is organized into the following sections:

- Overview of the approach used to construct the CDA, a statistically adjusted engineering model
- Derivation of electric and natural gas engineering estimates, the regression terms in the CDA
- Specification of the CDA model
- Model results

Statistically Adjusted Engineering Analysis

Household energy consumption was decomposed into the demand from various end uses using a regression-based SAE model. Engineering estimates of UECs (engineering UECs) were used as initial point estimates for each end use, such that all end uses in the household had an engineering UEC. The sum of the engineering UECs provided an initial estimate of total consumption for the household. The engineering UECs for each household served as the independent variables in a regression equation, where the dependent variable was the actual energy consumption for each household.

Equation 2 provides the general form of the SAE model used to estimate UECs. The team developed separate models for estimating consumption for electricity and natural gas end uses. Household energy consumption was equal to the sum of engineering UECs for all energy-consuming end uses multiplied by scalar adjustment factors (β_i) for each end use, plus residual unexplained error.

9 KEMA, Inc. 2010. *2009 California Residential Appliance Saturation Study*. California Energy Commission. Publication Number: CEC- 200-2010-004. http://web.archive.org/web/20190601194456/https://www.energy.ca.gov/appliances/rass/previous_rass.html

Equation 2: General Form of SAE Model

$$HHUEC_i = \sum_{j=1} (\beta j * ENGji) + \varepsilon i$$

where:

***HHUEC**$_i$* = Energy consumption for household i

***ENG** $_{ji}$* = Engineering UEC of electricity use for end-use j, for household i.

$\boldsymbol{\beta_j}$ = Estimated scalar adjustment parameter to the initial UEC for end-use j

ε_i = Error term

The scalar adjustments $\beta_{\mathbf{j}}$ were statistical adjustments made to each engineering UEC. These scalar adjustments were determined as coefficients from the linear regression. A scalar adjustment of 1 indicated that the engineering UEC provided an exact measure of the amount of energy used by a given end use. A scalar adjustment greater than 1 indicated the engineering UEC understated the actual consumption derived from the respective end use, that is, the initial estimate needed to be increased. Conversely, a scalar adjustment of less than 1 indicated that the engineering UEC overstated the actual consumption derived from the respective end use, in other words, the engineering UEC needed to be decreased. A negative coefficient implied that an end use reduced energy consumption.

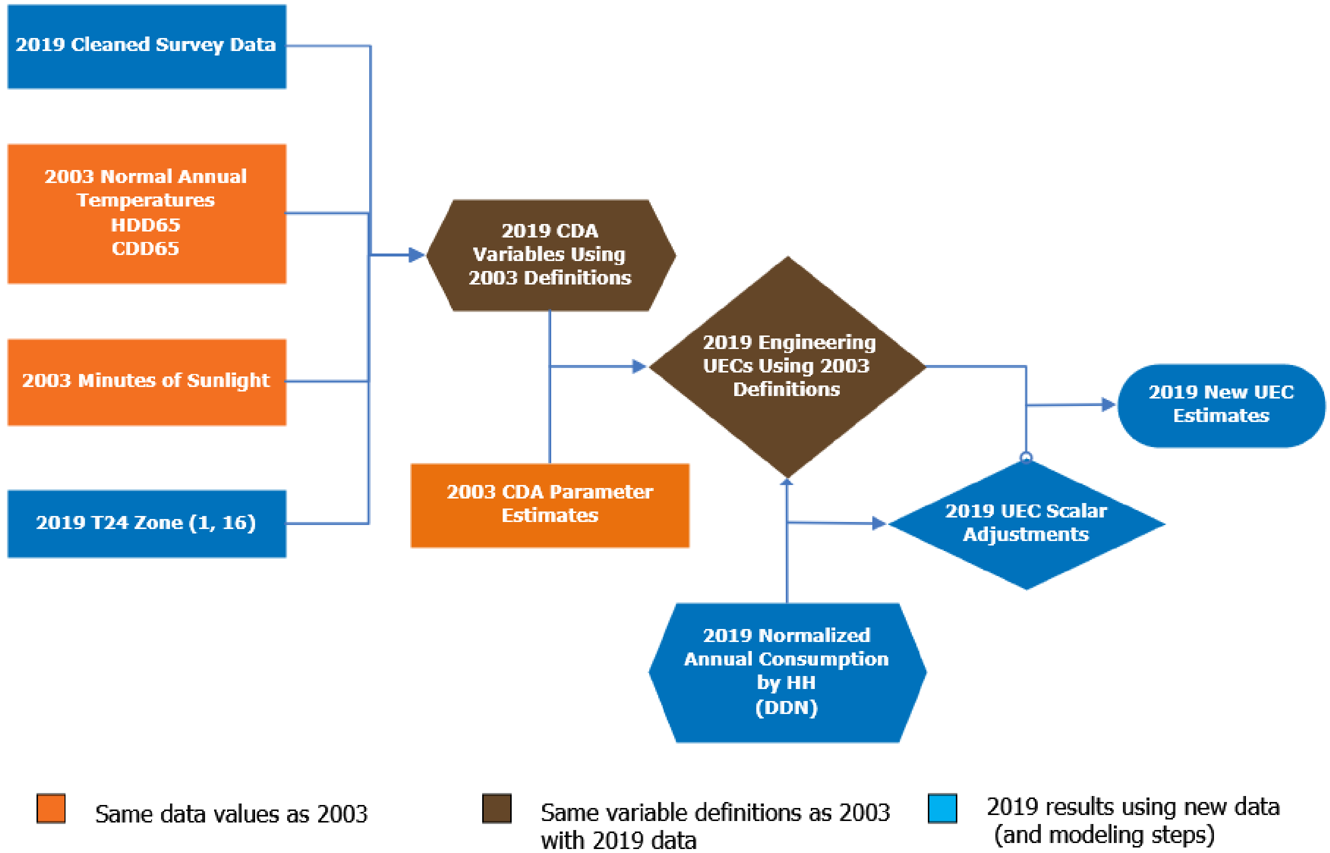
The study team developed the SAE model for RASS using the process illustrated in **Figure 6**. The data inputs on the left include survey data, normal-year temperatures along with degree-days, and minutes of sunlight. Also, a single indicator variable T24 identified households in Building Climate Zones 1 or 16, which have more restrictive building codes. Survey data provided information on end uses of each fuel, as well as demographic and housing characteristics at the household level.

The current data inputs were combined to create the CDA variables using the same calculations as for the CDA for the *2003 RASS*, as they were in 2009. The CDA variables consist of linear combinations of appliance and equipment stocks, structural features of the home, building shell and equipment efficiency factors, weather conditions, and utilization patterns.[10] The current CDA variables were then multiplied by the 2003 CDA parameter estimates and combined to yield initial UEC estimates for each end use. These initial UEC estimates served as the engineering estimates in the SAE model. The exception was the electric vehicle UEC estimates, which were added this year, based on new engineering estimates.

---

10 *California Statewide Residential Appliance Saturation Study Final Report*. June 2004. CEC 400-04-009. http://web.archive.org/web/20190601194456/https://www.energy.ca.gov/appliances/rass/previous_rass.html

Figure 6: Overview of SAE Process



Source: 2019 California Residential Appliance Saturation Survey

The SAE model used NAC as the dependent variable. The NAC was derived from monthly billing data using the degree-day normalization (DDN) model outlined in Chapter 4. The NAC values for the households were regressed against the respective engineering UECs to provide the scalar adjustments. The scalar adjustments for each end use were multiplied by the corresponding initial engineering estimates to provide adjusted UEC estimates for each end use.

While not shown in **Figure 6**, the final step of the process was to calibrate the adjusted UECs so that the sum of the final UECs was equal to the observed total NAC. This calibration was done at the sampling-strata level, which included information identifying the electric utility, presence of electric heat, and home type. Because of the sparseness of some of the sampling strata, the 151 strata were collapsed together into 30 strata to avoid adjustments based on small (or zero) sample sizes and provide more stability.

The team estimated the SAE model using only full-year residents, but the final 2009 UEC estimates contained both full-year and partial-year residents. The team calibrated the final new 2009 UEC estimates to average annual consumption by sampling strata from the combined series of full-year and partial-year residents.

Derivation of End-Use Engineering Estimates

As in the 2009 CDA, the CDA equations from the *2003 RASS* were used to develop the engineering estimates, which were then adjusted by the 2003 coefficients to get the 2009 inputs, which were used as a starting point for this analysis. Derivation of the UEC estimates is contained in the 2003 report.[11] This section presents the resulting UEC formulas that identify the source of the engineering estimates for the SAE model.

