CALMAC Note to Readers

This version of the RASS study was compiled by CALMAC in order to post a fairly good representation of the study on our database. It was created from three separate reports: An executive summary, a study methodology, and a results report.

These reports, and their appendices and attachments, along with an interactive database of the result are available (as of 9/15/04) on the Internet at:

http://www.energy.ca.gov/appliances/rass/index.html

The direct link to the searchable database is:

http://websafe.kemainc.com/RASSWEB/DesktopDefault.aspx

CALMAC is not responsible for maintaining these links to the reports or the database. If the link is no longer active at some future date the reader will have to search the Internet for it.

CALMAC Website Administrator.

CALIFORNIA STATEWIDE RESIDENTIAL APPLIANCE SATURATION STUDY

FINAL REPORT EXECUTIVE SUMMARY

Consultant Report

Prepared for:
California Energy Commission

Prepared by:
KEMA-XENERGY
Itron
RoperASW

June 2004
400-04-009

Prepared by:
KEMA-XENERGY
Itron
RoperASW

Contract No. 400-04-009

Prepared for:
California Energy Commission

Contract Manager: ***Glen Sharp***

Sponsors:
Pacific Gas and Electric (PG&E)
San Diego Gas and Electric (SDG&E)
Southern California Edison (SCE)
Southern California Gas Company (SoCalGas)
Los Angeles Department of Water and Power (LADWP)

1.1 Report Overview

This report highlights key findings from the California Energy Commission's 2003 Statewide Residential Appliance Saturation Study (RASS). This executive summary provides an overview of the results from the study including energy use and equipment saturations throughout the State of California.

The executive summary is a companion document to a comprehensive methodology and results report that includes energy consumption tables from the conditional demand analysis along with a series of “cross tabs” which display the RASS results in a comprehensive format.

The sections of this summary report include:

2. **Study Background.** An overview of the project approach.
3. **Unit Energy Consumption and Appliance Saturation Summaries.** Results from the Conditional Demand Analysis (CDA) that was performed on the RASS data. Results are provided for both electric and natural gas end uses.
4. **Fuel Shares.** Gas continued to be the predominant space heating and water heating fuel in the California marketplace. These tables show how the share of gas and electric appliances and equipment vary.
5. **Air Conditioning.** Air conditioning is the primary driver of peak energy demand in California and the saturation of central air conditioning systems is increasing.
6. **New Dwellings.** Newer dwellings (built after 1996) are larger, have a slightly higher average number of residents, and have higher average incomes than older dwellings. New dwelling electricity use has a corresponding increase although it is counteracted by higher incidences of energy efficient equipment.
7. **Income Effects.** Income strongly correlates to energy use because of the resulting larger dwellings and prevalence of more energy consuming equipment. However, this section also demonstrates that all income groups have customers who use above average amounts of energy.
8. **Energy Efficiency Actions.** The use of energy efficiency equipment and conservation actions continue to grow as evidenced by the increase in these items in new dwellings. However, there is still a large market segment that is not adopting these products and practices.
9. **Technology.** The prevalence of technology in the dwelling is increasing as more people work at home, have more equipment, and use their technology to do a wide range of activities. This information is important from the standpoints of energy use and future customer relations and communication vehicles.
10. **Data Comparisons.** The study results provide a reasonable match to Census data. The section also provides information on the effect the non-respondent study had on the final results.

1.2 Study Background

For the first time in California, the large Investor Owned Utilities (IOUs) pooled resources and performed a RASS and Unit Energy Consumption (UEC) Study as a team. The project was administered by the California Energy Commission and sponsored by Pacific Gas and Electric (PG&E), San Diego Gas and Electric (SDG&E), Southern California Edison (SCE), Southern California Gas Company (SoCalGas), and Los Angeles Department of Water and Power (LADWP). KEMA-XENERGY was the prime consultant. Itron provided data cleaning and performed the Conditional Demand Analysis. RoperASW fielded the non-response follow-up.

The RASS effort has resulted in a research product that provides both statewide and utility-specific results. The study was designed to allow comparison of results across utility service territories, climate zones and other variables of interest (i.e. dwelling type, dwelling vintage, and income). The study includes results for 21,920 residential customers that are weighted to the population represented by the sponsoring utilities. The saturation results capture both individual and master metered dwellings. This rich set of customer data includes information on all appliances, equipment, and general usage habits. The study also includes a detailed conditional demand analysis that calculates unit energy consumption (UEC) values for all individually metered customers.

The study was initiated in late 2002 and the sampling plans and survey implementation occurred throughout 2003. The data was collected using a two stage direct mail survey targeted to a representative sample of California residential customers. The survey requested customers to provide details on their energy equipment and behaviors. A non-response follow-up survey was implemented at the end of the double mailing phase to a sub-sample of non-respondents. The non-response follow-up included telephone and in-person interviews in an effort to minimize non-response bias by using alternative surveying techniques.

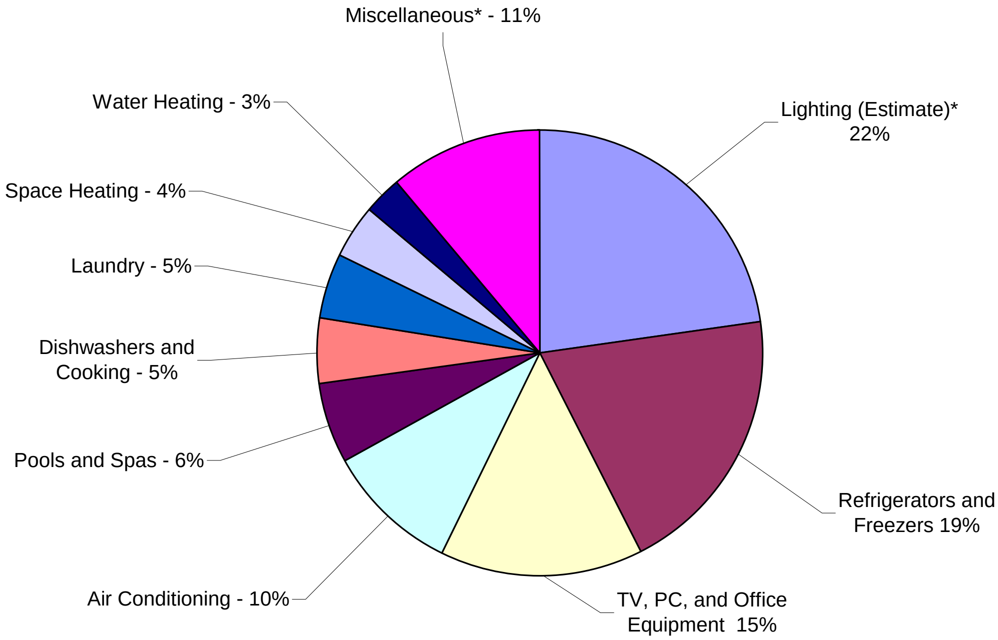
The results from the RASS study were used to develop a CDA model. This analytical method uses a combination of customer energy use with the responses from the customer survey to model end uses and develop unit energy consumption results for those end uses. The results of the CDA are included in summary form along with the general study results in this executive summary and are provided in further detail in the methodology section of the report.

The study also includes onsite metering for a sample of 180 RASS participants. The onsite metering sample was designed to over-sample air conditioning use, with the meters gathering both a whole-house and central air conditioning usage at each dwelling. The onsite meters are in the field at the time of publication and the final results from that portion of the project will be delivered as whole house and air conditioning load shapes after the 2004 cooling season has ended.

1.3 End Use Energy and Appliance Saturation Summaries

Using utility billing data from 2002 and normalized weather data for each climate zone in the state, the CDA was used to determine UEC values for end uses. This UEC section includes the individually metered customers only. As shown in Figure 1, annual electrical energy use in California is 5,914 kWh per household.

Figure 1
Statewide Electricity Use per Household
5,914 kWh per Household



*Note: An estimate of 1,200 kWh per household (20% of the total use) has been designated as interior lighting and was shifted from Miscellaneous to Lighting where it is combined with exterior lighting usage. This number comes from other lighting studies[1] that are better able to pinpoint this estimate than a conditional demand model as was used for the RASS.

The CDA model produced several results that varied from previous studies. The most notable are electric space heating and air conditioning, which are both lower than previous studies.[2] This is likely a result of the statewide electricity price increases and statewide 20/20 Program in effect during 2001 and 2002.[3] These two simultaneous effects combined to provide customers with a strong incentive to reduce their consumption. In the peak summer months, energy use dropped significantly, with roughly 30% of customers in PG&E's territory participating in the program.[4] While 2002 consumption was higher than that achieved in 2001, almost 50% of the conservation observed in 2001 persisted in 2002.[5] The CDA used 2002 billing data in the modeling process and thus was impacted by these effects.

The UECs presented in Table 1 and 2 show the full CDA results displayed first by utility and then by dwelling type.

Table 1
Electric UEC and Appliance Saturation Summaries by Utility

	PG&E		SDG&E		SCE		DWP	
	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.
All Households	6,265		5,445		6,102		4,071	
Primary Conventional Space Heating	1,113	10%	581	13%	734	6%	542	9%
Primary Heat Pump Space Heating	799	2%	458	3%	555	1%	201	3%
Auxilliary Space Heating	331	26%	156	24%	192	23%	103	17%
Furnace Fan (Gas Heat)	180	58%	91	60%	115	56%	71	26%
Attic Fan	102	12%	60	7%	159	10%	243	5%
Central Air Conditioning	1,108	39%	644	35%	1,494	48%	1,075	29%
Room Air Conditioning	181	14%	63	9%	202	20%	158	25%
Evaporative Cooling	469	5%	277	1%	797	5%	372	2%
Water Heating	2,585	9%	2,151	6%	2,342	5%	1,387	5%
Solar Water Heating	1,193	0%	1,501	1%	1,508	0%	0	0%
Dryer	652	45%	648	26%	717	18%	474	7%
Clothes Washer	97	78%	75	77%	129	77%	125	36%
Dish Washer	77	67%	69	71%	80	60%	73	27%
First Refrigerator	788	100%	780	100%	801	100%	754	100%
Additional Refrigerator	1,201	19%	1,054	19%	1,210	19%	933	6%
Freezer	928	23%	841	17%	983	15%	880	5%
Pool Pump	2,580	8%	2,557	12%	2,772	10%	3,096	2%
Spa	428	8%	445	12%	495	10%	423	2%
Outdoor Lighting	260	56%	268	53%	276	55%	218	42%
Range/Oven	268	61%	241	49%	271	27%	200	17%
Television	474	95%	446	94%	520	96%	479	94%
Spa Electric Heat	1,346	5%	903	6%	2,514	4%	895	1%
Microwave	131	95%	119	96%	139	96%	140	89%
Home Office Equipment	152	20%	159	19%	141	16%	134	18%
Personal Computer	602	72%	614	78%	515	66%	516	55%
Water Bed	787	2%	925	1%	818	2%	848	0%
Well Pump	829	8%	831	1%	952	2%	890	1%
Interior Lighting and Miscellaneous	1,840	100%	1,746	100%	1,896	100%	1,483	100%
Ave. Dwelling Size	*1,525*		*1,614*		*1,506*		*1,017*	
Ave. Residents	*2.89*		*2.75*		*3.12*		*2.86*	
Percent Single Family	*62.0%*		*59.4%*		*62.0%*		*25.6%*	
Percent of Population	*41.1%*		*11.5%*		*38.8%*		*8.6%*	

One important note on the results is that the LADWP population frame that was originally supplied for the study appears to have excluded a portion of the LADWP service area. It appears that the missing customers were predominantly single family homes which is part of the reason that the percentage of single family homes is so low for LADWP. The “missing” customers make up less than two percent of the total statewide population. However, the LADWP results need to take this into consideration when viewed individually.

Table 2
Electric UEC and Appliance Saturation Summaries by Dwelling Type

	All		Single Family		Multi Family		Mobile Home	
	UEC	**Sat.**	**UEC**	**Sat.**	**UEC**	**Sat.**	**UEC**	**Sat.**
All Households	5,914		7,105		3,953		5,662	
Primary Conventional Space Heating	871	9%	1,494	4%	646	17%	1,150	10%
Primary Heat Pump Space Heating	588	2%	1,077	1%	335	3%	1,031	3%
Auxilliary Space Heating	244	24%	296	28%	87	16%	298	31%
Furnace Fan (Gas Heat)	139	55%	162	68%	62	33%	118	58%
Central Air Conditioning	1,236	41%	1,423	46%	803	32%	1,143	39%
Room Air Conditioning	181	17%	227	15%	114	19%	227	34%
Evaporative Cooling	622	4%	688	5%	430	2%	537	27%
Water Heating	2,389	7%	3,079	5%	1,607	9%	3,258	17%
Solar Water Heating	1,345	0%	1,708	0%	344	0%	0	0%
Dryer	663	29%	713	34%	535	20%	549	42%
Clothes Washer	108	74%	127	95%	45	39%	11	86%
Dish Washer	77	61%	84	70%	62	48%	47	55%
First Refrigerator	789	100%	824	100%	731	100%	809	100%
Additional Refrigerator	1,178	18%	1,245	25%	673	6%	1,143	13%
Freezer	935	18%	937	24%	917	6%	951	30%
Pool Pump	2,671	9%	2,671	14%	0	0%	0	0%
Spa	460	8%	467	13%	270	1%	180	3%
Outdoor Lighting	264	54%	284	67%	201	33%	232	56%
Range/Oven	263	42%	301	41%	209	46%	208	27%
Television	490	95%	519	96%	442	94%	457	93%
Spa Electric Heat	1,704	4%	1,719	7%	694	0%	3,550	2%
Microwave	133	95%	140	97%	124	91%	113	96%
Home Office Equipment	148	18%	148	20%	148	16%	121	13%
Personal Computer	565	69%	578	75%	542	59%	458	45%
Water Bed	817	2%	840	2%	750	1%	773	3%
Well Pump	849	4%	862	5%	862	1%	724	18%
Interior Lighting and Miscellaneous	1,832	100%	2,146	100%	1,332	100%	1,463	100%

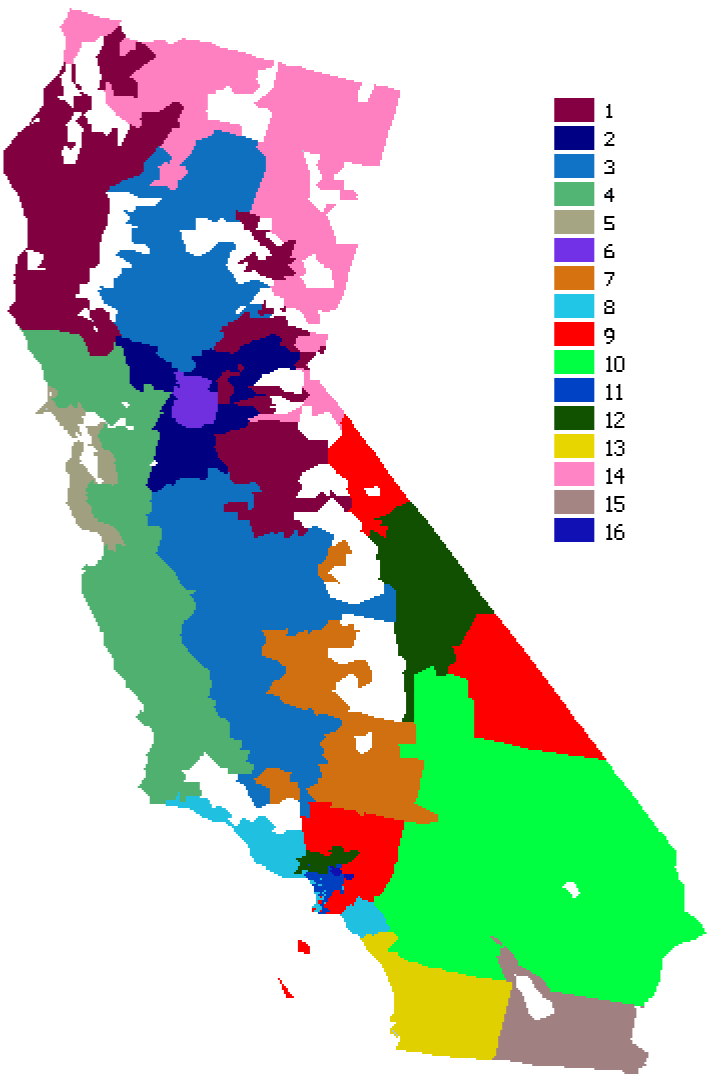
	All	Single Family	Multi Family	Mobile Home
Ave. Dwelling Size	*1,541*	*1,787*	*997*	*1,167*
Ave. Residents	*2.96*	*3.21*	*2.60*	*2.26*
Percent of Population	*100%*	*59%*	*37%*	*4%*

Figure 2 is a map of the Energy Commission forecast climate zones. These zones were used in the CDA modeling and provide regional summaries by climate. (A black and white version of this graph is available at the end of the report.)

Zones 1-5 are served by PG&E (Zones 3 and 4 have some SoCalGas overlap)
Zone 6 is served by SMUD and not included in the results
Zones 7-10 are served by SCE/SoCalGas
Zones 11-12 are served by LADWP/SoCalGas
Zone 13 is served by SDG&E (some SoCalGas overlap)
Zones 14-16 are served by other electric utilities and not included in the results

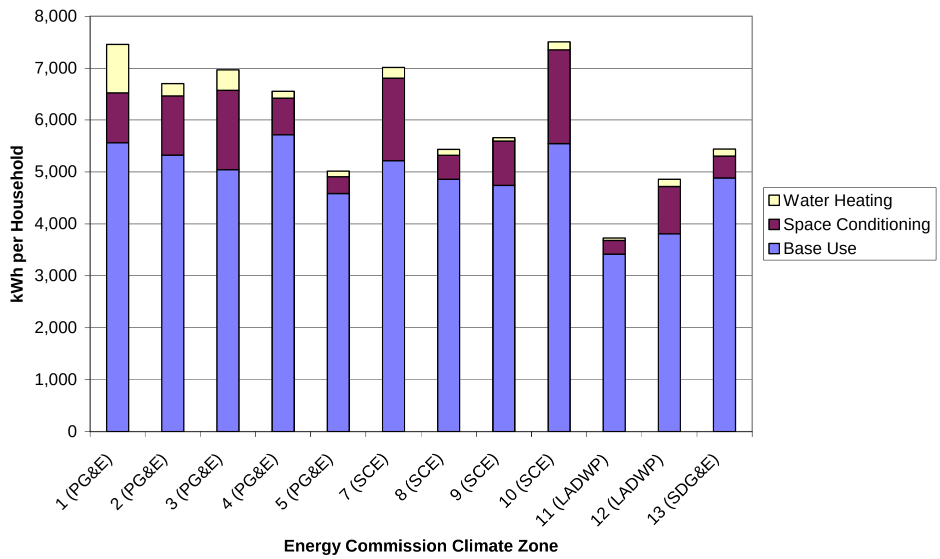
Figure 2

California Energy Commission Forecast Climate Zones



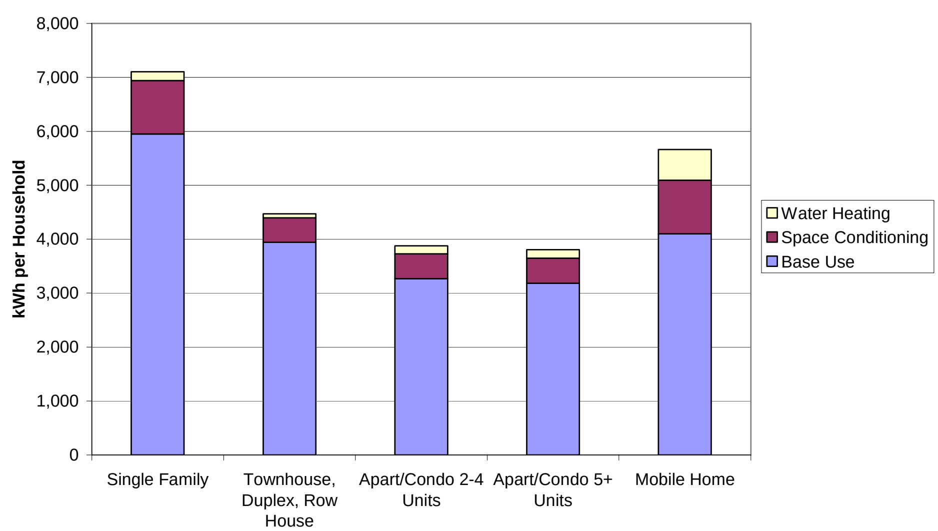
Both base energy use and space conditioning (heating and cooling) vary by climate zone (Figure 3). Climate Zone One has the lowest availability of gas, which is why its water heating UEC is so high.

Figure 3
Electric UECs by Climate Zone



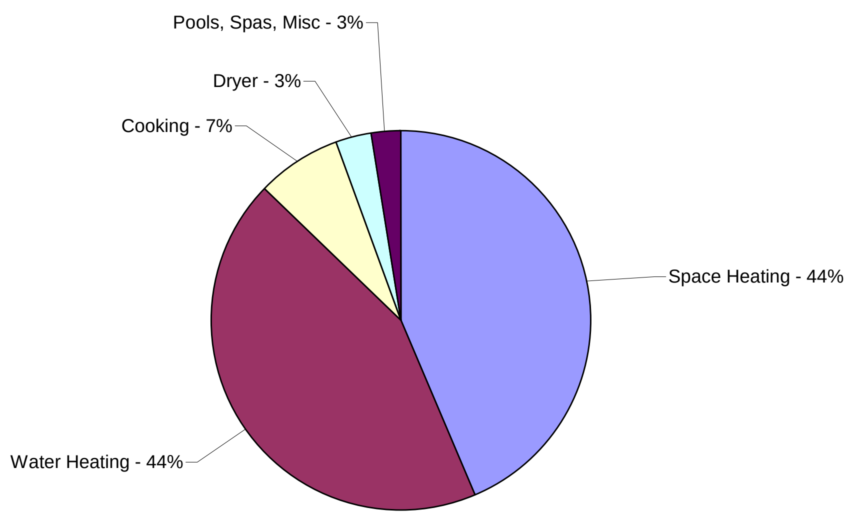
The mix of housing stock explains much of the difference in the base use shown in the climate zone table. Single family dwellings have the highest per dwelling electric use (Figure 4).

Figure 4
Electric UECs by Dwelling Type



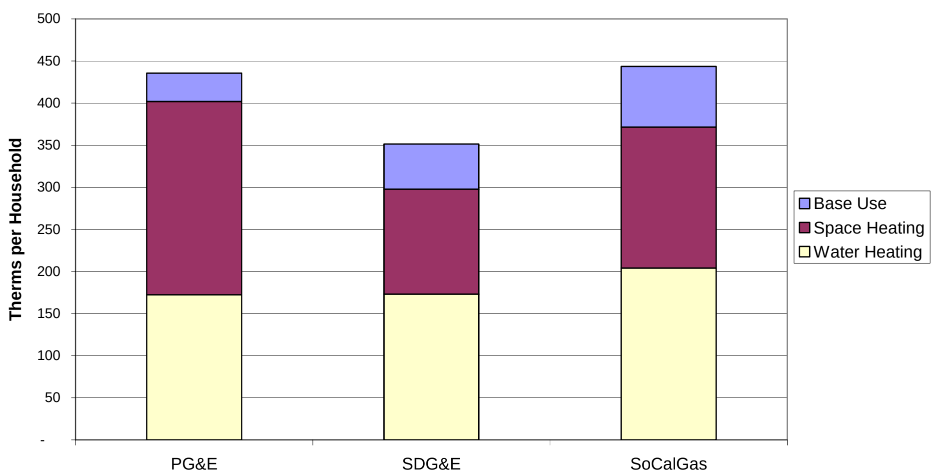
The annual energy consumption of the customers for whom we have gas bills (76% of the population) is 431 therms per household. Overall, 82% of the customers from the electrically based population were provided with gas UECs because they stated that they had a gas appliance. Figure 5 provides the gas consumption breakdown by end use.

Figure 5
Statewide Gas Energy Use



PG&E has the highest natural gas use with the biggest difference across utilities occurring in the heating end uses (Figure 6).

Figure 6
Gas UECs by Utility



Utility - Results include those customers who have a gas bill from the designated utility. 24% of the total electric based population does not have a gas account and is excluded from this table.

Natural gas end uses are listed in Table 3 and 4 for all homes with a gas account. For the combined gas and electric utilities as well as the statewide total, the final row in each table represents the total gas household consumption across the electrically based population. Because the sample was electrically based, this result is not fully representative of statewide gas use because of overlapping gas and electric service territories.

Table 3
Natural Gas UEC and Appliance Saturation Summaries by Utility

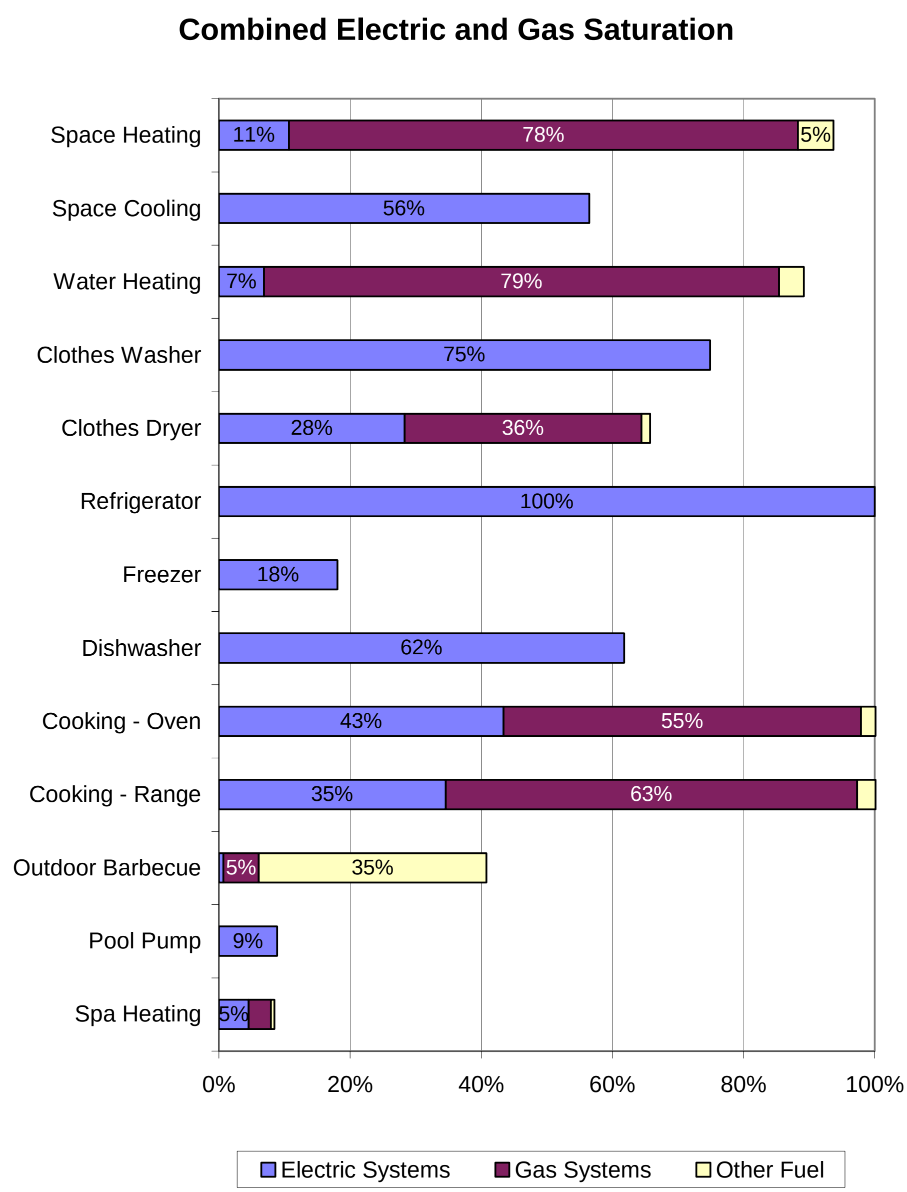
	All		PG&E		SDG&E		SoCalGas	
Homes with Gas Accounts	**UEC**	**Saturation of Homes with Gas Account**	**UEC**	**Saturation of Homes with Gas Account**	**UEC**	**Saturation of Homes with Gas Account**	**UEC**	**Saturation of Homes with Gas Account**
All Households	431		436		351		443	
Space Heating	202	93%	244	94%	135	92%	181	93%
Water Heating	201	94%	183	94%	181	96%	219	93%
Dryer	30	43%	25	28%	23	54%	33	53%
Range/Oven	43	72%	37	53%	35	71%	48	86%
Pool Heating	222	3%	225	2%	217	4%	222	3%
Spa Heating	81	5%	76	3%	86	7%	83	6%
Miscellaneous	2	100%	1	100%	2	100%	2	100%
Gas Use Across Electrically Based Utility Population	356		343		279		Not Applicable	

Table 4
Natural Gas UEC and Appliance Saturation Summaries by Dwelling Type

	Single Family		Multi Family		Mobile Home	
Homes with Gas Accounts	**UEC**	**Saturation of Homes with Gas Account**	**UEC**	**Saturation of Homes with Gas Account**	**UEC**	**Saturation of Homes with Gas Account**
All Households	508		270		433	
Space Heating	242	98%	102	83%	209	99%
Water Heating	206	99%	188	82%	193	99%
Dryer	31	55%	22	19%	13	39%
Range/Oven	46	73%	39	68%	28	90%
Pool Heating	222	4%	281	0%	0	0%
Spa Heating	81	7%	89	0%	114	3%
Miscellaneous	2	100%	1	100%	2	100%
Gas Use Across Electrically Based Utility Population	454		198		235	

Figure 7 provides a summary graph of the major saturation rates for all of the individually metered households in the state.

Figure 7
Combined Electric, Gas, and Other Fuel Saturations

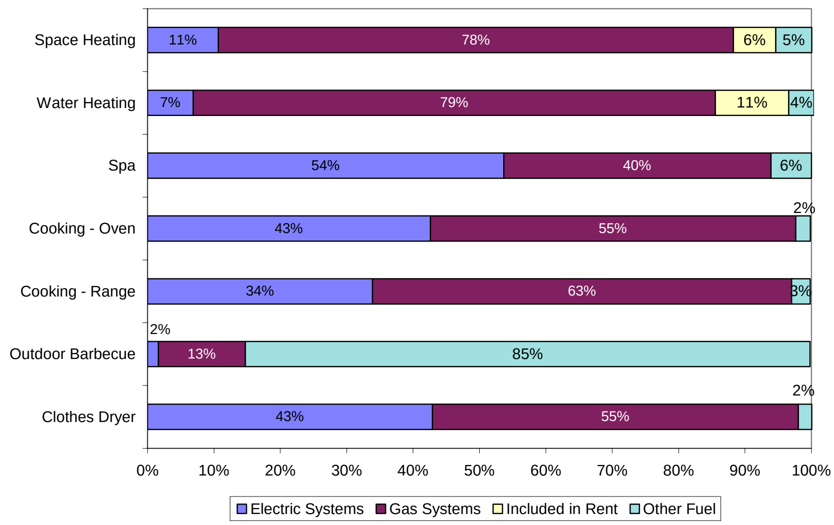


1.4 Fuel Shares

NOTE: *The remainder of the report (except where UECs are explicitly included) includes data from both individually and master metered dwellings. Master metered customers were not included in the CDA.*

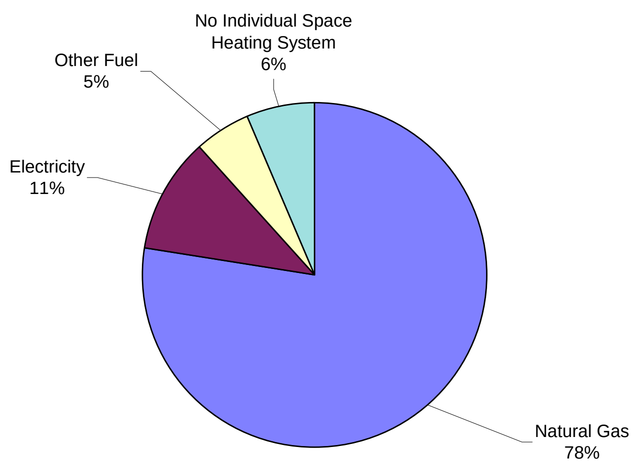
Overall fuel shares are included as Figure 8. Figures 8 and 9 include multi-unit systems, which are typically included in a tenant's rent. Shares represent the fuel share for customers who have the equipment.

Figure 8
Overall Shares of Electric and Gas Systems



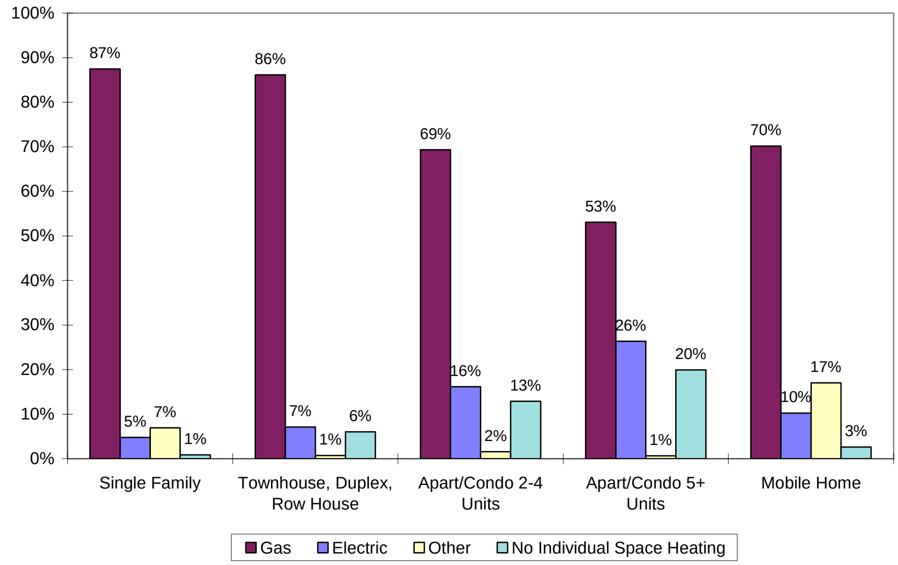
The vast majority of primary space heating systems are gas (Figure 9). The "No Individual Space Heating System" category includes people who have no space heating or a central building system that serves multiple apartments or dwellings.

Figure 9
Primary Space Heating Fuel



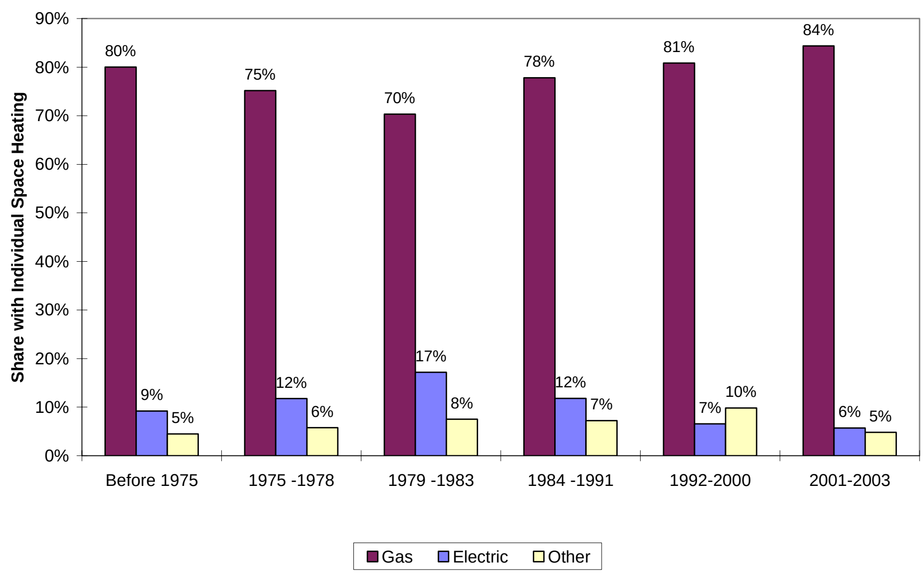
Electric heat is more common in apartments and condos than in single family dwellings (Figure 10). The “Other” fuel includes propane, wood, and other as reported by the customer.