Electric End-Use Engineering Estimates

Engineering estimates were derived for each of the electric end uses listed below.

- Primary space heating
- Secondary space heating
- Central air conditioning
- Room air conditioning
- Evaporative coolers
- Electric vehicles
- Water heating
- Primary refrigerators
- Secondary refrigerators
- Freezers
- Ranges and ovens
- Microwave ovens
- Dishwashers
- Clothes washers
- Dryers
- Outdoor lighting
- Televisions
- Home offices
- Personal computers
- Swimming pool pumps
- Spa pumps
- Spa heat
- Well pumps
- Forced air fans
- Miscellaneous

Each engineering UEC was the sum of one or more cross-product terms, times the corresponding 2003 CDA parameter estimates. The cross-product terms were products of binary variables that indicated the presence of each end use and basic quantitative variables,

11 KEMA-Xenergy, Inc. June 2004. *California Statewide Residential Appliance Saturation Study*. Prepared for the CEC
http://web.archive.org/web/20190601194456/https://www.energy.ca.gov/appliances/rass/previous_rass.html

such as surface area, heating or cooling degree-days, temperature setting, number of units, usage information, or preset engineering parameters. Additional continuous variables differentiated UECs for households according to income level and number of residents in households. Indicator variables were also used to provide separate UEC estimates by residence type, the presence of dual-paned windows, and seasonal effects. While cross-product terms for some of the simpler engineering equations reduced to just one variable, several of them consisted of multiple cross-product terms with several variables in each term. Because only a small number of the variables were relevant to a given household, the sum of these terms was essentially a basic multiplicative formula, with varying adjustments applied depending on the relevant variables.

For example, the general form of the formula for primary space heating is defined in Equation 3. The binary variable DEHEAT indicated whether electric heat was present in the household. In the 2009 CDA, the study team included a fixed term for the assumed efficiency of the heating technology, with a heat pump having a higher efficiency. This year, the study team added a coefficient to the actual CDA model that used the data to estimate the relative efficiency of heat pumps. The term A denotes a vector overall adjustment, depending on factors such as dwelling type, new construction, and thermostat settings. Finally, the term B represents the surface area of the home, the number of heating degree-days, or the product of the two.

Equation 3: General Form of Primary Space Heating

Space Heating = (DEHEAT) * A * B

Each of the electric end-use engineering UECs is presented below along with a brief description of some of the variables specific to each UEC. The following variables were used in multiple engineering UECs. Detailed descriptions for each of the individual variables used are in Chapter 4.

- HDD65 — Normal heating degree-days with a base of 65 degrees
- CDD65 — Normal cooling degree-days with a base of 65 degrees
- AREA — Surface area of the residence
- DPWIN — Indicator variable for dual-paned windows
- MF — Indicator variable for multifamily residence
- INC — Continuous variable for household income
- WINTER — Constant adjustment for proportion of winter months
- SUMMER — Constant adjustment for proportion of summer months
- (LOG NUMI + 1) — The number of people in the household, entering the equation in the form (1+log of the number)
- T24 — Indicator variable for household located in Building Climate Zone 1 or 16

Space Heating

The study team developed primary electric space-heating engineering UECs for both conventional electric (EHT_ENG) heat and electric heat with heat pump (EHP_ENG). In addition, engineering UECs were also developed for secondary (or auxiliary) space heating systems (AUXEH_ENG). **Table 44** presents the specific equations used to obtain engineering estimates for each of the space heating end uses.

Table 44: Electric Space Heating Engineering Estimates

	Cross Product Variable	Parameter from *2003 RASS*	Equation
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h	3.30E-05	DHEAT*HDD65*AREA
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_dwp	-8.39E-05	DEHEAT*HDD65*AREA*DPWIN
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_mf	-0.00112	DEHEAT*HDD65*AREA*MF
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_inc	-2.90E-10	DEHEAT*HDD65*AREA*INC
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_inc_dwp	1.77E-10	DEHEAT*HDD65*AREA*INC*DPWIN
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_inc_mf	2.01E-11	DEHEAT*HDD65*AREA*INC*MF
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_rm	-3.42E-05	DEHEAT*HDD65*AREA*ROOM
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_rm_dwp	2.35E-05	DEHEAT*HDD65*AREA*ROOM*DPWIN
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_rm_mf	1.54E-04	DEHEAT*HDD65*AREA*ROOM*MF
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_sbk	-7.48E-06	DEHEAT*HDD65*AREA*SETBK
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_sbk_dwp	-1.52E-05	DEHEAT*HDD65*AREA*SETBK*DPWIN
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_sbk_mf	5.88E-05	DEHEAT*HDD65*AREA*SETBK*MF
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_set	3.50E-06	DEHEAT*HDD65*AREA*HTTSET
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_set_dwp	-1.64E-07	DEHEAT*HDD65*AREA*HTTSET*DPWIN
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_set_mf	1.86E-05	DEHEAT*HDD65*ARE*HTTSET*MF
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_h_nonebu	4.83E-05	DEHEAT*HDD65*ARE*HTTSET*MF*NONELEBK
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_winter	0.18559	DEHEAT*AREA*WINTER
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_winter_minsun	-2.55E-04	DEHEAT*AREA*WINTER*MINSOFLIGHT
Primary heating (EHT_ENG and EHP_ENG)	eht_sq_T24_h	-4.06E-05	DEHEAT*AREA*HDD65*T24
Primary heating (EHT_ENG and EHP_ENG)	eth_H_Seasonal	-0.15854	DEHEAT*HDD65*SEASONAL
Secondary Heating (AUXHT_ENG)	eht_aux_h	0.01261	DEAUXHT*HDD65
Secondary Heating (AUXHT_ENG)	eht_aux_sq_h	3.40E-05	DEAUXHT*HDD65*AREA
Secondary Heating (AUXHT_ENG)	eht_aux_sq_h_mf	-1.02E-05	DEAUXHT*HDD65*AREA*MF
Secondary Heating (AUXHT_ENG)	eht_aux_sq_h_freq	1.78E-06	DEAUXHT*HDD65*AREA*ADDFREQ

Source: 2019 California Residential Appliance Saturation Survey

In the 2009 CDA, the primary heating engineering estimates identified households with conventional heating systems and those with systems with a heat pump by using an efficiency factor (1/EFFH), which was set to 0.5 for heat pumps, thereby reducing the space heating engineering UEC. This year, the study team separated the conventional systems and heat pumps and added another coefficient to the CDA regression model that will use the data to capture the actual difference in efficiency between heat pumps and conventional systems.

For primary and auxiliary heating systems, the main driver of the engineering estimates was heating degree-days with a base temperature of 65°F (HDD65). The study team used the updated weather station mappings to ensure the most accurate normal HDD65 values for each household. Primary heating system estimates contained additional terms used to adjust the effect of HDD65 on heating usage, depending on the minutes of sunlight, winter months, and whether the residence was a seasonal residence. Additional terms included in the primary heating system estimates allowed variation in the thermostat setting, building shell, dwelling type, and household income level. Variation in auxiliary heating system engineering estimates was limited to differences in surface area of the home, dwelling type, and thermostat setting (*ADDFREQ*).

Space Cooling

Space cooling engineering estimates were developed for central air conditioning (CAC_ENG), room air conditioners (RAC_ENG), and evaporative (swamp) coolers (SWAMP_ENG). **Table 45** presents the specific equations used to obtain engineering estimates for each of these end uses. Many of the terms used in space heating engineering estimates were also used for cooling, but in place of HDD65, the variable CDD65 was used to represent normal cooling degree-days, with a base temperature of 65°F. In addition, the central air conditioning equation included a term for new homes. As with the normal HDD65 series, the CDD65 series from the *2003 RASS* was used for this study.

Table 45: Electric Space Cooling Engineering Estimates

	Cross Product Variable	Parameter from *2003 RASS*	Equation
Central Air Conditioning (CAC_ENG)	cac_sq_c	0.00149	DCAC*CDD65*AREA
Central Air Conditioning (CAC_ENG)	cac_sq_c_new	4.85E-05	DCAC*CDD65*AREA*NEWHOME
Central Air Conditioning (CAC_ENG)	cac_sq_c_dwp	-1.20E-04	DCAC*CDD65*AREA*DPWIN
Central Air Conditioning (CAC_ENG)	cac_sq_c_mf	1.05E-03	DCAC*CDD65*AREA*MF
Central Air Conditioning (CAC_ENG)	cac_sq_c_inc	9.42E-11	DCAC*CDD65*AREA*INC
Central Air Conditioning (CAC_ENG)	cac_sq_c_inc_new	-1.68E-10	DCAC*CDD65*AREA*INC*NEWHOME
Central Air Conditioning (CAC_ENG)	cac_sq_c_inc_dwp	1.25E-10	DCAC*CDD65*AREA*INC*DPWIN
Central Air Conditioning (CAC_ENG)	cac_sq_c_inc_mf	-2.11E-09	DCAC*CDD65*AREA*INC*MF
Central Air Conditioning (CAC_ENG)	cac_sq_c_tset	-1.52E-05	DCAC*CDD65*AREA*TSETC
Central Air Conditioning (CAC_ENG)	cac_sq_c_tset_new	-2.14E-07	DCAC*CDD65*AREA*TESTC*NEWHOME
Central Air Conditioning (CAC_ENG)	cac_sq_c_tset_dwp	9.03E-07	DCAC*CDD65*AREA*TSETC*DPWIN
Central Air Conditioning (CAC_ENG)	cac_sq_c_tset_mf	-1.01E-05	DCAC*CDD65*AREA*TSETC*MF
Central Air Conditioning (CAC_ENG)	cac_sq_minsun_sum	1.00E-04	DCAC*AREA*MINSOFLIGHT*SUMMER
Central Air Conditioning (CAC_ENG)	cac_sq_evp_sum	0.01272	DCAC*AREA*DSWAMP*SUMMER
Central Air Conditioning (CAC_ENG)	cac_c_evp_sq	-1.69E-04	DCAC*CDD65*DSWAMP*AREA
Central Air Conditioning (CAC_ENG)	cac_sq_sum	-0.07495	DCAC*AREA*SUMMER
Room Air Conditioning (RAC_ENG)	rac_sq_c	5.15E-05	DRAC*CDD65*AREA
Room Air Conditioning (RAC_ENG)	rac_sq_c_dwp	-1.87E-05	DRAC*CDD65*AREA*DPWIN
Room Air Conditioning (RAC_ENG)	rac_sq_c_mf	1.13E-05	DRAC*CDD65*AREA*MF
Room Air Conditioning (RAC_ENG)	rac_sq_c_inc	-5.83E-10	DRAC*CDD65*AREA*INC
Room Air Conditioning (RAC_ENG)	rac_sq_c_tsetu	1.81E-05	DRAC*CDD65*AREA*TSETUSE
Room Air Conditioning (RAC_ENG)	rac_sq_c_rcnt	1.60E-05	DRAC*CDD65*AREA*RACCNT
Room Air Conditioning (RAC_ENG)	rac_c_evp_sq	-8.93E-05	DRAC*CDD65*DSWAMP*AREA
Evaporative coolers (SWAMP_ENG)	swamp_sq_c	6.35E-05	DSWAMP*AREA*CDD65
Evaporative coolers (SWAMP_ENG)	swamp_c	0.19156	DSWAMP*CDD65

Source: 2019 California Residential Appliance Saturation Survey

Table 46 shows two sets of thermostat settings used for space cooling. The variable TSETC referred to the average cooling temperature for central air conditioning, while TSETUSE was the frequency in which room air conditioners were used.

Water Heating

Engineering estimates were derived for conventional electric water heating (WHT_ENG) and solar water heating with electric backup (WHTS_ENG). **Table 46** presents the equation for the water heating engineering estimates, which were distinguished by the presence (or absence) of a tank with solar. The primary driver of the water heating UEC was the number of people in the household, shown by entering the equation in the form (1+log of the number). The number of people in the household also factored into dishwasher usage, clothes washer usage, and the number of showers taken per day.

The equation also included a measure of the average temperature difference from month to month. Because the *2019 RASS*, like the *2009 RASS*, used degree-day normalized **annual** consumption data as opposed to **monthly** consumption data, monthly temperature differences were not present in the dataset. Therefore, the average monthly temperature difference by climate zone from the *2003 RASS* was used for *WHTEMP_DIFF*. The FACTAWH term from the 2003 CDA was used to adjust for seasonal variation; this variable was equal to a constant for the 2019 CDA because the 2019 CDA, like the 2009 CDA, was based on annual consumption data.

Table 46: Electric Water Heating Engineering Estimates

	Cross Product Variable	Parameter From *2003 RASS*	Equation
WHT_ENG and WHTS_ENG	ewh_dwash	28.89343	DEWH*FACTAWH*DWASHU
WHT_ENG and WHTS_ENG	ewh_cwash	9.98225	DEWH*FACTAWH*CWASHU
WHT_ENG and WHTS_ENG	ewh_shw	18.4293	DEWH*FACTAWH*WHTSHWRS
WHT_ENG and WHTS_ENG	ewh_solar	-127.56103	DEWH*FACTAWH*DWHSOLAR
WHT_ENG and WHTS_ENG	ewh_add	15.96034	DEWH*ADDWHEL*FACTAWH
WHT_ENG and WHTS_ENG	ewh_num	42.08176	DEWH*FACTAWH*Log(NUMI+1)
WHT_ENG and WHTS_ENG	ewh_num_mf	-73.10609	DEWH*FACTAWH*Log(NUMI+1)*MF
WHT_ENG and WHTS_ENG	ewh_difftemp1	0.03581	DEWH*FACTAWH*WHTEMP_DIFF
WHT_ENG and WHTS_ENG	ewh	73.0256	DEWH*FACTAWH

Source: 2019 California Residential Appliance Saturation Survey

Electric Vehicles

Electric vehicle engineering estimates were based on the type of vehicles in the home, the total number of miles driven, and the proportion of charging done at home. The calculation for the engineering estimate of the consumption due to EV charging (ELEV_VEH) in the home is shown in Equation 4.

Equation 4: Energy Use Due to EV Charging

ELEC_VEH = DEVCHG*TypeofEVFactor*EVMiles*PropChgHome*0.3

DEVCHG is an indicator variable for the presence of EV charging in the home. TYPEOFEVFACTOR is a variable that is set to be 1.0*NUMBEV + 0.55*NUMHEV, since plug-in hybrid EVs get about 55 percent of the total energy from electricity.[12] EVMILES is the total miles driven in a year, and PROPCHGHOME is the proportion of charging done at home, as described in the previous chapter. The average EV uses 18.6 kWh/100 km,[13] which is equal to 18.6 kwh/62.1371 mi, which is 0.30 kWh/mi.

Refrigerators and Freezers

Refrigerator and freezer engineering estimates were primarily based on the DOE's ENERGY STAR refrigerator calculator as discussed in Chapter 4. **Table 47** presents the equations for refrigerators and freezers. The second refrigerator estimate also contained terms for variation during the summer months and for multifamily homes.

Table 47: Refrigerator and Freezer Engineering Estimates

	Cross Product Variable	Parameter from *2003 RASS*	Equation
First Refrigerator (RF1_ENG)	ref1_use	0.0833	DRF1*REFUSAGE1
Second Refrigerator (RF2_ENG)	ref2_use	0.1366	DRF2*REFUSAGE2
Second Refrigerator (RF2_ENG)	ref2_use_sum	-0.00404	DRF2*SUMMER*REFUSAGE2
Second Refrigerator (RF2_ENG)	ref2_use_mf	-0.053	DRF2*REFUSAGE2*MF
Freezer (FZ_ENG)	fz_use	0.12464	DFRZR*FZUSAGE

Source: 2019 California Residential Appliance Saturation Survey

Kitchen

Kitchen appliance engineering estimates were developed for ranges and conventional ovens (RNG_ENG), microwaves (MW_ENG), and dishwashers (DWH_ENG). **Table 48** presents the equations for ranges, microwaves, and dishwashers. The primary driver for each of these appliances was the number of people in the household, shown by entering the equation in the form (1+log of the number). Engineering estimates for ranges and ovens were allowed to also vary by income and by the presence of a microwave oven. The 2003 CDA used the FACTAMI term to adjust for seasonal variation. For the 2019 CDA, like the 2009 CDA, FACTAMI was equal to a constant because the 2019 CDA was based on annual consumption data.