Figure 10
Space Heating Fuel by Dwelling Type



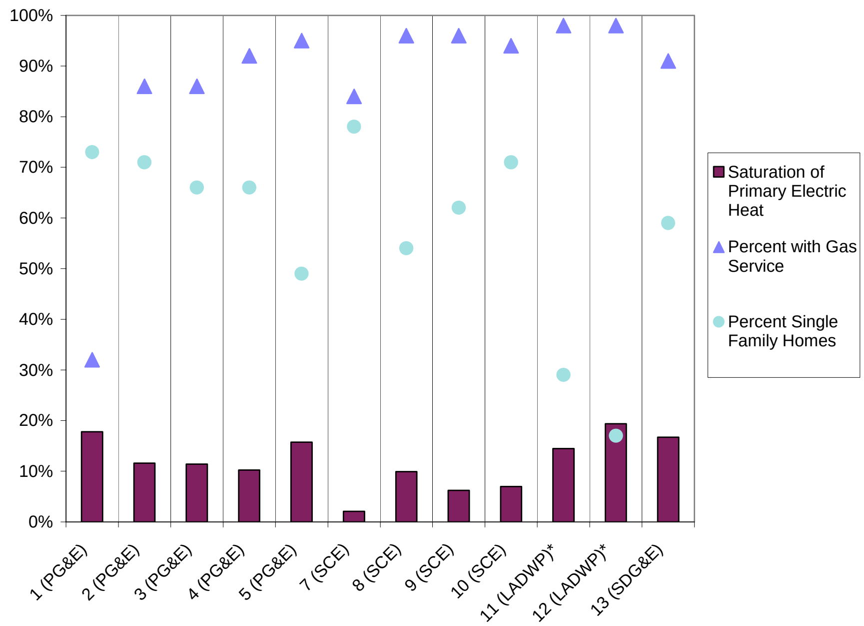
As shown in Figure 11, gas space heating is more common in newer dwellings. Dwellings built between 1979 and 1983 have the highest levels of electric heating. Figure 11 displays individually heated systems only.

Figure 11
Space Heating Fuel by Dwelling Age



Shares of electric space heating (Figure 12) are highest in Zone One where there is the least gas available and then in the more moderate southern climates (11, 12, 13). Zones 11 and 12 are high due to the high number of multifamily dwellings.

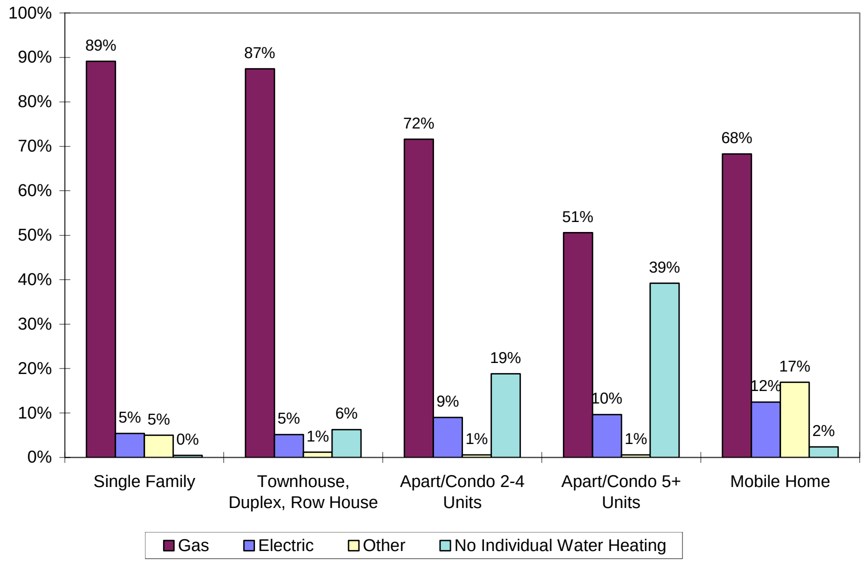
Figure 12
Shares of Electric Space Heating and HDD by Climate Zone



*Note that in Figure 12 the percentage of homes in LADWP's service territory is low. It appears that the original LADWP population file was missing a set of customers who are likely single family dwellings. LADWP's results are thus biased towards their multi-family population. Previous Energy Commission work shows single family rates more on the order of 50% in the LADWP territory as opposed to the 27% and 16% shown here.

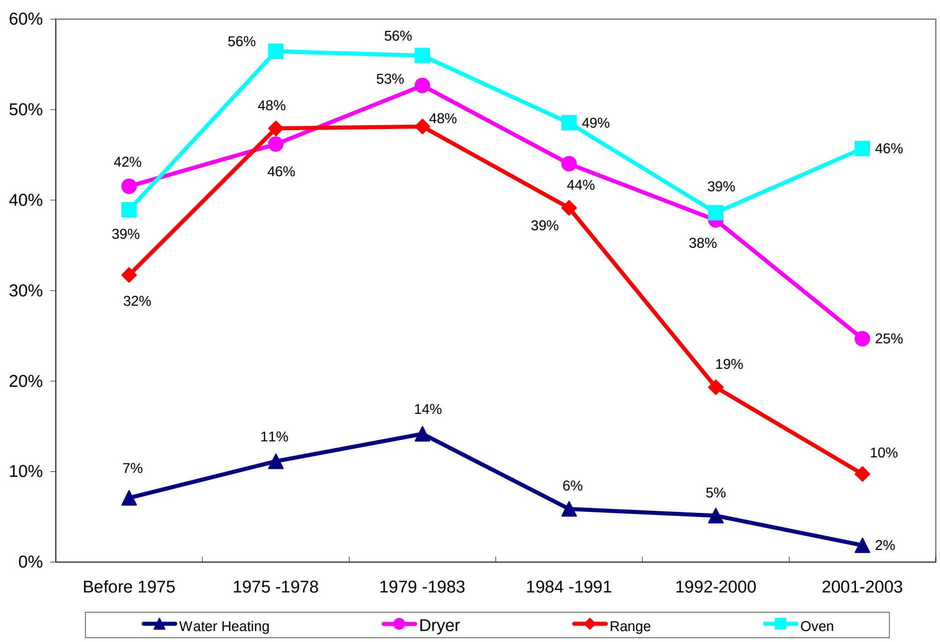
Water heating follows a similar fuel share pattern as space heating (Figure 13).

Figure 13
Water Heating by Dwelling Type



While electric shares are more prevalent in older buildings, it appears that many buildings that are more than 20 years old have been upgraded to natural gas systems and thus show lower shares of electric appliances (Figure 14). Electric ovens are still much more popular than electric ranges and continue to be installed extensively in newer dwellings.

Figure 14
Electric Appliances Share by Dwelling Age



As with most all other electric shares (Figures 15 through 17), the share in apartments is higher than in single family dwellings. Other fuels primarily represent propane, particularly in the mobile home market. All share tables represent the fuel share for customers who have the equipment.

Figure 15
Fuel Shares for Dryers by Dwelling Type

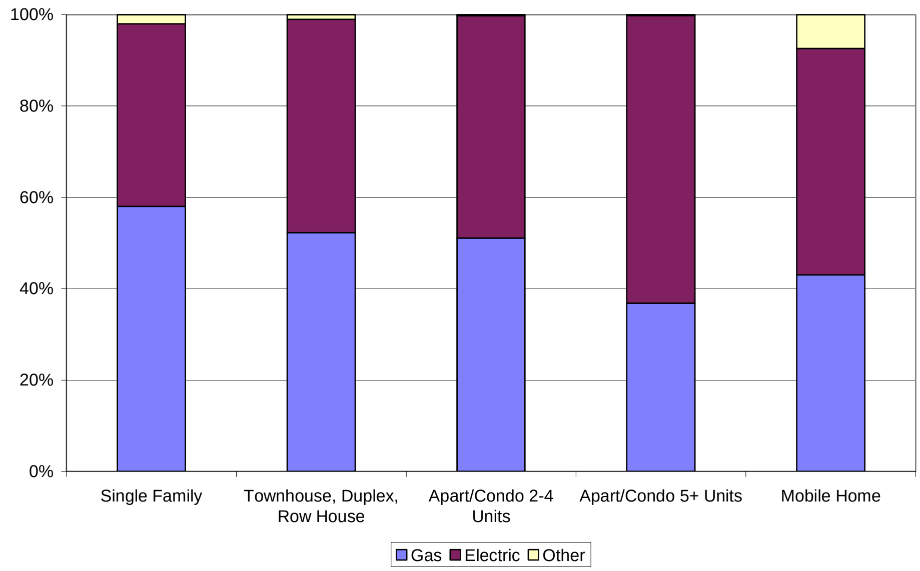


Figure 16
Fuel Shares for Ranges by Dwelling Type

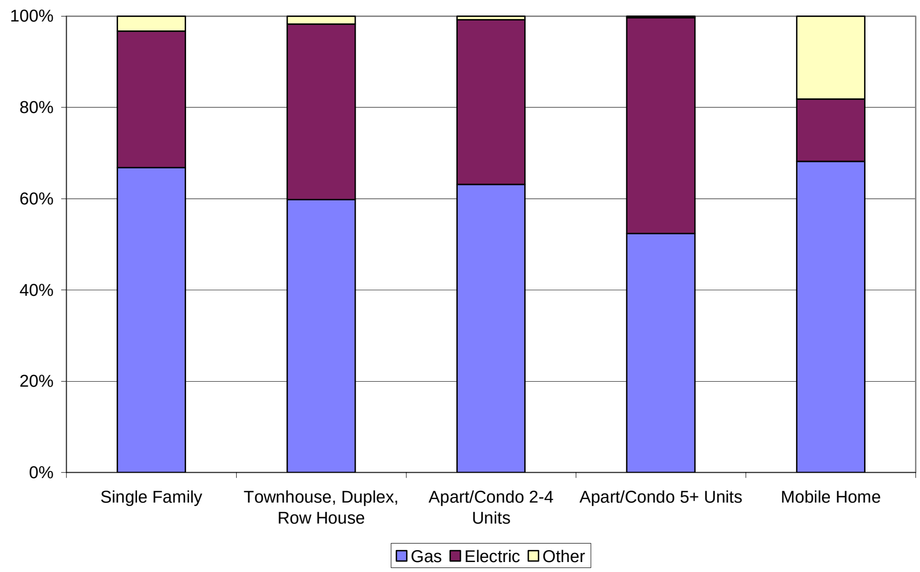
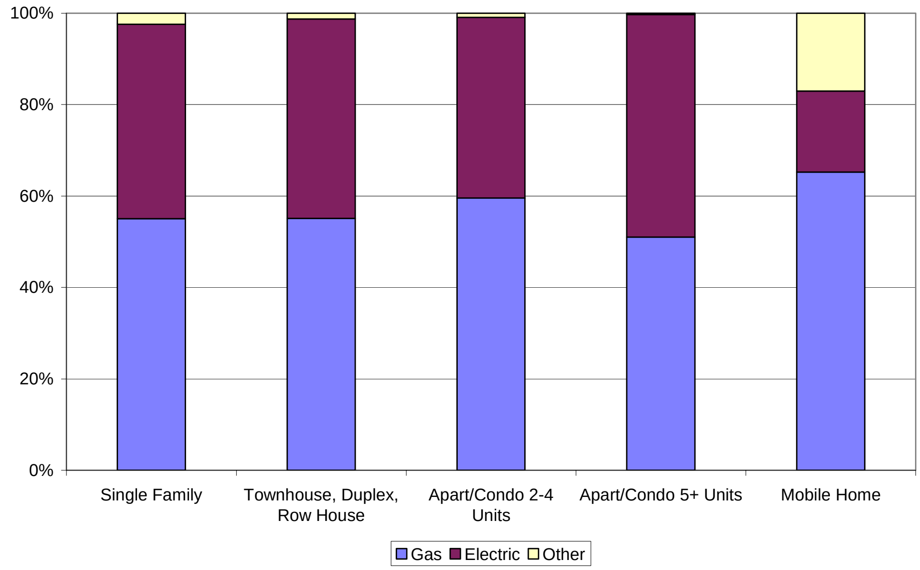


Figure 17
Fuel Shares for Ovens by Dwelling Type

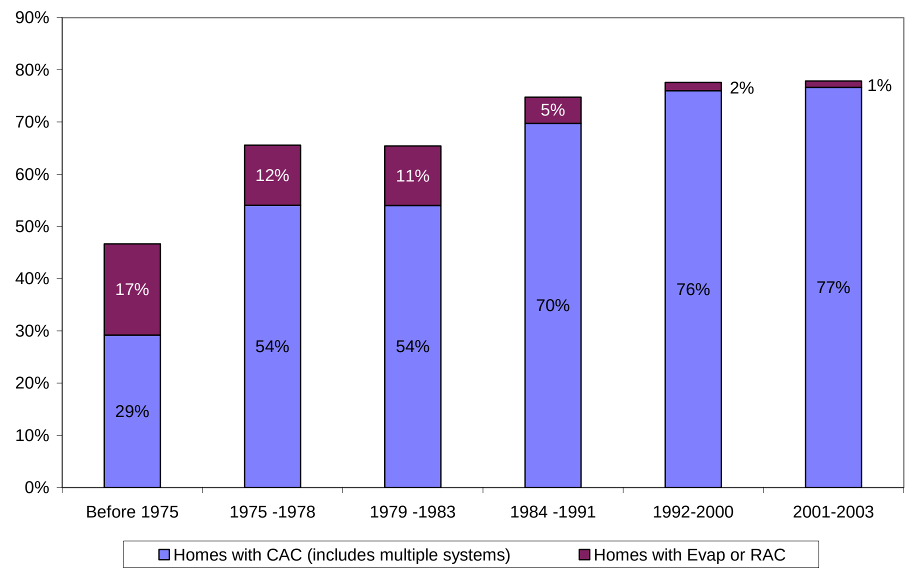


1.5 Air Conditioning

Air conditioning is the peak driver of energy use in California. The overall UEC for central air conditioning is 1,236 kWh per household. Room air conditioning has a UEC of 181 and evaporative systems 622. These values are somewhat lower than previous studies and forecasting values used at the Energy Commission. One possible reason for the lower than average use is attributed to the Statewide 20/20 Program.[6] Billing data for the CDA was from the second half of 2001, all of 2002, and the first part of 2003. UEC results have all been annualized and calibrated to 2002 service territory total usage. It is likely that the UECs reflect the 20/20 program impact and thus these air conditioning values should be considered conservative estimates.

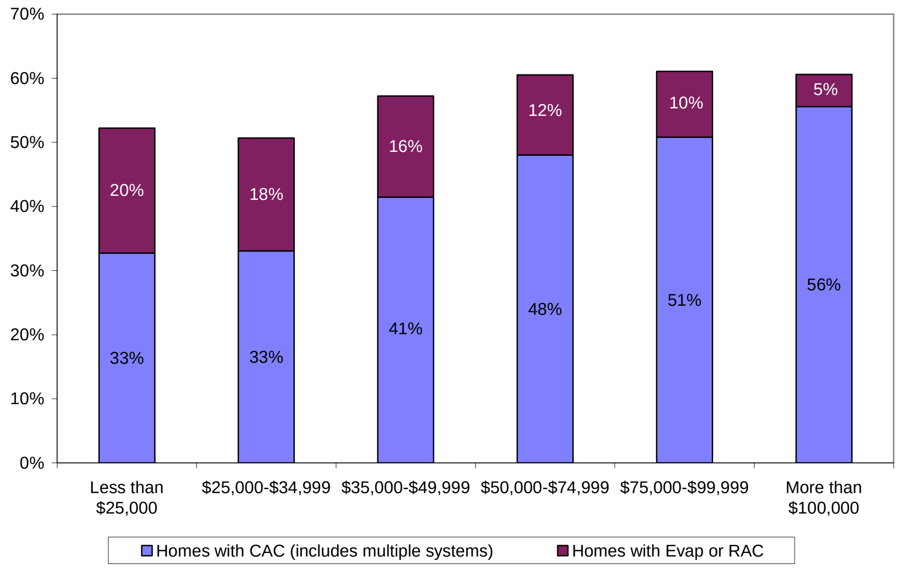
Air conditioning has grown overall with the biggest change in the type of systems installed. Room and evaporative units are going out of favor while central systems are present in 77% of the most recent dwellings (Figure 18).

Figure 18
Air Conditioning by Dwelling Age



Income plays a big role in air conditioning growth (Figure 19) as it is strongly correlated to the type and presence of air conditioning systems. However, dwelling age is a stronger driver of overall air conditioning usage.

Figure 19
Air Conditioning by Income

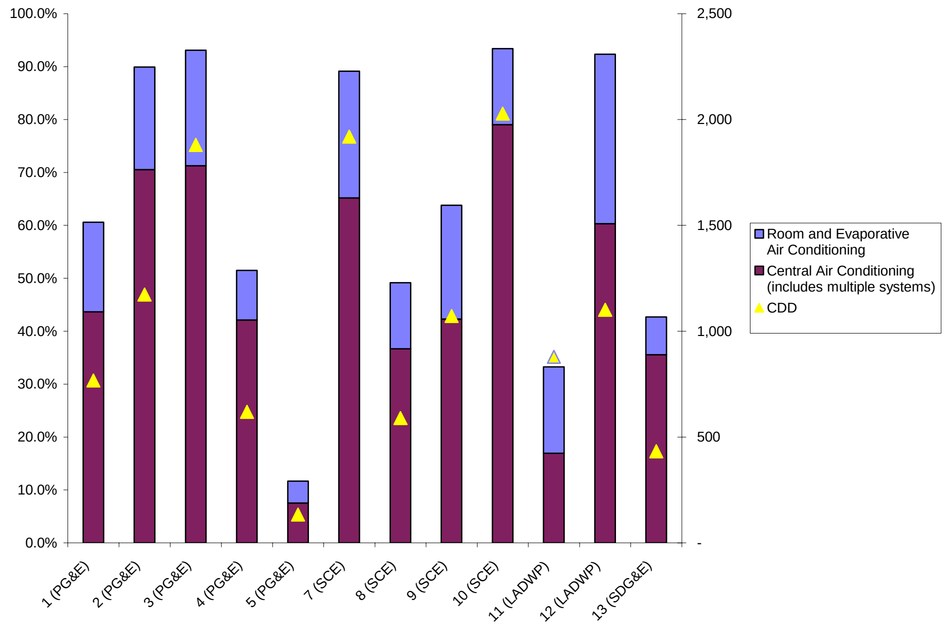


UECs for the state vary significantly by climate. The forecast zones and their respective cooling degree days (CDDs) in Table 5 justify the UECs for central air conditioning. Figure 20 which follows displays the saturations by type of air conditioning system along with the cooling degree days. All cooling degree days represent normalized weather. UECs throughout are based on normalized weather.

Table 5
Central Air Conditioning UECs by Climate Zone with CDDs

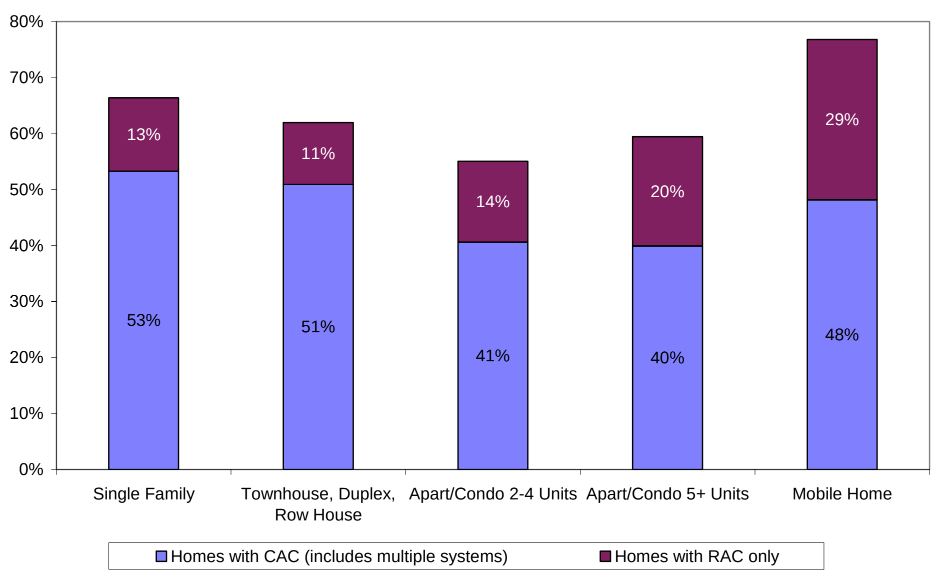
Energy Commission Forecast Climate Zone	Central AC UEC (kWh/Household)	CDD
Zone 1 (PG&E)	941	767
Zone 2 (PG&E)	1,082	1,173
Zone 3 (PG&E)	1,548	1,880
Zone 4 (PG&E)	885	619
Zone 5 (PG&E)	226	133
Zone 7 (SCE)	1,902	1,919
Zone 8 (SCE)	848	590
Zone 9 (SCE)	1,509	1,072
Zone 10 (SCE)	1,908	2,028
Zone 11 (LADWP)	915	879
Zone 12 (LADWP)	1,169	1,101
Zone 13 (SDG&E)	644	433

Figure 20
Saturation of Air Conditioning by Climate Zone



In order to see how the dwelling type affects air conditioning in hot climates, climate zones 5 and 11 were removed from Figure 21 because they had a combination of low air conditioning saturations and a high percentage of multi-family dwellings. The sub-sample better represents areas where air conditioning is more common.

Figure 21
Air Conditioning by Dwelling Type for All Zones Except 5 and 11



In Figure 21, single family dwellings make up 61% of the reported cases, townhouses 7%, apartments with 2-4 units 9%, apartments with more than 5 units 18%, and mobile homes 5%.

While newer dwellings represent the largest growth area for central air conditioning, about one third or 1.3 million of the central air conditioning units in operation are 14 years old or older (Figure 22).

Figure 22
Age Distribution of Central Air Conditioners

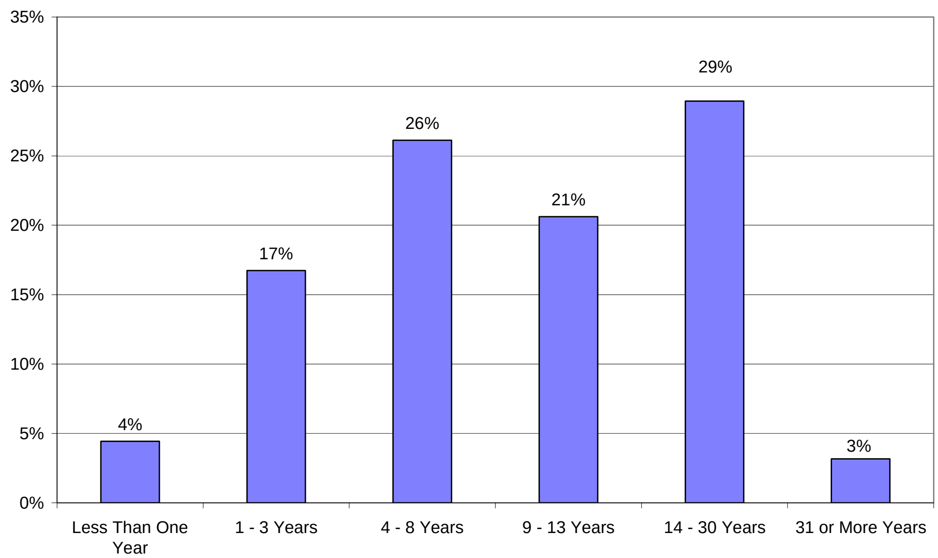
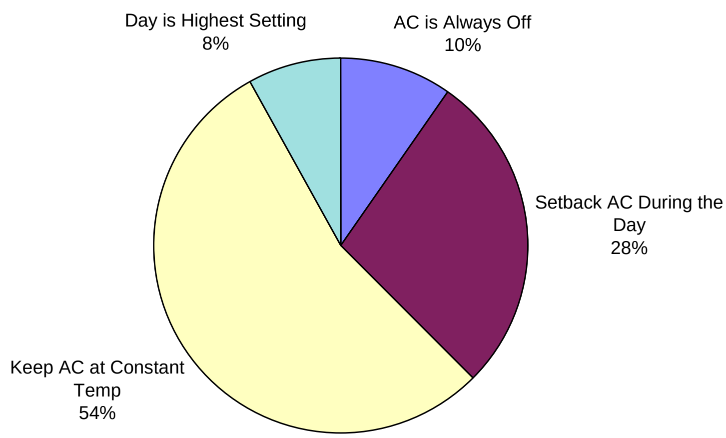


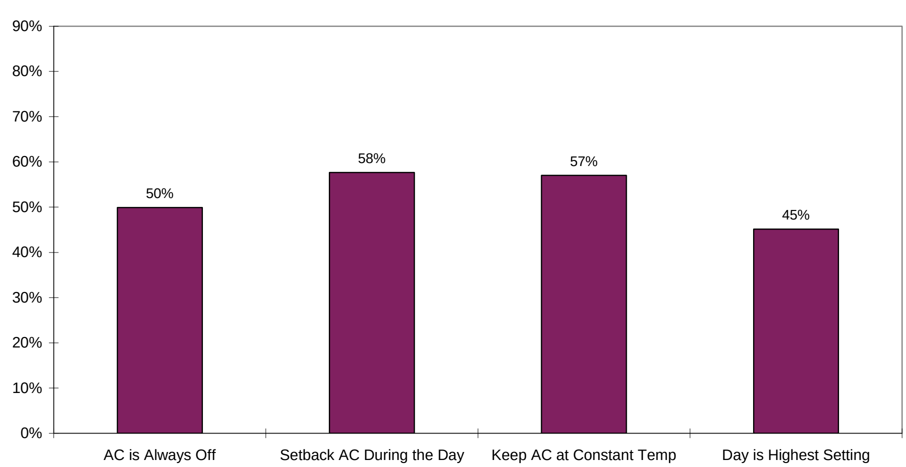
Figure 23 shows the breakdown of how customers with central air conditioning set their thermostats. Over half of all respondents reported keeping their thermostats set at a constant temperature throughout the day.

Figure 23
Air Conditioning Setback Habits



The presence of programmable thermostats slightly increases amongst those who actively setback (58%). However, the results illustrate that the presence of programmable thermostats does not appear to dramatically affect setback behaviors. Overall, 54% of dwellings have programmable thermostats (Figure 24). The average temperature setting using the midpoint of the survey ranges provided is 79.4°F in the morning, 77.4°F degrees during the day, 76.6°F in the evening, and 79.6°F at night.

Figure 24
Presence of Programmable Thermostats by Setback Habits



1.6 New Dwellings

The definition of new dwellings in this section is dwellings that are built after 1996. While the survey asked for the actual year the dwelling was built and included options for 2002 and 2003, the sample was drawn in mid to late 2002 so it best represents new construction that was in place through 2001 and into the first part of 2002. The RASS surveys were sent to customers starting in April 2003. There are a small number of dwellings reported as built in 2002 and 2003 and these are included in the new category. However, the new trends are not fully reported for 2002 and 2003 due to the sampling and surveying timelines. There are just over half a million dwellings built after 1996 which translates into five percent growth for this five year building period.

Almost two thirds of the total residential housing growth falls in just four climate zones (Figure 25). Refer to Figure 2 at the start of the report to view the geographic placement of each of these zones.

Figure 25
Distribution of New Dwellings by Energy Commission Forecast Climate Zone

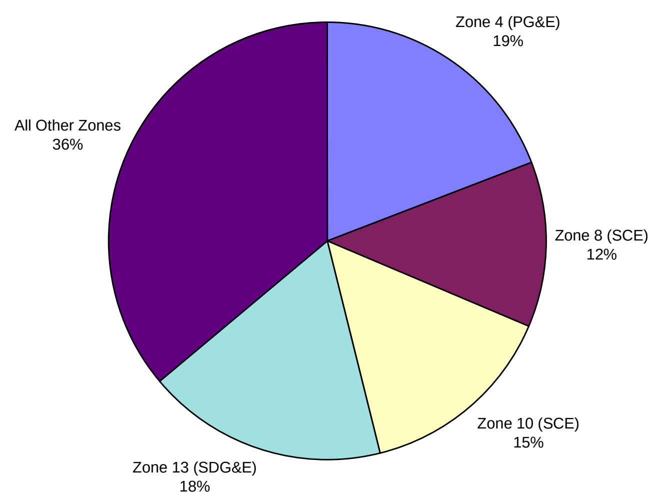
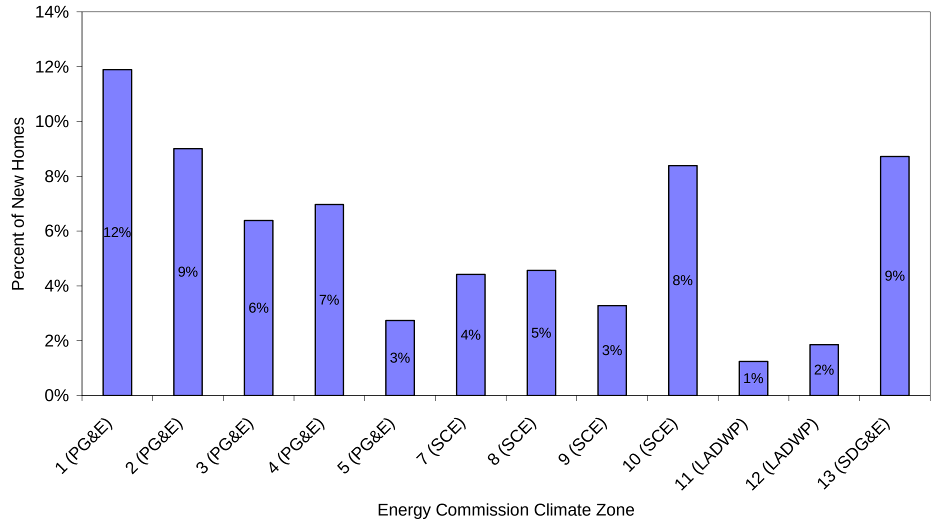


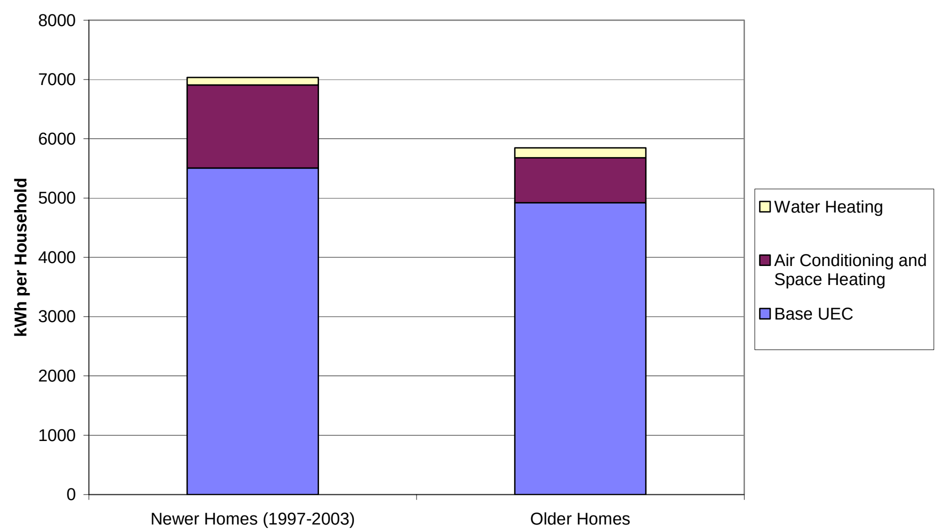
Figure 26 shows housing growth by zone as a percentage of the population in each zone. Zone 1 has the highest relative growth mostly because it is a large area with a relatively low base population that has seen solid growth in recent years.

Figure 26
Housing Growth Rate by Climate Zone



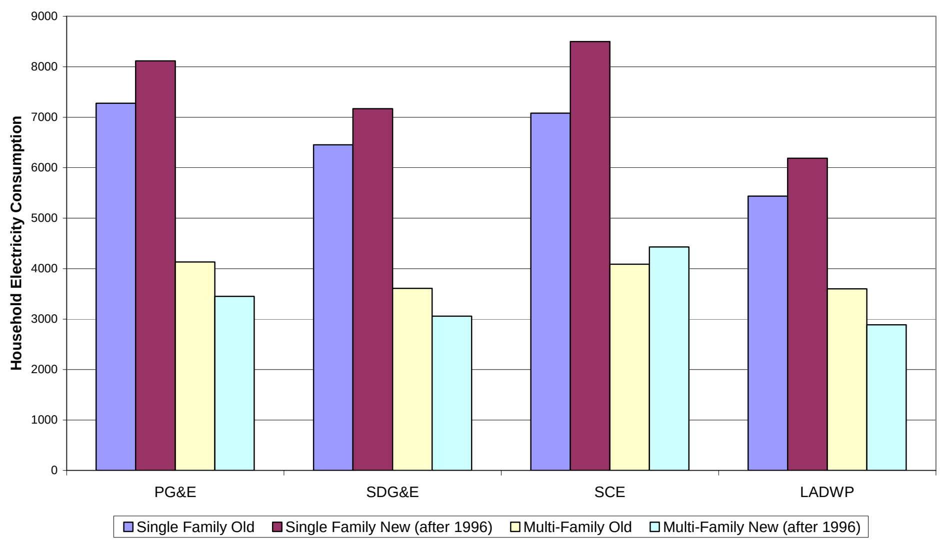
As shown in Figure 27, average electricity use in newer dwellings is 7,035 kWh per year compared to 5,846 in older dwellings. There are several factors affecting the increased usage including larger dwellings, more occupants per home, and more affluent occupants. Space conditioning shows the biggest increase because the saturation of central air conditioning in new dwellings (78%) is higher than that in older dwellings (41%).

Figure 27
Electric UECs for Newer and Older Dwellings



While the overall usage is shifting upwards, the increase is only occurring in single family dwellings (Figure 28). In general, new multi-family dwellings are using less energy than existing buildings with the exception of the SCE service territory.

Figure 28
Electric UECs for Newer and Older Dwellings by Dwelling Type

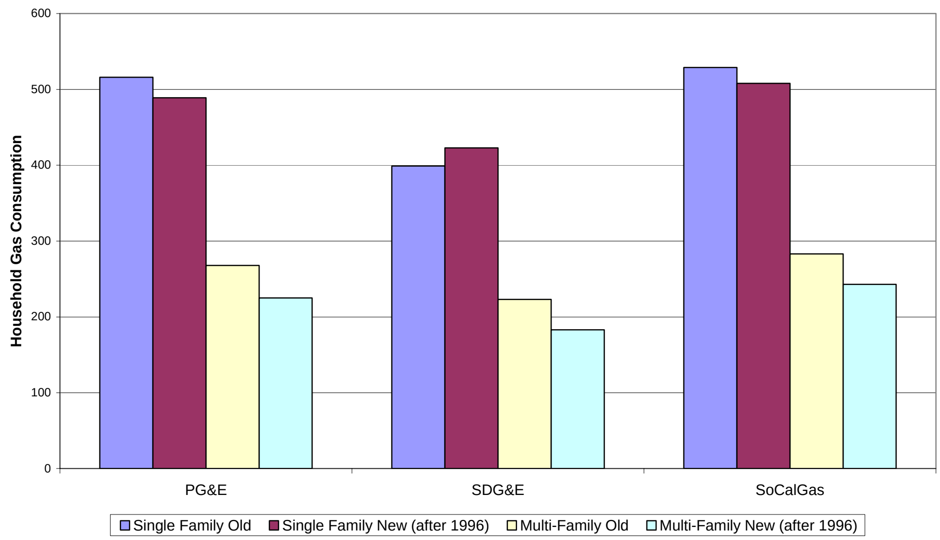


Gas shares are increasing as shown in the fuel share section (1.4). Despite this, new homes are using approximately the same amount of energy as older homes (Figure 29).

Figure 29
Natural Gas UECs for Newer and Older Dwellings

While the average gas use for new dwellings is slightly higher than older dwellings, this can be a little misleading. If you examine usage by utility and dwelling type, the average use is declining for all groups with the exception of single family homes in SDG&E (Figure 30)[7]. A higher portion of new homes are single family dwellings which in turn increases the overall statewide average gas use for new dwellings.[8]

Figure 30
Natural Gas UECs for Newer and Older Dwellings by Dwelling Type



In order to review all of the factors affecting new dwellings, Table 6 provides a comparison of the characteristics of newer and older dwellings. New dwellings are 42% larger than the average existing stock and occupied by homeowners with higher incomes. While newer dwellings have slightly lower cooling degree days than older dwelling, they have central air conditioning installed at almost double the rate of existing dwellings. The overall usage increase from older to newer dwellings is lower than might be expected using these facts alone. New dwellings use 20% more electricity and about the same amount of gas. As a counter to these upward trends, conservation equipment is going into newer dwellings at higher rates which is helping to control the rate of energy consumption growth.