---

12 Source: U.S. DOE Alternative Fuels Data Center, https://afdc.energy.gov/vehicles/electric_emissions_sources.html.

13 Source: Electric Vehicle Database, https://ev-database.org/cheatsheet/energy-consumption-electric-car.

Table 48: Electric Kitchen Appliance Engineering Estimates

	Cross Product Variable	Parameter from *2003 RASS*	Equation
Range oven (RNG_ENG)	ecook_num	37.1557	DERNGOV*Log(NUMI+1)
Range oven (RNG_ENG)	ecook_num_inc	5.20E-05	DERNGOV*Log(NUMI+1)*INC
Range oven (RNG_ENG)	ecook_num_micor	-5.78601	DERNGOV*Log(NUMI+1)*MICRO
Range oven (RNG_ENG)	erngov	-22.0967	DERNGOV
Micro-wave (MW_ENG)	micwv	8.33	DMWV*FACTAMI*Log(NUMI+1)
Dishwasher (DWH_ENG)	edwash_num	9.89775	DDW*Log(NUMI+1)*FACTADW
Dishwasher (DWH_ENG)	Edw	-6.41515	DDW*FACTADW

Source: 2019 California Residential Appliance Saturation Survey

Laundry

Laundry included engineering estimates for clothes washers (CWS_ENG) and clothes dryers (EDY_ENG). Equations for these appliances appear in **Table 49**. Like kitchen appliances, the primary driver for clothes washers and dryers was the number of people in the household, entering the equation in the form (log of the number+1). The clothes dryer estimates also included a term for the number of loads per day. The FACTACW term was used in the 2003 CDA to adjust for seasonal variation; this variable was equal to a constant for the 2019 CDA because the 2019 CDA was based on annual consumption data.

Table 49: Electric Laundry Engineering Estimates

	Cross Product Variable	Parameter from *2003 RASS*	Equation
Clothes washer (CWS_ENG)	ecwash_num	37.09798	DCW*FACTACW*Log(NUMI+1)
Clothes washer (CWS_ENG)	Ecw	-40.09798	DCW*FACTACW
Clothes Dryer (EDY_ENG)	edry_use	16.78199	DEDRY*FACTADR*EDRYU
Clothes Dryer (EDY_ENG)	edry_num	5.5022	DEDRY*FACTADR*Log(NUMI+1)
Clothes Dryer (EDY_ENG)	Edry	-27.02423	DEDRY*FACTADR

Source: 2019 California Residential Appliance Saturation Survey

Outdoor Lighting

The engineering estimate for outdoor lighting (OLT_ENG) was derived using the equation in **Table 50**. The binary variable DOLT indicated whether outdoor lighting was present, while the variable OLTFIX provided a count of the number of outdoor lighting fixtures. The formula allowed for differentiation based on the number of outdoor fixtures that used CFLs and high-intensity discharge (HID) bulbs, or fixtures on sensors and timers. The variable HRDK indicated the number of hours of darkness in the climate zone.

Because of multicollinearity problems, the *2003 RASS* and this study did not estimate a separate indoor lighting UEC. Indoor lighting was assumed to be part of the Miscellaneous UEC. Indoor lighting was treated the same way in the 2009 CDA and the 2019 CDA.

Table 50: Outdoor Lighting Engineering Estimates

	Cross Product Variable	Parameter From *2003 RASS*	Equation
OLT_ENG	olit_cfl	-5.65594	DOLT*OLTFIX*ONOCFL
OLT_ENG	olit_hid	5.26879	DOLT*OLTFIX*OPROPHID
OLT_ENG	olit_sen	-4.17967	DOLT*OLTFIX*OPROPSENS
OLT_ENG	olit_tim	11.10408	DOLT*OLTFIX*OPROPTIM
OLT_ENG	olit_hrdk	2.11248	DOLT*OLTFIX*HRDK
OLT_ENG	olt	-20.00278	DOLT*OLTFIX

Source: 2019 California Residential Appliance Saturation Survey

Home Electronics and Office

While separate engineering estimates were derived in the *2009 RASS* for televisions (CTV_ENG), home offices (OFF_ENG), and personal computers (PCS_ENG), they are presented together in **Table 51**. The equations for each of these engineering estimates contained terms to include the number of hours of use. In addition, the estimate for televisions assumed a value to differentiate the energy consumption per hour between large- and small-screen televisions. Personal computers included a term for the total number of desktop and laptop computers in the household. The 2019 CDA used these same estimates, using the parameters from the 2003 CDA.

Table 51: Home Electronics and Office Equipment Engineering Estimates

	Cross Product Variable	Parameter from *2003 RASS*	Equation
TV's (CTV_ENG)	tvhrs_kw	36.48776	DTV*TVKW*TVHRS
TV's (CTV_ENG)	tv_kw	99.84392	DTV*TVKW
Home office (OFF_ENG)	ehmoffuse	0.80713	DHMOFF*HMOFFHRS
Home office (OFF_ENG)	hmoff	-0.712	DHMOFF
Personal computers (PCS_ENG)	pc_num	16.48716	DPC*PCNUM
Personal computers (PCS_ENG)	pc_num_hrs	1.68823	DPC*PCNUM*PCHRS 1
Personal computers (PCS_ENG)	epc	6.52058	DPC

Source: 2019 California Residential Appliance Saturation Survey

Swimming Pool and Spa

Engineering estimates were developed for swimming-pool filter pumps (PMP_ENG), spa filter (SPA_ENG) pumps, and spa heaters SPH_ENG). Each of these engineering estimates was a function of the frequency of use, as shown in **Table 52**. Swimming-pool filter use was indicated by PLFILT, while spa filter and spa heat use were indicated by SPAFREQ and SPAHTFRQ, respectively. Additional terms were added to the spa heat estimate to account for a cover or a combined electric and solar spa heating system.

Table 52: Swimming Pool and Spa Engineering Estimates

	Cross Product Variable	Parameter from *2003 RASS*	Equation
Pool filter pump (PMP_ENG)	plpmp_flt	-17.9017	DPLPMP*PLFILT
Pool filter pump (PMP_ENG)	plpmp_flt_sz	0.00116	DPLPMP*PLFILT*PLSIZE
Pool filter pump (PMP_ENG)	plpmp	177.43949	DPLPMP
Spa filter pump (SPA_ENG)	spa_pmp	1.8575	DSPA*SPAFREQ
Spa filter pump (SPA_ENG)	spa_pmp_sz	0.6434	DSPA*SPAFREQ*SPASIZE
Spa heat (SPH_ENG)	espa_ht_freq	4.11848	DEHTSPA*SPAEHTFREQ
(SPH_ENG)	espa_ht_freq_sz	-0.19491	DEHTSPA*SPAEHTFREQ*SPASIZE
Spa filter pump (SPA_ENG)	espa_ht_sz_cov	7.22828	DEHTSPA*SPASIZE*SPCOV
Spa filter pump (SPA_ENG)	espa_ht_solar	6.29138	DEHTSPA*SPASOLAR

Source: 2019 California Residential Appliance Saturation Survey

Well Pump

Well pumps (WPM_ENG) are used in areas that do not have municipal water. The well-pump engineering estimate was based entirely on the number of people in the household, entering the equation in the form (log of the number+1), as shown in **Table 53**.

Table 53: Well-Pump Engineering Estimates

	Cross Product Variable	Parameter from *2003 RASS*	Equation
WPM_ENG	wellpuse	55.41209	DWELLP*Log(NUMI+1)
WPM_ENG	wellp	0.64884	DWELLP

Source: 2019 California Residential Appliance Saturation Survey

Forced Air Fan

The CDA for the *2003 RASS* included an estimate for forced-air furnace fans (VENT1_ENG). **Table 54** shows that this engineering estimate was based solely on HDD65 and the surface area of the home. This same estimate was used for the 2019 CDA.

Table 54: Forced-Air Fan Engineering Estimates

	Cross Product Variable	Parameter from *2003 RASS*	Equation
VENT1_ENG	fafan_sq_h	2.30E-05	DFFAN*HDD65*AREA

Source: 2019 California Residential Appliance Saturation Survey

Miscellaneous Uses

The engineering estimate for the miscellaneous UEC (MISC_ENG) was calculated using the equation presented in **Table 55**. It accounted for all energy consumption not captured by the other UECs. The terms used in this engineering estimate include a combination of demographic, structural, and seasonal variables. In addition, parameters for attic or ceiling fans were included to avoid collinearity with the cooling terms.

Table 55: Miscellaneous Electric Engineering Estimates

	Cross Product Variable	Parameter From *2003 RASS*	Equation
MISC_ENG	miss_inc	3.09E-04	INC
MISC_ENG	miss_sq	0.04769	SQFT
MISC_ENG	miss_numi	43.11824	Log(NUMI+1)
MISC_ENG	miss_newh	-42.01492	NEWHOME
MISC_ENG	miss_mf	-8.54592	MF
MISC_ENG	miss_seasonal	-142.36973	SEASONAL
MISC_ENG	miss_ceil	19.19172	DCEILF
MISC_ENG	fat_c	0.35164	DATTFAN*CDD65
MISC_ENG	fat_sq_c	-7.05E-05	DATTFAN*CDD65*AREA
MISC_ENG	miss_epl_ht	88.18653	EPLHT

Source: 2019 California Residential Appliance Saturation Survey

Natural Gas End-Use Engineering Estimates

Engineering estimates were derived for each of the natural gas end uses listed below. While a general description of the formulas used to create estimates for each end use is presented below, derivation of these formulas can be found in the *2003 RASS* report.[14]

- Primary space heating
- Secondary space heating
- Water heating
- Ranges and ovens
- Clothes dryers
- Swimming pools and spas
- Miscellaneous

Space Heating

Table 56 presents the equations for primary (GHT_ENG) and secondary (auxiliary) (GAUXHT_ENG) natural gas space heating. The terms used to estimate primary natural gas heating were like those used for electric heating estimates but included terms for system age and whether the residence was a new home or mobile home.

14 The *2003 RASS* report misprinted parameters for some of the terms in the 2003 natural gas CDA. Estimates presented reflect the correct CDA results.

Table 56: Gas Space-Heating Engineering Estimates

	Cross Product Variable	Parameter From *2003 RASS*	Equation
Primary heating (GHT_ENG)	ght_sq_T24_winter	0.000238	DGHEAT*AREA*WINTER*T24
Primary heating (GHT_ENG)	ght_sq_T24_h	-1.60E-05	DGHEAT* HDD65* AREA *T24
Primary heating (GHT_ENG)	ght_sq_h	-2.68E-07	DGHEAT*HDD65* AREA
Primary heating (GHT_ENG)	ght_sq_h_new	-6.70E-06	DGHEAT*HDD65* AREA *NEWHOME
Primary heating (GHT_ENG)	ght_sq_h_age	-1.90E-06	DGHEAT*HDD65* AREA *GHTAGE
Primary heating (GHT_ENG)	ght_sq_h_dwp	-2.50E-06	DGHEAT*HDD65* AREA *DPWIN
Primary heating (GHT_ENG)	ght_sq_h_mf	-4.00E-05	DGHEAT*HDD65* AREA *MF
Primary heating (GHT_ENG)	ght_sq_h_inc	4.73E-11	DGHEAT*HDD65* AREA *INC
Primary heating (GHT_ENG)	ght_sq_h_inc_new	6.42E-12	DGHEAT*HDD65* AREA *INC*NEWHOME
Primary heating (GHT_ENG)	ght_sq_h_inc_age	-6.31E-13	DGHEAT*HDD65* AREA *INC*GHTAGE
Primary heating (GHT_ENG)	ght_sq_h_inc_dwp	-1.97E-11	DGHEAT*HDD65* AREA *INC*DPWIN
Primary heating (GHT_ENG)	ght_sq_h_inc_mf	-1.11E-11	DGHEAT*HDD65* AREA *INC*MF
Primary heating (GHT_ENG)	ght_sq_h_rm	2.26E-06	DGHEAT*HDD65* AREA *GROOM
Primary heating (GHT_ENG)	ght_sq_h_rm_age	-3.13E-07	DGHEAT*HDD65* AREA *GROOM*GHTAGE
Primary heating (GHT_ENG)	ght_sq_h_rm_dwp	4.56E-06	DGHEAT*HDD65* AREA *GROOM*DPWIN
Primary heating (GHT_ENG)	ght_sq_h_rm_mf	2.27E-06	DGHEAT*HDD65* AREA *GROOM*MF
Primary heating (GHT_ENG)	ght_sq_h_sbk	-5.18E-07	DGHEAT*HDD65* AREA *SETBK
Primary heating (GHT_ENG)	ght_sq_h_sbk_age	-1.32E-07	DGHEAT*HDD65* AREA *SETBK*GHTAGE
Primary heating (GHT_ENG)	ght_sq_h_sbk_dwp	1.73E-06	DGHEAT*HDD65* AREA *SETBK*DPWIN
Primary heating (GHT_ENG)	ght_sq_h_sbk_mf	4.95E-06	DGHEAT*HDD65* AREA *SETBK*MF
Primary heating (GHT_ENG)	ght_sq_h_set	5.36E-07	DGHEAT*HDD65* AREA *HTTSET
Primary heating (GHT_ENG)	ght_sq_h_set_age	3.04E-08	DGHEAT*HDD65* AREA *HTTSET*GHTAGE
Primary heating (GHT_ENG)	ght_sq_h_set_dwp	-6.13E-08	DGHEAT*HDD65* AREA *HTTSET*DPWIN
Primary heating (GHT_ENG)	ght_sq_h_set_mf	5.96E-07	DGHEAT*HDD65* AREA *HTTSET*MF
Primary heating (GHT_ENG)	ght_sq_h_nonebu	-1.70E-06	DGHEAT*HDD65* AREA *NONGBU
Primary heating (GHT_ENG)	ght_sq_winter	0.01694	DGHEAT* AREA *WINTER

	Cross Product Variable	Parameter From *2003 RASS*	Equation
Primary heating (GHT_ENG)	ght_sq_winter_minsun	-2.30E-05	DGHEAT* AREA *WINTER*MINSOFLIGHT
Primary heating (GHT_ENG)	ght_h_age	-0.00847	DGHEAT*HDD65*GHTAGE
Primary heating (GHT_ENG)	ght_h_new	0.00104	DGHEAT*HDD65*NEWHOME
Primary heating (GHT_ENG)	gth_h_seasonal	-0.00771	DGHEAT*HDD65*SEASONAL
Primary heating (GHT_ENG)	ght_sq_h_mh	5.23E-06	DGHEAT*HDD65*AREA *MH
Primary heating (GHT_ENG)	ght_sq_h_inc_mh	-4.42E-11	DGHEAT*HDD65*AREA*INC*MH
Secondary Heating (GAUXHT_ENG)	ght_aux_h	0.022054	DNGAUXHT*HDD65
Secondary Heating (GAUXHT_ENG)	ght_aux_sq_h	0.000003812	DNGAUXHT*HDD65* AREA
Secondary Heating (GAUXHT_ENG)	ght_aux_sq_h_mf	-0.000001903	DNGAUXHT*HDD65* AREA *MF

Source: 2019 California Residential Appliance Saturation Survey

Water Heating

Table 57 presents the equation used to produce engineering estimates of natural gas water heating. The primary differences between the electric and natural gas estimates were the natural gas equation terms for seasonal variation and new homes, while the electric estimate contained terms for multifamily households and whether an electric water heater was added in the past year.

Kitchen

Natural gas kitchen appliances were limited to ranges and ovens (GRNG_ENG). **Table 58** shows the equations used to create engineering estimates for these appliances. The same terms were used as for electric ranges and ovens.

Table 57: Gas Water-Heating Engineering Estimates

	Cross Product Variable	Parameter From *2003 RASS*	Equation
GWHT_ENG and GWHTS_ENG	gwh_num	-1.17111	DGWH*FACTAWH* Log(NUMI+1)
GWHT_ENG and GWHTS_ENG	gwh_dwash	0.65463	DGWH*FACTAWH*DWASHU
GWHT_ENG and GWHTS_ENG	gwh_cwash	0.45847	DGWH*FACTAWH*CWASHU
GWHT_ENG and GWHTS_ENG	gwh_solar	-2.67182	DGWH*FACTAWH*DWHGSOLAR
GWHT_ENG and GWHTS_ENG	gwh_num_new	-3.13922	DGWH*FACTAWH*LOG(NUMI+1)*NEWHOME
GWHT_ENG and GWHTS_ENG	gwh_num_seasonal	-9.0196	DGWH*FACTAWH* Log(NUMI+1)*SEASONAL
GWHT_ENG and GWHTS_ENG	Gwh	1.40E+01	DGWH*FACTAWH
GWHT_ENG and GWHTS_ENG	gwh_difftemp1	0.009662794	DGWH*FACTAWH*WHTEMP_DIFF
GWHT_ENG and GWHTS_ENG	gwh_shw	0.21075	DGWH*FACTAWH*TOTAL_SHTSHWRS

Source: 2019 California Residential Appliance Saturation Survey

Table 58: Gas Range or Oven Engineering Estimates

	Cross Product Variable	Parameter From *2003 RASS*	Equation
GRNG_ENG	gcook_num	6.31481	DGRNGOV* Log(NUMI+1)
GRNG_ENG	gcook_num_inc	-3.11E-06	DGRNGOV* Log(NUMI+1)*INC
GRNG_ENG	gcook_num_micor	-1.24E+00	DGRNGOV* Log(NUMI+1)*MICRO
GRNG_ENG	dgrngov	-3.18E+00	DGRNGOV

Source: 2019 California Residential Appliance Saturation Survey

Laundry

Engineering estimates for natural gas clothes dryers were estimated using the equation shown in Table 59. The terms were the same as the ones used for electric clothes dryers.