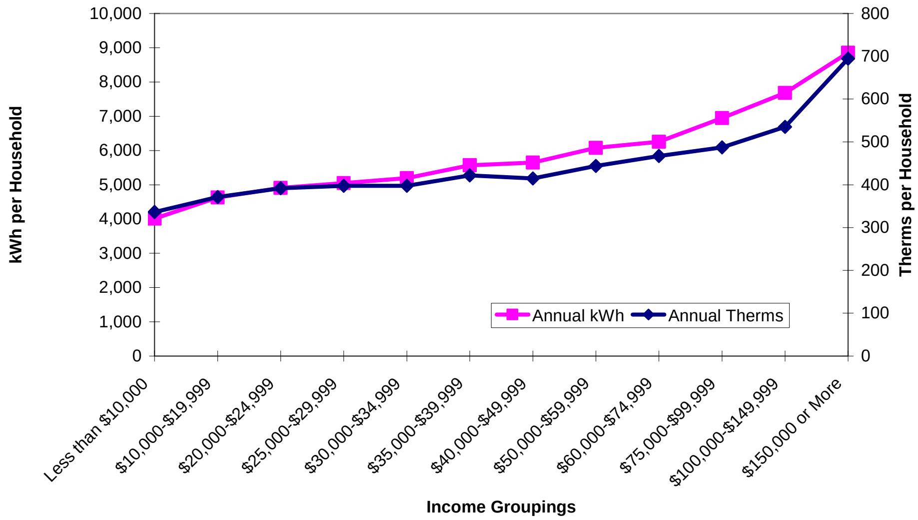
Table 6
Comparison of Newer and Older Dwellings

	Newer Dwellings (Built after 1996)	Older Dwellings	Percent Difference
Annual Electric Household Consumption	7,159	5,960	20%
Annual Gas Household Consumption	468	459	2%
Dwelling Size	2039	1,434	42%
Number of Residents	3.14	2.93	7%
Average Income	86,276	58,082	49%
Percent Single Family	74%	58%	28%
Owners	83%	62%	35%
Saturation of Central AC	78%	41%	93%
Cooling Degree Days	962	900	7%
Cooling Degree Days (those with CAC)	1,119	1,279	-13%
Programmable Cooling Thermostat	85%	47%	83%
Pool Saturation	13%	8%	59%
Average Number of Computers per Home	1.21	0.93	30%
Gas Primary Heating	86%	83%	5%
Heating Degree Days	2,050	2,023	1%
Exterior Wall Insulation Throughout	91%	51%	77%
Attic Insulation	91%	66%	38%
Double Pane Windows Throughout	79%	31%	157%
Low Flow Showerheads Throughout	71%	54%	32%
Average Number of CFLs per Home	2.29	1.74	32%
Horizontal Access Washers	13%	9%	43%

1.7 Income Effects

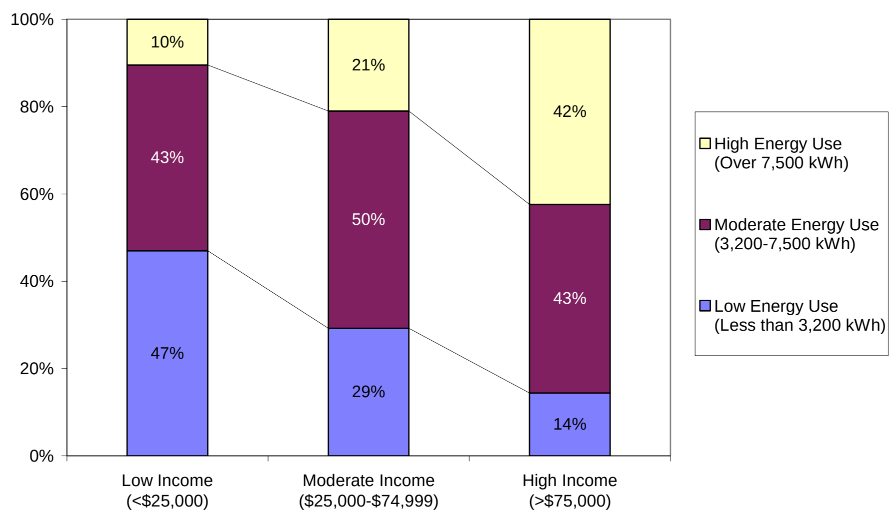
As shown in Figure 31, both electricity and natural gas usage increase as income levels increase.

Figure 31
Average Electricity and Natural Gas Use by Income



While income is strongly correlated with energy use, low usage does not imply that customers are low income (see Figure 32). By breaking electricity usage into quartiles (moderate includes the two middle quartiles for each case), it follows that 12% of the low income group has the highest energy use (over 7,500 kWh per year) while 13% of high income families use less than 3,200 kWh per year.

Figure 32
Electricity Usage Compared with Income



Overall, the income breakdown follows expected trends with respect to the fact that higher income households use more energy. This is indicated in Table 7 by the larger dwellings, increase in central air conditioning, more pools, and more computers.

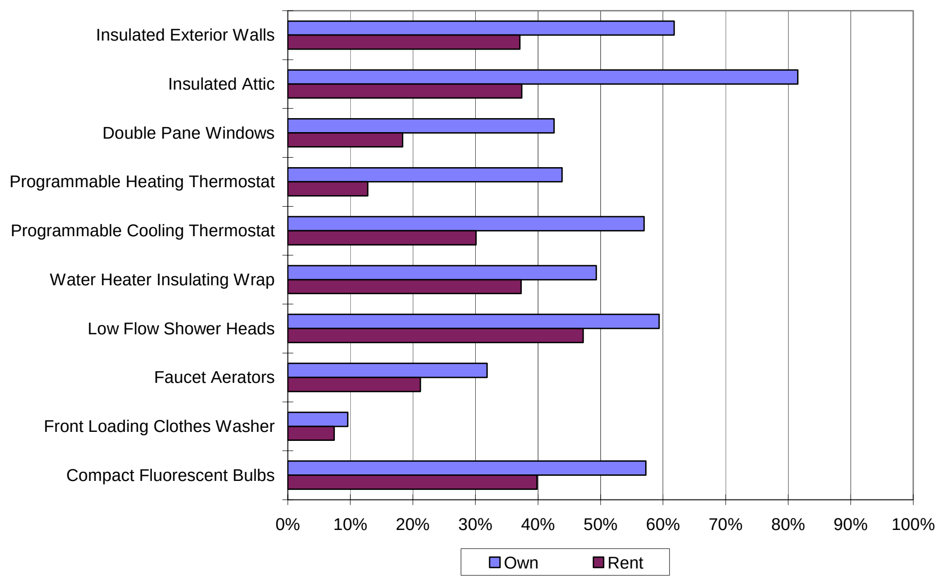
Table 7
Comparison of Households by Income

	Low Income (<$25,000)	Moderate Income ($25,000-$74,999)	High Income (>$75,000)
Percent of Population	24%	50%	26%
Dwelling Size	1,009	1,369	2,062
Dwelling Age	36.3	34.0	29.4
Percent Single Family	37%	59%	78%
Percent Own	37%	63%	86%
Number of People	2.80	2.92	3.11
Annual Electric Household Consumption	4,552	5,683	7,895
Annual Gas Household Consumption	370	430	575
Central Air Conditioning Saturation	32%	42%	54%
Gas Heating Saturation	78%	83%	86%
Pool Saturation	2%	6%	19%
Average Number of Computers per Home	0.46	0.90	1.47
Work at Home	15%	17%	27%
Programmable Heating Thermostat	14%	29%	55%
Dwellings with CFLs	42%	50%	60%

1.8 Energy Efficiency Actions and Opportunities

Energy efficiency actions are present in increasing numbers as technologies become more popular and more readily available or are required by changes in building codes. Figure 33 shows that people who own their dwelling are more likely to take energy efficiency actions than renters. Note that all actions represent the number of homes with a given efficiency improvement in place. In the case of low cost “portable” measures such as compact fluorescent bulbs, which could benefit renters directly and have a very short payback period, there is still a large relative difference in the adoption rates between owners (57%) and renters (40%).

Figure 33
Energy Efficiency Actions/Equipment by Ownership



Owners make up 63% of the population and renters the remaining 37%. Owners are predominantly in single family dwellings (79%) while renters make up 9% of townhouses, 20% of apartments with two to four units, 46% of apartments with more than five units, and 1% of mobile homes.

Figure 34 compares these same energy efficiency actions and equipment across newer and older dwellings. This comparison highlights the fact that participant knowledge of efficiency details is somewhat limited. Saturations of major measures such as insulation and double pane windows should be 100% based on building standards. The fact that they appear lower in Figure 34 is indicative of the fact that not all participants were aware of what they have in their dwellings. Personally driven efficiency actions that are not tied to a new dwelling standard such as front loading clothes washers and compact fluorescent bulbs show a much closer comparison between newer and older dwellings.

Figure 34
Energy Efficiency Actions/Equipment by Dwelling Age

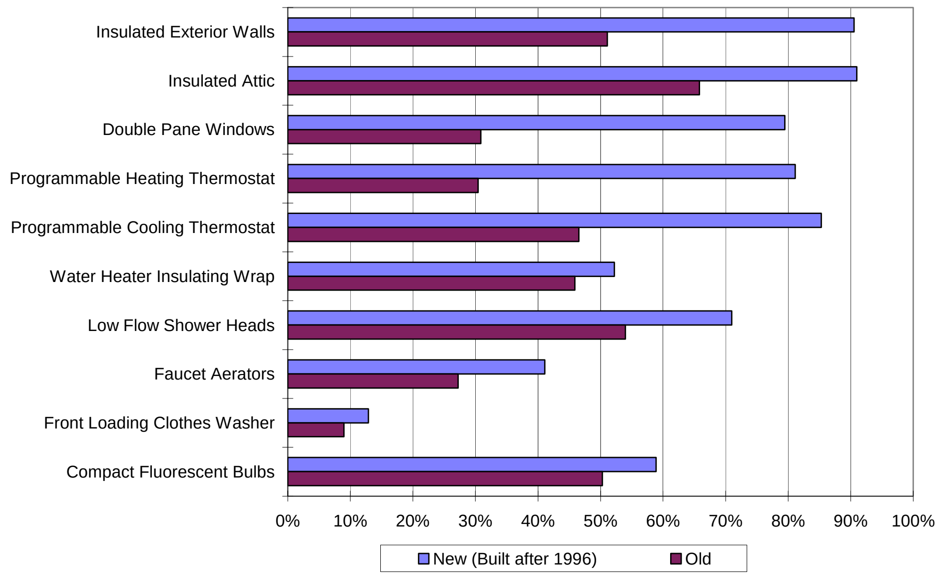
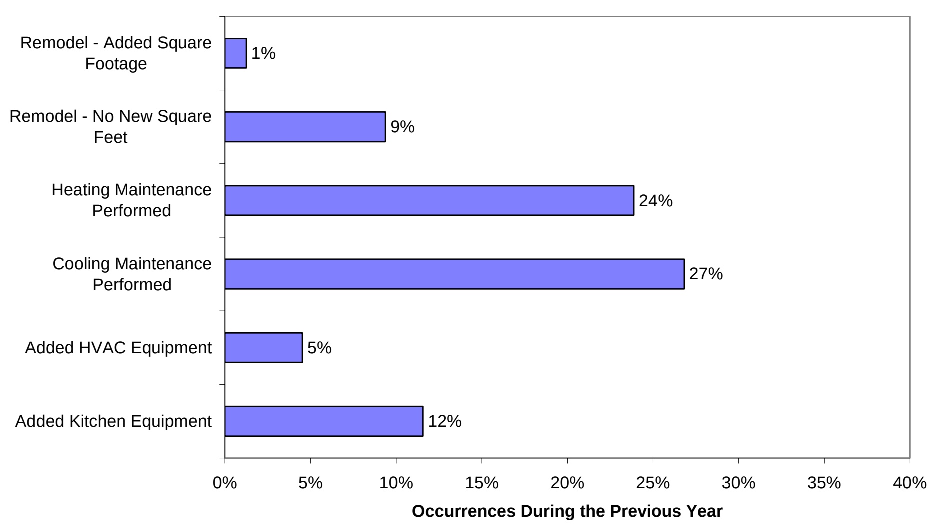


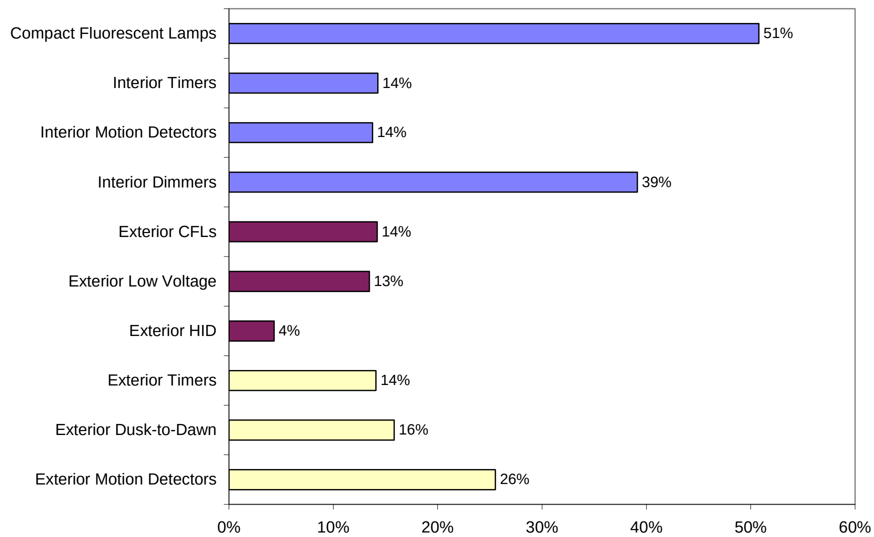
Figure 35 provides examples of opportunities for energy efficiency communication or sales with customers. On average, one in ten dwellings was remodeled in the previous 12 months. Ten percent of those dwellings included the addition of square footage. Maintenance, major equipment replacement, and kitchen appliance remodels also raise opportunities for households to increase efficiency.

Figure 35
Remodeling and Repair Opportunities



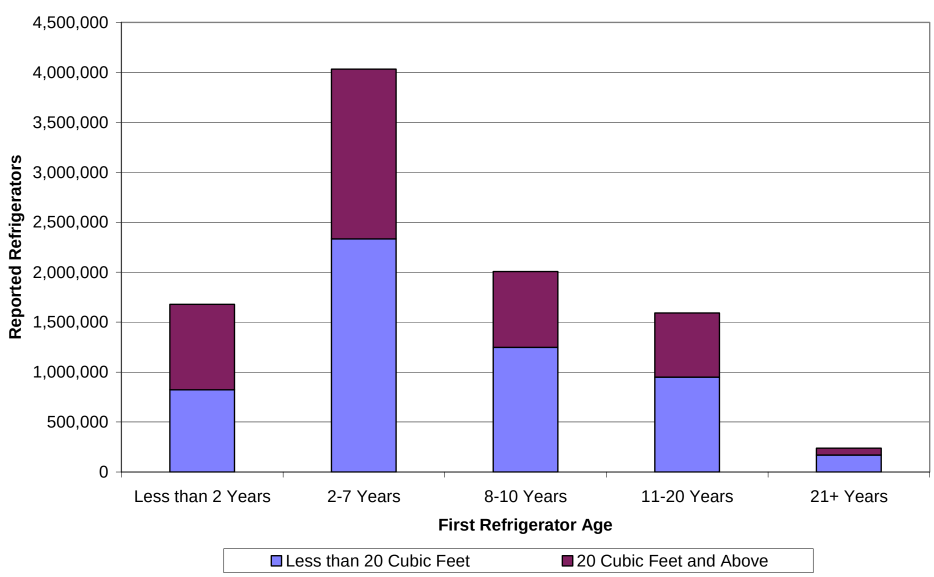
Compact fluorescent lamps (CFLs) have been heavily marketed through various program initiatives throughout the state. Interior CFLs can be found in 51% of all dwellings (Figure 36).

Figure 36
Penetration of Various Lighting Equipment and Devices



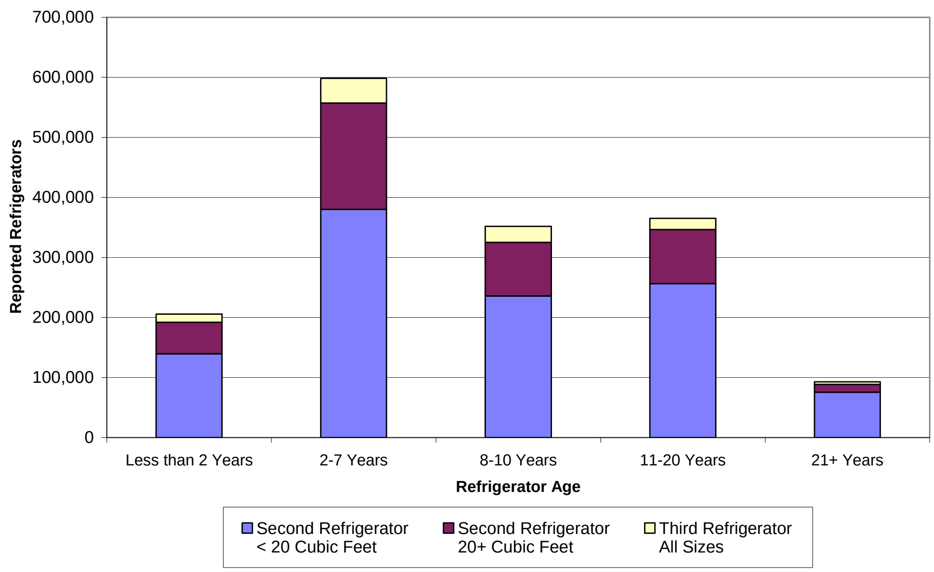
The UEC for first refrigerators is 789 kWh per household. From Figure 37, there are a total of 1.8 million refrigerators that are 11 years or older and will likely need to be replaced in the next five years. Currently, 42% of all refrigerators are over 20 cubic feet in size, however, 51% of new refrigerators fall in the over 20 cubic foot category. Six percent of all customers reported that they discarded a refrigerator in the prior twelve months.

Figure 37
First Refrigerators by Size and Age



Second and third refrigerators use an average of 1,178 kWh per unit. 18% of dwellings report at least one additional refrigeration unit. While there are almost 460 thousand additional units that are 11 years or older, there is a relatively strong market for new additional units as well (Figure 38).

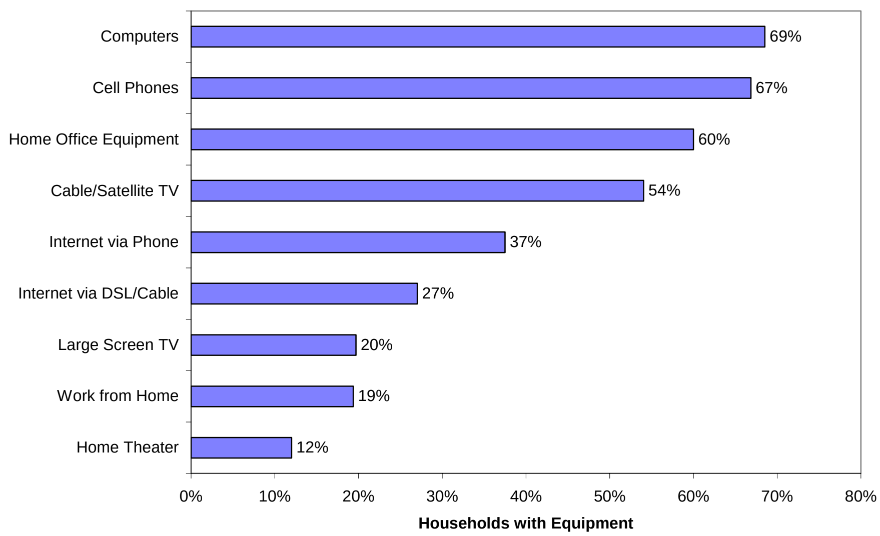
Figure 38
Second and Third Refrigerators by Size and Age



1.9 Technology

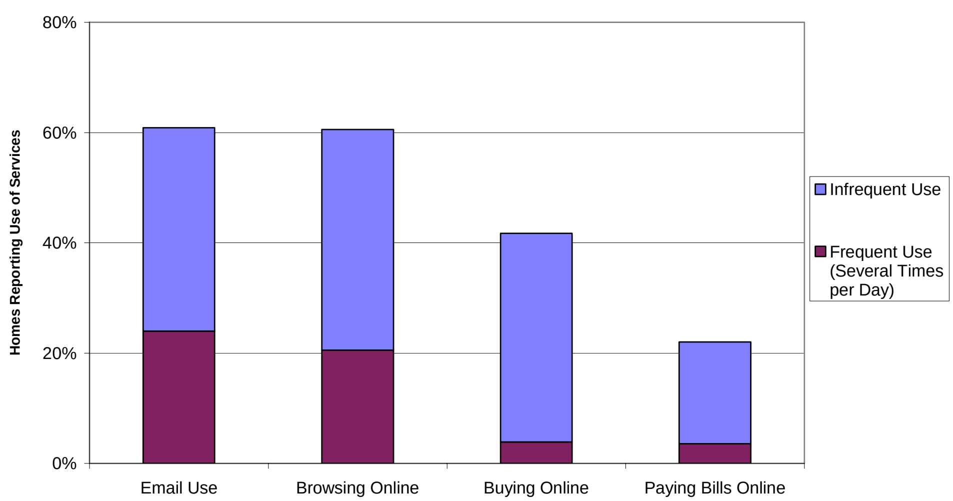
While the number of dwellings with more than three computers is just under 6%, there is a computer in 69% of all dwellings (Figure 39). Other entertainment, general technology, and communication services are also appearing in numerous dwellings.

Figure 39
Penetration of Technology Equipment



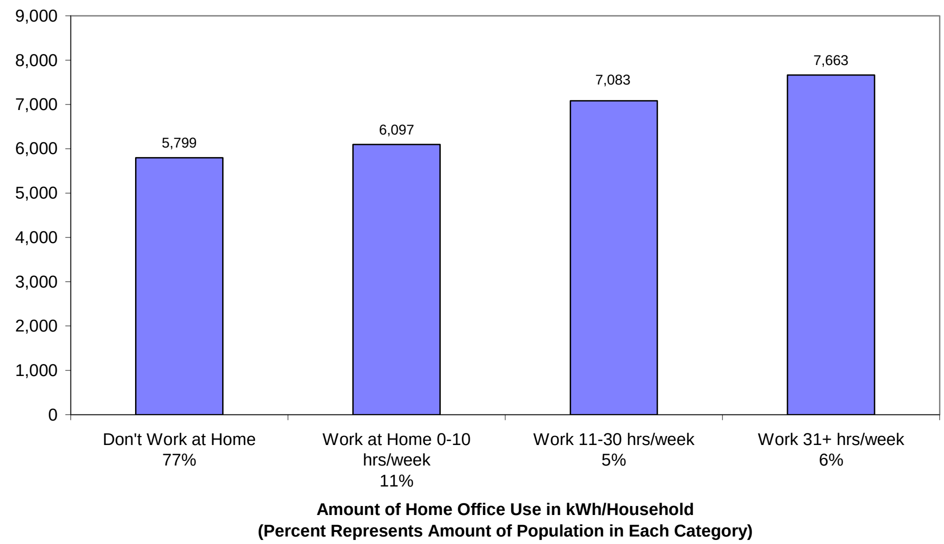
As people have more PCs, they are spending much more time on the PC and using it for a range of other services (Figure 40).

Figure 40
Use of Online Computer Services



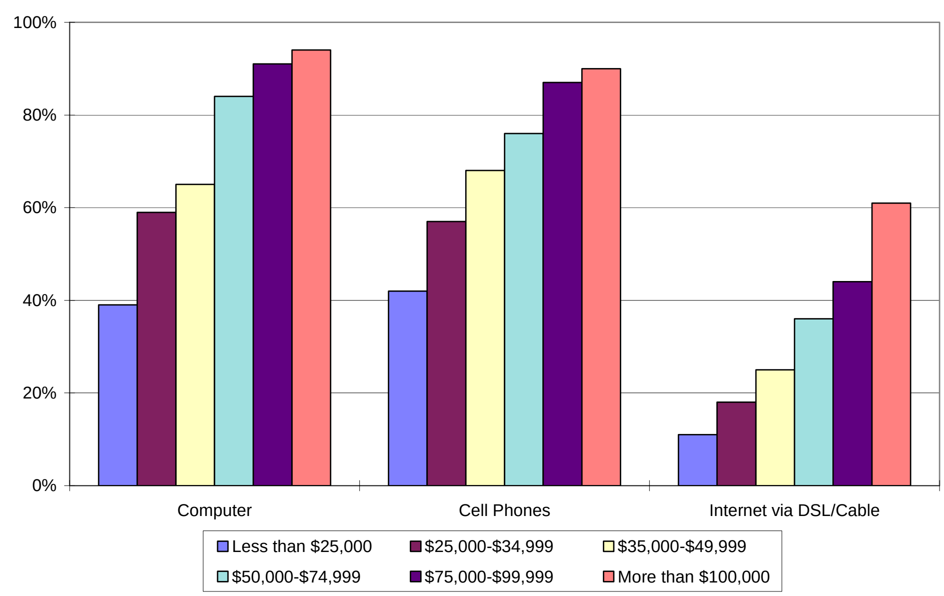
Home offices are currently found in 23% of all dwellings. While home offices add to energy use, they occur in all energy use categories. As home offices are used more regularly, average consumption per household increases (Figure 41).

Figure 41
Electricity Use by Amount of Home Office Use



Many discretionary end uses have a strong income correlation. Figure 42 provides three examples of that trend.

Figure 42
Technology Services by Income



1.10 Data Comparisons

Effect of Combining the Main Sample and Non-response Follow-Up Sample

To combine the results from the main sample and the follow-up efforts, the study combined the weights from both components to create a set of individual weights that represents the number of households that each participant represents. Instead of fully weighting the non-respondent results to represent all non-respondents, the follow-up sample weights were reduced in a systematic approach. This assumed that the follow-up sample represents only those customers who would respond to the follow-up survey but not to the main survey, rather than assuming the follow-up respondents represent all non-respondents to the main survey. This approach improved overall precision and reduced the likelihood of individual outlier cases in the non-respondent sample from skewing overall results. The non-response follow-up proved to be a successful way to capture a segment of the population underserved by the direct-mail campaign. Table 8 shows several key results for customers by dwelling type and survey method.

In general, non-respondents had similar energy usage and major equipment holdings as direct-mail participants but differed significantly in that they were less likely to be property owners, less likely to be using energy-efficient lighting, more likely to be non-English speaking, more likely to be ethnically diverse, and less educated overall. It follows from this that the direct-mail campaign was most successful with individuals who were more aware of energy efficiency, were more motivated because of their ownership, more educated, and more capable of handling an English survey. The non-response follow-up was able to get to more Spanish-speaking customers. While the non-response follow-up adds significant cost to a project of this magnitude, the fact that customers differ in these ways indicates that it is a wise step to take to minimize non-response bias found in a single-method survey approach.

Table 8

Comparison of Results by Surveying Method and Dwelling Type

	Single Family		Multi-Family (2-4 Units)		Multi-Family (5+ Units)		Mobile Homes	
	Initial Mail	Non-Response	Initial Mail	Non-Response	Initial Mail	Non-Response	Initial Mail	Non-Response
Completed Surveys	12,599	1,225	2,979	409	2,866	512	526	37
Weighted to Population	2,363,823	3,693,704	524,317	1,155,001	513,069	1,463,655	95,691	103,602
Average Electric Consumption	7,248	7,160	4,429	4,201	3,689	3,969	6,271	6,531
Average Gas Consumption	547	538	341	338	215	216	491	478
Average Dwelling Size	1,837	1,755	1,156	1,061	925	914	1,258	1,083
Average Dwelling Age	14.5	18.9	24.0	24.8	28.4	34.6	19.4	27.9
Average Number of People	2.88	3.42	2.53	2.74	2.10	2.68	2.30	2.22
Average Number of Seniors	0.53	0.30	0.38	0.13	0.37	0.15	0.74	0.42
Average Income	73,389	68,714	54,246	47,346	45,388	41,702	30,971	28,807
Owners	91%	81%	50%	26%	26%	13%	87%	89%
Central Cooling	50%	47%	40%	33%	41%	31%	60%	38%
Gas Space Heating	85%	89%	77%	75%	46%	54%	57%	56%
All Exterior Walls Insulated	56%	61%	45%	48%	43%	44%	65%	59%
CFL Penetration	63%	50%	55%	42%	51%	37%	57%	51%
Primary Language English	92%	80%	85%	67%	87%	69%	95%	81%
Head of Household Hispanic	12%	26%	17%	36%	13%	33%	9%	20%
College Grad or Higher	53%	44%	47%	39%	50%	36%	23%	18%

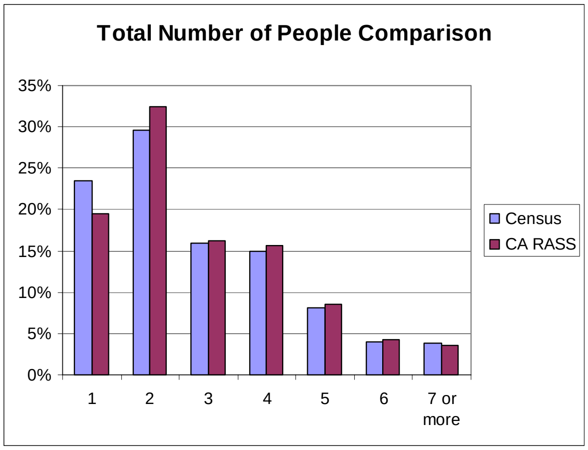
Comparison to Census Data

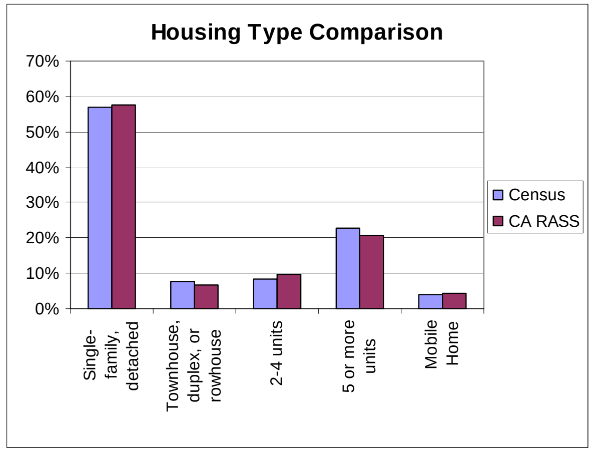
To understand how the results correspond to the population of California, we compared 2000 census data to the RASS results.[9] Overall, the comparison of the RASS demographic information to the 2000 Census data is reasonable, and the sampling plan yielded a set of customer respondents that closely mirrors the population at large. The most notable area where the study appears to fall short is in the single-occupant rental market. The shortfalls occur predominantly in the young-adult age groups. Because the results aligned with census data, the study group decided to keep the initial sample weights and not post-stratify the results.

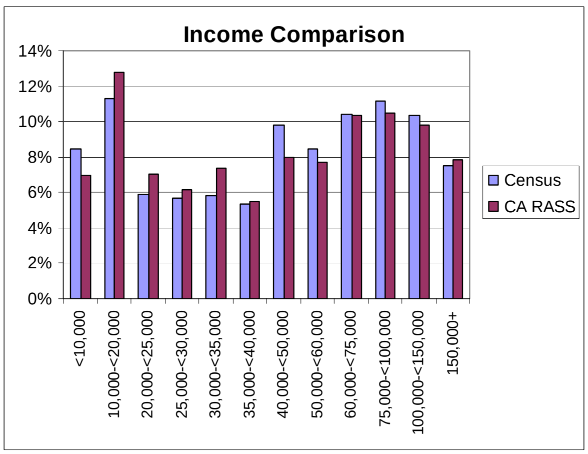
A few of the Census-to-RASS comparison values (most notably ethnicity and language) were asked in a different format from the Census so comparisons are not directly relevant. Despite language results that differ in form enough that a comparison is not meaningful, the fact that our Hispanic ethnicity numbers come out very close to the Census helps to confirm that we were able to capture results from that population segment. As noted above, this is in large part because of the non-response follow-up efforts. A series of Census comparison tables is included below as Figure 43.

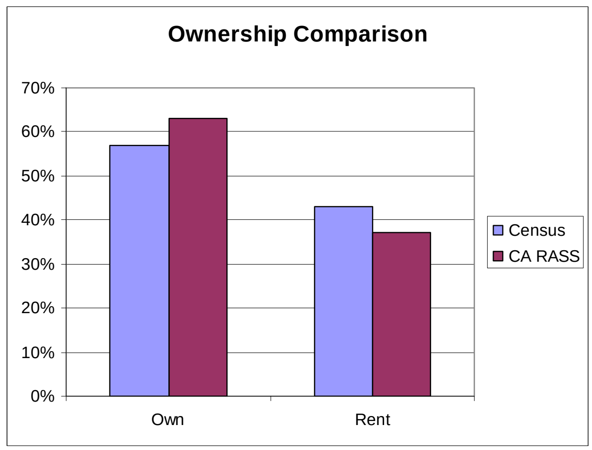
Figure 43

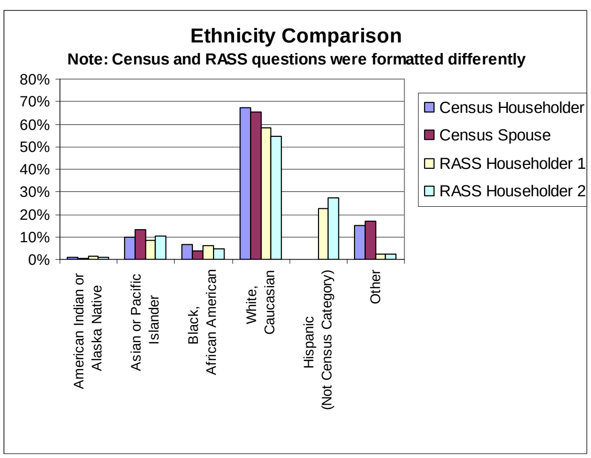
Comparison of RASS Results to 2000 Census Results

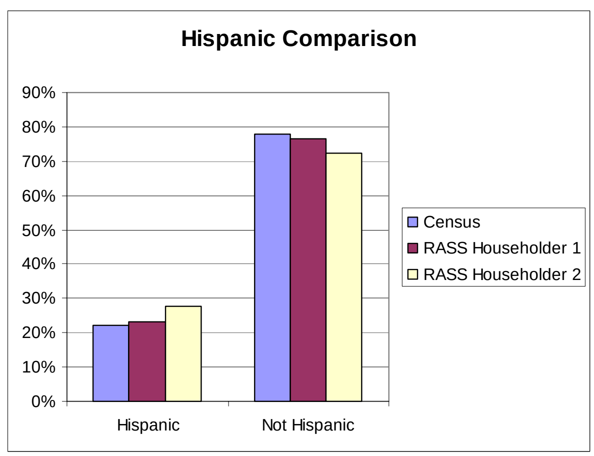












Appendix: Black and White Copy of Figure 2 from Page 8

ENDNOTES

[1] Lighting numbers triangulated from Baseline Energy Use Characteristics, Technology Energy Savings, Volume I, California Energy Commission, May 1994, publication p300-94-006 as well as various KEMA-XENERGY RECAP Program results.

[2] Previous RASS studies were performed by SCE in 1995, PG&E in 1995, and SDG&E in 1993.

[3] Details on the 20/20 program can be found at the Energy Commission web site: http://www.energy.ca.gov.

[4] PG&E press release dated 8/31/2002 which discusses 20/20 program savings in the residential market (http://www.pge.com/news/archived_news_releases/006a_news_rel/020831.shtml).

[5] Energy Commission Forecast Demand Office, April 2003, settlement-quality metered load data from the California Independent System Operator (CAISO) and revised employment data from the California Employment Development Department. Further detail is also available in the Public Interest Energy Strategy Report (Energy Commission Publication #100-03-012F).

[6] This is attributed to the fact that during the course of the study, the statewide 20/20 program was in effect. This program offered customers an opportunity to reduce their total bill by 20% if they reduced their usage 20% from the previous year's usage. As an example of the impact of this program, roughly 30% of PG&E customers qualified for this program in 2001 and 2002.

[7] The SDG&E increase for single family homes is attributable to the fact that new buildings are much larger than older buildings in that service territory and increasing at a much higher rate than in other service territories.

[8] SoCalGas performed an internal re-weighting of their data to account for the customers who were not served by the electrically based population. While the housing type trends are similar to those displayed in Figure 29, the re-weighted values show an overall usage for older homes at 453 therms and new homes at 430 therms. By re-weighting, SoCalGas was able to adjust the balance of single family and multi-family dwellings to better match their population. This resulted in declining energy use overall as well as by housing type for the SoCalGas new home population.

[9] Census Data Source: Census 2000 5% PUMS for California

CALIFORNIA STATEWIDE RESIDENTIAL APPLIANCE SATURATION STUDY

FINAL REPORT

Consultant Report

Prepared for:
California Energy Commission

Prepared by:
KEMA-XENERGY
Itron
RoperASW

June 2004
400-04-009

Prepared by:
KEMA-XENERGY
Itron
RoperASW

Contract No. 400-04-009

Prepared for:
California Energy Commission

Contract Manager: **_Glen Sharp_**

Sponsors:
Pacific Gas and Electric (PG&E)
San Diego Gas and Electric (SDG&E)
Southern California Edison (SCE)
Southern California Gas Company (SoCalGas)
Los Angeles Department of Water and Power (LADWP)

CONTENTS

ILLUSTRATIONS

1: PROJECT METHODOLOGY INTRODUCTION

For the first time in California, the large Investor Owned Utilities (IOUs) pooled resources and performed a RASS and Unit Energy Consumption (UEC) Study as a team. The project was administered by the California Energy Commission and sponsored by Pacific Gas and Electric (PG&E), San Diego Gas and Electric (SDG&E), Southern California Edison (SCE), Southern California Gas Company (SoCalGas), and Los Angeles Department of Water and Power (LADWP). KEMA-XENERGY was the prime consultant. Itron provided data cleaning and performed the Conditional Demand Analysis. RoperASW fielded the non-response follow-up.

The RASS effort has resulted in a research product that provides both statewide and utility-specific results. The study was designed to allow comparison of results across utility service territories, climate zones and other variables of interest (i.e. dwelling type, dwelling vintage, and income). The study includes results for 21,920 residential customers that are weighted to the population represented by the sponsoring utilities. The saturation results capture both individual and master metered dwellings. This rich set of customer data includes information on all appliances, equipment, and general usage habits. The study also includes a detailed conditional demand analysis that calculates unit energy consumption (UEC) values for all individually metered customers.

The study was initiated in late 2002 and the sampling plans and survey implementation occurred throughout 2003. The data was collected using a two stage direct mail survey targeted to a representative sample of California residential customers. The survey requested customers to provide details on their energy equipment and behaviors. A non-response follow-up survey was implemented at the end of the double mailing phase to a sub-sample of non-respondents. The non-response follow-up included telephone and in-person interviews in an effort to minimize non-response bias by using alternative surveying techniques.

The results from the RASS study were used to develop a CDA model. This analytical method uses a combination of customer energy use with the responses from the customer survey to model end uses and develop unit energy consumption results for those end uses. The results of the CDA are included in summary form along with the general study results in this executive summary and are provided in further detail in the methodology section of the report.

The study also includes onsite metering for a sample of 180 RASS participants. The onsite metering sample was designed to over-sample air conditioning use, with the meters gathering both a whole-house and central air conditioning usage at each dwelling. The onsite meters are in the field at the time of publication and the final results from that portion of the project will be delivered as whole house and air conditioning load shapes after the 2004 cooling season has ended.

Because of the need to serve a wide range of users, the study was designed to produce multiple products:

- A high level summary of key findings;
- Detailed saturation tables for all appliances and equipment holdings;
- Detailed UEC tables for 25 electric and 8 gas end uses;
- Whole house and air conditioning load shapes; and
- An Internet-based tool providing customized data filtering and viewing.

The concept of using a statewide survey instrument provided the Energy Commission and other parties with a consistent set of questions and study results to use for statewide planning and cross utility comparisons. In addition, the sample includes sufficient data enabling utility specific analyses. The project required a cooperative effort among the sponsors as they came together to create a unified research plan, program materials, and implementation strategy. The sponsors all shared project costs and final results. Each utility provided the data necessary to create a unified sampling plan. Each utility also provided customer specific information for customers who were selected for the sample. In order to insure individual customer anonymity, the study participants were assigned a generic identification number that includes details about sampling their strata. Respondent zip codes are the only other information that is generally available in the final study database as to the customer's location. In addition to the "non-confidential" data, each utility received a "confidential" dataset of results for their service territory with customer identification information as provided by the utility initially. This key allows the utility to match up the RASS data with their own account information.

This report is split into two volumes because of the size. Volume One describes the study design and implementation methods while Volume Two details the results in the form of UEC banners and saturation banners.

2: STUDY DESIGN AND IMPLEMENTATION

2.1 OVERVIEW OF STUDY DESIGN

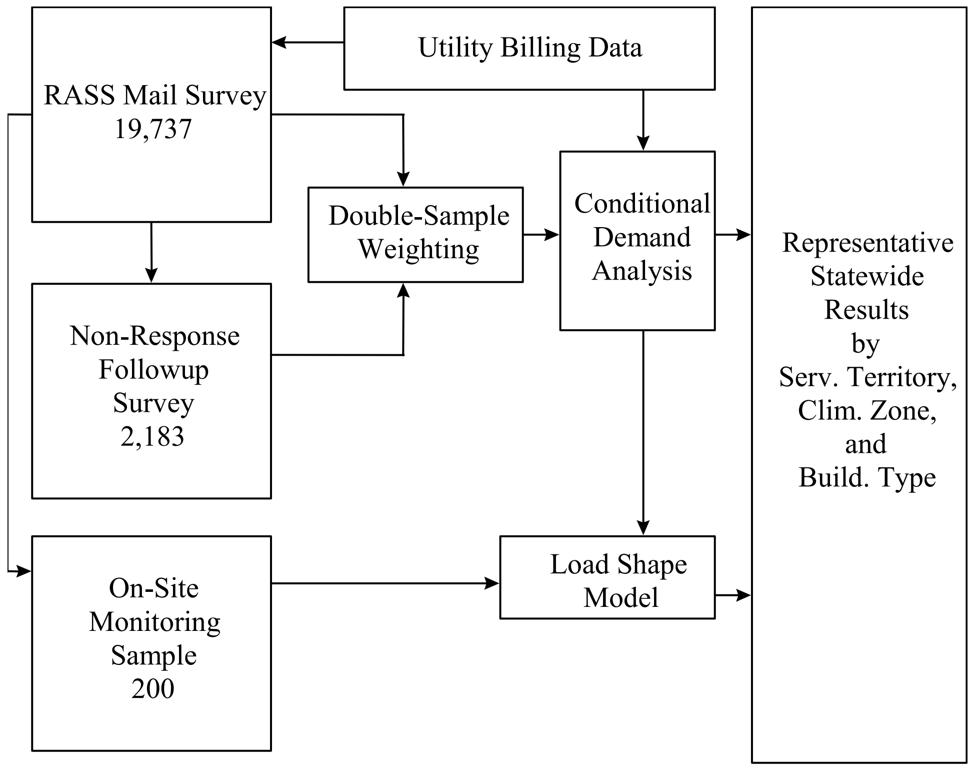
A hybrid data collection strategy involving four different data collection methods was employed for the RASS. An overview of our approach is shown in Figure 2-1. The numbers indicated on Figure 2-1 are the number of completed surveys received through each method. This section first discusses the sample design for each of the surveying components and then describes responses for each surveying type, the process of weighting results to the population, and the study's resulting precision.

Most of the survey data was collected using a mail survey. Telephone interviewing was used to gather data initially from master-metered electric accounts and for collecting survey data from a sample of non-respondents to the mail survey. An in-person interview was used in a similar fashion as part of the non-response study for cases where telephone numbers could not be obtained or attempts by phone proved unsuccessful. Finally, we collected hourly electric load data from the total home and a central cooling system for a small sample of homes. Detailed on-site surveys were conducted on the homes in the hourly metering sample. The onsite metering sample count includes results for the 180 sites metered, which are also included in the mail survey results total.

The RASS study included sending out two rounds of mail surveys to approximately one hundred thousand homes that are served by an individual electric meter. We obtained survey responses from 18,970 of these individually metered homes via the two rounds of mail surveys.

To reduce the non-response bias that was likely to occur from a mail survey alone, a second step of surveying efforts was pursued on a sample of 5,000 non-respondents to the mail survey. We ultimately surveyed a total of 2,183 of these non-respondents using either a third mail survey with an incentive, a telephone interview, or an in-person interview at the home.

Figure 2-1
Overview of Approach



Master metered electric accounts that serve between 2 and 4 units were surveyed similar to individually-metered accounts, except the cover letter was written with special instructions for the tenant(s) to fill out the survey for only one of the units in the building. Master-metered accounts that serve more than 4 units were surveyed using a two-stage method. In the first stage, a telephone survey was conducted with a facility manager of the master-metered complex or mobile home park to obtain data on the common area equipment and to obtain mail addresses for specific dwelling units served by the account. The second stage involved selecting a sample of the dwelling units that were identified in stage one. Mail surveys were sent to each sampled unit with phone follow-up activities to non-respondents to help maximize response. We completed 616 stage one interviews with 5,593 surveys being sent out in the second stage to addresses collected through the telephone process. A total of 767 master metered mail surveys were completed and returned.

The on-site metering component collected hourly load data for the total home and the central cooling system. For each onsite-metered site, we have the responses to the standard RASS questions so that the metering information can be leveraged using the larger set of RASS responses. In addition, we have collected detailed housing shell, lighting, plug load, and nameplate data for use in analyzing the variation of hourly demand for the metered participants. The metering data will be presented in the form of whole house and air conditioning load shapes. In order to capture a full air conditioning cycle, meters are being left in through the summer of 2004 and the load shapes will be provided as an addendum to this report in the fall of 2004.

Using a detailed sample plan and subsequent weighting of the data to the population, the data collection activities resulted in a representative database containing 21,920 responses to the RASS survey. These data have been combined with electric and gas billing data to estimate unit energy consumption using conditional demand analysis. The full RASS data and conditional demand analysis provide saturation and end use shares that are statistically reliable for all of the segments of interest (e.g., by service territory, climate zone, and dwelling type). The second, smaller, database of 180 homes will contain the hourly load shape data along with the expanded set of survey information.

2.2 SAMPLING APPROACH

The sampling approach for RASS was based on residential population sample frames obtained from the sponsoring electric utilities (PG&E, SCE, SDG&E and LADWP). A multi-step approach was used to obtain the data. The first data transfer included all of the residential utility records with variables that were required to create the sample. Once the sample was drawn, a second request gathered the customer specific contact information for the sample subset. This multi-staged approach was preferred to maintain confidentiality of each utility's population frame.

A third round of data transfer occurred closer to the analytical piece of the study. This included a request for transaction level billing data for all customers targeted in the sample. In addition to gathering billing histories from the electric utilities, this step included a billing data matching process that located gas records for customers in SoCalGas' service territory as well as individuals served by other gas utilities from whom we could request bills. The bill matching process and cleaning is detailed in Section 4.2. By completing this step after the mailings were sent, we were able to obtain a more current set of bills to use in the conditional demand analysis.

The billing data used in the RASS study are shown in Table 2-1.

Table 2-1
Utility Billing Data Requirements

Electric Population Frame Data • Premise/Control # • Service city and 5-digit service zip code • Average Daily kWh consumption for premise over previous 12 months • Dwelling type indicator • Geo-demographic indicator (if available) • Electric rate schedule with baseline allowance codes • Gas service indicator (PG&E and SDG&E) • Other geographic indicators (division, forecast climate zone) • Service description field • Meter set date or Premise establishment date • Customer service start date
Contact information for sampled accounts • Service address • Mail address, city, state, and zip • Customer name
SoCalGas Population Frame, and PG&E gas-only accounts or SDG&E gas-only accounts • Premise/control # • Service address, city and 5-digit service zip code • Mail address, city, state, and zip (for supplemental matching) • Customer name (for supplemental matching)
Transaction billing data (for all sampled accounts) (One year of transactions for customer that resides in the dwelling at the time of the first survey mailing). • kWh and/or therm usage • Current and prior read dates • Transaction type (regular bill, adjustment, estimate) • Applicable rate schedule with baseline allowance • Any information describing unique characteristics of account, e.g., employee code.

The remainder of Section 2.2 outlines the sample design and provides further detail on how the variables included in Table 2-1 were used for the study sample. The individually metered sample design is discussed first, followed by the designs for the master metered customers, the non-response follow-up and finally the onsite metering sites.

2.2.1 Individually Metered Sample Design

We used a stratified random sample design for individually metered customers. We worked with Energy Commission staff to determine the appropriate stratification variables. The total population[1] was split into 105 strata based on electric utility, age of home, presence of electric heat, home type, and Energy Commission forecast climate zone.

Stratification Variables

The first four of the five stratification variables were constructed using utility billing records for the residential population provided by the utilities during the first phase of data transfer. Age of the home was determined by the meter set date variable. Presence of electric heat was determined based on a flag in the billing system that indicates whether the home was likely to have electric heat. The electric heating variable is typically recorded by the utility at the time of the meter set date and is often used to determine the customer's baseline energy usage allotment. Home type was determined for PG&E and SCE using a combination of the annual usage variable and the dwelling type variable. Both PG&E and SCE use a rate code variable that indicates whether the home is likely single- or multi-family. SDG&E and LADWP do not have a variable in their billing records that indicates the likely dwelling type, and as such the home type variable for these two utilities was constructed using the annual usage variable only. The Energy Commission climate zone variable was constructed using the utility billing record variable "service zip code" mapped to a lookup table of Energy Commission forecast climate zones by zip code. The Energy Commission provided this lookup table.

The five stratification variables were assigned the values listed below. Each of the strata variables is used in the designation of the SFCODE which is the first six digits of each individually metered customers' unique identifier. The position each value takes up in the SFCODE is noted next to the variable and the number in parentheses following the description is the value used in the SFCODE creation. Note that some of the strata sub-groups for sample frames were too small and were combined with other groupings. In these cases a simpler SFCODE is used which designates the utility and a simple number as the last digit.

- Electric utility (1st position of SFCODE): 1=PG&E, 3=SDG&E, 4=SCE, or 5=LADWP;
- Age of home (2nd position): 0=old (prior to 1997) or 1=new (1997 or newer);

- Presence of electric heat (3rd position): 0=no or 1=yes based on utility billing records;
- Home type (4th position): combination of dwelling type (where available from utility billing records) and usage
- for PG&E and SCE: 1=single family "high" (= 15 kWh/day), 2=single family "low" (< 15 kWh/day), or 3=multi-family (all usage values); and
- for SDG&E and LADWP: 4=low (< 10 kWh/day), 5= medium (10 - 20 kWh/day), or 6=high (> 20 kWh/day).
- Energy Commission forecast climate zone (5th and 6th position).

Sample Frame

Table 2-2 presents the individually metered sample frame. The columns of the table include the following information for each of the stratum.

- Columns A through E indicate the strata variables;
- Column F shows the designated SFCODE prefix used in the database for each customer assigned to this strata;
- Column G shows the proportion of the population for each strata, where the total population comprises households in PG&E, SCE, SDG&E, and LADWP electric service territories;
- Column H shows the target number of completes using a modified proportional allocation method;
- Column I gives the expected response rates per strata;
- Column J contains the minimum mailout, which was determined by the expected response rates along with the target number of completes; and
- Column K shows the actual mailout, which is 1% higher than the minimum mailout to account for potential turnover of households.

Table 2-2
Individually Metered Sample Design

A	B	C	D	E	F	G	H	I	J	K
Electric Utility	Home Age	Electric Heat Presence	Home Type	CEC Forecast Climate Zone	SFCODE Prefix	Proportion	Target Completes	Expected Response Rate	Minimum Mailout	Actual Mailout
LADWP	Old	No	Low	11	500411	3.76%	1,644	40%	4,111	4,152
				12	500412	1.18%	517	40%	1,293	1,306
			Medium	11	500511	1.63%	713	55%	1,296	1,309
				12	500512	0.95%	413	55%	750	758
			High	11	500611	0.40%	173	55%	314	318
				12	500612	0.44%	190	55%	346	350
		Yes	All	All	500001	0.10%	150	50%	300	303
	New	No	Low	11	510411	0.23%	150	40%	375	379
		All others	All others	All others	500002	0.18%	150	50%	300	303
PGE	Old	No	SF-Low	1	100201	0.53%	230	45%	511	516
				2	100202	0.57%	251	45%	558	563
				3	100203	1.62%	708	45%	1,573	1,588
				4	100204	3.05%	1,333	45%	2,962	2,991
				5	100205	4.17%	1,822	45%	4,048	4,088
			SF-High	1	100101	0.51%	222	55%	403	407
				2	100102	1.18%	515	55%	936	945
				3	100103	3.21%	1,403	55%	2,551	2,577
				4	100104	3.77%	1,647	55%	2,995	3,025
				5	100105	4.38%	1,915	55%	3,482	3,516
			MF	1	100301	0.13%	150	40%	375	379
				2	100302	0.47%	206	40%	516	521
				3	100303	1.13%	493	40%	1,232	1,244
				4	100304	2.05%	896	40%	2,239	2,261
				5	100305	4.31%	1,882	40%	4,704	4,751
		Yes	SF-Low	1	101201	0.37%	160	45%	356	360
				3	101203	0.18%	150	45%	333	337
				4	101204	0.20%	150	45%	333	337
				5	101205	0.17%	150	45%	333	337
			SF-High	1	101101	0.90%	394	55%	716	723
				2	101102	0.30%	150	55%	273	275
				3	101103	0.75%	327	55%	595	601
				4	101104	0.49%	215	55%	391	395
				5	101105	0.26%	150	55%	273	275
			MF	3	101303	0.15%	150	40%	375	379
				4	101304	0.86%	374	40%	935	944
				5	101305	1.41%	617	40%	1,543	1,558
			All others	All others	100001	0.22%	150	50%	300	303

Table 2-2
Individually Metered Sample Design
(Continued)

A	B	C	D	E	F	G	H	I	J	K
Electric Utility	Home Age	Electric Heat Presence	Home Type	CEC Forecast Climate Zone	SFCODE Prefix	Proportion	Target Completes	Expected Response Rate	Minimum Mailout	Actual Mailout
PGE (cont.)	New	No	SF-Low	2	110202	0.12%	150	45%	333	337
				3	110203	0.19%	150	45%	333	337
				4	110204	0.29%	150	45%	333	337
				5	110205	0.31%	154	45%	343	346
			SF-High	2	110102	0.24%	150	55%	273	275
				3	110103	0.35%	154	55%	280	283
				4	110104	0.30%	150	55%	273	275
				5	110105	0.35%	153	55%	278	280
			MF	3	110303	0.18%	150	40%	375	379
				4	110304	0.17%	150	40%	375	379
				5	110305	0.24%	150	40%	375	379
			All others	All others	100002	0.20%	150	50%	300	303
		Yes	MF	4	111304	0.11%	150	40%	375	379
			All others	All others	100003	0.32%	160	50%	320	323
	All	All	All	14	100004	0.12%	150	50%	300	303
SCE	Old	No	SF-Low	7	400207	0.52%	227	45%	504	509
				8	400208	3.59%	1,566	45%	3,480	3,515
				9	400209	4.30%	1,878	45%	4,174	4,215
				10	400210	2.29%	999	45%	2,220	2,242
				11	400211	0.71%	311	45%	692	699
				All others	400299	0.18%	150	45%	333	337
			SF-High	7	400107	0.83%	362	55%	659	666
				8	400108	3.96%	1,730	55%	3,146	3,177
				9	400109	3.91%	1,707	55%	3,103	3,134
				10	400110	3.58%	1,562	55%	2,840	2,868
				11	400111	0.40%	173	55%	315	318
				All others	400199	0.36%	158	55%	287	290
			MF	7	400307	0.37%	163	40%	407	411
				8	400308	2.60%	1,134	40%	2,834	2,862
				9	400309	2.52%	1,103	40%	2,757	2,785
				10	400310	1.59%	695	40%	1,738	1,756
				11	400311	0.84%	367	40%	918	928
				All others	400399	0.10%	150	40%	375	379

Table 2-2
Individually Metered Sample Design
(Continued)

A	B	C	D	E	F	G	H	I	J	K
Electric Utility	Home Age	Electric Heat Presence	Home Type	CEC Forecast Climate Zone	SFCODE Prefix	Proportion	Target Completes	Expected Response Rate	Minimum Mailout	Actual Mailout
SCE (cont.)	Old (cont)	Yes	SF-Low	8	401208	0.17%	150	45%	333	337
				9	401209	0.12%	150	45%	333	337
				10	401210	0.12%	150	45%	333	337
			SF-High	8	401108	0.27%	150	55%	273	275
				9	401109	0.16%	150	55%	273	275
				10	401110	0.29%	150	55%	273	275
			MF	8	401308	1.19%	519	40%	1,298	1,311
				9	401309	0.63%	277	40%	693	700
				10	401310	0.35%	152	40%	379	383
				11	401311	0.25%	150	40%	375	379
			All others	All others	400004	0.26%	150	50%	300	303
	New	No	SF-Low	8	410208	0.26%	150	45%	333	337
				9	410209	0.15%	150	45%	333	337
				10	410210	0.31%	155	45%	344	348
			SF-High	8	410108	0.29%	150	55%	273	275
				9	410109	0.15%	150	55%	273	275
				10	410110	0.50%	220	55%	400	404
			MF	8	410308	0.15%	150	40%	375	379
		All others	All others	All others	400001	0.38%	164	50%	328	331
	All	All	All	15	400002	0.16%	150	50%	300	303
				16	400003	0.10%	150	50%	300	303
SDGE	Old	No	Low	9	300409	0.44%	193	40%	483	488
				13	300413	3.48%	1,518	40%	3,795	3,833
			Medium	9	300509	0.42%	185	55%	336	340
				13	300513	3.40%	1,484	55%	2,698	2,725
			High	9	300609	0.18%	150	55%	273	275
				13	300613	1.48%	648	55%	1,178	1,190
			All others	All others	300001	0.11%	150	50%	300	303
		Yes	Medium	13	301513	0.17%	150	55%	273	275
			High	13	301613	0.28%	150	55%	273	275
			All others	All others	300002	0.14%	150	50%	300	303
	New	No	Low	13	310413	0.46%	201	40%	503	508
			Medium	13	310513	0.40%	175	55%	319	322
			High	13	310613	0.19%	150	55%	273	275
		All others	All others	All others	300003	0.24%	150	50%	300	303
Total						100.00%	46,807		100,000	100,999

Initial Mail Sample Allocation

We used a modified proportional allocation to assign the RASS sample to each of the strata. This approach ensures sufficient sample is allocated to all strata to achieve a minimum specified precision level for each stratum. Once the minimum sample sizes are determined for each stratum, the remaining sample points are assigned in a manner proportional to the population distribution.

Columns G and H in Table 2-2 show the sample proportions and target number of completes using a modified proportional allocation method.

Proportional allocation gives the best precision for the population as a whole for estimates of saturations or other proportions. At the same time, assigning target completes by cell ensures representation in the sample of each of these population segments. The stratification also allows higher mailout rates for groups that are likely to have lower response rates based on experience from prior RASS studies.

In Table 2-2, column K shows the actual mailout. The actual mailout is somewhat higher than the minimum mailout (Table 2-2, column J) because we expected a limited number of households to "turnover" between the time that the sample is pulled and the surveys are mailed.

The total sample mailout was set at 100,999 (approximately 1% greater than the minimum mailout of 100,000) and the total target number of completes was 46,807 based on an average response rate of 47%.

2.2.2 Master Metered Sample Design

Master metered electric accounts were surveyed differently depending on the type of units the account serves. All master-metered accounts were assigned sample based on a proportional sample design that approximates the ratio of target completes to the number of units or dwellings (not accounts). For this study, we stratified master-metered accounts by utility and by type of account: master-metered accounts serving 2-4 units, mobile home parks with 5 units or more, multi-family complexes with 5-20 units, and multi-family complexes with more than 20 units. Each type of account was associated with a unique survey approach to most effectively solicit a response.

Accounts serving 2-4 units were surveyed similarly to individually metered accounts, in that one survey was mailed to the service address associated with the account. The cover letter instructed the account holder to fill out the survey for one of the units.

Master-metered accounts serving more than 4 dwelling units were surveyed using a two-stage method. In the first stage, we conducted telephone surveys with a facility manager of the multi-family complex or mobile home park to obtain data on the common area equipment and to obtain mail addresses for the dwelling units served by the account. The number of calls that were made was determined by considering both the desired target completes (unit level) for a given stratum and the number of surveys to send for each account for which a telephone survey has been completed. For account types with little variation within a particular account (e.g., medium-sized multi-family complexes), it was preferable to send a smaller number of surveys per account to obtain more variation by surveying more accounts. For account types with more variation, such as mobile homes parks, sending a larger number of surveys per account was appropriate.

The second stage involved selecting a sample of units based on information provided by the facility manager. For multi-family accounts with 5-20 units, we sent four surveys each to the complexes surveyed in stage one. For multi-family accounts with more than 20 units, we sent eight surveys each. For the mobile home parks, we sent ten surveys each to the parks surveyed in stage one.

Stratification Variables

We used a stratified random sample design for the master metered customer base. The total population of master-metered homes[2] was split into 16 strata based on two variables. Each of the strata variables is used in the designation of the SFCODE which for master metered sites is the first two digits of each customers' unique identifier. The position each value takes up in the SFCODE is noted next to the variable and the number in parentheses following the description is the value used in the SFCODE creation.

- Electric utility (1st position in SFCODE): 1=PG&E, 3=SDG&E, 4=SCE or 5=LADWP; and
- Type: 1=2-4 unit, 2=multi-family (5-20 units), 3=multi-family (>20 units), or 4=mobile home (greater than 4 units).

Sample Frame

Table 2-3 presents the individually metered sample frame. The columns of the table include the following information for each of the stratum.

- Columns A and B indicate the strata;
- Column C shows the SFCODE prefix for each strata;

- Column D shows the total number of units for each strata;
- Column E shows the total number of meters or accounts for each strata;
- Column F shows the proportion of the population for each strata, where the total population comprises master-metered units in PG&E, SCE, SDG&E, and LADWP electric service territories;
- Column G shows the target number of phone (stage one) surveys;
- Column H shows the target number of mail survey completes using a proportional allocation method;
- Column I gives the expected response rates per strata; and
- Column J contains the target mailout, which was determined by the expected response rates along with the target number of completes.

Table 2-3
Master Metered Sample Design

A	B	C	D	E	F	G	H	I	J
						Target Completes			
Electric Utility	Type	SFCODE Prefix	Number of Units	Number of Meters	Proportion of Population	Phone Survey	Mail Survey	Expected Response Rate	Target Mailout
	2-4 units	11	44,411	20,128	10%	0	225	33%	675
	Multi-family 5-20 units	12	18,507	2,061	4%	100	100	25%	400
	Multi-family >20 units	13	39,171	700	9%	100	200	25%	800
	Mobile Home >4 units	14	101,305	1,638	23%	100	500	50%	1,000
PG&E	Subtotal		203,394	24,527	47%	300	1025	36%	2,875
	2-4 units	41	13,025	5,609	3%	0	65	33%	195
	Multi-family 5-20 units	42	14,139	1,533	3%	75	75	25%	300
	Multi-family >20 units	43	16,080	324	4%	40	75	25%	300
	Mobile Home >4 units	44	110,710	1,592	25%	110	550	50%	1,100
SCE	Subtotal		153,954	9,058	35%	225	765	40%	1,895
	2-4 units	31	8,630	3,883	2%	0	45	33%	135
	Multi-family 5-20 units	32	4,449	467	1%	25	25	25%	100
	Multi-family >20 units	33	6,821	131	2%	15	35	25%	140
	Mobile Home >4 units	34	41,500	645	10%	40	200	50%	400
SDG&E	Subtotal		61,400	5,126	14%	80	305	39%	775
	2-4 units	51	3,782	1,739	1%	0	20	33%	60
	Multi-family 5-20 units	52	1,269	134	0%	5	5	25%	20
	Multi-family >20 units	53	10,010	152	2%	25	50	25%	200
	Mobile Home >4 units	54	1,137	8	0%	1	5	50%	10
LADWP	Subtotal		16,198	2,033	4%	31	80	28%	290
	2-4 units		69,848	29,620	16%	0	355	33%	1,065
	Multi-family 5-20 units		38,364	4,061	9%	205	205	25%	820
	Multi-family >20 units		72,082	1,155	17%	180	360	25%	1,440
	Mobile Home >4 units		254,652	3,875	59%	251	1,255	50%	2,510
Total	Total		434,946	38,711	100%	636	2,175	37%	5,835

Initial Master Metered Mail Sample Allocation

We used a proportional allocation to assign the RASS sample to each of the stratum. The sample was assigned based on units or dwellings. Columns F, G and H in Table 2-3 show the population proportions and target number of stage one and two survey completes using a proportional allocation method. In Table 2-3, column J shows the target mailout.

Following the proportions assigned to the individually metered sample, we assigned sample on a proportional basis where the sample size was equal to one-half a percent of the population. Different from the individually metered sample, we did not use a modified proportional approach, and as such did not increase the sample for certain strata based on a pre-determined "minimum" amount of sample. The reason for not using the modified approach for the master-metered sample is that the variable "type" was included in the sample design only because the mailing strategy differs across the different types of units. This dimension of the sample design was not added to obtain a certain level of precision for estimates by the variable "type". Thus, a minimum number of sample points were not required for the strata and the proportional allocation method was most appropriate.

The total number of stage one surveys was set at 636 and the number of stage 2 surveys at 5,835. The total target number of completes was expected to be 2,175 based on an average response rate of 37%.

2.2.3 Non-Response Follow-Up Sample Design

KEMA-XENERGY worked with RoperASW to perform the non-response portion of the project. The objective was to obtain survey responses from a portion of the customers who did not respond to the mail survey to help reduce non-response bias by using multiple targeting approaches. A subsample of 5,000 customers from the original RASS sample who had not responded to the initial mailings was selected in 465 Zip Codes.

The more densely populated areas of the state were clustered by Zip Code. Clustering allowed more efficient data collection by in-person contact. Customers in these areas were contacted by First Class Mail with a small incentive and those that did not respond were contacted by telephone and/or in person.

The more sparsely populated areas of the state did not provide the opportunity for clustering that could result in efficient in-person contact. Customers in these areas were contacted by Priority Mail with a larger incentive. Those that did not respond were contacted by telephone only.

Clustered sample cases were designated as belonging to Group A. Non-clustered sample cases were designated as belonging to Group B.

Sample Selection

Step 1- Separate customers by sample group

The first step in the sample selection was to separate the households into clustered and non-clustered sample groups. The following 3-digit Zip Codes were allocated to the non-clustered sample group.

- 934, 935, 939, 949, 954, 955, 960, and 961

In addition, the two-stage selection procedure for the clustered sample resulted in inadequate sample sizes for certain strata. Customers in the following strata were also allocated to the non-clustered sample group.

- 100002, 100004, 101102, 110102
- 300001, 300003, 301513, 301613
- 400001, 400002, 400003, 400004, 400199.

Step 2 - Select the sample members from the non-clustered group

The second step in the sample selection was to select customers from within the non-clustered sample group. The non-clustered Zip Codes and strata (11.6% of the overall non-response households) contained records for 9,513 customers. The group was allocated 580 of the 5000 sample cases.

We selected a stratified random sample of households using the following procedure:

- Allocated sample fraction to strata: We computed a sample fraction as the total number of cases in a stratum to the total number of cases in the non-clustered sample.

Allocated sample cases to strata: We multiplied the sample fraction times the sample allocation (580) to get the number of sample cases allocated to each stratum. We used a statistical rounding procedure to allocate an integer number of cases to each stratum.

We selected a systematic sample from each stratum to control the final number of cases selected from the stratum. The KEMA-XENERGY sample ID was used as the sort key.

The sampling rate for cases selected from the non-clustered sample was 0.0610. The highest number of cases selected in a stratum in the non-clustered sample was 64. Two of the strata that had cases in the non-clustered area had no selections.

Step 3 - Select sample members from the clustered group

The third step in the sample selection was to select customers from within the clustered sample group.

The clustered Zip Codes (88.4% of the overall non-response households) contained records for 72,740 customers. The group was allocated 4420 of the 5000 sample cases. We allocated 20 sample cases to each Zip Code group. We used a two-stage procedure in which we first selected Zip Codes and then selected cases from within the Codes.

The initial sample had households in 1689 Zip Codes. The clustered sample had cases in 1150 Zip Codes. The non-clustered sample had cases in 539 Zip Codes.

Zip Code groups were selected using the following procedure:

Zip Codes were grouped numerically so that there was a minimum of 20 cases per group. The Zip Codes were collapsed into 830 Zip Code Groups.

A systematic sample of 221 Zip Code Groups was selected (the sort key was Zip Code number). The probability of selection of a Zip Code Group was the number of cases in the Zip Code Group divided by 329.01. Since the largest number of cases in a Zip Code Group was 259 cases, there were no certainty selections among the Zip Code Groups.

Within each selected Zip Code Group, we selected a systematic sample of households. The sampling rate within each Zip Code Group was the 20 divided by the number of households in the Zip Code Group. In 180 of the 221, all 20 selected cases were located within a single Zip Code.

The sampling rate for cases selected from the clustered sample was 0.0608.

Step 4 - Reassign cases to "rationalize" survey procedures

Group A (clustered) cases were those that received First Class Mail followed by telephone and in-person data collection procedures. Group B (non-clustered) cases were those that received Priority Mail followed by telephone data collection procedures.

Some cases from the sparsely populated regions fell into Zip Codes that were included in a clustered Zip Code Group. To rationalize procedures, these cases (n=70) were reassigned to Group A.

Some cases from the densely populated regions fell into Zip Codes that contained fewer than 5 selected cases. To rationalize procedures, those cases (n=95) were reassigned to group B.

2.2.4 On-Site Metering Design

A sub-sample of the initial sample was used for on-site metering. On-site meter installations were done on homes in the general population with an over-weighting of homes with air conditioning. Thus, the initial sub-sampling effort took into consideration the need to target air conditioning users. The target number of installed on-site meters was 200. The metering sample called for 50 homes without air conditioning and 150 with air conditioning.

In order to achieve the results in a cost effective manner, we developed a grid that split the state into six categories and parceled out the targets in such a way that we picked 20 large geographic areas and targeted 10 customers in each area. The strategy attempted to capture a ratio of AC to non-AC customers in each area in a way that mirrors the split in that climate zone with an excess of targets to air conditioning customers. We also tried to spread the surveys around the state and amongst the utility sponsors and Energy Commission climate zones. Table 2-4 shows the planned target onsite areas and air conditioning breakdown within the various climates.

The targeting by the climactic areas was done roughly by three digit zip codes with some refinement in large areas. Table 2-5 provides the target areas used for recruiting onsite participants.

Table 2-4
Target Areas for the Onsite Metering Sample

	Hot Climate (AC most likely)	**Moderate Climate (AC likely)**	**Cool Climate (AC least likely)**
No Air Conditioning	**12 sites** (max. of 1 non-AC site per area)	**24 sites** (max. of 4 non-AC sites per area)	**20 sites**
Air Conditioning	**108 sites**	**36 sites**	**0 sites**
Geographic Target Areas	**12 areas** (**120** total sites)	**6 areas** (**60** total sites)	**2 area** (**20** total sites)
Recruitment Strategy	Recruit with goal of meeting AC targets. Recruiters may get more AC customers than listed, but cannot exceed maximum number of non-AC households by target area.		Recruit whatever customers we get since there is a low probability we will get an AC customer.

Table 2-5
Target Groups for Onsite Metering

Target group	**Climate category (from Table 2-4)**	**Proposed three digit zip code of target area**	**CEC climate zones covered**	**Major town(s) in segment**	**Utilities with customers in target area**
1	Hot	917	9,10	Covina, Pomona, Ontario, Upland	SCE
2	Hot	919	(9)** 13	La Mesa, Spring Valley, Lemon Grove	SDG&E
3	Hot	920	13	El Cajon, Poway, Escondido	SDG&E
4	Hot	922	10	Cathedral City, Desert Hot Springs, Palm Springs, 29 Palms, Yucca Valley	SCE
5	Hot	924	10	San Bernardino	SCE
6	Hot	925	10	Riverside, Hemet, Moreno Valley, Sun City	SCE
7	Hot	933	3	Bakersfield	PG&E
8	Hot	937	3	Fresno	PG&E
9	Hot	952	1,2	Stockton, French Camp, Valley Springs	PG&E
10	Hot	956	2,4	Davis, Vacaville, Winters	PG&E

Table 2-5
Target Groups for Onsite Metering
(continued)

11	Hot	959	1,3	Chico, Marysville, Grass Valley	PG&E
12	Hot	960	1,3	Redding, Red Bluff	PG&E
13	Moderate	900	11	Los Angeles	LADWP, SCE
14	Moderate	910/911	9,12,16	S. Pasadena, Sunland, Tujunga, Altadena, Monrovia	LADWP, SCE
15	Moderate	919	(9)** 13	Chula Vista	SDG&E
16	Moderate	921	13	San Diego	SDG&E
17	Moderate	927	8	Santa Ana, Fountain Valley, Tustin	SCE
18	Moderate	945	4,5	Concord, Hayward, Livermore	PG&E
19	Cool*	941/946	5	Oakland and East Bay Hills, San Francisco	PG&E
20	Cool*	952/953	1	Foothill areas (sites over 2,500 feet elevation, more electric): Areas above Sonora, Angel's Camp	PG&E

Note: Because of the limited availability of recruits in the foothill areas and long travel distances between sites, we shifted the number of targets per area to 14 in the Oakland / SF area and 6 in the foothills. This yielded the 20 targeted cool area sites.
*SDG&E had some customers listed in zone 9 in the final sample. This was eventually changed to zone 13 where all SDG&E customers are located.