Table 59: Gas Clothes Dryer Engineering Estimates

	Cross Product Variable	Parameter From *2003 RASS*	Equation
GDRY_ENG	gdry_use	0.6391	DGDRY*FACTADR*GDRYU
GDRY_ENG	gdry_num	0.50575	DGDRY*FACTADR* Log(NUMI+1)
GDRY_ENG	Gdry	-1.53717	DGDRY*FACTADR

Source: 2019 California Residential Appliance Saturation Survey

Swimming Pool and Spa

The equations used to produce engineering estimates for natural gas heat for pools and spas are presented in **Table 60**. Both sets of equations were functions of size and frequency of use, as well as an adjustment for whether the pool or spa had a cover.

Table 60: Gas Heating Swimming Pool and Spa Engineering Estimates

	Cross Product Variable	Parameter From *2003 RASS*	Equation
GPHT_ENG)	gpl_ht	-1.30781	DGPLHT
GPHT_ENG)	gpl_ht_freq	2.76838	DGPLHT*GPLHTFREQ
GPHT_ENG)	gpl_ht_sz	0.00046	DGPLHT*PLSIZE
GPHT_ENG)	gpl_ht_sz_cov	0.000234	DGPLHT*PLSIZE*DPLCOV
Spa Heat (GSPA_ENG)	gspa_ht	3.5606	DGHTSPA
Spa Heat (GSPA_ENG)	gspa_ht_freq	0.81287	DGHTSPA*SPAGHTFREQ
Spa Heat (GSPA_ENG)	gspa_ht_freq_sz	0.00161	DGHTSPA*SPAGHTFREQ*SPASIZE
Spa Heat (GSPA_ENG)	gspa_ht_sz_cov	-0.12805	DGHTSPA*SPASIZE*SPCOV
Spa Heat (GSPA_ENG)	gspa_ht_solar	1.64078	DGHTSPA*SPAGSOLAR

Source: 2019 California Residential Appliance Saturation Survey

Miscellaneous Uses

Table 61 shows the engineering estimate for the natural gas miscellaneous UEC (GMISC_ENG) that contained terms for medical equipment and natural gas barbecues.

Table 61: Gas Miscellaneous Engineering Estimates

	Cross Product Variable	Parameter From *2003 RASS*	Equation
GMISC_ENG	miss_gmedical	2.70E+01	DGMED
GMISC_ENG	miss_gbbq	2.22319	DGBBQ

Source: 2019 California Residential Appliance Saturation Survey

Specification of CDA Models

The study team used the engineering estimates presented above to construct separate electric and natural gas SAE models. The basic model consisted of linear combinations of the respective electric and natural gas engineering estimates for each household. The intercept term was excluded from each model, thereby constraining household consumption to equal the sum of the individual engineering estimates plus residual error.

Because of collinearity among end-use terms, several end uses were combined in the electric and natural gas models. For combined terms, a scalar adjustment was estimated and applied to each UEC.

The derivations of the electric and natural gas SAE models are presented separately below.

Derivation of the Electric SAE

This section presents the specification of the electric SAE model. While many of the engineering estimates entered the model directly or as a binary variable multiplied by the engineering estimate, some end uses required additional manipulation. The study team used interaction terms to adjust primary electric heat and central air-conditioning estimates for households without heating and cooling DDN terms. In addition, several engineering estimates were combined because of multicollinearity. The treatment of each engineering estimate in the SAE model is discussed below.

Space Heating

The SAE model estimated scalar adjustments for primary and secondary electric heating systems. Primary electric heat entered the SAE model as three terms, with separate scalar adjustments for each term. The first term was the engineering estimate for conventional electric heat, and the second term was the engineering estimate for electric heat with a heat pump. In the 2009 CDA, these two were combined, with an assumed efficiency adjustment for the heat pumps of 0.5. Because there are many more heat pumps in use now, the study team estimated the adjustment factors for the two technologies separately, which allowed the data to determine the efficiency difference for heat pumps. The conventional electric heat engineering estimate was interacted with a binary variable (*DEHEAT*), identifying whether a conventional electric heating system was present in the household. A scalar adjustment was estimated for the conventional electric heat term *EHT_ENG_NEW,* as shown in Equation 5.

Equation 5: Electric Space Heating

$$EHT\_ENG\_NEW = DEHEAT * EHT\_ENG$$

Because the efficiency of heat pumps is being calculated in the CDA model based on the data, the heat pump engineering estimate is multiplied by a binary variable indicating the presence of a heat pump for primary electric heat in the household. A scalar adjustment was estimated for the conventional electric heat term *EHT_ENG_NEW,* as shown in Equation 6.

Equation 6: Electric Heat Pump

$$EHP\_ENG\_NEW = DEHP * EHP\_ENG$$

A separate term was added to the SAE model for secondary (or auxiliary) space heating. As shown in **Equation 7**, this term was set as equal to the engineering estimate (*AUXHT_ENG*) times the binary variable (*DEAUXHT*), indicating the presence of electric auxiliary heat.

Equation 7: Electric Auxiliary Space Heating

$$AUXHT\_ENG\_NEW = DEAUXHT * AUXHT\_ENG$$

Space Cooling

Space cooling end uses consisted of central air conditioning, room air conditioning, and evaporative (swamp) coolers. The central air conditioning SAE adjustment is shown in **Equation 8**. This term was an interaction of engineering estimates for central air conditioning with an indicator variable for people who had central air conditioning.

Equation 8: Central Air Conditioning

$$CAC\_ENG\_NEW = CAC\_ENG * DCAC$$

In addition, separate terms were also added for room air conditioning and evaporative cooler. As seen in Equation 9 and Equation 10, each of these terms simply consisted of the respective engineering estimate times a binary variable indicating the presence of either a room air conditioner or evaporative cooler.

Equation 9: Room Air Conditioning

$$RAC\_ENG\_NEW = RAC\_ENG * DRAC$$

Equation 10: Evaporative Cooling

$$SWAMP\_NEW = SWAMP\_ENG * DSWAMP$$

Electric Vehicles

Because the EV charging energy was calculated as described in the previous chapter, the engineering estimate for this is just a binary variable indicating the presence of EV charging multiplied by the charging energy estimate, as shown in Equation 11.

Equation 11: Electric Vehicles

$$EV\_NEW = EVCHGKWH * DEVCHG$$

Water Heating

The engineering estimates for stand-alone electric water heating and solar water heating with electric backup were combined, as seen in Equation 12.

Equation 12: Electric Water Heating

$$WHT\_ENG\_NEW = WHT\_ENG * DEWHT + WHTS\_ENG * DEWHTSOLAR$$

Kitchen Appliances

The range/oven, microwave, and dishwasher engineering estimates were collapsed into a single variable, *KITCHEN*. The estimated scalar adjustment for *KITCHEN* was applied to engineering estimates for each respective appliance, as represented in Equation 13.

Equation 13: Electric Kitchen Appliances

*KITCHEN = RNG_ENG * DERNGOV +MW_ENG * DMW + DWH_ENG * DWH*

Laundry Equipment

A new variable *LAUNDRY* was derived from the sum of clothes washer and electric clothes dryer engineering estimates, as seen in Equation 14.

Equation 14: Electric Laundry Equipment

*LAUNDRY = CWS_ENG * DCWS + EDY_ENG * DEDRY*

Spas

Estimates for spa filter pumps (*SPA_ENG*) and electric spa heaters (*SPH_ENG*) were combined into the single variable *SPA*. The new variable *SPA* was the sum of spa filter and spa heating engineering estimates multiplied by the respective binary variables. A single scalar adjustment was estimated for the term *SPA*.