2.3 PROJECT IMPLEMENTATION

Throughout the time the sample frame was being developed and finalized, we created all of the program materials and planned out the overall project implementation. This section details the results of that planning and implementation effort.

2.3.1 Materials Design, Pretest, and Direct Mailings

All materials for the program were designed with input from all program sponsors. While this was a more complicated task than working with a single client, the group worked well together and was able to accommodate the needs of the group while maintaining a survey that was user-friendly and comprehensive.

The direct mail surveys were pretested with a sample of 20 energy customers. Results of the pretest were shared with the Energy Commission and participating utilities, and all parties agreed on modifications to be made. The results from the survey pretest are included as Appendix Y.

The direct mail package consisted of:

- An outgoing envelope (7.5 x 10.5 inches) with a window opening
- A business reply envelope (7 x 10 inches)
- A 20 page scannable survey (6.75 x 9.75 inches)
- A cover letter - several different types of letters were used:
    - Standard first mailing letter
    - First mailing letter for sites with 2-4 units
    - First mailing letter for master metered sites
    - Second mailing letter (same for all customers).

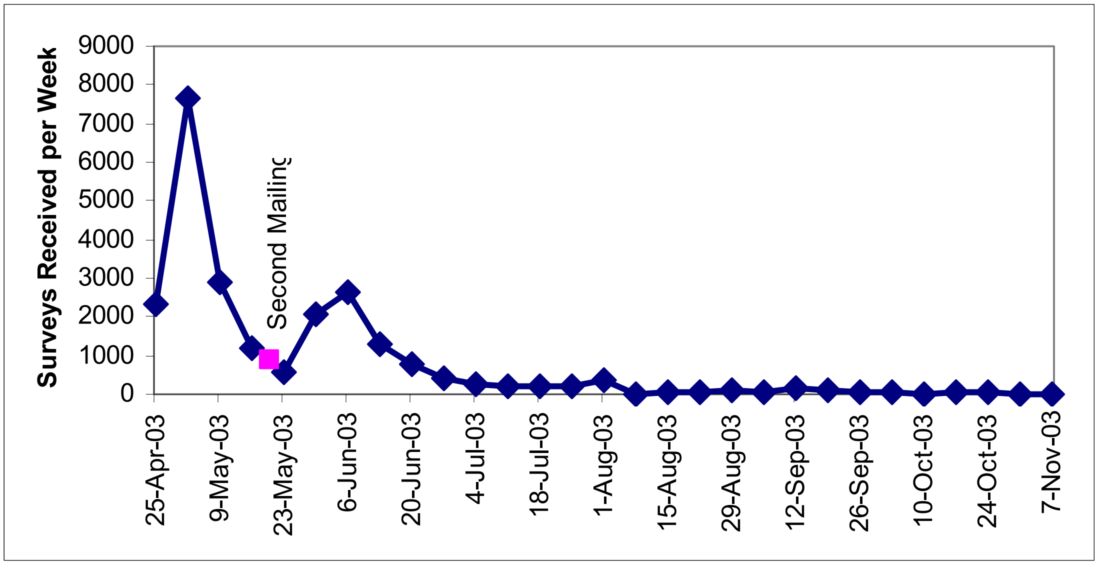
Copies of all direct mail materials are included as Appendix X.

A bar code, containing the tracking number (SFCODE), along with the respondent name, mailing address, and service address, was printed on the survey and was designed to show through the window of the main envelope. The service address was also provided on the survey to direct the respondent to fill out the survey for the dwelling that was targeted by the sampling plan. A specific cover letter depending on the type of respondent identified the sponsor(s), provided the motivation for completing the survey, and gave survey instructions. The survey instrument also included various instructions.

All packages were compiled at a mailing house where they were sorted to obtain the most favorable postage rate. Direct mail solicitation packages were mailed third class to all customers on April 16, 2003. As responses came into KEMA-XENERGY's office, we barcoded the surveys and created a list of completed surveys. Three weeks after the first mailing, we extracted the names of all participants who had replied to date and on May 20th sent a second solicitation package. This step saved on postage costs and reduced unwanted mail to those who had replied. The cover letter of the second mailing stressed the importance of the study and repeated the request for customer participation. In all other regards, the second package was identical to the first.

Figure 2-2 illustrates the influx of surveys as a result of the two mailings.

Figure 2-2
Weekly Count of Surveys Received



Appendix E includes the final Data Collection Protocols document which details the survey processing steps, training information, and phone scripts for master metered electric accounts and onsite metered site recruitment.

2.3.2 Toll Free Customer Support

A key component of the mail survey process is a survey support hot line. This hot line was set up for respondents to ask clarifying questions or to obtain assistance in determining correct responses. The hot line also provided a means for non-English-speaking respondents to complete the survey by phone, if they desire. Operators had a Spanish version of the survey translated and ready to use to assist Spanish-speaking callers with any survey questions.

The RASS survey hot line accepted calls with a live operator from 8:30 AM to 7 PM. At all other times, we had an answering machine available to accept messages and returned calls the following business day or as requested by the customer. Several operators were native Spanish speakers.

In total there were 302 calls received by the toll free line for the duration of the RASS project. Just over half of the calls (160) were English questions. There were 141 Spanish calls and one in Russian. We used a translation firm to assist us with the Russian call.

2.3.3 Individually Metered Survey Completes

We received a total of 18,970 responses to the mail survey. As mentioned above, we expected a total of approximately 47,000 responses assuming a 47% response rate. The actual response rate was 19%. There are several factors that may have caused the lower than expected response rate including:

- Direct mail solicitations (first and second mailings) did not include a monetary incentive

- Customers may not have identified with the Energy Commission logo on the envelope and thus did not feel compelled to open the package and read the letter and other materials;

- Customers may have been adversely affected by the 2001-2003 politically charged energy issues (blackouts, high prices, utility bankruptcy) and thus less willing to participate;

- Consumers are ever more targeted with direct mail and various surveys so may be less willing than in years past to spend their free time participating in this type of research; and

- The survey may have appeared to be too long for some customers.

While the response rate was lower than expected, the overall scope of the study was large enough to capture a large quantity of responses and allow for reasonably precise results. Section 2.5 provides more detail on the study precision.

Table 2-6 presents the number of completes and response rate for each of the individually metered sample strata. The columns of the table include the following information for each strata.

- Columns A through E indicate the strata variables;
- Column F indicates SFCODE prefix for the strata grouping;
- Column G indicates the number of target completes;
- Column H indicates the actual number of mail surveys returned;
- Column I indicates the expected response rate; and
- Column J indicates the actual response rate.

Table 2-6
Individually Metered Mail Survey Response

A	B	C	D	E	F	G	H	I	J
Electric Utility	Home Age	Electric Heat Presence	Home Type	CEC Forecast Climate Zone	SFCODE Prefix	Target Completes	Actual Completes	Expected Response Rate	Actual Response Rate
LADWP	Old	No	Low	11	500411	1,644	487	40%	12%
				12	500412	517	150	40%	11%
			Medium	11	500511	713	197	55%	15%
				12	500512	413	118	55%	16%
			High	11	500611	173	50	55%	16%
				12	500612	190	75	55%	21%
		Yes	All	All	500001	150	30	50%	10%
	New	No	Low	11	510411	150	37	40%	10%
		All others	All others	All others	500002	150	30	50%	10%
PGE	Old	No	SF-Low	1	100201	230	130	45%	25%
				2	100202	251	117	45%	21%
				3	100203	708	323	45%	20%
				4	100204	1,333	745	45%	25%
				5	100205	1,822	1,056	45%	26%
			SF-High	1	100101	222	124	55%	30%
				2	100102	515	227	55%	24%
				3	100103	1,403	592	55%	23%
				4	100104	1,647	705	55%	23%
				5	100105	1,915	836	55%	24%
			MF	1	100301	150	68	40%	18%
				2	100302	206	64	40%	12%
				3	100303	493	148	40%	12%
				4	100304	896	337	40%	15%
				5	100305	1,882	741	40%	16%
		Yes	SF-Low	1	101201	160	120	45%	33%
				3	101203	150	83	45%	25%
				4	101204	150	79	45%	23%
				5	101205	150	58	45%	17%
			SF-High	1	101101	394	188	55%	26%
				2	101102	150	80	55%	29%
				3	101103	327	173	55%	29%
				4	101104	215	107	55%	27%
				5	101105	150	69	55%	25%
			MF	3	101303	150	42	40%	11%
				4	101304	374	105	40%	11%
				5	101305	617	201	40%	13%
			All others	All others	100001	150	55	50%	18%

Table 2-6
Individually Metered Mail Survey Response
(continued)

A	B	C	D	E	F	G	H	I	J
Electric Utility	Home Age	Electric Heat Presence	Home Type	CEC Forecast Climate Zone	SFCODE Prefix	Target Completes	Actual Completes	Expected Response Rate	Actual Response Rate
PGE (cont.)	New	No	SF-Low	2	110202	150	76	45%	23%
				3	110203	150	50	45%	15%
				4	110204	150	78	45%	23%
				5	110205	154	67	45%	19%
			SF-High	2	110102	150	59	55%	21%
				3	110103	154	57	55%	20%
				4	110104	150	50	55%	18%
				5	110105	153	41	55%	15%
			MF	3	110303	150	41	40%	11%
				4	110304	150	38	40%	10%
				5	110305	150	49	40%	13%
			All others	All others	100002	150	53	50%	17%
		Yes	MF	4	111304	150	38	40%	10%
			All others	All others	100003	160	45	50%	14%
	All	All	All	14	100004	150	84	50%	28%
SCE	Old	No	SF-Low	7	400207	227	102	45%	20%
				8	400208	1,566	773	45%	22%
				9	400209	1,878	739	45%	18%
				10	400210	999	428	45%	19%
				11	400211	311	128	45%	18%
				All others	400299	150	64	45%	19%
			SF-High	7	400107	362	161	55%	24%
				8	400108	1,730	706	55%	22%
				9	400109	1,707	650	55%	21%
				10	400110	1,562	565	55%	20%
				11	400111	173	58	55%	18%
				All others	400199	158	64	55%	22%
			MF	7	400307	163	68	40%	17%
				8	400308	1,134	411	40%	14%
				9	400309	1,103	314	40%	11%
				10	400310	695	281	40%	16%
				11	400311	367	129	40%	14%
				All others	400399	150	48	40%	13%

Table 2-6
Individually Metered Mail Survey Response
(continued)

A	B	C	D	E	F	G	H	I	J
Electric Utility	Home Age	Electric Heat Presence	Home Type	CEC Forecast Climate Zone	SFCODE Prefix	Target Completes	Actual Completes	Expected Response Rate	Actual Response Rate
SCE (cont.)	Old (cont)	Yes	SF-Low	8	401208	150	67	45%	20%
				9	401209	150	62	45%	18%
				10	401210	150	58	45%	17%
			SF-High	8	401108	150	63	55%	23%
				9	401109	150	72	55%	26%
				10	401110	150	48	55%	17%
			MF	8	401308	519	179	40%	14%
				9	401309	277	79	40%	11%
				10	401310	152	44	40%	11%
				11	401311	150	65	40%	17%
			All others	All others	400004	150	75	50%	25%
	New	No	SF-Low	8	410208	150	58	45%	17%
				9	410209	150	37	45%	11%
				10	410210	155	79	45%	23%
			SF-High	8	410108	150	48	55%	17%
				9	410109	150	46	55%	17%
				10	410110	220	72	55%	18%
			MF	8	410308	150	31	40%	8%
		All others	All others	All others	400001	164	50	50%	15%
	All	All	All	15	400002	150	71	50%	23%
				16	400003	150	69	50%	23%
SDGE	Old	No	Low	9	300409	193	95	40%	19%
				13	300413	1,518	705	40%	18%
			Medium	9	300509	185	81	55%	24%
				13	300513	1,484	621	55%	23%
			High	9	300609	150	66	55%	24%
				13	300613	648	257	55%	22%
			All others	All others	300001	150	65	50%	21%
		Yes	Medium	13	301513	150	73	55%	27%
			High	13	301613	150	84	55%	31%
			All others	All others	300002	150	52	50%	17%
	New	No	Low	13	310413	201	72	40%	14%
			Medium	13	310513	175	55	55%	17%
			High	13	310613	150	57	55%	21%
		All others	All others	All others	300003	150	52	50%	17%
Total						46,807	18,970	46%	19%

2.3.4 Master Metered Mail Implementation

The master metered market is a segment that is often excluded from research studies because they are a difficult market to contact and survey. For this study, master metered customers were targeted using a two-phased approach. Customers were solicited for phase one using a phone survey script which can be found in the Data Collection Protocols (Appendix E).

Calls were tracked in utility specific databases so that one caller could focus on a given utility and unit type to facilitate data collection. Each phone surveyor entered customer names and addresses as provided by the facility manager into a central address spreadsheet for each utility/unit type grouping. Address files were then combined to create a central mailing database for each of the specific types.

Master metered homes with 2-4 units were not included in the stage one phone calls, but were sent customized letters requesting that they fill out the survey for only one of the units. Mobile homes were surveyed using the standard individually metered mailing with the same cover letter. Mobile homes were included in the stage one process primarily to obtain correct addresses for residents of a particular mobile home park.

Once customer names were obtained using the phase one screening, a mailing list was sent to the mailing house and solicitation packages were made up for each of the master metered customers. KEMA-XENERGY staff then transposed the survey responses gathered in the phase one calls onto the physical RASS surveys. This step assisted master metered customer with their survey responses and improved accuracy on questions where the property manager/landlord was able to assist with technical information. The technical information included the type of building, heating system, and other common equipment. This manual step insured that the master metered tenants provided appropriate answers to the questions for which they were unlikely to know the answers while allowing the tenants to provide details on the other items in the survey which they did control. Once the surveys were hand prepared, the mailing house sent them out to the customers with a special master meter letter explaining the process.

2.3.5 Master Metered Survey Completes

We conducted a total of 616 stage one phone surveys, and mailed out 5,593 mail surveys to master metered customers. Table 2-7 below provides the number of phone surveys conducted and mail surveys sent out by strata. The columns of the table include the following information for each of the stratum.

- Columns A and B indicate the strata;
- Column C shows the SFCODE prefix for each strata;
- Column D shows the target number of phone survey completes for each strata;
- Column E shows the actual number of phone survey conducted for each strata;
- Column F shows the target mailout; and
- Column G shows the actual mailout.

There are several reasons why the actual completes differed from the targets. In some cases, we were unable to reach any additional customers after repeated calls. In other cases we ran out of valid phone numbers and had minimal success with alternative methods for looking up phone numbers. Phone numbers were initially targeted using a phone number matching service with utility supplied phone numbers added in as an additional contact number.

We had varied results reaching customers depending on their location and type of dwelling. On average, it took 8.1 phone calls to reach each phase one customer. Mobile homes had the lowest call rate with 5.3 calls per completed phone survey. Master meters with 5-20 units took 9.4 calls per complete and multi-family dwellings with over 20 units were the hardest to reach with 10.7 calls per complete. Overall, we were able to complete 97% of the targeted phone surveys.

Table 2-7
Master Metered Phone Survey Response and Actual Mailout

A	B	C	D	E	F	G
Electric Utility	Type		Phone Survey		Mailout	
		SFCODE Prefix	Target Completes	Actual Completes	Target	Actual
PG&E	2-4 units	11	na	na	675	672
	Multi-family 5-20 units	12	100	101	400	407
	Multi-family >20 units	13	100	91	800	690
	Mobile Home >4 units	14	100	100	1,000	954
	Subtotal		300	292	2,875	2,723
SCE	2-4 units	41	na	na	195	194
	Multi-family 5-20 units	42	75	75	300	300
	Multi-family >20 units	43	40	40	300	320
	Mobile Home >4 units	44	110	110	1,100	1100
	Subtotal		225	225	1,895	1,914
SDG&E	2-4 units	31	na	na	135	135
	Multi-family 5-20 units	32	25	26	100	104
	Multi-family >20 units	33	15	7	140	64
	Mobile Home >4 units	34	40	40	400	398
	Subtotal		80	73	775	701
LADWP	2-4 units	51	na	na	60	57
	Multi-family 5-20 units	52	5	3	20	12
	Multi-family >20 units	53	25	22	200	176
	Mobile Home >4 units	54	1	1	10	10
	Subtotal		31	26	290	255
Total	2-4 units		na	na	1,065	1,058
	Multi-family 5-20 units		205	205	820	823
	Multi-family >20 units		180	160	1,440	1,250
	Mobile Home >4 units		251	251	2,510	2,462
	Total		636	616	5,835	5,593

We received a total of 767 responses to the master metered mail survey. As mentioned above, we expected a total of approximately 2,175 responses assuming a 37% response rate. The lower response rate is similar to that found in the individually metered mailouts and is attributed to similar factors as detailed in Section 2.3.3 above. While response in the two to four unit sites and mobile home parks was reasonable, the response in the multi-family sites with five or more units was extremely low. We expect this is a result of the study-wide response rate issues, plus was impacted by the challenge of identifying unit addresses and the fact that many surveys were sent generically addressed to the California Energy Customer and were not personalized.

Table 2-8 presents the mail survey response for master metered customers. The columns of the table include the following information for each of the stratum.

- Columns A and B indicate the strata;
- Column C shows the SFCODE prefix for each strata;
- Column D shows the target number of mail survey completes for each strata;
- Column E shows the actual number of mail survey completes for each strata;
- Column F shows the expected response rate; and
- Column G shows the actual response rate.

Table 2-8
Master Metered Mail Survey Response

A	B	C	D	E	F	G
Electric Utility	Type	SFCODE Prefix	Target Completes	Actual Completes	Expected Response Rate	Actual Response Rate
PG&E	2-4 units	11	225	139	33%	21%
	Multi-family 5-20 units	12	100	10	25%	2%
	Multi-family >20 units	13	200	16	25%	2%
	Mobile Home >4 units	14	500	217	50%	23%
	Subtotal		1025	382	36%	14%
SCE	2-4 units	41	65	33	33%	17%
	Multi-family 5-20 units	42	75	7	25%	2%
	Multi-family >20 units	43	75	10	25%	3%
	Mobile Home >4 units	44	550	211	50%	19%
	Subtotal		765	261	40%	14%
SDG&E	2-4 units	31	45	24	33%	18%
	Multi-family 5-20 units	32	25	3	25%	3%
	Multi-family >20 units	33	35	0	25%	0%
	Mobile Home >4 units	34	200	93	50%	23%
	Subtotal		305	120	39%	17%
LADWP	2-4 units	51	20	2	33%	4%
	Multi-family 5-20 units	52	5	0	25%	0%
	Multi-family >20 units	53	50	2	25%	1%
	Mobile Home >4 units	54	5	0	50%	0%
	Subtotal		80	4	28%	2%
Total	2-4 units		355	198	33%	19%
	Multi-family 5-20 units		205	20	25%	2%
	Multi-family >20 units		360	28	25%	2%
	Mobile Home >4 units		1,255	521	50%	21%
	Total		2,175	767	37%	14%

2.3.6 Non-Response Follow-up Implementation

Customer information for the non-response follow-up sample was provided by the four sponsoring electric utilities: PG&E, SDG&E, SCE, and LADWP. Information included the identifier SFCODE, customer name, service address, mailing address, and phone number. The initial utility-provided file only included a few phone numbers. Additional phone numbers were provided mid-way through the data collection period by three of the energy providers. No additional numbers were provided by the fourth (LADWP).

In order to maximize telephone surveying effectiveness, RoperASW also sought telephone numbers from an electronic cross-directory service for the sampled customers on the list. Roughly 30% of those requested were returned with a telephone number.[3] In total 17% of the customers had at least one number initially and close to 90% had at least one number after cross-directory and utility company lists were added. However, many of the contact numbers were outdated, disconnected, or otherwise unproductive.

In order to track survey progress, a sample management database was created from the sample data. The database held all identifying information for the 5,000-member sample. The database was used to control all phases of the survey. The database was used to prepare mailing lists, interview lists and assignment materials. It was also used as a log to track interviewer assignments and final status codes as well as to generate various status reports.

Non-Response Follow-Up Materials

The questionnaire used in the non-response follow-up was the same form used for all of the initial customer mailings. KEMA-XENERGY provided copies of the questionnaire for all mailings and interviewer-gathered responses.

Group A (clustered) mailings included: (Samples of all materials appear in the Appendix.)

- Outside envelope, of the same shape as those used by KEMA-XENERGY (and provided by KEMA-XENERGY), but printed by RoperASW to look different from the initial mailout in an effort to increase recipient interest. Colored border triangles and the word First Class Mail were added to the envelope. RoperASW modified the return address by adding a tag line requesting customer participation as well as adding the Energy Commission logo.

- Cover letter, revised from those used by KEMA-XENERGY in earlier mailings, and printed on RoperASW letterhead.

- Questionnaire, provided by KEMA-XENERGY and with a RoperASW-provided label showing the customer name and the address of the household.

- A postage paid, business reply envelope, provided by KEMA-XENERGY and addressed to KEMA-XENERGY's questionnaire processing center.

- A $1 bill as a thank you incentive.

Each packet also included a white 3.5" X 5.5" card with a message in Spanish providing a number to call at KEMA-XENERGY to complete an interview if the person could not complete the questionnaire in English.

Group B (non-clustered) mailings included: (Samples of all materials appear in the Appendix.)

- U.S. Postal Service Priority Pack outside envelope.

- Cover letter, revised from those used by KEMA-XENERGY in earlier mailings and slightly revised from the Group A letter with respect to the incentive.

- Questionnaire.

- The postage paid, business reply envelope addressed to KEMA-XENERGY.

- A $5 bill as a thank you incentive.

- A white 3.5" X 5.5" card with a message in Spanish inviting the customers to call the toll-free survey center to complete their survey in Spanish.

- A blue 3.5" X 5.5" card promising an additional incentive for a completed questionnaire if the blue card was returned to RoperASW with the name and mailing address of the person who completed the questionnaire. These cards, when confirmed against a list of completed questionnaires reported by KEMA-XENERGY, were used to authorize payment of an additional $15 incentive.

A separate advance letter was sent to all customers that did not respond to the mailed requests. The advance letter, on RoperASW letterhead, notified the recipient that a RoperASW interviewer would be attempting contact by phone or in person within the next few weeks. The advance letter carried the same study information as that included in the questionnaire packets. Advance letters were mailed roughly a week before interviewers began working their assignments.

Non- Response Follow-Up Data Collection

Initial mailings were sent out over several days, beginning July 18. All 5,000 sample members were included, with the Group A members receiving the First Class packet with the $1 incentive and Group B members receiving the Priority Mail packet with the $5 (+$15 promised) incentive. All returned completed questionnaires were mailed directly from the customer to KEMA-XENERGY. Every few days, KEMA-XENERGY provided an update file listing the identification numbers of returned questionnaires. Those so identified were marked as complete on the sample management database and excluded from all subsequent data collection efforts.

Those customers that did not respond to the mailing within three weeks were designated for interviewer (telephone or in-person) follow-up. Advance letters indicating that the customer would be contacted by phone or in person were sent out to all such customers on August 13. A total of 4,596 advance letters were mailed using First Class postage.

In the two months prior to the advance letter mailing, telephone and in-person interviewers were recruited for the study. Telephone interviewers were recruited from among those used by RoperASW in the past. These interviewers work from their homes. Physical location in California was not required for this staff. A total of 16 telephone interviewers were hired for the study.

Many of the in-person interviewers were also recruited from among those used by RoperASW in the past. The number needed to staff this study required that additional recruiting be carried out. Word of mouth among interviewers, recommendations by supervisors from other data collection companies, and advertisements were used. A total of 34 in-person interviewers were eventually hired and trained to work the study. An additional 5 interviewers were provided by KEMA-XENERGY in the last few weeks of data collection.

Both telephone and in-person interviewers were trained by conference call in the week prior to sending out assignments to interviewers on August 14 and 15. Each trainee received a set of training materials describing their job in advance of the training call. RoperASW field managers also sent each interviewer a packet with all pertinent program materials. Telephone staff was trained separately from in-person staff because of the additional procedures that were required of the in-person staff in arranging their work and in making trips to the Zip Code areas.

Interviewers (both telephone and in-person) were provided with a customer list for each Zip Code assigned. The list provided customer name, address of the housing unit served, and, for some customers, a telephone number. A labeled questionnaire was provided for each customer listed. Interviewers typically had four or more Zip Code lists to work.

In-person interviewers received a letter of introduction for themselves to be presented to the customer at the time of in-person contact. In-person packets also included an identification badge, generic cover letters similar to the letters used in the initial mailings, and magnet thermometers to be handed out as a gift to those customers that agreed to participate. Finally, in-person interviewers received a supply of plastic door-hanger bags, business reply envelopes, and "Sorry I missed you" letters. The door-hangers were used on the final (third) trip to the Zip Code if a successful interview could not be conducted. The door hanger bag was left with a full survey package including the business reply envelope and cover letter.

The day after the advance letters were mailed, the assignments began being shipped to interviewers. Work was assigned to phone interviewers for all of Area B (which was to have no in-person follow-up), and roughly 55 Zip Codes from Area A. The Area A Zip Codes were identified because no in-person staff was yet on board to cover those codes. In all, 1,470 cases were assigned to telephone interviewers.

All remaining work from Area A was assigned directly to in-person interviewers. In-person assignments were made based on the interviewer's proximity to the Zip Code(s). Roughly 3,130 were assigned to in-person interviewers. In-person interviewers could not be identified in proximity to a handful of Zip Codes. Special procedures, discussed below, were followed for these few.

Telephone interviewers were directed to make an unlimited number of calls to the phone number of record on the assignment list. For those without phone numbers, directory assistance was consulted. The result of each call was to be recorded on the assignment sheet. Work progressed on each Zip Code until each customer finished an interview, refused, or was discovered to have no usable phone number. A limit of six weeks was applied to telephone interviewer work. As Zip Codes were completed or the time limit was reached, materials were returned to RoperASW. All cases with non-final dispositions from Area A were reassigned to in-person interviewers (with a few exceptions for Zip Codes in which no in-person interviewer was available).

In-person interviewers were directed to make several attempts to complete each interview by telephone before making a trip to the Zip Code in person. Interviewers were directed to make up to three visits to each customer in a Zip Code in an attempt to complete an interview. If no interview could be completed by the third trip, a questionnaire hanger bag was left on the door. Work proceeded for in-person interviewers until the work was completed or the field period expired (mid November, roughly 12 weeks after it began).

Interviewers were paid for each complete interview returned to RoperASW. In-person interviewers were also paid a small amount for each hanger bag that was left on the third trip.

As the final six weeks of data collection began, a few Area A Zip Codes (roughly two dozen) had not been assigned to in-person interviewers. About half of these were identified as being very up-scale and comprised of gated communities where the probability of in-person success was felt to be limited. Due to its earlier success as a data collection mode, customers in these zips were contacted by Priority Mail packet with a $5 (+ $15) incentive. The other half of the unassigned Zip Codes were in areas where RoperASW had been unable to recruit in-person interviewers. Those were turned over to interviewers located at or managed by KEMA-XENERGY. The KEMA-XENERGY interviewers were trained by RoperASW supervisors.

Work from all interviewers was reviewed upon receipt at RoperASW. The first three interviews were checked for correct administration. Interviewers were retrained as needed. Checking was continued for those requiring additional training until it was clear that the questionnaires were being administered correctly. All questionnaires were reviewed for completeness before they were checked in on the sample management database.

Interviewers were supervised by telephone. A supervisor contacted each interviewer several times each week. The telephone interviewer supervisor worked in-house at RoperASW. The in-person interviewers managed by RoperASW were supervised by two off-site supervisors.

The number of completed surveys in each Zip Code was regularly reviewed to keep the interviewer from providing too many or too few interviews within each sample point. Each assignment sheet showed the number of interviews desired from that Zip Code. When that goal was achieved, the remaining work on that Zip Code was halted and the interviewer was directed to continue work on the next Zip Code in the assignment. Data collection by interviewers concluded on November 15, 12 weeks after it had begun.

Although KEMA-XENERGY had planned on completion of data collection by the end of October, the effort was extended in an effort to bolster the response rate. In the final weeks, several additional steps were taken. The additional efforts in November resulted in an increase in the response rate of 7 percentage points. These efforts included:

1) RoperASW reassigned unworked Zip Codes to interviewers who traveled outside their designated area to collect interviews, often staying overnight to accomplish their assignment.

2) KEMA-XENERGY arranged for or provided interviewers to supplement several Zip Codes that RoperASW was unable to staff.

3) RoperASW offered a bonus to all working interviewers to finish interviews with 50% of the customers or complete the third trip to the Zip Code and leave hanger bags if an interview was not completed.

4) Zip Codes that could not be staffed by either RoperASW or KEMA-XENERGY were mailed Priority Mail Packets with a $5 incentive. (No additional $15 was offered to this final group.)

Non-Response Follow-Up Response Rate

Table 2-9 shows the final overall response to the non-response follow-up effort. Since interviews that were completed by mail were sent to KEMA-XENERGY directly while phone and in-person interviews were returned to RoperASW, both companies logged in completed surveys.

Table 2-9
Final Status for All Customers

	Frequency	Percentage of Total	Percentage of Eligible
Complete reported by KEMA-XENERGY	746	14.9%	15.6%
Complete logged at Roper	1514	30.3%	31.6%
Refusal	228	4.6%	4.8%
Gated community, access denied	114	2.3%	2.4%
Business, ineligible	24	0.5%	
Vacant	108	2.2%	
Insufficient address	73	1.5%	
No Final Status	2193	43.9%	45.7%
Total	5000	100.0%	100.0%

Of the 5,000 sampled customers, 45% (2,260) completed interviews. If ineligible households (i.e. those that were actually businesses, were vacant, or had an address insufficient to locate the household) were removed from the base, the overall rate rose to 47%.

Of those reported as complete, a third were reported as received directly at KEMA-XENERGY. Almost all of these (roughly 660) resulted from the initial mailed requests. That is, roughly 13% of all customers responded to the initial RoperASW mailing. The return for the $1 incentive First Class packet was not as high as the return for the $5 (+$15) Priority Mail packet. Return for the First Class packet was 10.6%; return for the Priority Mail packet was 32.4%.

While the incentive was responsible for much of the difference between the two mailing options, part was also due to the unique physical characteristics of the Priority Mail envelope. The mailing envelope used for prior KEMA-XENERGY mailings was very plain with a simple Energy Commission return address on its face. RoperASW modified the First Class envelope in an attempt to make it more interesting and thus more likely to be opened. While the First Class envelope resulted in a higher return than was expected (10.6% v 7.5%), the Priority Mail envelope indicated much more clearly an important document that should be opened. The return to the Priority Mail pack was also higher than anticipated (32.4% v 25%).

The response rate resulting from telephone interviewer efforts alone was 12% (175/1469). This rate was lower than would normally be expected by telephone if all sample had telephone numbers. The original files provided by KEMA-XENERGY had phone numbers for only 17% of the sample, and the later supplemental list of phone numbers was not received until most of the telephone interviewer work had been returned to RoperASW at the end of the first six weeks of field work. Efforts to obtain phone numbers from credit search services resulted in matches for only 30% of the lookups and directory assistance provided fewer numbers than had been anticipated. As a result, the telephone interviewers were handicapped by the lack of readily available customer telephone numbers.

The response rate resulting from in-person interviewers was roughly 34%.[4] The count of completed interviews excludes 324 completed questionnaires received directly by KEMA-XENERGY after assignments were made to interviewers. Those 324 included late responses to the initial mailing, customer responses by mail after being contacted by an interviewer, responses to special mailings to selected Zip Codes, returns from hanger-bag questionnaires, and returns by interviewers managed by KEMA-XENERGY. The response received from in-person efforts does not allow a direct comparison of in-person rates with telephone rates because in-person interviewers were encouraged to complete interviews by phone if possible.

Response rates for the two sampling groups (based on clustering of Zip Codes) were essentially the same. Group A returned 45% overall. Group B returned 47% overall. Response rates for all eligible customers grouped by the four utility providers ranged from 40% to 50% (Table 2-10). San Diego, serviced by SDG&E, was a difficult area to staff initially; the Zip Codes selected there contained many gated communities; and, the fires that raged in the area late in the study interfered with interviewer efforts.

Table 2-10
Non-Response Follow-up Response by Utility

Provider	% Response Eligible Customers	% Response All Customers	Number of Zip Codes	Number of Customers
PG&E	46	44	239	2001
SDG&E	40	36	49	543
SCE	50	49	148	1976
LADWP	49	45	29	480
TOTAL	47	45	465	5000

Response rates by Zip Code ranged from 0% to 100%. Rates that were extremely high or low tended to come from small (Group B) Zip Codes, where, when the sample was only 1 case, the only response rates possible were 0% or 100%. Table 2-11 shows the distribution of Zip Codes by response rate for eligible customers. As can be seen, the two extremes of response were comprised predominately of Zip Codes that contained few sample units.

Table 2-11
Distribution of Zip Codes by Response Rate

% Response from Eligible Customers	Number of Zip Codes	% of Zip Codes	Average Customers Sampled per Zip
0-10%	79	17	2.6
11-20	12	3	17.1
21-30	24	5	14.0
31-40	41	9	10.8
41-50	130	28	14.7
51-60	67	14	18.4
61-70	31	7	13.8
71-80	12	3	10.1
81-90	3	1	10.7
91-100	65	14	1.3
TOTAL	464	100	10.8

Note: One Zip Code is excluded in this count. It had a sample size of 1 and the unit was ineligible.

The field effort was managed in an attempt to ensure that Zip Codes ended near the 50% target response. The effort was moderately successful. Roughly 42% of Zip Codes finished within 10 points of the 50% response target. Another 34% finished at 40% or below, and 24% finished above 60%. The range of response rates was affected by those Zip Codes that had small numbers of sample members; a small sample size was more likely to yield extreme results. If all Zip Codes with 3 or fewer cases were removed, 62% of the Zip Codes had a final response rate within 10 points of 50%. A list of Zip Codes showing group designation and response rate is in Appendix G.

2.3.7 Onsite Implementation

On-Site Metering. The on-site metering data collection provided valuable detailed data for use in understanding hourly demand issues and additional collection of data that cannot be obtained though a mail or other interview based survey process. The primary reason for this data collection activity is to gather hourly load shape data for a sample of homes and report on whole house and central air conditioning systems. The onsite data collection and metering activities collected the following information on a targeted sample of 200 homes, 150 of which have central cooling systems while the remaining 50 do not:

- Responses for all mail survey questions;
- One year of hourly load data for the total dwelling unit and for the central cooling system;
- Detailed housing shell characteristics such as insulation levels and window areas by type; and
- Nameplate data on major appliances (i.e. heating, cooling, water heater).

The responses to the mail RASS survey are required to facilitate the estimation of load shapes by region and market segment through the leveraging of the RASS data and various analytical methods.

The monitoring of the cooling and total home hourly loads will be performed using four-channel true-RMS current loggers. One channel captured the current draw on the main service, with a second channel measuring the branch circuit to the air conditioning unit. If the air conditioning unit is fed from more than one circuit, as may be the case with a split system air conditioning, two additional channels were available to measure those circuits as well. Due to the small size of the metering equipment, it was typically connected and secured in the main circuit breaker panel and was completely non-intrusive to the resident. The meters operate on 10-year battery power with no power connection required. Trained field technicians performed all meter installation and recorded meter readings throughout the study period.

Data was retrieved three times per year, or every four months, by field technicians. Although the meters could store well over a year's worth of data, preliminary data retrieval was performed to ensure the quality and operation of the installation. At the time of the first meter read, all meters were changed to read every 15 minutes instead of the initial hourly reads. This change improved data quality while maintaining the proposed quarterly read schedule. Data from the meters is easily exported to comma-separated-values format. Automated data cleaning and analysis procedures were developed to prepare and process the data.

The technical information collected on-site reflects the types of data that can only be reliably collected by a trained on-site surveyor. Lighting, shell, and nameplate data for this sample provides a complete set of very rich data that is useful in understanding the factors that drive hourly demand.

The process for on-site metering data collection consisted of the following activities:

- Telephone recruitment;
- Metering installation visit and survey data collection; and
- Up to three follow-up visits to obtain data from loggers (loggers are removed on the final visit).

A $50 incentive was offered to each home in the on-site metering sample. Half of the incentive was paid at the time of the meter installation. The second half will be paid after the logger has been removed. The loggers are being kept in place through the 2004 air conditioning season to unsure a full air conditioning cycle for load shape development.

All field personnel were trained in KEMA-XENERGY's Oakland office. Training consisted of program overview training as well as detailed metering installation instructions (including a live demonstration at an employees home), detailed review of the onsite survey, and electrical safety training. Field personnel, with the exception of two CEUS auditors who served as field trainers, were accompanied on their first few site visits to complete the training process and insure that they were adequately trained to perform meter installations. Details from the program training materials are outlined in the Data Collection Protocols, which is Appendix E. These protocols also include copies of the onsite survey instrument and training guide for completing the survey.

2.4 SURVEY WEIGHTS

This section discusses the process of assigning sample weights for both the individually metered and master metered samples. To minimize potential bias in the saturation and CDA results, the various components of the project were merged together and results were weighted to the initial population frame. This process adjusts the results so that they are representative of the population at large.

The individually metered sample contains both the initial mail respondents and the follow-up non-respondent sample. Thus, separate weights were created for each sampling cell to account for the different sampling approaches. Basic weights were developed for the master metered sample equal to the ratio of the population count to the completed sample count for the cell.

Individually Metered Sample Weights

We considered two different approaches to combining the follow-up survey results with the initial mail survey sample. The first approach that was considered would have treated the main survey respondents as representing only those customers who would respond to this survey if they had received it, and the follow-up respondents as representing all other customers. Thus, since the main survey response rate was 19 percent, the main survey would represent 19 percent of the population and the follow-up respondents 81 percent. Estimates from the two surveys would then be combined by taking the weighted average of the two, with these proportions as weights.

The effect of this weighting approach would be to increase the variance, or the widths of the confidence intervals, for the combined sample. If the confidence interval widths for the main and follow-up samples are w_m and w_f, respectively, and the main survey response rate is r, the confidence interval width for the combined sample is:

$$w_c = \sqrt{r^2 w_m^2 + (1-r)^2 w_f^2},$$

where:

w_c = width of confidence interval of combined sample,

r = survey response rate

w_m = width of confidence interval of main sample; and

w_f = width of confidence interval of of follow - up sample.

With a sample size approximately 1/10th that of the main sample, the follow-up sample would have confidence interval widths w_f almost three times as large as the main sample. Thus, the confidence intervals for the combined sample would be almost one and a half as big as those for the main sample alone. Despite this result, the difference in confidence interval widths does not imply that including the follow-up sample degrades our representation of the underlying population. The main sample alone, while more precise (having less variability) is a more precise estimate for a poorly defined population. By surveying the non-response group using different data collection approaches, we were able to target elements of the population who would not have replied to the mail survey. This therefore decreases (improves) the non-response bias that occurs in a single surveying method survey and produces a more precise combined estimate.

The second approach to weighting the combined sample was the one that was ultimately used. It resulted in a more modest effect on the precision of the combined sample results while still gaining the benefits of a multi-pronged surveying approach and its resultant reduction of non-response bias. Essentially, the follow-up sample is weighted less heavily. The justification for this approach is to assume that the follow-up sample represents only those customers who would respond to the follow-up survey but not to the main survey, rather than assuming the follow-up respondents represent all non-respondents to the main survey. In effect, the combined sample is treated as representing only those customers who would respond to one or the other stage of the survey. Thus, the combined sample using this weighting approach strictly represents 54 percent of the population, since the main survey response rate was 19 percent and the follow-up response rate was 44 percent. Using this approach, the main sample reflects approximately 35 percent of the covered population and the follow-up sample 65 percent.

The equations for the initial mail sample stratum weights (w_1) and the follow-up sample stratum weights (w_2) are presented below.

$$w_1 = \frac{\frac{N}{n_1} \times \frac{n_1}{n_s}}{\frac{n_1}{n_s} + \frac{n_2}{n_f} \times \left(1 - \frac{n_1}{n_s}\right)} \quad \text{and} \quad w_2 = \frac{\frac{N}{n_2} \times \frac{n_2}{n_f} \times \left(1 - \frac{n_1}{n_s}\right)}{\frac{n_1}{n_s} + \frac{n_2}{n_f} \times \left(1 - \frac{n_1}{n_s}\right)}$$

where:

N = population
n_1 = response to initial mail survey
n_2 = response to follow-up survey
n_s = initial mail sample (number of initial surveys mailed)
n_f = follow-up sample

for each stratum.

Table 2-12 provides the sample weights by strata for the individually metered sample. Individually metered weights range from a low of 47 to a high of 375 for the direct mail responses and 656 to 7,292 for the non-response follow-up surveys. The overall individual sample represents just over 9.9 million customers throughout the state.

Master Metered Sample Weights

The process of creating weights for the master metered sample consisted of implementing standard sampling procedures. The basic weights were developed for each sampling cell as the ratio of the population count to the completed sample count for the cell. The population counts used to calculate the weights for the master metered sample were based on the sample frame counts developed from the initial utility billing system extracts. The completed sample counts (number of respondents) were derived directly from the RASS sample.

Table 2-13 provides the sample weights by strata for the individually metered sample. There were a few strata where we were unable to get responses. Those were combined with other strata from their respective utilities to allow us to create weights for all customers. The strata without responses were SDG&E multi-family with over 20 units, LADWP multi-family with 5-20 units, and LADWP mobile homes. Table 2-13 includes the combined dwelling types.

Table 2-12
Individually Metered Weights

A	B	C	D	E	F	G	H	I	J	K
Electric Utility	Home Age	Electric Heat Presence	Home Type	CEC Forecast Climate Zone	SFCODE Prefix	Population	Initial Mail Completes (Sample 1)	Follow-Up Completes (Sample 2)	Weight 1	Weight 2
LADWP	Old	No	Low	11	500411	373,175	487	98	179.7	2914.7
				12	500412	117,405	150	27	190.6	3289.2
			Medium	11	500511	161,809	197	31	222.5	3805.9
				12	500512	93,677	118	15	208.6	4604.0
			High	11	500611	39,246	50	6	249.7	4460.5
				12	500612	43,236	75	11	132.2	3029.3
		Yes	All	All	500001	9,703	30	5	71.9	1509.4
	New	No	Low	11	510411	22,673	37	8	130.5	2230.8
		All others	All others	All others	500002	18,077	30	7	121.0	2064.0
PGE	Old	No	SF-Low	1	100201	52,179	130	14	164.9	2195.5
				2	100202	56,948	117	12	137.4	3405.6
				3	100203	160,601	323	27	194.6	3620.2
				4	100204	302,496	745	67	164.1	2690.2
				5	100205	413,401	1,056	77	163.2	3131.1
			SF-High	1	100101	50,357	124	6	230.6	3626.2
				2	100102	116,818	227	16	227.2	4078.0
				3	100103	318,444	592	44	236.8	4051.7
				4	100104	373,897	705	87	175.0	2879.5
				5	100105	434,596	836	82	212.0	3138.8
			MF	1	100301	13,378	68	11	61.7	834.6
				2	100302	46,867	64	12	160.2	3051.0
				3	100303	111,838	148	25	211.6	3220.9
				4	100304	203,268	337	48	165.9	3069.8
				5	100305	427,028	741	101	182.4	2890.2
		Yes	SF-Low	1	101201	36,361	120	4	206.1	2908.9
				3	101203	17,776	83	7	80.9	1580.4
				4	101204	19,860	79	6	109.0	1874.8
				5	101205	16,585	58	7	99.6	1544.0
			SF-High	1	101101	89,323	188	24	186.2	2263.5
				2	101102	29,246	80	6	164.8	2677.5
				3	101103	74,287	173	10	224.1	3552.1
				4	101104	48,810	107	6	233.9	3963.2
				5	101105	26,133	69	3	237.1	3256.7
			MF	3	101303	15,008	42	8	103.0	1335.2
				4	101304	84,892	105	14	246.3	4216.7
				5	101305	140,040	201	37	190.5	2750.0
			All others	All others	100001	21,360	55	6	119.3	2466.2
	New	No	SF-Low	2	110202	11,769	76	6	67.7	1104.1
				3	110203	18,887	50	7	125.5	1801.5
				4	110204	28,442	78	6	150.5	2784.0
				5	110205	30,591	67	10	157.8	2001.6
			SF-High	2	110102	23,550	59	6	155.4	2397.2
				3	110103	34,984	57	9	163.9	2849.2
				4	110104	29,547	50	7	181.8	2922.2
				5	110105	34,663	41	7	227.3	3620.8
			MF	3	110303	17,439	41	10	79.7	1417.3
				4	110304	17,043	38	6	121.5	2071.2
				5	110305	24,013	49	5	182.6	3013.0
			All others	All others	100002	19,960	53	5	146.4	2440.1
		Yes	MF	4	111304	11,391	38	3	158.0	1795.8
			All others	All others	100003	31,706	45	5	219.7	4363.5
	All	All	All	14	100004	11,912	84	7	61.6	963.0

Table 2-12
Individually Metered Weights
(continued)

A	B	C	D	E	F	F	G	H	I	J
Electric Utility	Home Age	Electric Heat Presence	Home Type	CEC Forecast Climate Zone	SFCODE Prefix	Population	Initial Mail Completes (Sample 1)	Follow-Up Completes (Sample 2)	Weight 1	Weight 2
SCE	Old	No	SF-Low	7	400207	51,498	102	7	233.3	3956.9
				8	400208	355,440	773	85	158.1	2743.8
				9	400209	426,252	739	97	170.2	3097.6
				10	400210	226,709	428	60	147.4	2727.4
				11	400211	70,693	128	10	193.2	4596.4
				All others	400299	17,854	64	19	84.1	656.3
			SF-High	7	400107	82,319	161	3	375.4	7291.9
				8	400108	392,688	706	62	216.2	3871.6
				9	400109	387,282	650	82	204.7	3100.4
				10	400110	354,469	565	79	177.5	3217.9
				11	400111	39,275	58	4	296.9	5513.8
				All others	400199	35,824	64	5	237.4	4126.5
			MF	7	400307	36,910	68	5	258.9	3861.0
				8	400308	257,250	411	71	177.8	2594.0
				9	400309	250,289	314	85	160.0	2353.5
				10	400310	157,814	281	35	177.7	3082.7
				11	400311	83,381	129	15	197.9	3856.8
				All others	400399	9,697	48	9	52.9	795.5
	Old (cont)	Yes	SF-Low	8	401208	16,833	67	6	107.2	1608.3
				9	401209	11,681	62	6	65.0	1275.7
				10	401210	12,005	58	3	122.7	1629.9
			SF-High	8	401108	26,940	63	4	232.2	3077.2
				9	401109	16,003	72	7	111.9	1135.5
				10	401110	28,921	48	18	141.0	1230.8
			MF	8	401308	117,859	179	23	224.0	3381.0
				9	401309	62,890	79	18	188.8	2665.1
				10	401310	34,437	44	11	143.3	2557.3
				11	401311	25,029	65	6	152.5	2519.8
			All others	All others	400004	25,628	75	7	135.6	2208.3
	New	No	SF-Low	8	410208	25,796	58	11	122.0	1701.8
				9	410209	14,937	37	9	83.4	1316.8
				10	410210	30,717	79	6	187.4	2652.6
			SF-High	8	410108	28,745	48	3	232.4	5862.6
				9	410109	15,073	46	4	155.6	1979.2
				10	410110	49,891	72	12	215.6	2863.7
			MF	8	410308	15,085	31	4	133.6	2735.6
		All others	All others	All others	400001	37,206	50	11	160.5	2652.9
	All	All	All	15	400002	15,710	71	6	92.2	1527.5
				16	400003	10,331	69	9	47.1	786.9
SDGE	Old	No	Low	9	300409	43,840	95	13	161.7	2190.9
				13	300413	344,515	705	51	221.1	3698.6
			Medium	9	300509	41,982	81	4	288.0	4662.7
				13	300513	336,787	621	49	237.1	3867.8
			High	9	300609	17,535	66	5	108.9	2069.4
				13	300613	147,090	257	21	237.3	4100.1
			All others	All others	300001	10,891	65	5	72.6	1234.3
		Yes	Medium	13	301513	17,091	73	7	89.6	1507.6
			High	13	301613	27,669	84	4	187.4	2982.4
			All others	All others	300002	13,923	52	5	98.3	1762.3
	New	No	Low	13	310413	45,658	72	10	205.4	3087.3
			Medium	13	310513	39,778	55	8	237.6	3338.8
			High	13	310613	18,415	57	3	197.3	2389.6
		All others	All others	All others	300003	23,632	52	7	139.7	2338.1
Total						9,912,862	18,970			

Table 2-13
Master Metered Weights

Utility	Home Type	SFCODE Prefix	Population	Completes	Weight
PGE	2-4 units (du-tri-quadplex)	11	44,411	139	319.5
	Multi-family 5-20 units	12	18,507	10	1850.7
	Multi-family >20 units	13	39,171	16	2448.2
	Mobile home >4 units	14	101,305	217	466.8
SDGE	2-4 units (du-tri-quadplex)	31	8,630	24	359.6
	Multi-family 5+	32/33	11,270	3	3756.7
	Mobile home >4 units	34	41,500	93	446.2
SCE	2-4 units (du-tri-quadplex)	41	13,025	33	394.7
	Multi-family 5-20 units	42	14,139	7	2019.9
	Multi-family >20 units	43	16,080	10	1608.0
	Mobile home >4 units	44	110,710	211	524.7
LADWP	2-4 units (du-tri-quadplex)	51	3,782	2	1891.0
	MF and MH 5+	52/53/54	12,416	2	6208.0
TOTALS			434,946	767	

2.5 COMPARISON OF RESULTS ACROSS SAMPLING AND STUDY GROUPS

2.5.1 Non-Response Follow-Up Comparison

The non-response follow-up proved to be a successful way to capture a segment of the population underserved by the direct-mail campaign. Table 2-14 shows several key results for customers by dwelling type and survey method.

Table 2-14
Comparison of Results by Surveying Method

	Single Family		Multi-Family (2-4 Units)		Multi-Family (5+ Units)		Mobile Homes	
	Initial Mail	Non-Response	Initial Mail	Non-Response	Initial Mail	Non-Response	Initial Mail	Non-Response
Completed Surveys	12,599	1,225	2,979	409	2,866	512	526	37
Weighted to Population	2,363,823	3,693,704	524,317	1,155,001	513,069	1,463,655	95,691	103,602
Average Electric Consumption	7,248	7,160	4,429	4,201	3,689	3,969	6,271	6,531
Average Gas Consumption	547	538	341	338	215	216	491	478
Average Dwelling Size	1,837	1,755	1,156	1,061	925	914	1,258	1,083
Average Dwelling Age	14.5	18.9	24.0	24.8	28.4	34.6	19.4	27.9
Average Number of People	2.88	3.42	2.53	2.74	2.10	2.68	2.30	2.22
Average Number of Seniors	0.53	0.30	0.38	0.13	0.37	0.15	0.74	0.42
Average Income	73,389	68,714	54,246	47,346	45,388	41,702	30,971	28,807
Owners	91%	81%	50%	26%	26%	13%	87%	89%
Central Cooling	50%	47%	40%	33%	41%	31%	60%	38%
Gas Space Heating	85%	89%	77%	75%	46%	54%	57%	56%
All Exterior Walls Insulated	56%	61%	45%	48%	43%	44%	65%	59%
CFL Penetration	63%	50%	55%	42%	51%	37%	57%	51%
Primary Language English	92%	80%	85%	67%	87%	69%	95%	81%
Head of Household Hispanic	12%	26%	17%	36%	13%	33%	9%	20%
College Grad or Higher	53%	44%	47%	39%	50%	36%	23%	18%

In general, non-respondents had similar energy usage and major equipment holdings as direct-mail participants but differed significantly in that they were less likely to be property owners, less likely to be using energy-efficient lighting, more likely to be non-English speaking, more likely to be ethnically diverse, and less educated overall. It follows from this that the direct-mail campaign was most successful with individuals who were more aware of energy efficiency, were more motivated because of their ownership, more educated, and more capable of handling an English survey. The non-response follow-up was able to get to more Spanish-speaking customers. While the non-response follow-up adds significant cost to a project of this magnitude, the fact that customers differ in these ways indicates that it is a wise step to take to minimize non-response bias found in a single-method survey approach.

2.5.2 Master Metered Comparison

The master metered population has traditionally been difficult to survey. In order to attempt to capture master meter responses, this study used the two-phased approach in an effort to gather additional information about the master metered segment from property managers and thus minimize the amount of information that customers had to provide directly. While this allowed the study to target master metered homes, it still proved to be difficult to capture the market in a comprehensive way. Overall responses to the master metered survey were low and particularly low in multi-family facilities with over five units. While the market characteristics of master metered customers appear different from their corresponding housing group in the individually metered sample, it is difficult to draw strong conclusions from these results because of the relatively low number of responses. Table 2-15 provides a comparison of these two groups.

In general, it appears that the master metered mobile homes act fairly similarly to the individually metered mobile homes. Many of the direct mail based master meter results appear to have similar bias issues as were seen in the individually metered mail only study results (see previous section). This includes a higher number of senior citizens, higher education levels, and lower ethnicity variation.

2.5.3 Energy Usage Comparison

RASS results were also compared against the overall population for an energy usage bias. This involved using the original population sample frame and comparing the results received to the population usage on the basis of average energy use by strata. As is common with this type of study, the highest usage strata reported slightly lower than average use for respondents and the lowest energy group has slightly higher than average results. Table 2-16 shows the results by usage category and by utility.

Table 2-15
Comparison of Individually and Master Metered Customer Results

	Single Family	Multi-Family (2-4 Units)		Multi-Family (5+ Units)		Mobile Homes	
	Individual Metered	Individual Metered	Master Metered	Individual Metered	Master Metered	Individual Metered	Master Metered
Completed Surveys	13,824	3,388	200	3,378	46	563	521
Weighted to Population	6,057,528	1,679,318	73,475	1,976,724	107,955	199,293	253,514
Average Dwelling Size	1,787	1,090	1,817	917	617	1,167	992
Average Dwelling Age	17.2	24.6	10.8	33.0	4.0	23.8	18.6
Average Number of Seniors	0.39	0.21	0.54	0.21	0.56	0.57	0.72
Average Income	70,538	49,500	75,745	42,659	24,747	29,846	27,947
Owners	85%	33%	87%	16%	10%	88%	87%
Central Cooling	48%	35%	21%	34%	8%	49%	47%
Gas Space Heating	87%	76%	78%	52%	69%	56%	79%
All Exterior Walls Insulated	59%	47%	31%	44%	10%	62%	50%
Clothes Washer	96%	54%	87%	27%	23%	86%	68%
Primary Language English	85%	73%	87%	74%	81%	88%	96%
Head of Household Hispanic	21%	30%	15%	28%	7%	15%	11%
College Grad or Higher	47%	42%	58%	39%	42%	20%	21%

Table 2-16
Comparison of Energy Use For Respondents and the Target Population

Usage by Household		Dwelling Type and Usage Strata Definition							
Utility		6 High	5 Med	4 Low	1 SF-High	2 SF-Low	3 MF	all	Utility Totals
PG&E	Population Count				1,684,655	1,165,896	1,112,205	84,938	4,047,694
	Respondent kWh/Year				9,640	3,629	3,995	7,188	6,306
	Population kWh/Year				9,815	3,536	3,926	6,878	6,327
	Average Error				-1.8%	2.6%	1.7%	4.5%	-0.3%
SCE	Population Count				1,455,364	1,260,415	1,050,641	88,824	3,855,244
	Respondent kWh/Year				9,112	3,730	4,146	7,497	5,962
	Population kWh/Year				9,427	3,611	4,063	8,120	6,034
	Average Error				-3.3%	3.3%	2.0%	-7.7%	-1.2%
SDG&E	Population Count	210,709	435,638	434,013				48,446	1,128,806
	Respondent kWh/Year	12,106	5,277	2,343				4,831	5,404
	Population kWh/Year	11,267	5,158	2,297				6,125	5,240
	Average Error	7.4%	2.3%	2.0%				-21.1%	3.1%
LADWP	Population Count	82,482	255,486	513,253				27,780	879,001
	Respondent kWh/Year	10,432	4,869	2,257				5,046	3,872
	Population kWh/Year	11,865	4,991	2,227				5,588	4,041
	Average Error	-12.1%	-2.4%	1.4%				-9.7%	-4.2%
Strata Totals	Population Count	293,191	691,124	947,266	3,140,019	2,426,311	2,162,846	249,988	9,910,745
	Respondent kWh/Year	11,635	5,126	2,297	9,395	3,681	4,068	6,603	5,853
	Population kWh/Year	11,435	5,096	2,259	9,635	3,575	3,993	7,030	5,886
	Average Error	1.7%	0.6%	1.7%	-2.5%	3.0%	1.9%	-6.1%	-0.6%

The "All" strata column includes customers who were grouped together into composite strata because there were not enough of them with similar characteristics to create individual strata. Because they are a composite of multiple types of homes, their usage varies much more widely than the defined strata groups. However, these "All" customers represent a relatively small segment of the overall study population.

The largest differences in usage (indicated by the error percent which is the difference between the respondent usage and the population usage divided by the population usage) occur in SDG&E and LADWP's service territory. After comparing LADWP results with previous results for their territory, it appears that the single family market is underrepresented. Since single family customers use more energy than multi-family customers, it appears that this compounds the fact that the high use area was underrepresented. Section 2.5.6 below further discusses the LADWP shortfall.

SDG&E's results are in part affected by the fact that with their relatively small sample population they had a higher relative number of customers grouped into "All" strata. Some of this was caused by a misclassification of the climate zone in the SDG&E service territory (discussed further in Section 2.5.5).

Across the board, PG&E's results were underestimated by 0.3 percent of the population's energy use and SCE's results were under estimated by 1.2 percent. These two utilities together display the phenomena of under-representing the highest users and over-estimating in the lowest use strata.

2.5.4 Census Data Comparison

The weighting procedures for the individually and master metered samples are only appropriate if there is no basis for identifying differential response rates within sampling cells associated with customer characteristics that may relate to parameters of interest. Furthermore, our experience is that not all groups respond to surveys such as the RASS at the same rate. Of particular relevance to a RASS study is the tendency of response rates to vary among income levels and the elderly to respond at higher rates. Because neither households of various income levels nor elderly households can be identified reliably from utility billing information, they cannot be associated with specific sampling cells. Consequently, differential response rates from these groups may distort or bias the results for each cell. Adjustment for this type of differential response is accomplished by post-stratification weighting.

To determine whether post-stratification weighting was necessary, we compared the distribution of responses gathered from the RASS project with US Census Bureau data from 2000. Overall, the comparison of the RASS demographic information to the 2000 Census data is reasonable, and the sampling plan yielded a set of customer respondents that reasonably mirror the population at large. The most notable area where the study appears to fall short is in the single-occupant rental market. The shortfalls occur predominantly in the young-adult age groups. Because the results aligned well with census data for other comparison segments, the study group decided to keep the initial sample weights and not post-stratify the results.

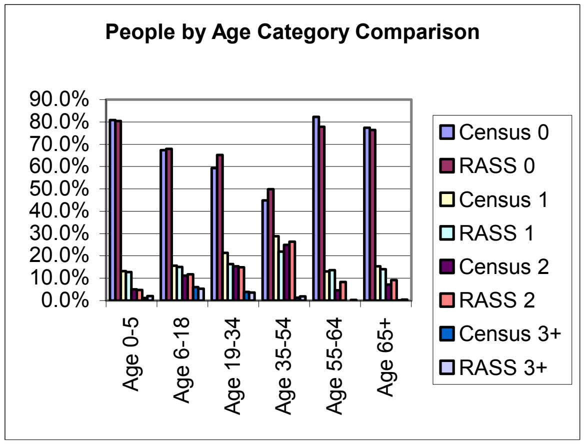
A few of the Census-to-RASS comparison values (most notably ethnicity and language) were asked in a different format from the Census so comparisons are not directly relevant. Despite language results that differ in form enough that a comparison is not meaningful, the fact that the RASS' Hispanic ethnicity numbers come out very close to the Census helps to confirm that we were able to capture results from that population segment. As noted above, this is in large part because of the non-response follow-up efforts. A series of comparison tables is included below as Figure 2-3.

2.5.5 Reassignment of Energy Commission Climate Zones

During the process of reviewing the final results, KEMA-XENERGY discovered that the original climate zone assignment file had some errors in it. The assignment file is a link that ties customer zip code to the appropriate Energy Commission forecast climate zone. While most of the problematic zip codes were in the PG&E service territory where zones 4 and 5 had a section of zip codes that had to be switched, there were other smaller areas that needed refinement. In addition, there were several zip codes in the SDG&E area that had been changed or added and had been erroneously assigned to climate zone 9. Since all SDG&E customers should fall in climate zone 13, these were reassigned.

All reassignments occurred prior to the final reporting of results and all results and tables reporting values by climate zone use the corrected climate zones. However, due to the nature of sampling and the need to maintain the existing sample frame when assigning weights, the customers remain in their initial strata.

Figure 2-3
Comparison of RASS Results to 2000 Census Results



2.5.6 Calibration Issue with LADWP Totals

It appears from reviews of the LADWP results that the population provided by LADWP at the outset of the study may have excluded a number of higher consumption, single family customers. This is indicated by the fact that the overall energy use and population counts for the residential population appears lower than numbers the Energy Commission gathered from LADWP FERC filings.[5] In addition, this is further affirmed by the fact that the number of single family homes in the LADWP service territory as reported in the study appears to be significantly lower than results obtained on other studies. Previous Energy Commission information points to single family rates in the 40-50% range. The RASS results for climate zone 11 show 27% and climate zone 12 show 17%. Both of these zones are served by LADWP.

Because the missing customers represent a small overall number, the study group has decided that it is important to caveat the results, but that there is not a significant impact on the overall statewide results.

2.5.7 Calibration of SoCalGas' Results

Because of the fact that the study was electrically focused and served the IOUs plus LADWP, a portion of SoCalGas' customers who are served by other electric providers (i.e. The City of Anaheim) were not included in the sample. In addition, the fact that a section of LADWP customers were missing and likely to be single family dwellings with higher than average use left the SoCalGas population underrepresented. As discussed in Volume 2 - Section 2.1 (Calibration Results), SoCalGas' calibration factor was 1.121. This indicates that the sample underrepresented their population by approximately 12%. When we compared the new and old responses, this difference came to light in that the ratio of new single family homes in the sample was high enough to skew the SoCalGas new/old results such that overall new homes used slightly more energy than older homes. See Section 1.6 in the Executive Summary for a detailed discussion of the new/old home issue. SoCalGas went back and recalibrated the values using their own weighting values and came up with a revised set of weighted average annual therm values that more closely represents their actual new/old population splits. Table 2-17 displays both the RASS and SoCalGas revised totals. While the RASS values provide a good proxy of the overall use, cutting the data into smaller groups such as the new/old split can create big enough differences small segments of the population (new homes in this case) which can skew the overall totals. Please take note of this fact when reviewing SoCalGas results.

Table 2-17
Revised Weighting of SoCalGas Customers for New/Old Dwellings

Dwelling Type	Vintage	SCG Re-Weighted Average Annual Therms	SCG Re-Weighted Customer Count	CA RASS Sample Count	CA RASS Weighted Average Annual Therms	CA RASS Weighted Customer Count
All Homes	All	451.5	4,981,668	7686	441.1	3,743,921
	Pre 1997	452.9	4,678,961	7211	440.6	3,535,623
	Post 1996	430.2	302,707	475	450.4	208,297
Single Family	All	535.6	3,346,603	5352	521.8	2,475,867
	Pre 1997	540.0	3,111,348	4981	524.8	2,306,366
	Post 1996	478.1	235,255	371	480.2	169,500
Multiple Family	All	275.8	1,635,065	2334	275.9	1,268,054
	Pre 1997	276.8	1,567,613	2230	274.8	1,229,257
	Post 1996	251.2	67,451	104	310.3	38,797

2.6 PRECISION OF RASS ESTIMATES

This section discusses the sampling variability associated with the individually and master metered samples. We provide 90 percent confidence intervals for various percentage estimates based on the RASS sample.

2.6.1 Individually Metered Sample Precision

Table 2-18 presents the precision of estimates from the individually metered RASS sample for the individual utility service territories and for the population. The last three columns in the table 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. At worst, which corresponds to an estimate of 50 percent, the overall population estimate generated from the RASS individually metered sample has a precision of +/-1.2 percentage points at 90 percent confidence.

Table 2-18
Precision of Estimates for the Individually Metered Sample

			90% Confidence Bounds (+/-) For Estimated Responses		
Utility	Population	Total Completes	50/50%	20/80%	10/90%
PG&E	4,047,694	9,265	1.9%	1.5%	1.1%
SCE	3,857,361	7,979	2.0%	1.6%	1.2%
SDG&E	1,128,806	2,527	3.7%	2.9%	2.2%
LADWP	879,001	1,382	4.5%	3.6%	2.7%
Total	9,912,862	21,153	1.2%	1.0%	0.7%

By way of example, 50% of all PG&E's customers answered that all of their walls have exterior wall insulation. The actual value for this response includes the +/- 1.9% shown on the table or between 48.1 and 51.9%. Ten percent of SCE customers report that they have remodeled their home in the past 12 months. Using the 10 or 90% estimate column for SCE, the actual value falls in the range of 8.8 to 11.2%.

It should be noted that these confidence intervals assume a design effect equal to one. (The design effect impacts confidence intervals due to its impact on the effective sample size, since the effective sample size is equal to the sample size divided by the design effect.) That is, we have effectively assumed that the variance within the follow-up sample clusters is the same as the variance across the state.

2.6.2 Master Metered Sample Precision

Table 2-19 presents the precision of estimates from the master metered RASS sample for the individual utility service territories and for the population. The last three columns in the table 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. At worst, which corresponds to an estimate of 50 percent, the overall population estimate generated from the RASS individually metered sample has a precision of +/- 4.6 percentage points at 90 percent confidence.

Table 2-19
Precision of Estimates for the Master Metered Sample

			90% Confidence Bounds (+/-) For Estimated Responses		
Utility	Population	Total Completes	50/50%	20/80%	10/90%
PG&E	203,394	382	5.7%	4.6%	3.4%
SCE	153,954	261	6.0%	4.8%	3.6%
SDG&E	61,400	120	12.4%	9.9%	7.4%
LADWP*	16,198	4	-	-	-
Total	434,946	767	4.6%	3.6%	2.7%

* We did not calculate confidence bounds individually for LADWP since the number of completes was so low.

3: DATABASE PREPARATION

3.1 Introduction

This section provides a description of the databases that will be generated as part of the Residential Appliance Saturation Survey (RASS) performed for the California Energy Commission.

The RASS Survey was a scannable survey form. The form asked participants to fill in their best answer to each question. Since the vast majority of the surveys were mailed to participants, the responses were for the most part self-reported. The non-response follow-up effort did include some surveys that were completed by trained interviewers. Participants did have access to a toll-free survey help-line if they needed assistance in completing their form.

Following is a discussion of the construction of the databases that were used in the project and how these databases were populated, checked for data quality, and how missing values were filled for the purpose of estimating the CDA model. This section provides a brief description of the contents of the three databases and a schematic of the database preparation process. Section 4 discusses the cleaning tools used to create the databases. There are three core databases created from the RASS effort, the raw survey results, the cleaned survey and CDA results, and the billing data. In addition, each sponsoring utility received a copy of their own sample frame information so that they can link RASS responses with their specific customers.

3.2 Database Formats

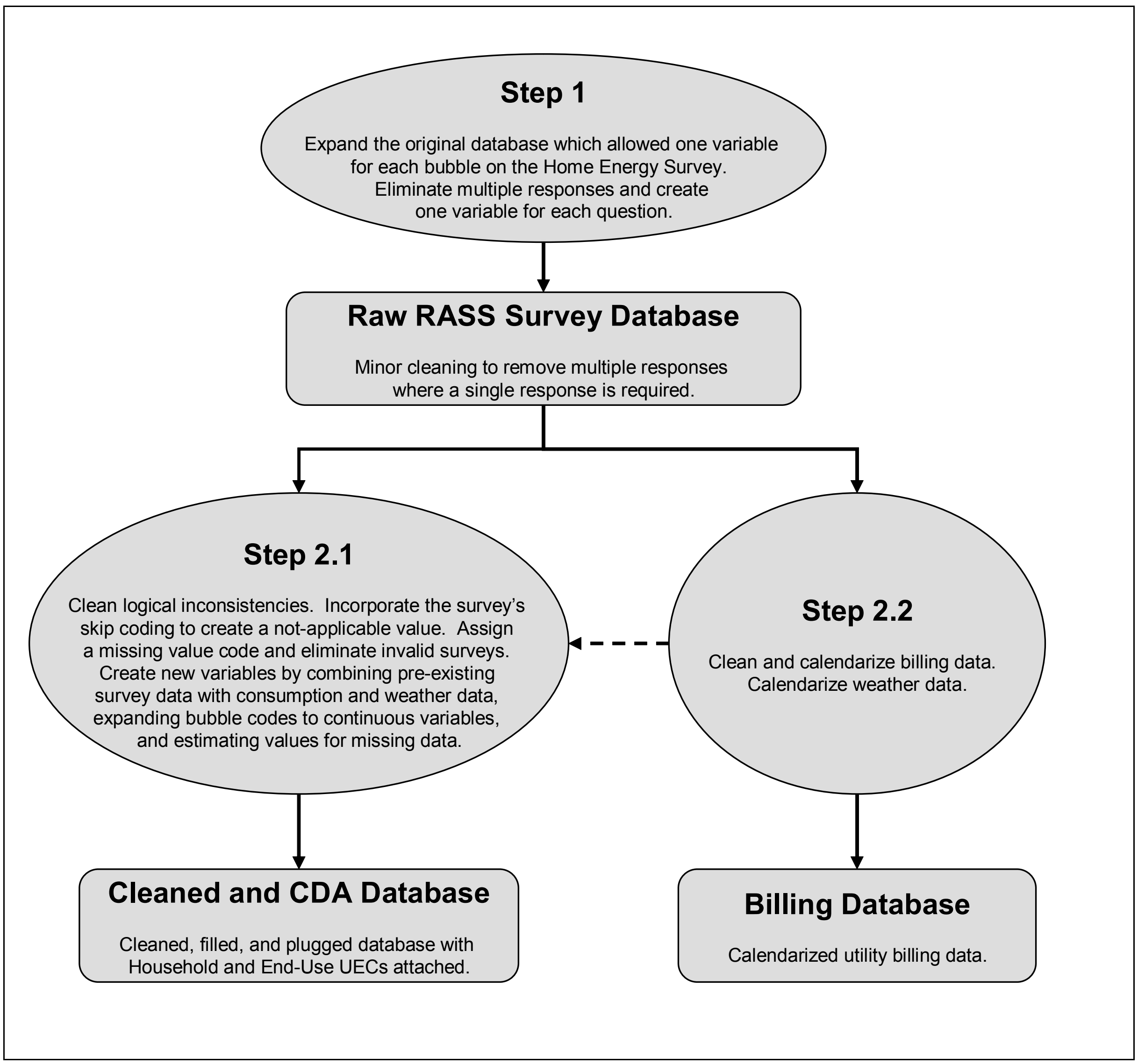
The tool that was used for analysis in this project is created from The SAS System (SAS). To facilitate the use of SAS, the data was stored in SAS datasets and SAS was used to perform all the tasks described in this document. The SAS System allows for large, fanned out databases that are easily manipulated and SAS supports the analytical processes needed in the Conditional Demand Analysis.

All final databases were provided in a series of output formats – SAS datasets for both PC and mainframe SAS, and a flat comma delimited file that can be imported into other database platforms. The study team has also developed a web interface to provide access to the data in a user friendly manner. The web interface allows users to subset saturation results and view them by a variety of crossed variables.

3.3 Overview of Database Preparation

An overview of the database preparation process is presented in Figure 3-1. As presented in Figure 1, three databases were generated to hold the data in various stages of the process. The three databases contain the raw survey data, a cleaned version of the survey data along with the CDA results, and the billing data. All final datasets were provided in confidential format where all identifying information about the customers' address, name, and utility identification numbers were removed.