Equation 15: Spa Filter Pumps and Electric Spa Heating

*SPA = SPA_ENG * DSPA +SPH * DESPAH*

Miscellaneous and Ventilation

The *2003 RASS* estimated separate UECs for forced-air fans and attic/ceiling fans but forced-air fans were estimated separately. The SAE model also combined the forced-air fan (*VENT1_ENG*) with the miscellaneous engineering estimate (*MISC_ENG*), providing a single scalar adjustment for the combined term.

Equation 16: Electric Miscellaneous and Ventilation

*MISC_ENG_NEW = MISC_ENG * DMISC + VENT1_ENG * DVENT1*

Home Office and PC

The two home office end uses were combined into one estimate. The home office (OFFUSE_ENG) and the PC (PCS_ENG) engineering estimates were combined, providing a single scalar adjustment for the combined term.

Equation 17: Electric Miscellaneous and Ventilation

*OFFPCS_NEW = OFFUSE_ENG * DHMOFF + PCS_ENG * DPC*

Other Electric SAE Terms

The remaining electric end uses were included in the SAE model by simply multiplying the engineering estimates by indicator variables that identified the presence of each end use. Each of the terms listed **Table 62** received a separate scalar adjustment.

Table 62: Electric End-Use Terms for SAE Model

Appliance	SAE Term
Outdoor Lights	*OLTUSE_NEW = OLTUSEeng * DOLT*
Televisions	*TVUSE_NEW = TVUSEeng * DTV*
Pool filter pump	*PMP_ENG_NEW = PMP_ENG * DPLPMP*
Well pump	*WPM_ENG_NEW = WPM_ENG * DWELLP*
First refrigerator	*RF1_ENG_NEW = RF1_ENG * DRF1*
Second refrigerator	*RF2_ENG_NEW = RF1_ENG * DRF2*
Freezer	*FZ_ENG_NEW = FZ_ENG * DFZ*

Source: 2019 California Residential Appliance Saturation Survey

Summary of SAE Electric Model

In summary, the final SAE electric model was expressed as the following equation:

Equation 18: SAE Electric Model

$$NAC_{kwh} = \beta_1 * EHT\_ENG\_NEW + \beta_2 * EHP\_ENG\_NEW +$$
$$\beta_3 * AUXHT\_ENG\_NEW + \beta_4 * CAC\_ENG\_NEW +$$
$$\beta_5 * RAC\_ENG\_NEW + \beta_6 * SWAMP\_NEW + \beta_7 * EV\_NEW +$$
$$\beta_8 * WHT\_ENG\_NEW + \beta_9 * KITCHEN + \beta_{10} * LAUNDRY +$$
$$\beta_{11} * SPA + \beta_{12} * OLTUSE\_NEW + \beta_{13} * TVUSE\_NEW$$
$$\beta_{14} * OFFPCS\_NEW + \beta_{15} * PMP\_ENG\_NEW$$
$$\beta_{16} * WPM\_ENG\_NEW + \beta_{17} * RF1\_ENG\_NEW + \beta_{18} * RF2\_ENG\_NEW +$$
$$\beta_{19} * FZ\_ENG\_NEW + \beta_{20} * MISC\_ENG\_NEW$$

Derivation of the Natural Gas SAE

The natural gas SAE was limited to three terms: space heating, water heating, and base load. The derivation of these terms is presented below.

Space Heating

Engineering estimates for primary and secondary natural gas space heating were combined to develop the natural gas space heating term, as seen in Equation 19. The equation shows engineering estimates for primary and secondary natural gas heating each multiplied by the respective binary variable.

Equation 19: Natural Gas Primary and Auxiliary Space Heating

$$GHT\_ENG\_NEW = GHT\_ENG * DGHEAT + GAUXHT\_ENG * DNGAUXHT$$

Water Heating

Similar to electric water heat, the engineering estimates for stand-alone natural gas water heat with natural gas backup for solar water heat were combined, as seen in Equation 20. As presented in Chapter 4, the variables *DGWHT* and *DGWHTSOLAR* were binary variables indicating whether a household had each respective water heater type.

Equation 20: Natural Gas Water Heating

$$GWHT\_ENG\_NEW = GWHT\_ENG * DGWHT + GWHTS\_ENG * DGWHTSOLAR$$

Natural Gas Base Load

The natural gas range/oven, clothes dryer, spa heat, pool heat, and miscellaneous (including gas BBQ and medical equipment) engineering estimates were combined to make up the BASE term, as seen in Equation 21. This term received a single scalar adjustment in the SAE model.

Equation 21: Natural Gas Base Load

$$GBASEUSE_{new} = GRNGOVUSE_{eng} * DGRNGOV + GDRYUSE_{eng} * DGDRY + GSPAHUSE_{eng} * DGSPAH + GPLHUSE_{eng} * DGPLH + GMISSUSE_{eng} * DGMISS$$

Summary of the SAE Natural Gas Model

In summary, the final SAE natural gas model was expressed as the following equation.

Equation 22: SAE Natural Gas Model

$$NACtherms = \beta_1 * GHT\_ENG\_NEW + \beta_2 * GWHT\_ENG\_NEW + \beta_3 * GBASE\_ENG\_NEW$$

Estimated Model Results

The electric and natural gas SAE models were estimated using an ordinary least squares method for households with a fitted DDN model. Compared with the 2009 study, the model performed better, without a need for as many manipulations or special coefficients. This is most likely due to the use of daily energy data from AMI data in the consumption normalization process. The relationship between daily energy and degree days is much stronger and more direct than the relationship between monthly billing energy and degree days.

All parameter estimates in the final electric model, shown in **Table 63**, were significant within 0.05 percent.

Table 63: Electric SAE Model Final Parameter Values

Variable	Parameter Estimate	Standard Error	t Value
Conventional Electric Heat	0.91550	0.07183	12.74
Heat Pump Electric Heat	0.72744	0.06968	10.44
Auxiliary Heat	1.74777	0.56960	3.07
Central Air	0.94826	0.02193	43.24
Room AC	2.83316	0.15975	17.73
Swamp Cooler	1.45163	0.08541	17.00
Electric Vehicles	0.33764	0.02542	13.28
Kitchen	1.14267	0.08878	12.87
Spa	0.60142	0.03311	18.17
Laundry	0.61531	0.04205	14.63
Water Heat	0.76845	0.03283	23.41
Outdoor Lights	0.15727	0.01632	9.64
TV	0.48347	0.02427	19.92
Home Office/PCs	0.28315	0.01983	14.28
Pool Pump	1.53627	0.11384	13.49
Well Pump	0.88076	0.02621	33.60
Misc	1.74284	0.06567	26.54
Refrigerator 1	0.95145	0.03460	27.49
Refrigerator 2	1.39691	0.07621	18.33
Freezer	0.91550	0.07183	12.74

Source: 2019 California Residential Appliance Saturation Survey

The three parameters in the natural gas model, seen in **Table 64**, were also highly significant. The natural gas heating parameter indicates a 20 percent reduction in the natural gas heating UEC from the initial engineering estimate. The adjusted estimate of water heating use was reduced by about 44 percent from the initial engineering estimate. The base consumption parameter indicates a 24 percent increase in the initial engineering estimates of all terms in the base load.

Table 64: Gas SAE Model Final Parameter Values

Variable	Parameter Estimate	Standard Error	t Value
Gas Heat	0.79930	0.01045	76.50
Water Heat	0.55879	0.01816	30.77
Base	1.24194	0.01288	96.39

Source: 2019 California Residential Appliance Saturation Survey

CHAPTER 6: Load Profiling Method

Overview

Load profiles are a valuable source of information for a variety of energy analysis objectives that range from cost allocation and energy efficiency program design to load forecasting and system planning. Load profiles improve understanding of how customers use electricity and provide concrete information about the strengths and weaknesses of energy conservation measures. Load usage patterns also assist in analyzing growth or stagnation within specific customer groups. Load forecasting uses data to estimate future load magnitudes. System planners generally use peak-demand data and average daily load curve data to assist in planning future production, transmission, and distribution requirements. Planning and evaluation of conservation and demand-side management strategies also depend heavily on load research data. With concerns over electric supply adequacy and market price volatility, load analysis is a key element in assessing the potential effectiveness of load management strategies for planners and determining actual effects of load management programs for implementers and evaluators.

Data Preparation

The data were prepared as follows:

- Excluded RASS households with less than 200 days of interval data in the 13 months comprised between October 1, 2018, and November 20, 2019.
- Excluded households that displayed sudden very large increases in hourly energy use that are commonly associated with interval load measurement errors.
- In addition, given the lead time necessary for the utilities to provide the interval data for the CDA and the load profiles, it was not possible to request interval data for all RASS households. Only households that had surveys ready for analysis on April 1, 2020, were included in load profiling, if the participating utilities provided their interval data.
- Four of the five participating electric utilities (PG&E, SCE, SDG&E and SMUD) provided electric interval data for the RASS.

Method

Most load profiles are estimated with one of two methods: mean-per-unit, or ratio.[15] Each of these can be applied to samples that are stratified or unstratified. The RASS load profiles were estimated using a very specific variant of the stratified mean-per-unit method, where individual weights vary within each stratum. In other words, two sample records in the same stratum can have different weights

15 For a thorough description of these methods, see Chapter 7: Data Analysis. *Load Research Manual, Third Edition*. Association of Edison Illuminating Companies. 2017.

The weights developed for RASS warrant a separate discussion, which is presented in Chapter 2. Using the weights for the entire set of survey respondents as a starting point, the weights used in load profiling were adjusted ("raked") to account for the explicit and implicit stratification variables of the sample and a reduced sample size (the number of survey respondents for which the study team had interval data lower than the number of survey respondents).

For RASS, the mean-per-unit (MPU) load profile estimates are stratified, weighted average demands (means) per household. These estimates can be expanded to the target population by multiplying by the number of households in the population. For any given hour, the mean estimated demand per household for the RASS stratified random sample is the sum of the individual household demands at that hour in that stratum, multiplied by their individual weights, divided by the sum of weights for the stratum. The sum of weights approximates the number of all households in the stratum in the population.

The formulas for calculating the RASS MPU stratified sample mean and variance are:

Equation 23

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} W_{ih}\, y_{ih}}{N_h}$$

Equation 24

$$s_{\bar{y}_{st}}^2 = \sum_{h=1}^{L} \frac{W_{ih}^2 s_{y_h}^2}{n_h}\left(1 - \frac{n_h}{N_h}\right)$$

where

W_{ih}	=	the weight of household *i* in stratum *h*
y_{ih}	=	the demand of household *i* in stratum *h*
N	=	the size of the total population at the time the sample was drawn
N_h	=	the number of the population customers in stratum *h* at the time the sample was drawn
n_h	=	the number of the sample customers in stratum *h*
$s_{y_h}^2$	=	the sample standard deviation of y within stratum *h*
L	=	the number of strata

LIST OF ACRONYMNS AND RELATED DEFINITIONS

Acronym	Definition
AC	Air conditioning — cooling system to control the humidity, ventilation, and temperature in a building.
ACS	American Community Survey — a survey conducted by the U.S. Census Bureau.
AMI	Advanced metering infrastructure — an integrated system of smart meters and other equipment that support two-way communication between the utility and the customer. Smart meters can record energy usage in short intervals throughout the day.
CAC	central air conditioning — a system where air is cooled at a central location and distributed to and from rooms by one or more fans and ductwork.
California ISO	California Independent System Operator — entity that oversees the operation of California's bulk electric power system, transmission lines, and electricity market generated and transmitted by its member utilities in California.
CARE	California Alternative Rates for Energy — is a program that provides discounts on electric and natural gas bills to low-income households.
CDA	conditional demand analysis — a statistical technique that combines utility consumption data with weather information and household survey data to produce energy consumption estimates by end use or equipment.
CDD	Cooling degree days — are a measure of how much (in degrees) and for how long (in days) the air temperature was above a certain reference temperature (i.e. 65°F). CDD are used in calculations of energy consumption for cooling a building.
CEC	California Energy Commission — established in 1975 and based in Sacramento, the CEC is primary energy policy and planning agency for California. It is committed to reducing energy costs, curtailing greenhouse gas emissions, and ensuring a safe, resilient, and reliable supply of energy.
CFL	compact fluorescent lamp — a fluorescent bulb designed to fit into a standard household light fixture. CFLs use less energy than the predecessors, incandescent bulbs.
DDN	degree-day normalization — statistical method of estimating annual energy consumption for normal weather conditions.

Acronym	Definition
DEER	Database for Energy Efficient Resources — database that provides information on the incremental energy savings associated with installing energy-efficient measures or equipment compared to what equipment is commonly installed.
DLP	digital light processing — the use of micromirrors to reflect light and color onto a screen. These micromirrors are positioned in a semiconductor chip and are very small.
DVR	digital video recorder — a consumer electronics device designed for recording video in a digital format within a mass storage device such as USB flash drive, hard disk drive, or any other storage device.
End Use	A category of equipment or appliance that uses energy and provides a benefit or a service to the user, (for example, space heating, space cooling, refrigerators).
EV	electric vehicle — a vehicle, often an automobile, that uses one or more electric motors to create movement.
F	Fahrenheit — a temperature scale based on 32 degrees for the freezing point of water and 212 degrees for the boiling point of water.
FCZ	forecasting climate zones — geographic areas defined by the CEC to assist energy forecasting and planning and sometimes are also called electricity demand forecast zones. The FCZs are specific to electricity providers.
FERA	Family Electric Rate Assistance Program —– provides discounts on energy bills to income qualified households. FERA income allowances are slightly higher than CARE allowances.
HDD	Heating degree days — are a measure of how much (in degrees) and for how long (in days) the air temperature was below a certain reference temperature (i.e. 65°F). HDD are used in calculations of energy consumption for heating a building.
IOU	investor-owned utilities — private electricity and natural gas providers whose stock is publicly traded. IOU energy rates are regulated, usually by the state's utility commission.
LADWP	Los Angeles Department of Water and Power — a publicly owned electric and water utility serving residential and commercial customers in Los Angeles and surrounding communities.
LCD	liquid crystal display — a type of electrically generated image shown on a thin, flat panel. LCD screens are found in consumer electronics like laptops, tablets, and smartphones.

Acronym	Definition
LED	Light-emitting diode — an electronic device that glows when a voltage is applied. Energy-saving LED bulbs are often used instead of CFLs or other light fixtures.
NAC	normalized annual consumption — an estimate of yearly energy consumption that has variations in weather effects removed.
NEM	Net-Energy Metering — billing mechanism that allows customers to generate energy onsite to meet their energy needs and receive a financial benefit for any excess energy sent to their utility.
OLED	organic light-emitting display — a display technology based on the use of an organic substance to produce light. OLED screens are found in consumer electronics like TVs, smartphones, tablets, and watches.
PC	personal computer — a multipurpose computer whose size, capabilities, and price make it feasible for individual use.
PG&E	Pacific Gas and Electric Company — an investor-owned electric and natural gas utility serving residential and commercial customers in Northern and Central California.
POU	Publicly owned utility — are publicly-run electric and natural gas providers. POUs include government-run (federal, state, or municipal) and public utility districts that operate independently of city or county government. Unlike IOUs, publicly owned utilities do not issue stock or have shareholders.
PV	Photovoltaic — PV devices, like those found in solar power panels, generate electricity directly from sunlight via an electronic process that occurs naturally in certain types of materials.
RAC	room air conditioning — cooling provided to rooms rather than the entire home or business.
RASS	Residential Appliance Saturation Study — a comprehensive survey of California residents to collect information about characteristics of their homes, their appliances and heating and cooling equipment, use of solar or electric vehicles, and general energy use.
SAE	statistically adjusted engineering — a method of analyzing energy savings that uses statistical modeling and engineering estimates of energy savings.
SAS	statistical analysis system — a software suite that can manipulate, manage, and retrieve data from a variety of sources and perform statistical analysis on it.
SCE	Southern California Edison Company — an investor-owned electric utility serving residential and commercial customers in Southern California.

Acronym	Definition
SDG&E or SDGE	San Diego Gas & Electric Company — an investor-owned electric and natural gas utility serving residential and commercial customers in San Diego and surrounding areas.
SMUD	Sacramento Municipal Utility District — a community-owned electric utility serving Sacramento County and parts of Placer County.
SoCalGas	Southern California Gas Company — an investor-owned natural gas utility based in Los Angeles serving residential and commercial customers. SoCalGas is a subsidiary of Sempra Energy, based in San Diego.
T24	Title 24 — California building standards code, a set of standards for new construction and existing buildings.
UEC	unit energy consumption — the amount of energy a single appliance is estimated to use in a year.
USPS	United States Postal Service
VEE	Validation, editing, and estimation — processing information to assess the quality, edit information, and estimate missing values.