Figure 3-1
Schematic of the Database Preparation Process



- **Database 1: Raw RASS Survey Database.** This database contains the RASS questionnaire responses. Minimal cleaning was be undertaken to eliminate multiple responses such that there is a single answer for all questions that no not allow multiples and multiple responses as appropriate throughout the survey.

- **Database 2: Cleaned Survey and CDA Database.** The cleaned RASS survey results and CDA database is a result of implementing a number of data cleaning and quality control techniques on the raw survey results. These steps are outlined in detail in Section 4. The database contains both the cleaned survey input and household and end-use UECs calculated from the CDA and UEC models. The cleaning techniques used to create the database included quality control procedures to check for and correct logical inconsistencies, the definition of missing and not applicable values, the elimination of invalid surveys, and the plugging of fuel types. In some cases, where the database indicates that the respondent did not reply to a question, a response may be estimated from other customer information provided in the survey. This database also contains the normalized monthly heating and cooling degree days and household and end-use UEC for each respondent.

- **Database 3: Billing Database.** This database contains the billing data that was used in the CDA modeling process. To create this database, the team used the billing data provided by the utility companies. These data were cleaned for missing and inconsistent values and were calendarized and normalized to a 30.4 day month.

3.4 Database 1: Raw RASS Survey Database

The original RASS results were scanned electronically into fixed format text files. The data in these files represents the respondents' answers to the RASS Survey.

The first step of the database preparation process, Step 1 in Figure 1, used SAS to read in the original data and create variables for each bubble on the RASS Survey. For example, a variable from the raw data file, which has 10 possible answers, will be expanded to create 10 new fields in the SAS dataset. Expanding the fields in this manner allowed for a rigorous analysis of multiple answers provided by survey respondents.

For those questions where the respondent answered the question with only one response, a single variable was created to hold the answer. For questions where respondents answered multiple values, the project team determined an algorithm to collapse the multiple response fields down into a single field. For most variables, the team chose either the largest or the smallest response the respondent provided. For example, if the respondent provided multiple answers for the year their home was built, the team chose to accept the oldest possible housing age. Several questions that eventually end up with a single response such as primary heating or water heating fuel were kept with an individual variables for each survey option so that the cleaning process could more comprehensively evaluate other survey responses before making a decision on the appropriate single response for the question.

3.5 Database 2.1: Cleaned and CDA Database

A multi-step cleaning process was used to create the Cleaned and CDA Database. The cleaning process began with a process to count missing values and checking for and cleaning logical inconsistencies. Surveys were eliminated if they were invalid due to too many multiple responses, incomplete surveys, or too many logical inconsistencies. Section 4 outlines cleaning code details.

Prior to estimating the CDA models, algorithms were designed to fill and plug missing variables. A careful review was undertaken to validate and check fuel and system types. Once the values were used in the CDA model, the household and end-use UECs and the normalized monthly heating and cooling degree days for each site were appended to the cleaned survey data. Pre and Post-cleaned annualized electricity and therm consumption variables were also added to these data. Section 4 includes an in-depth description of the CDA data filling and consumption cleaning processes. Section 5 provides a detailed description of the CDA modeling process.

3.5.1 Non Response Indicator

After the initial data were cleaned of multiple responses, the team differentiated between non-responses due to the survey's skip coding and simple respondent non-response. During the cleaning process, skip coding non-response was assigned a value of 99, which meant not applicable. Simple respondent non-response was assigned a value of 97. Surveys that contained an excessive amount of non-responses were eliminated as invalid.

3.5.2 Logical Response Inconsistencies

Throughout the survey, respondents were asked several questions where their response should have naturally influenced their response to later questions. When the respondent's answers to these types of questions was not consistent, the team either filled the answer with the most appropriate response, or if no response was obvious, the respondent was given a missing value code (97) and a logical inconsistency flag. The flag's value was increased by one each time the respondent answered in a manner that was logically inconsistent. A large percentage of the sample (slightly over 60%) never responded in a logically inconsistent fashion, and many of the observed inconsistencies were so minor that they were handled in the cleaning process. Only 1.7% of the sample responded in a manner that was logically inconsistent more than five times during the survey.

3.5.3 Filling Missing Values

For a variety of reasons, the Cleaned RASS Survey contains a number of missing values. Simply allowing these missing values to disqualify an observation from the regression dataset would create non-response bias in the estimation of model parameters. A multi-step approach was used to fill missing values. This approach is discussed in detail in Section 4.3, the CDA Data Filling Process.

3.5.4 Refining Fuel and System Types

It is fairly well known that survey respondents often misreport fuels and system types. This kind of misreporting creates significant challenges when using results to predict overall endues consumption levels using the CDA approach. As a result, a considerable amount of care was taken to validate (and, where justified, to override) reported systems and fuels. The team's approach to fuel checking and plugging is discussed in detail in the Data Cleaning and Processing Section.

3.5.5 Normalized Weather and Estimated UECs

The Cleaned and CDA Database contains estimates of each site's normalized electric and gas whole household UEC and UECs for all end-uses. The creation of these estimates required the creation of calendarized energy consumption and weather data. Energy consumption was used as the dependent variable and weather was used as one of the independent variables in the UEC models. The normalized weather, used to create the UECs was also appended to the Cleaned and CDA Database.

3.6 Database 2.2: Billing Database

The Billing Database holds the monthly energy consumption for each survey site. The consumption data includes information on the monthly electricity and natural gas usage, the year and month, the presence of a gas master meter, and two indicators for the utilities serving each customer. During the creation of the Billing Database, the information on energy consumption and the meter reading date was used to calendarize the site's energy consumption for the month standardized to a fixed number of days per month.

3.7 Data Delivery

Data was delivered to each study sponsor on CD. The CD contained the files noted below.

Survey Data:
SURVDATA.csv (unformatted)
SURVDATF.csv (formatted)
SURVDATA.xpt
SURVDATA.sas7bdat
SURVCONT.xls (contents)
FORMATS.txt (format statements)
ApplyFormats.txt (applies formats to specific variables)

Billing Data:
BILLDATA.csv
BILLDATA.xpt
BILLDATA.sas7bdat
BILLCONT.xls (contents)

Raw Survey Data:
sampledata_all.zip (individually metered)
sampledata_mm.zip (master metered)

Electronic Version of Survey:
Statewise-RASS-var-final.doc

SAS Files: See description of files included in Appendix L

For all of the datasets, we have created a comma-delimited version, a SAS export file for mainframe users, and a PC SAS dataset. In addition, we have included an Excel copy of the contents of the database which should allow users to sort the data as you need and access labels in a central location. We have provided an electronic version of the survey as well.

*As a reminder, each sponsoring utility received a dataset with their customer specific information tied to the generic SFCODE, which serves as the primary key for the RASS databases. The SFCODE is a generic value that provides embedded information about the utility and strata from which the customer was drawn. All study participants receive the full statewide survey data set, which includes not only the cleaned survey responses, but also UEC results for each of the individually metered customers. The cleaned survey data is provided in both formatted and unformatted form. There is a copy of the format statements on the CD as well so that users can use the raw data and apply formats dynamically.

The CD also included a copy of the billing data (again, generically labeled) that was used for the conditional demand analysis. It also included a raw survey file that includes results for each possible response on the survey.

4: DATA CLEANING AND PROCESSING

4.1 RASS Survey Data Cleaning

The section outlines the criteria used to eliminate surveys that were determined to have excessive amounts of invalid data, the cleaning done on RASS Survey variables, and the creation of new variables from the cleaning process and the combination of survey variables.

4.1.1 Overview

When the surveys were scanned, each bubble response was initially read as if it were its own variable. In the SAS program *min_max.sas*, if multiple bubbles were filled for the same question (where the question did not allow multiple responses), the project team developed a hierarchical procedure to decide which response to select. For most questions, either the highest or the lowest response was selected as appropriate. The resulting dataset (*survdata_short*) contains one variable per question (except for questions where multiple responses were allowed).

A SAS program called TooManyResponses.SAS was run on the initial SAS data set, SampleData_All1 to identify problematic responses and correct them. The program counted the number of multiple responses (where the question did not allow multiple responses) to assess possible problems with the survey data provided by each customer.

The systematic approach to data validation and cleaning performed in this program concentrated on the following issues:

- Multiple question response
- No response indicated
- Logical response inconsistencies
- Missing values
- Fuel misreporting

Multiple Question Responses

Most questions in the RASS Survey were designed to have one response. However, many respondents provided multiple answers for at least one question on the survey. For each question, an algorithm to collapse the multiple response fields into a single field was developed. The algorithm provides a systematic method for handling multiple responses in a consistent and logical manner. The field retained from this cleaning process was given the original field name from the RASS Survey. For most variables, the team chose either the largest or the smallest response the respondent provided. For example, if the respondent provided multiple answers for the year their home was built, the team chose to accept the oldest possible housing age. Several questions that eventually end up with a single response such as primary heating or water heating fuel were kept with an individual variable for each survey option so that the cleaning process could more comprehensively evaluate other survey responses before making a decision on the appropriate single response for the question. Details of those subsequent steps are provided in this section.

Non-Response Indicator

After the initial data were cleaned of multiple responses, non-responses due to the survey's skip coding and simple respondent non-response were identified. The raw text non-response data coding for the RASS Survey did not distinguish skip coding non-response from simple respondent non-response. These two types of non-response, however, are not equivalent. During the cleaning process, skip coding non-response was assigned a value of 99, meaning not applicable. Simple respondent non-response was given a value of 97.

Logical Response Inconsistencies

Survey respondents were asked several questions where their response should have naturally influenced their response to later questions. When the answers to these questions were inconsistent, an attempt was made to fill the answer with the most appropriate response or, if no response was obvious, the response was given a missing value code of 97.

To assess this potential problem further, a flag for logical inconsistencies was created. The flag's value was increased by one each time the respondent answered in a logically inconsistent manner. For example, the survey questioned the respondent about the number of computers in the residence, the number of hours the household computers are used, and the types of activities undertaken on the computers. If a respondent stated that they had no computers or failed to answer this question, but proceeded to list hours of usage and tasks undertaken, their first response to the number of household computers would be counted as logically inconsistent. The logical inconsistency flag would be augmented by one. The total number of logical inconsistencies was checked to determine surveys that were answered poorly. The logical inconsistency flags indicate that most respondents answered the survey's questions in a logically consistent manner.

Missing Values

As discussed above, the Cleaned RASS Survey contained a number of missing values. Simply allowing these missing values to disqualify an observation from the regression dataset would create non-response bias in the estimation of model parameters. Replacing these missing values with overall means for the variables in question would also lead to biased estimates insofar as question-specific non-respondents tend to be different from respondents. To minimize non response bias, the team created a new "filled" variable for the purposes of the conditional demand analysis. The process for filling missing values used a multi-step approach that relied on correlations between the question with the missing response and other questions that contained valid responses. The team used this approach to fill missing values for household income, square footage of the home, number of residents, and the age of the home. These variables are in addition to the cleaned survey data and are developed primarily for use in the conditional demand analysis.

In addition to the four primary plugged variables, there were surveys with missing values for residence type (no response or "other") and surveys whose chosen residence type did not match the utility's residence type designation. To fill these missing values and check the discrepancies between survey response and utility records, a multi-step procedure was developed. The algorithm relied on a series of logistical checks with other pertinent information supplied from the respondent. Details on the CDA filling process are included as Section 4.3.

Refining Fuel and System Types

It is common for survey respondents to misreport fuels and system types. This kind of misreporting can be troublesome for the process of disentangling end-use consumption levels through the conditional demand approach. As a result, considerable care was taken to validate (and, where justified, to override) reported systems and fuels. The Data Cleaning section of this report discusses the algorithm used to fill for missing and incorrectly specified fuel type and the number of observations effected by this issue.

4.1.2 Invalid Surveys

The cleaning process eliminated unwanted multiple responses, coded missing responses, and checked for logical inconsistencies. Individual surveys needing extensive cleaning could represent respondents whose limited understanding of their system types or the survey format invalidated their survey responses. This section details the criteria developed to eliminate survey observations containing data deemed too unreliable for use in the survey saturation tables or the CDA analysis. In particular, surveys were eliminated if they contained an excessive number of multiple responses, were incomplete, or contained too many logically inconsistent responses.

In all cases, the team did initial physical checks to customers' physical surveys to insure that problems identified using a systematic computer based screen were correctly representing the issue identified. This included checking for surveys answered in pen (the scanner requires pencil and most pen surveys are caught upfront before they are scanned) and insuring that invalid surveys were in fact highly problematic. The extra systematic check with manual backup identified a small number of surveys that had made it into the database in pen. Surveys that were "fixable" were manually adjusted as necessary. This includes coloring over pen marks with pencil to insure readability on the scanner.

Multiple Responses. Respondents who provided more than 15 multiple responses were deleted from the second (cleaned survey) database.

Incomplete Surveys. To help determine if a survey was incomplete, 20 variables were chosen that all respondents should have answered. The 20 variables included household type, age of residence, system types, education, and income. None of the 20 variables was included in a skip coding sequence. The 20 variables were chosen to represent all areas of the survey (beginning, middle, and end). Households not responding to at least 10 of the key survey questions were eliminated from the database.

Four survey pages were also examined to determine incomplete surveys, pages 1 (Home and Lifestyle), 2 (Home and Lifestyle), 8 (Laundry, Food Preparation), and 9 (Refrigerators). Surveys with no responses at all on page 1 and 2 or page 8 and 9 were eliminated because these represent areas of the survey that most everyone should have answered and which make estimation nearly impossible for other missing data. As a third check for incomplete surveys, surveys with no responses on page 8 or 9, that were also missing at least 5 of the 20 variables chosen above, were also eliminated from the database.

Logical Inconsistencies. Logical inconsistencies in a respondent's answers also indicated a lack of understanding of either their system types or the survey format.

Flags for logical inconsistencies were created during the data cleaning process. Respondents with more than 10 logically inconsistent responses were eliminated from the second database, however, no surveys were found to have 10 or more logical inconsistencies

Table 4-1 summarizes the total number of invalid surveys identified in the cleaning process. This is in addition to 315 that were returned with all customer information removed making it impossible to process them for inclusion in the study results. The number of surveys eliminated due to incomplete surveys is not surprising given the length of the survey. The survey's length required a reasonable time commitment from respondents, increasing the probability that some respondents would not have the time to fully complete the survey.

Table 4-1
Summary of Invalid Surveys

Reasons for Eliminating Survey	Number Eliminated
Too many multiple responses	13
Incomplete survey	146
Too many logical inconsistencies	0

4.1.3 Survey Specific Cleaning

The remainder of this section describes the data cleaning efforts for the individual survey responses. Included is a discussion on how invalid surveys were identified and a description of the data cleaning for individual questions organized by the following survey sections:

- Your Home and Lifestyle
- Space Heating
- Space Cooling
- Water Heating
- Laundry
- Food Preparation
- Refrigerators
- Freezers
- Spas and Hot Tubs
- Pools
- Entertainment and Technology
- Lighting
- Miscellaneous Appliances
- Household Information

Numerous sections include additional filling in of missing data for the CDA which is detailed as a second cleaning step in Section 4.3.

4.1.4 Your Home and Lifestyle

The Your Home and Lifestyle section of the survey contained 20 questions. Table 4-2 summarizes the percent of missing responses for nine of the major questions in this section. The section continues with details on the method for cleaning and filling these nine questions.

Table 4-2
Missing Home and Lifestyle Responses

Question	Missing Pre Cleaning	Missing Post Cleaning
A1. Type of Building (*DWLTYPE*)	11.3%	11.1%
A2. Own or rent home (*OWNRENT*)	1.8%	1.6%
A3. How long at address (*YRS_RES*)	1.8%	2.1%
A4. Seasonal Occupancy (SEASOCC)	2.2%	0.0%
A6. Year home built (*BUILTYR*)	7.7%	8.8%
A7. Number of bedrooms (*NUMROOM*)	1.1%	5.5%
A8. How many square feet of living space (*SQFT*)	10.0%	11.0%
A18. Is natural gas available (*NGSERV*)	6.1%	1.6%
A19. Natural gas hookup in home (*NGLINE*)	13.3%	1.1%

A1 – Type of Building (DWLTYPE)

The following describes the *DWLTYPE* variable and the process to develop a residence type variable to be used in the CDA model (*RESIDENCE – note this variable was concatenated to RES in the final database files*). Note that in this case, the original survey response to type of building (*DWLTYPE*) was unchanged from the original responses.[6] Instead, a new variable RESIDENCE was constructed. The process used the individual's survey response, the utility's residence type description (fourth digit of *SFCODE* for individual metered customers and the second digit of *SFCODE* of master meter customers), the residence street address, the survey response for payment of heating, cooling, water heating, and laundry systems, and the survey response for the dwelling's square footage.

For the following discussion, *DWLTYPE* is the individual's response to the survey question, *RESTYPE* is the fourth digit of *SFCODE* provided by the utilities, and *RESIDENCE* is the new variable created in the following process. If there was no problem with the individual's original *DWLTYPE* response, the value for *RESIDENCE* is their original *DWLTYPE* value.

On the survey, *DWLTYPE* has the following coding:

- 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, and
- 6 is other.

DWLTYPE was coded 97 if the survey respondent did not answer the question.

The utility codes for *RESTYPE* for individual metered customers are as follows:

- 1 or 2 are single family residences,
- 3 is a multifamily residence,
- 4 is a low usage residence,
- 5 is a medium usage residence,
- 6 is a high usage residence, and
- 0 is unknown.

The following are the rules for developing the *RESIDENCE* variable for individual metered customers.

- If *DWLTYPE* is equal to 2, 3, or 4 and the utility's *RESTYPE* code is equal to 1, 2, 4, 5, 6, leave *RESIDENCE* at the individual's response for *DWLTYPE*. In this situation the survey response overrides the utility's *RESTYPE* code (codes 1 and 2 are single family; 4, 5, and 6 are usage codes).

- If *DWLTYPE* is equal to 1 and the utility's *RESTYPE* is equal to 3 (utility code for multifamily), proceed through the following checks:

    Review the service address. If address ends in a number 1-4 or the letter A, B, C, or D, set *RESIDENCE* to 3.

Review the service address. If the address ends in a number larger than 4 or a letter later than D, set *RESIDENCE* to 4.

Review the service address. If the service address does not end in a letter or a number, check if the property owner pays for a major system and if the survey response to square footage is less than 1,500.2 If both checks are satisfied, then set *RESIDENCE* to 2.

If none of the above conditions is met, set *RESIDENCE* to 1.

- If *DWLTYPE* is equal to 6 and the utility's *RESTYPE* is equal to 3, proceed through the following checks:

  Review the service address. If the address ends in a number 1-4 or a letter A-D, set *RESIDENCE* to 3.

  Review the service address. If the address ends in a number larger than 4 or a letter later than D, set *RESIDENCE* to 4.

  If the address does not end in a number or a letter, set *RESIDENCE* to 2.

- If *DWLTYPE* is equal to 6 and *RESTYPE* is equal to zero (the utility does not know the *RESTYPE*), proceed through the following checks:

  Review the service address. If the address ends in a number 1-4 or a letter A-D, set *RESIDENCE* to 3.

  Review the service address. If the address ends in a number larger than 4 or a letter later than D, set *RESIDENCE* to 4.

  If the address does not end in a number or a letter and the survey response to square footage is greater than or equal to 1,500, set *RESIDENCE* to 1.

  If the address does not end in a number or a letter and the survey response to square footage is less than 1,500, set *RESIDENCE* to 2.

- If *DWLTYPE* is equal to 6 and *RESTYPE* is equal to 4, 5, or 6 (utility codes are based on usage, not a residence type indicator), proceed through the following checks.

  Review the service address. If the address ends in a number 1-4 or a letter A-D set *RESIDENCE* to 3.

Review the service address. If the address ends in a number larger than 4 or a letter later than D set *RESIDENCE* to 4.

If the address does not end in a number of a letter, set RESIDENCE to 1.

- If *DWLTYPE* is equal to 6 and *RESTYPE* is equal to 1 or 2 (utility codes for single family), proceed through the following checks.

  Review the service address. If the address ends in a number 1-4 or a letter A-D, set *RESIDENCE* to 3.

  Review the service address. If the address ends in a number larger than 4 or a letter later than D, set *RESIDENCE* to 4.

  If the address does not end in a number or a letter, set *RESIDENCE* to 1.

- If *DWLTYPE* is equal to 97 and *RESTYPE* is equal to 1 or 2 (utility codes for single family), proceed through the following checks.

  Review the service address. If the address ends in a number 1-4 or a letter A-D, set *RESIDENCE* to 3.

  Review the service address. If the address ends in a number larger than 4 or a letter later than D, set *RESIDENCE* to 4.

  If the address does not end in a number or a letter, set *RESIDENCE* to 1.

- If *DWLTYPE* is equal to 97 and *RESTYPE* is equal to 3 (utility code for multifamily), proceed through the following checks:

  Review the service address. If the address ends in a number 1-4 or a letter A-D, set *RESIDENCE* to 3.

  Review the service address. If the address ends in a number larger than 4 or a letter later than D, set *RESIDENCE* to 4.

  If the address does not end in a number of a letter, set *RESIDENCE* to 2.

- If *DWLTYPE* is equal to 97 and *RESTYPE* is equal to zero (utility code for unknown), proceed through the following checks:

Review the service address. If the address ends in a number 1-4 or a letter A-D, set *RESIDENCE* to 3.

Review the service address. If the address ends in a number larger than 4 or a letter later than D, set *RESIDENCE* to 4.

If the address does not end in a number or a letter and the property owner does not pay for any major system, set *RESIDENCE* to 1.

If the address does not end in a number or a letter, the survey response to square footage is less than 1,500, and the property owner pays for at least one of the major systems, set *RESIDENCE* to 2.

- If *DWLTYPE* is equal to 97 and *RESTYPE* is equal to 4, 5, or 6 (utility code for usage strata), proceed through the following checks:

Review the service address. If the address ends in a number 1-4 or a letter A-D, set *RESIDENCE* to 3.

Review the service address. If the address ends in a number larger than 4 or a letter later than D, set *RESIDENCE* to 4.

If the address does not end in a number or a letter and the property owner does not pay for any major systems, set *RESIDENCE* to 1.

If the address does not end in a number or a letter and the property owner pays for at least one of the major systems, set *RESIDENCE* to 2.

If at the end of this process *RESIDENCE* is still equaled 6 or 97, an attempt was made to match the observation with a telephone number from the survey. Team members then telephoned the survey respondents to determine the appropriate residence type. If the residence type could not be determined, *RESIDENCE* type was set to missing for the cleaned survey dataset.

The resulting *RESIDENCE* variable *(which was shortened to RES in the final database)* has the following definitions.

- 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

Creating a *RESIDENCE* variable for master metered customers followed a similar, but simplified process. The strata codes for *RESTYPE* for master meter customers are as follows:

- 1 is a 2 to 4 unit duplex, triplex or quadplex,
- 2 is a 5-20 unit multifamily residence,
- 3 is a 20 + unit multifamily residence,
- 4 is a mobile home park.

During the survey process all individuals living in master meter units with a *RESTYPE* of 2-4 received a phone call to insure the identification of the individual. Given that each of these individuals received a call, if *DWLTYPE* and *RESTYPE* differ, the cleaning code for master customers assumes that the *RESTYPE* coding is correct.

The following are the rules for developing the *RESIDENCE* variable for master meter customers.

- If *DWLTYPE* is equal to 1, the individual's survey response is assumed to be incorrect. The master meter file only contains individuals living in multifamily residences. In this situation, the *RESTYPE* code is maintained.

- If *DWLTYPE* is equal to 2, and *RESTYPE* is not equal to 1, the strata *RESTYPE* code is preserved.

- If *DWLTYPE* is equal to 3 or 4, and *RESTYPE* equals 2 or 3, the survey response is maintained, otherwise the strata *RESTYPE* code is preserved.

- If *DWLTYPE* is equal to 5, and *RESTYPE* is not equal to 4, the strata *RESTYPE* code is maintained.

Due to slight inconsistencies between the *RESTYPE* code and the *DWLTYPE* and *RESIDENCE* code, the master meter customers with problems in their *DWLTYPE* variable must have their *RESTYPE* variable transformed. If a master meter customer's *RESTYPE* code was preserved, the following rules were used to transform *RESTYPE* to *RESIDENCE*:

- If *RESTYPE* is equal to 1, *RESIDENCE* is set equal to 2.
- If *RESTYPE* is equal to 2 or 3, *RESIDENCE* is set equal to 4.
- If *RESTYPE* is equal to 4, *RESIDENCE* is set equal to 5.

If there were no problems in master meter customers' survey response to *DWLTYPE*, their survey response is carried over into their *RESIDENCE* variable.

A2 – Own or Rent Home (*OWNRENT*)

Responses to the *OWNRENT* question are unchanged.

- 1 indicates respondent owns or is buying the residence,
- 2 indicates respondent rents or leases the residence, and
- 97 indicates respondent did not answer this question.

A3 – Length of Time at this Address (*YRS_RES*)

Responses for how long have you lived at this residence (*YRS_RES*) and what year was the home built (*BUILTYR*) were cross-referenced. In particular, if the number of years at the address response was longer than the number of years the house has been in existence based on the *BUILTYR* response, then both *YRS_RES* and *BUILTYR* were set to missing.

A4/A5 – Seasonal Occupancy (*SEASOCC*)

The description of your residence (*SEASOCC*) responses were cleaned based on the following criteria.

- If the months the home is typically occupied (*SEASJAN*... *SEASDEC*) totaled two or fewer, then *SEASOCC* was set to 4 (vacation or rental home),

- If the months the home is typically occupied (*SEASJAN*... *SEASDEC*) totaled three or more, then *SEASOCC* was set to 2 (partial year or seasonal residence),

- If the months the home is typically occupied (*SEASJAN... SEASDEC*) was left blank, then *SEASOCC* was set to 1 (year-round residence).

A6 – Year Home Built (*BUILTYR*)

The responses for how long have you lived at this residence (*YRS_RES*) and what year was the home built (*BUILTYR*) were cross-referenced. In particular, if the number of years at the address response was longer than the number of years the house has been in existence based on the *BUILTYR* response, then both *YRS_RES* and *BUILTYR* were set to missing.

A further check of *BUILTYR* was made by comparing the age of the major heating (*HTSYSAGE*) and water heating (*PRWHAGE*) systems to the age of the home based on the response to *BUILTYR*. If the ages of the water heating or space heating systems were greater than the age of the home, then the *BUILTYR* was set to missing.

BUILTYR is the basis for the derivation of the home's age (*AGEHOME*) and the NEWHOME variables which are used in the conditional demand analysis model. The construction of the *AGEHOME* and NEWHOME variables are discussed further in Section 4.3 - CDA Data Filling Process.

A7 – Number of Bedrooms (*NUMROOM*)

The number of bedrooms responses were screened based on the following criteria.

- If *DWLTYPE* equaled 1, 2, 3, or 4, SQFT was less than 2,500, and NUMROOM is greater than 9, then NUMROOM was set to missing.
- If *DWLTYPE* equaled 3 or 4, SQFT was less than 1,500, and NUMROOM is greater than 5, then NUMROOM was set to missing.
- If *DWLTYPE* equaled 5 and NUMROOM was greater than 5, then NUMROOM was set to missing.

A8 – Square Feet of Living Space (*SQFT*)

The responses to the number of square feet of living space in the home (*SQFT*) are unchanged.

The *SQFT* variable is the basis for the number of square feet variable (SQFT_A) variable derived for use in the CDA model. The *SQFT_A* variable is a continuous variable derived from the *SQFT* responses and plugged when the response to *SQFT* is missing. A discussion of the derivation of *SQFT_A* is discussed further in Section 4.3 - CDA Data Filling Process.
.

A16 – Number of Occupants by Age Group (NR0-5, NR6-18, NR19_34, NR35-54, NR55-64, NR65-99)

The responses to this question are unchanged and are used to construct the following variables.

- 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*), and
- Number of people living in the household 19-64 (*ADULTS*).

For the CDA analysis *RESCNT* will be plugged when missing to the new variable *NUMI*. The *NUMI variable* will be discussed later in Section 4.3 - CDA Data Filling Process.

A18 – Natural Gas Availability (*NGSERV*)

The responses to whether natural gas was available were changed if the cleaned data indicated the presence of a natural gas line, but the respondent indicated there was no natural gas service.

A19 – Natural Gas Hookup in the Home (*NGLINE*)

An initial clean of the survey question *NGLINE*, the existence of a natural gas line, was undertaken to improve the accuracy of the database for space heat and water heat fuel choice. If a household does not have a natural gas line, it cannot have natural gas appliances. To insure the accuracy of the *NGLINE* survey question, surveys were matched with billing data. If the household's survey response to *NGLINE* indicated that they did not have a natural gas line, and billing data were available, the response to *NGLINE* was cleaned to agree with the household's billing data.

Other Cleanings

A1 (subset) – Number of Stories (*STORIES*)

The responses to this question were changed if respondents answered the question when the skip coding instructed them to skip the variable.

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

The responses to *ACEILINS* were changed if the respondent indicated that they had no insulation, *ACEILINS* was missing, and they provided a value for *CEILINCH*. For these observations, *ACEILINS* was recoded as "yes."

A14/A15 – **Remodeling (*REMOD*)**

The response to *REMOD* was changed if the individual listed a type of remodel, but *REMOD* was "no" or missing. For these observations, *REMOD* was set to "yes."

4.1.5 Space Heating

Data cleaning in the space heating section of the survey consisted of two processes.

- **Cleaning the Space Heating Survey Responses.** This process cleaned the existing raw survey responses to eliminate multiple responses, survey inconsistencies, and illogical responses. In addition, variables were constructed to indicate the primary and secondary heating fuel and primary and secondary heating system types.

- **Accounting for Fuel Misreporting.** In this process, space heating survey responses were compared to electric and gas billing data and other fuel-related survey responses to determine the consistency of the fuel type survey responses. Corrections were made to the primary system type and primary heating system type variables.

Cleaning Space Heating Survey Responses

The following steps were taken to clean the survey responses in the space heating section of the survey. The steps are listed in the order that they were undertaken during the cleaning process. Note that during the cleaning phase, no fuel switching is performed. The fuel switching analysis comes during the process of creating the variables used in the CDA.

- Survey data were read from the text files and a variable was created for each bubble on the survey form.

- The field representing how the home pays for heating (*PAYHEAT*) was evaluated. If there were multiple answers for this question, then the first answer was taken (i.e., if pay heat was indicated as both "yes" and "no, it is part of my rent," then pay heat is set to "yes").

- The cleaned field indicating if there is natural gas service at the home (*NGLINE*) was checked. If there was no natural gas service at the residence and the heating type indicated that there was natural gas heating, then these fields were set to missing (97).

- If a residence had natural gas, then no propane heating systems were allowed at the home. Any propane heating systems were set to missing.

- If the survey responses indicated that a residence had central heat pump heating and central forced air heating, then the heat pump was determined to be the primary heating source and the central forced air heating response was set to missing.

- If five or more electric space heating sources were indicated at the site, then it was determined that the answers for type of heating were erroneous and all were set to missing.

- If the survey indicated that there are three or more "other" space heating systems at the site, then it was determined that the answers for type of heating were erroneous and all were set to missing.

- The primary heating system was determined by selecting the first system in the list of primary heating systems as presented on the survey. All other primary heating systems were moved to the additional heater column. If no primary heating system was indicated on the survey, then the first additional heating system was moved to the primary heating system survey field.

- If the survey response to *PAYHEAT* was "yes" and there was no system indicated, then all system variables were set to missing.

- If the *PAYHEAT* response was "no" and all system variables were missing, then the system variables were set to not applicable (99).

- If *PAYHEAT* was missing and all of the system variables were missing, then the system variables were coded as missing (97).

- If *PAYHEAT* was missing or "no" and at least one system variable was provided, then a new variable *PAYHEAT1* was created and the system variables were maintained. These residences were then analyzed with their billing data to determine the correct value for *PAYHEAT*.

Once this process was completed, a primary heating fuel indicator variable (*PHTFUEL*) was constructed with the following definitions.

- 1 = natural gas
- 2 = electric

- 3 = bottled gas
- 4 = wood
- 5 = solar
- 6 = other
- 97 = respondent failed to answer question
- 99 = respondent does not pay for heat or does not have a primary heating system.

In addition to this space heating fuel type indicator, a number of primary space heating system type variables were constructed for use in the CDA model and provide a higher level of cleaning. These variables are defined in Section 4.3 - CDA Data Filling Process.

Accounting for Fuel Misreporting for Space Heating

The determination of fuel misreporting for space heating is a manual process. In particular, criteria were specified to identify survey responses with likely fuel misreporting. These survey responses were stored in a worksheet and reviewed. Included in this workbook were electric and gas billing records, if available, and selected fuel-related variables that might help in determining the presence of and appropriate value for the space heating system fuel type.

Following are the criteria used to select the residences that were reviewed for fuel misreporting.

- All surveyed residences where respondents did not indicate if they paid for heating and all surveyed single family residences where the respondent indicated that the heating bill is included in the rent were saved to a worksheet for review. In this worksheet, electric and gas consumption fields were stored along with other appropriate fuel-related variables to help decide what type of system was likely to be present at the residence. If all indicators in the bills showed no sign of the end use, then no change was made. Otherwise, the pay heat field was set to "yes," a flag was set to indicate that a manual fuel switch had been made for the site, and an indication was made in the worksheet as to what fuel type was found at the site. This decision was very generic in nature. Indicators of gas or electric systems were set, and for heating, a determination as to whether it is a room or central system was also made if possible from the other variables available on the survey.

- All surveyed residences that had gas consumption and *PAYHEAT* was "yes" but the respondent indicated that their main heating system was not gas were stored in a worksheet for review. As with the previous criterion, all appropriate variables were also included in the worksheet. The billing data were re-examined to determine if gas or electric systems were present and, if possible, if the system was a room or central system. If this could not be determined, the systems were set to central for gas systems and room for electric systems.

- All surveyed residences that had no gas consumption and where the respondent indicated that there was no gas service, yet indicated the presence of natural gas appliances, were stored to a third worksheet and manually reviewed. Through this inspection it was determined if these appliances should be changed to either electric or possibly propane service.

- All sites where there were responses in the space heating system type (B2) questions but the pay heat or pay water heat was not answered positively (*PAYHEAT1*) were stored to a fourth worksheet and manually reviewed. Using all applicable survey questions and the utility billing data, the team determined whether the resident paid for the use of these systems on their utility bill. If so, the survey variables were changed appropriately to reflect the billing status of the systems. If the systems listed were wood, propane, or solar, billing records provided no additional data. These individuals were assumed to pay for their heat.

The primary heating fuel type indicator variable (*PHTFUEL*) was refined based on this analysis and recorded in a new variable (*PHTFUEL2*). By not overriding the initial *PHTFUEL* variable, the initial primary space heating fuel responses were preserved.

In addition, the primary space heating system types were also refined. This process is discussed in Section 4.3 - CDA Data Filling Process.

All sites where the respondents indicated that they pay for their heat (*PAYHEAT was set equal to 1*) but did not provide information about the system type (B2) were stored on a fifth worksheet and manually reviewed. Using all applicable survey data and billing records, it was determined if the system type was electric, gas, or other. If it was not possible to determine if the system was room or central air conditioning, gas systems were assumed to be central and electric systems were set to room.

4.1.6 Space Cooling

This process cleaned the existing raw survey responses to eliminate multiple responses, survey inconsistencies, and illogical responses. In addition, variables were constructed to indicate the type and number of central and room air conditioning system types. The following steps created the cleaned variables for the space cooling section of the survey.

- Survey data were read from the text files and a variable was created for each bubble on the survey form.

- The field representing how the home pays for central cooling (*PAYCOOL*) was reviewed. If multiple answers were indicated, then the first answer was taken (i.e., if *PAYCOOL* was indicated as "yes" and "no, in rent," then *PAYCOOL* was set to "yes").

- The central heat pump heating variable was tested. If true and the survey indicated the presence of central air conditioning, it was changed to central heat pump.

- If the survey indicated the addition of a central air conditioning unit in the past 12 months, the home is owner occupied, yet there are no central air conditioning units specified, one was added.

- For room air conditioning, the three fields that represent units 1, 2, and 3 were checked for multiple answers. The maximum value was selected from the list of three systems offered on the survey form.

- If the type of room air conditioning field was missing but the age field indicated the presence of a room air conditioning unit, then window/wall air conditioner was set for that unit number.

- If the survey indicated the addition of a room air conditioning unit in the past 12 months but no room air conditioning units were specified, one was added.

- Room air conditioning units were counted by adding units 1, 2, and 3. If this count was greater than zero, then the no room air conditioner indicator was set to false.

- If type of room air conditioner was indicated and age was not, then age was set to missing (97).

- If age of room air conditioner was indicated and type of room air conditioner was not, the type of room air conditioner was set to missing (97).

- If no room air conditioner unit was specified, the variable was set to not applicable (99).

Manual comparison of central air conditioning survey responses and electric consumption data was undertaken on a limited basis. This analysis was limited to surveys whose CDA results and consumption profiles supported the presence of central air. The results of this plugging process are listed in Section 4.3 - CDA Data Filling Process.

4.1.7 Water Heating

Data cleaning in the water heating section of the survey consists of two processes.

- **Cleaning the Water Heating Survey Responses.** This process cleaned the existing raw survey responses to eliminate multiple responses, survey inconsistencies, and illogical responses.

- **Accounting for Fuel Misreporting.** In this process, water heating survey responses were compared to electric and gas billing data and other fuel-related survey responses to determine the consistency of the fuel type survey responses.

Cleaning Water Heating Survey Responses

The following steps were taken to clean the survey responses in the water heating section of the survey. The steps are listed in the order they were undertaken during the cleaning process. Noted that during the initial cleaning phase, fuel misreporting is not analyzed, but is reviewed in Section 4.3 - CDA Data Filling Process.

- Survey data were read from the text files and a variable was created for each bubble on the survey form.

- The field representing how the home pays for heating (*PAYWH*) was evaluated. If there were multiple answers for this question, then the first answer was taken (i.e., if *PAYWH* was indicated as both "yes" and "no, it is part of my rent," then *PAYWH* was set to "yes").

- The cleaned field indicating whether there is natural gas service at the home (*NGLINE*) was checked. If there was no natural gas service is at the residence and the water heating type indicated that there was natural gas water heating, then these fields were set to missing (97).

- If a residence had natural gas, then no propane water heating systems were allowed at the home. Any propane water heating systems were set to missing.

- If five or more water heating sources were indicated at the site, then it was determined that the answers for type of water heating are erroneous and all are set to missing.

- If the survey indicated that there were three or more "other" water heating systems at the site, then it was determined that the answers for type of water heating were erroneous and all were set to missing.

- The primary water heating system was determined by selecting the first system in the list of primary water heating systems as presented on the survey. All other primary water heating systems were moved to the additional water heater column. If no primary water heating system was indicated, then the first additional water heating system was moved to the primary water heating system survey field.

- If the survey response to *PAYWH* was "yes" and no system was indicated, then all system variables were set to missing.

- If the *PAYWH* response was "no" and all system variables were missing, then the system variables were set to not applicable (99).

- If *PAYWH* was missing and all of the system variables were missing, then the system variables were coded as missing (97).

- If *PAYWH* was missing or "no" and at least one system variable was provided, then a new variable *PAYWH1* was created and the system variables were maintained. These residences were then analyzed with their billing data to determine the correct value for *PAYWH*.

- The number of showers/baths taken per day (*SHWRDAY* and *BATHDAY*) was conditionally checked against the cleaned number of residents (*RESCNT*). For the number of showers/baths, two per day per person was the maximum value allowed; responses over that were set to "no response."

Accounting for Fuel Misreporting for Water Heating

Determining fuel misreporting for water heating was a manual process. In particular, criteria were specified to identify survey responses with likely fuel misreporting. The survey responses for these residences were stored in a worksheet and reviewed. Included in this workbook were electric and gas billing records, if available, and selected fuel-related variables that might help in determining the presence of and appropriate value for the water heating system fuel type.

The following criteria were used to select the residences that were reviewed for water heating fuel misreporting.

- All surveyed residences where respondents did not indicate if they paid for water heating (*PAYWH* is missing) and all surveyed single family residences where the respondent indicated that their water heating bill is included in their rent were saved to a worksheet for review. In this worksheet, the electric and gas consumption fields were stored along with other appropriate fuel-related variables to help determine what type of system was likely present at the residence. If all indicators in the bills showed no sign of the end use, then no change was made. Otherwise the pay water heat field was set to "yes," a flag was set to indicate that a manual fuel switch had been made for the site, and an indication was made in the worksheet as to what type of water heating system was found at the site. This decision was very generic in nature. Indicators of gas or electric systems were set.

- All surveyed residences that had gas consumption, where *PAYWH* was "yes," but the respondent indicated that the water heating system was not natural gas were stored in a worksheet for review. As with the previous criterion, all appropriate variables were included in the worksheet. The team re-examined the billing data to determine if gas or electric systems were present.

- All surveyed residences that had no gas consumption, where the respondent indicated that there was no gas service, but indicated the presence of natural gas appliances were stored to a third worksheet and manually reviewed. Through this analysis it was determined if these appliances should be changed to either electric or possibly propane service.

- All sites where there were responses in the system water heating system type (D2) questions but the pay water heat response was missing were stored to a worksheet and manually reviewed. Using all applicable survey questions and the utility billing data, a determination was made on whether the resident paid for the use of these systems on their utility bill. If so, the survey variables were changed appropriately to reflect the billing status of the systems.

- All sites where the resident indicated they paid for their water heat but did not provide information on the water heating system type were stored on a worksheet and manually reviewed. Using all applicable survey questions and utility billing data, it was determined if the water heater was gas, electric, or other.

4.1.8 Laundry

There are six questions (E1-E6) in this section of the survey. The raw responses were cleaned to account for multiple and inconsistent responses.

The consistency checks for clothes washers included the following. The number of loads washed during an average week (sum of *CWHWLD*, *CWWWLD*, *CWCWLD*) was conditionally checked for an out-of-range response against the number of residents (A16), after the number of residents was cleaned. In particular, up to five loads per week per person was assigned as the outer limit of reasonableness; any responses over that value were set to "missing."

Responses to the clothes washer type (*CWTYP*), clothes washer age (*CWAGE*), and the clothes dryer type (*CDTYP*) were used to confirm or override the response to the presence of laundry equipment in the respondent's home (*LNDRYEQP*). In particular, if the respondent described their clothes washer's type or age, it was assumed they had laundry equipment for the private use of the home (*LNDRYEQP = 1*); if not, *LNDRYEQP* was unchanged. Note that *LNDRYEQP* is only overridden if the individual provided system information but did not state that they had laundry equipment in their home.

The consistency checks for clothes dryers included the following. The number of dryer loads during an average week (*DRYLDS*) was conditionally checked for an out-of-range response against the number of residents (A16), after the number of residents was cleaned. In particular, up to five dryer loads per week per person was assigned as the outer limit of reasonableness; any responses over that value were set to "missing."

In addition, for clothes dryers, if natural gas service was determined not present (*NGLINE=0*), but a respondent reported a natural gas clothes dryer (*CDTYP*), the clothes dryer fuel was set to "missing." Finally, manual comparison of clothes dryer fuel types was undertaken on a limited basis. This analysis was limited to surveys whose CDA results and consumption profiles supported the presence of a dryer with an alternative fuel source. The results of this plugging process are listed in Section 4.3 - CDA Data Filling Process.

4.1.9 Food Preparation

There are five questions (F1-F5) in this section of the survey. The raw responses were cleaned to account for multiple and inconsistent responses.

The food preparation section cleaning deals with one possible inconsistency. If natural gas is determined not present (*NGLINE=0*) but the respondent reports a natural gas range or oven, that response (*CKRNTYP* and/or *CKOVTYP*) was set to "no response."

4.1.10 Refrigerators

There are two sets of questions (*REFNUM* and the series of characteristics in G2) in this section of the survey. The raw responses were cleaned to account for multiple and inconsistent responses.

These following are consistency checks for refrigerators. These checks were developed given that the columnar design of the questions related to the characteristics of each refrigerator.

- If a there is a missing response to the characteristics of the first refrigerator, the respondent indicated that they have a first refrigerator, and a response was provided in the second refrigerator column, than this response was assume to apply to the first refrigerator. A similar approach was used for the second refrigerator relative to the third refrigerator. This process leads to a reduction in percentage of first refrigerators with missing values but will lead to an increase in missing values for second and third refrigerators.

- If a respondent indicated a number of refrigerators less than the number of refrigerators for which they provide responses, then the number of refrigerators was increased to be consistent with the characteristics data. For example, if the respondent indicated they had only one refrigerator in *RFNUM* but provided characteristics detail in G2 for two refrigerators, then the response to *RFNUM* was changed to indicate the ownership of two refrigerators.

- If *RFNUM* was larger than the set of refrigerator characteristics provided, the characteristics were set to "missing." The characteristics will be filled during the CDA analysis.

- If *RFNUM* was missing and characteristics were provided in G2, *RFNUM* was set to be consistent with the number of characteristics. If no characteristics were provided, *RFNUM* remains missing.

- If *RF1OTH* was missing and a door style was provided for *RF1STY* (the refrigerator variable with the fewest missing observations), *RF1OTH* was set to zero, indicating that the refrigerator did not have an ice maker. A similar procedure was followed for *RF2OTH* and *RF3OTH*.

Note that none of the data cleans results in a number of refrigerators less than the number indicated in *RFNUM*. The number can increase depending on the amount of information provide in the characteristics question G2. The increase in the number of refrigerators may lead to an increase in the percent or refrigerator characteristics missing after cleaning.

Table 4-3 presents a summary of the percent of missing values for each refrigerator after the preliminary data cleaning was performed. Note these missing values for respondents who indicated that they have a first, second or third refrigerator.

Table 4-3
Missing Refrigerator Number and Characteristics

Refrigerator Questions	Missing Prior to Cleaning	Missing Post Cleaning
G1. How many refrigerators do you have plugged in? (*RFNUM*)	1.2%	0.1%
G2. Refrigerator 1		
Door Style (*RF1STY*)	3.6%	2.3%
Size in Cubic Feet (*RF1SZ*)	7.3%	6.1%
Frost-Free or Manual Defrost (*RF1DEF*)	7.1%	5.9%
Age (*RF1AGE*)	4.4%	3.6%
Other Features (*RF1OTH*)	3.6%	1.7%
G2. Refrigerator 2 (*RFNUM* = 2, 3)		
Door Style (*RF2STY*)	11.9%	16.5%
Size in Cubic Feet (*RF2SZ*)	13.5%	17.9%
Frost-Free or Manual Defrost (*RF2DEF*)	14.5%	19.3%
Age (*RF2AGE*)	11.8%	14.5%
Other Features (*RF2OTH*)	11.9%	16.4%
G2. Refrigerator 3 (*RFNUM* = 3)		
Door Style (*RF3STY*)	26.1%	29.6%
Size in Cubic Feet (*RF3SZ*)	23.5%	26.6%
Frost-Free or Manual Defrost (*RF3DEF*)	28.0%	31.0%
Age (*RF3AGE*)	20.8%	23.2%
Other Features (*RF3OTH*)	26.1%	29.3%

In addition to these checks, algorithms to fill missing values for each of the refrigerators' characteristics were developed. The filled variables were used in the CDA model to develop an engineering estimate of monthly kWh usage. Details of the filling algorithms and the development of the engineering estimate of usage are described in Section 4.3 - CDA Data Filling Process.

4.1.11 Freezers

This section of the survey contains two questions (*FZNUM* and the series of characteristics in H2). The raw responses were cleaned to account for multiple and inconsistent responses.

In addition, a few other consistency checks were used for freezers. These checks were developed given that the columnar design of the questions related to the characteristics of each freezer.

- If a there is a missing response for the characteristics of the first freezer, the respondent indicated that they have a first freezer, and a response is provided in the second freezer column, then this response was assumed to apply to the first freezer.

- If a respondent indicated a number of freezers less than the number of freezers for which they provided responses, then the number of freezers was increased to be consistent with the characteristics data. For example, if the respondent indicated they had only one freezer in *FZNUM* but provided characteristics detail in H2 for two freezers, then the *FZNUM* response was changed to indicate the ownership of two freezers.

- If *FZNUM* was missing and characteristics were provided in H2, *FZNUM* was set to be consistent with the number of characteristics. If no characteristics were provided and *FZNUM* was missing, *FZNUM* was set to zero.

Note that none of the data cleans resulted in a number of freezers less than the number indicated in *FZNUM*. The number could increase depending on the amount of information provide in the characteristics question H2. The possible increase in the number of freezers may lead to an increase in the percent of freezer characteristics missing.

Table 4-4 presents a summary of the percent of missing values for each freezer after the preliminary data cleaning was performed. Note the pre-cleaned missing were for survey respondents whose survey response indicated that they had one or two freezers. The post-cleaned missing were for respondents whose cleaned *FZNUM* was equal to one or two.

Table 4-4
Missing Freezer Number and Characteristics

Freezer Questions	Missing Prior to Cleaning	Missing Post Cleaning
H1. How many freezers to you have plugged in? (*FZNUM*)	4.9%	4.0%
H2. Freezer 1 (*FZNUM* = 1, 2)		
Style (*FZ1STY*)	5.5%	5.6%
Size in Cubic Feet (*FZ1SZ*)	8.2%	8.6%
Age (*FZ1AGE*)	6.5%	6.8%
H2. Freezer 2 (*FZNUM* = 2)		
Door Style (*FZ2STY*)	17.8%	23.9%
Size in Cubic Feet (*FZ2SZ*)	23.8%	27.4%
Age (*FZ2AGE*)	22.7%	26.4%

In addition to these checks, the team developed algorithms to fill missing values for each of the freezers' characteristics. The filled variables were used in the CDA model to develop an engineering estimate of monthly kWh usage. Details of the filling algorithms and the development of the engineering estimate of usage are described in Section 4.3 - CDA Data Filling Process.

4.1.12 Spas and Hot Tubs

There are seven questions (I1-I7) in this section of the survey. The raw responses were cleaned to account for multiple and inconsistent responses.

In addition, the following checks were made.

- If the respondent provided answers to one or more of the spa or hot tub characteristics (I2-I7) and indicated that they did not have the use of a spa or hot tub (*SPTYP*=No spa or hot tub), then the response to the use of a spa or hot tub was set to "yes" (*SPATYP* = Yes, I pay for its energy use).

- Natural gas spa or hot tub heating (*SPHT*) was only permitted if a gas line was present at the residence (*NGLINE* = Yes). If there was no natural gas line to the residence, then *SPHT* was set to "missing."

- Spas and hot tubs were only permitted in households designated as single family, town homes, or mobile homes.

4.1.13 Pools

This section contains six questions (J1-J6). The raw responses were cleaned to account for multiple and inconsistent responses.

In addition, the following checks were made.

- If the respondent provided responses to one or more of the pool characteristics (J2-J6) and indicated that they did not have the use of a pool (*PLTYP*= No pool), then the response to the use of a pool was set to "yes" (*PLTYP* = Yes, I pay for its energy use).

- Natural gas pool heating (*PLHT*) was only permitted if a gas line was present at the residence (*NGLINE* = Yes). If there was no natural gas line at the residence and the cleaned survey response to *PLHT* was natural gas, then *PLHT* was set to "missing."

- If *PLTYP* equaled 1, Yes I have use of a swimming pool at my home and I pay for it, a pool was only permitted in single family homes. If *PLTYP* equaled 1 and the cleaned value of *RESIDENCE* is not single family, then *PLTYP* was set equal to 2 (pool is in the common area and I do not pay for the energy use). Ninety-six pools were changed from *PLTYP* equal to 1 to *PLTYP* equal to 2 due to this restriction.

4.1.14 Entertainment and Technology

There are eight questions (K1-K8) in this section. The raw responses were cleaned to account for multiple and inconsistent responses.

The entertainment and technology section consistency checks primarily deal with tabular formatting in questions relating to the presence of a television and accessories (K1) and use of appliances in the home (K8). In particular, if the responses in the tables were entirely blank, then the responses to the individual questions were coded as "no response." Otherwise, if at least one response was recorded in the table, then the remainder of the "missing" answers were considered "none" or "no," as opposed to "missing." This process leads to the same percentage missing for all entertainment and technology appliances after cleaning. Table 4-5 summarizes the percent of responses with missing values for the variables used to develop the entertainment and technology appliance ownership indicator variables.

In addition to the initial cleaning process, a number of indicator variables were constructed for the CDA analysis. This process involved cleaning the raw survey responses to account for missing values. The derivation of these variables is discussed in Section 4.3 - CDA Data Filling Process.

Table 4-5
Missing Entertainment and Technology Appliances

Entertainment and Technology Question (K1)	Missing Prior to Cleaning	Missing Post Cleaning
Home Theater (*THEATER*)	46.2%	0.9%
Large Screen Televisions (*BSTV*)	44.7%	0.9%
Standard Size Televisions (*CLTV*)	7.0%	0.9%
DVD Player (*DVD*)	31.6%	0.9%
VCR (*VCR*)	16.4%	0.9%
Personal Video Recorder (TiVo)	51.9%	0.9%
Stereo (music)	27.0%	0.9%

4.1.15 Lighting

There are two questions (L1 and L2) in this section of the survey. These questions were subject to the raw cleaning algorithms designed to account for inconsistent and multiple responses.

The lighting section consistency checks primarily deal with tabular formatting in questions relating to the presence of indoor (L1) and outdoor (L2) lighting technologies. In particular, if the responses in the tables were entirely blank, then responses to the individual questions were coded as "missing." Otherwise, if at least one response was recorded in the table, then the remainder of the "missing" answers were considered "none" or "no," as opposed to "missing."

Additionally, outdoor lighting variables used in the CDA model were developed. Development of these variables is discussed in Section 4.3 - CDA Data Filling Process.

4.1.16 Miscellaneous Appliances

There are 12 questions (M1-M12) in the miscellaneous appliances section. The raw responses were cleaned to account for multiple and inconsistent responses.

The miscellaneous appliances section consistency checks primarily deal with tabular formatting in the question relating to the presence of miscellaneous appliances (M1). In particular, if the responses in the tables were entirely blank, then the responses to the individual questions were coded as "no response." Otherwise, if at least one response was recorded in the table, then the remainder of the "missing" answers were considered "none" or "no," as opposed to "missing."

The responses to the addition of appliances in the last 12 months (M9) were also used as a check against the presence of any of the appliances included in this question. That is, if the respondent indicated that any of the covered appliances were added in the last twelve months, but did not indicate the presence of this appliance in the previous sections of the survey, then the responses in the earlier sections were overridden to indicate the presence of the appliance. If the respondent indicated the addition of a central heating system (*CHADD*), it was a single family home, and the cleaned response to *PAYHEAT* did not indicate the household paid for their heat, then *PAYHEAT* was changed to show the household paid for their heat.

If the resident indicated the addition of a microwave oven (*MWADD*) and the cleaned response to the presence of a microwave oven (*MWUSE*) indicated no microwave, then *MWUSE* would be set to 2, a seldom used microwave.

In addition to the initial cleaning process, a number of indicator variables were constructed for the CDA. This process involved cleaning the raw survey responses to account for missing values. Table 4-6 presents the percent of responses with missing values for the variables used to develop the appliance ownership indicator variables for the CDA. The derivation of the variables used in the CDA analysis is discussed in Section 4.3 - CDA Data Filling Process.

Table 4-6
Missing Miscellaneous Appliances

Miscellaneous Appliances (M1)	Missing Prior to Cleaning	Missing Post Cleaning
Portable Fan (*FNPORT*)	11.4%	3.0%
Ceiling Fan (*FNCEIL*)	11.5%	3.0%
Attic Ventilator (*WNDATV*)	22.5%	3.0%
Electric Attic Fan (*FNATTIC*)	21.9%	3.0%
Whole House Fan (*FNWHOLE*)	22.7%	3.0%
Electric Air Cleaner (*AIRCLEAN*)	22.2%	3.0%
Humidifier (*HUM*)	22.5%	3.0%
Dehumidifier (*DEH*)	23.2%	3.0%
Water Purification (*WHPURIFY*)	21.6%	3.0%
Heated Waterbed (*WBED*)	22.9%	3.0%
Electric Blanket (*ELBLNKET*)	19.1%	3.0%
Aquarium (*AQUAR*)	22.2%	3.0%
Trash Compactor (*TRSHCOMP*)	22.0%	3.0%
Sauna – Electric (*SAUNA*)	23.2%	3.0%
Electronic Security System (*SCRTYSYS*)	21.2%	3.0%
Pool or water garden Pump (*POND*)	21.9%	3.0%
Electric Garage Door Opener (*GRGDROPN*)	14.6%	3.0%
Lawn Mower – electric (*LAWNMOWR*)	22.5%	3.0%

4.1.17 Household Information

There are seven questions (N1 - N7) in the household information section of the survey. Table 4-7 shows the percent of missing responses for these questions. Raw responses were maintained for household information variables N3-N7. Differences in the pre and post-cleaning percentages are due to a decline in the number of respondents due to invalid surveys.

Responses to the survey questions concerning second homes in California (*PTHME* (N1), *PTHMELOC*, and *PTHMEUTL* (N2)) were cleaned for logical consistency. If the respondent provided a location (*PTHMELOC*) and/or a utility (*PTHMEUTL*) for their second home, then *PTHME* was set to one. If the respondent stated that they did not have a second home in California (*PTHME* = 2), then both *PTHMELOC* and *PTHMEUTL* were set to not applicable (99). If the respondent did not answer *PTHME*, or stated that they had a second home, and did not provide information about the second homes location (*PTHMELOC*) or utility (*PTHMEUTL*), these variables were set to missing.

Responses to the total household income question were further refined to create a continuous variable and to infer missing values. The resulting variable (*AVGINC*) was used in the CDA analysis. The derivation of *AVGINC* is discussed in Section 4.3 - CDA Data Filling Process.

Table 4-7
Missing Household Information

Household Information	Missing Pre Cleaning	Missing Post Cleaning
N1. Own vacation home (PTHME)	4.1%	3.7%
N2. Location for vacation home (PTHMELOC)	4.9%	4.6%
N2. Electric utility provide for vacation home (PTHMEUTL)	5.1%	4.8%
N3. Highest level of education (*EDUC*)	3.9%	3.5%
N4. Primary spoken language (*ETHNIC*)	2.8%	2.5%
N5. Number of occupants of home disabled (DISABLED)	4.0%	3.7%
N7. Household total annual income (*INCOME*)	14.1%	13.8%

4.2 Billing and Weather Data

This section discusses the development of the billing data that is stored in the RASS billing database. This section includes a description of the billing databases provided by each utility, the calendarization routines employed to standardize the bills, the methods used to clean errors in the billing data, and the merging of the billing data with the survey data.

4.2.1 Billing Databases

The three California IOUs and the Los Angeles Department of Water and Power were responsible for delivering their billing data to the study team. The data transaction requested included billing records for all customers sampled. Southern California Gas Company provided gas consumption data for households that were identified as SCG customers using an account matching process. The following discusses the variables included in the billing databases and the timeframe for various steps in the billing data process.

Pacific Gas & Electric Company

The gas and electric billing data for PG&E customers was provided in two separate data sets. The variables included in the data sets were similar, with the electric file containing the kWh consumption and the gas file containing the therm consumption. Both data sets contained a PG&E control identification number while neither data set contained a premise or customer identification number.

The billing frame for PG&E customers contained information from 41,111 residential electric meters and 29,833 gas meters. PG&E's billing data included the following set of information: an old PG&E identification number (CNTL) that was identical for gas and electric customers, a new PG&E electric identification number, a new PG&E gas identification number, kWh and therm consumption, gas and electric tariffs, and a start and end date for the bills. The electric billing data covered the period October 2001 through June 2003 and the gas billing data covered the period October 2001 through August 2003.

Southern California Edison

The billing frame contains information on 39,276 residences within SCE's territory. SCE's billing data includes the following set of information: a customer number, a premise number, kWh consumption, bill date, number of billing days, tariff, Energy Commission weather zone, and SCE weather stations. The billing data covers the period November 2001 through May of 2003.

San Diego Gas & Electric Company

The SDG&E billing frame contains data for 11,179 residences within SDG&E's territory. The SDG&E gas and electric data was provided in a single data set. The data set had billing data for 11,179 residences with electric usage and 7,063 residences with gas usage. The data included a premise identification number, customer number, customer name, read date, electricity tariff, gas tariff, an indicator that the bill was an estimate or a regular read, bill date, electricity consumption, therm consumption, and the number of billing days in the cycle. The billing data covered the period November 2001 through May 2003.

Los Angles Department of Water & Power

The LADWP billing frame contained data for 9,073 customers within LADWP's territory. The LADWP data included a reference number that was unique to the dwelling, a service number which was used in the first RASS data request, annual consumption, an electricity tariff code, an all electric flag, electricity consumption, a read date, and the number of billing days. The billing data covered the period October 2001 through July 2003.

Southern California Gas Company

The sample frame for the RASS survey is based on information on electric customers from the three electric serving IOUs and LADWP. As such, collecting natural gas billing data for customers served by SoCalGas involved a customer matching procedure between the RASS sample frame data and the SoCalGas customer records. This multi-step procedure is discussed below.

- **Step 1—Identify SoCalGas 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 saved for further analysis. The frequency of cases by utility before and after the merging is as follows:

Table 4-8
Counts of Residences by Utility and by SoCalGas Zip Codes

Utility	Number of Residences in the Sample Frame	Number of Residences with SoCalGas Zip Codes
LADWP	9,073	9,034
PG&E	41,111	2,156
SCE	39,276	35,047
SDG&E	11,179	1,098
Total	100,639	47,335

- **Step 2—Disaggregate Customer Address:** The service address variable in the sample frame was disaggregated into the following pieces:

  Street Number
  Street Number Fraction
  Street Direction
  Street Name

Apartment/Unit Number
Zip code

Code was developed for each utility that created the 6 pieces of the address. These pieces along with the customer name and survey ID were matched against SoCalGas' population data for further analysis.

- **Step 3—Customer Address Merging with SoCalGas Master File:** The merging of sample addresses with SoCalGas master file data to capture account number and rate information involved several phases. The two files were first merged 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, master metered accounts were located along with addresses that may have a missing street direction or different apartment/unit number designation (e.g. D instead of 4). The customer name appearing in the sample frame as well as the SoCalGas master file was utilized 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, and the first 6 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. This phase generated only a few (less than 75) additional matches.

  For the remaining unmatched sample after phases 1 through 3, the final 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.

  There were 8,621 returned RASS respondent surveys in SoCalGas service territory with 7,836 being matched and an account number and rate designation extracted. All of the targeted non-respondent surveys were also matched since the final non-response survey was not complete at the time of data transfer. This led to an additional 1,583 account numbers and rate designations.

- **Step 4—Download SoCalGas Account Number File:** The 9,419 merged records were placed into a SAS transport file and downloaded from the SoCalGas mainframe. Of the 784 un-matched records, 155 indicated

SoCalGas was there gas utility and those records were included for further investigation.

- **Step 5—SoCalGas Billing Data:** The SoCalGas Account Number File was merged with the billing data provided by SoCalGas. The billing data contained information on the addresses associated with the 9,419 RASS respondent and non-respondent surveys. The billing data included an account start data and termination data, a bill start data and a bill end date, a premise identifier and an account number, the gas tariff, and the therm consumption. SoCalGas's billing database included natural gas consumption from October 2001 to October 2003.

4.2.2 Billing Data Calendarization and Weather Data Incorporation

Calendarization of the billing data transformed billing cycle data into monthly data for the five utilities. Minor differences in the original database formats and the variables included in the billing databases led to slight differences in the calendarization routines used for each utility. The following steps were used to calendarize the data.

- Billing histories were merged with the survey identifier, selecting only those bills associated with a survey. This process dramatically reduced the size of the billing databases since the initial billing requests targeted bills from all sampled customers.

- Weather data were merged onto the billing databases using the CEUS climate zones and the meter read end dates from the billing records. If the billing data had both a bill start and a bill end date, weather data were merged on for both the start and the stop dates. Heating and cooling monthly degree days were created using either the start and stop dates or the stop date of the current bill and the stop date of the previous bill.

- If the utility provided a customer identification code, the customer code was checked to determine if the customer identification was constant during the billing period. If there was a change in customers during the billing period, billing records for the final customer were retained and the bills for previous customers were dropped from the billing database[7].

- A daily database was created from the billing cycle data. To create the daily database, the first step was to determine the number of days in the billing period. The length of the billing period was calculated either as the difference between the start date and the end date, or the difference between the end of the previous bill and the current end date.[8] Using the calculated number of billing days, monthly consumption and monthly heating and cooling degree days were divided equally into daily consumption and daily heating and cooling degree days. The daily consumption and degree days were deposited into a data set by their calendar day, month, and year variable that was augmented from the start of the billing period. This process spread the billing data into calendar days.[9]

- The daily database was summed over the calendar months to create a data set with calendar monthly consumption and degree days.

- The calendarized consumption and degree data was normalized to a 30.4-day month. If the billing data contained less than 10 calendar days in the month, the consumption was set to missing.

- The RASS survey data used to generate the CDA models was based on an electric individual metered residential sample. The gas data matched to these data included natural gas master meters. During the calendarization routine, the gas master meter accounts are identified using the gas tariffs. These accounts were given a master meter flag and the residences' therm consumption was set to missing.

4.2.3 Merging SCG Billing Data into Billing Database

After the SCG billing data was calendarized, it was divided into four databases based on the survey's electric utility identification code. The SCG billing data was then merged onto the appropriate electricity billing record based on the survey code and the calendar month. During this process, special attention was paid to the labeling of the heating and cooling degree days associated with the SCG database. The heating and cooling degree days associated with SCG, PG&E gas data, and SDG&E gas data were carefully labeled as gas heating and cooling degree days. This terminology insured that the degree days associated with gas and electricity data would correctly represent the weather conditions during the billing period.

4.2.4 Billing Data Cleaning

The consumption data was derived directly from the utility billing files. Billing records, while reasonably accurate, contained some anomalies that can be very troublesome in the application of conditional demand analysis. Billing records were inspected closely for the following problems:

- Erroneous billing days and/or read dates.
- Abnormal monthly consumption.
- Missing or zero electricity usage (the latter may indicate an inactive account).

These errors were corrected, or the observation's consumption was set equal to missing. To limit problems with short billing months that were a result of the calendarization routine, the first and last calendar month for each billing record were deleted.

During the cleaning process close attention was paid to PG&E's billing data. During the survey period, PG&E changed customer identification numbers. The new identification process created two issues. First, examination of the billing records indicated that several customers had a missing bill during the change-over period. These bills were identified, given a missing kWh or therm value, and assigned the start and stop date associated with the missing timeframe. Second, the change of identification numbers made it more difficult for PG&E to correctly match gas and electric records for a given residence. Unfortunately, the final billing data does not have gas records for some PG&E customers who receive both electric and gas service from PG&E.[10]

Electric bills are provided in the final survey and CDA database in annualized summary format to allow for analysis using the final billing values. The electric annual pre-cleaned value is *ELEMN12*. The cleaned annualize electric usage is *ELEMNCDA*. The corresponding pre-cleaned annual gas usage is *THMMN12* and the cleaned value is *THMMNCDA*

4.2.5 Integrating the Cleaned RASS Data with Billing Data

In order to run the CDA, the billing data was merged with the survey data using the individual identification code SFCODE. The resulting database contains 21,153 unique individuals with 365,864 individual monthly observations (Table 4-9). If an individual has both a gas and an electric bill, the consumption and monthly weather information for both bills were contained on a single monthly observation.

Table 4-9
Summary of Billing Data Availability

Utility	Individual Monthly Observations	Unique Individuals	Average Number of Monthly Observations
PG&E	172,982	9,265	19
SDG&E	40,878	2,527	16
SCE	126,818	7,979	16
LADWP	25,186	1,382	18

4.3 CDA Data Filling Process

This section describes the data filling used for variables included in the CDA. The data filling consists of four processes:

- Filling and estimating missing values
- Creating new fuel variables
- Creating indicator variables and continuous variables
- Creating refrigerator and freezer usage variables

It is important to recall that the RASS study included both individually metered and master metered customer results. Because of the need to tie responses to a customer bill, only individually metered customers were included in the CDA model development process and the UEC simulation process.

The following data filling process only includes survey responses from the individually metered electric customer frame.

4.3.1 Filling Missing Values

The cleaned data from the RASS Survey contained a substantial number of missing values. Simply allowing these missing values to disqualify an observation from the CDA would create non-response bias in the estimation of model parameters. Replacing these missing values with overall means for the variables in question would also lead to biased estimates insofar as question-specific non-respondents tend to be different from respondents. In order to minimize non-response bias, a multi-step approach was used and can be seen in Figure 1.

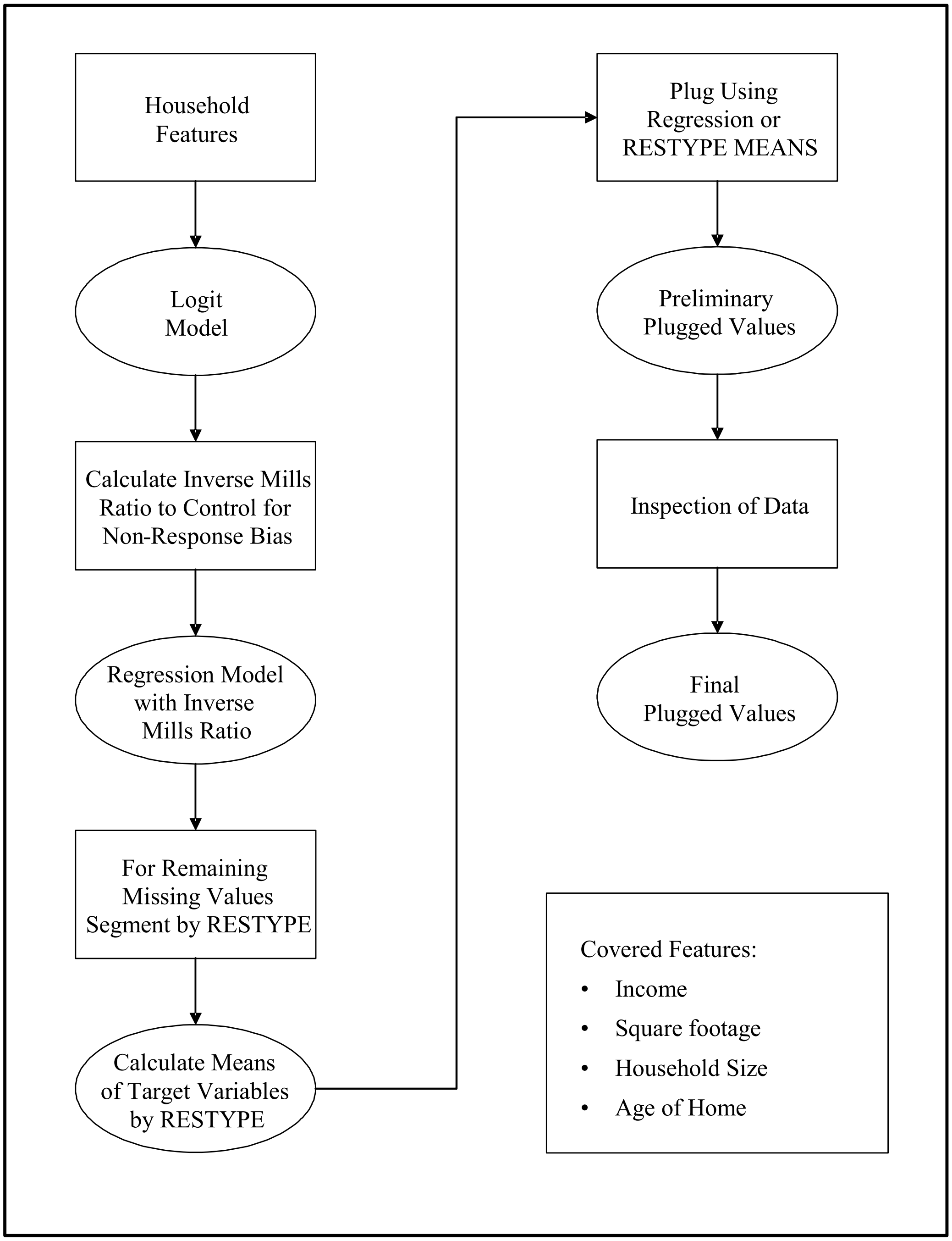
- First, a set of logit equations, each explaining the likelihood of responding to a specific question, was estimated. Once estimated, these equations were used to calculate an inverse Mills' ratio.

- Second, a regression model was used to calculate the predicted value of the missing variable. The inverse Mills' ratio was used in this regression model, as an independent variable, to control for non-response bias.

- Third, remaining missing responses were replaced with means drawn from the specific housing segment into which the household in question falls.

- The following survey elements were covered by this "plugging" routine: income, square footage, household size, and age of home.

Figure 4-1
Filling Missing Values



Square Footage and Surface Area

The survey variable *SQFT* was transformed into the continuous variable *SQFT_A* using the mid-point of the survey range for the variable *SQFT* for all but the three following values.

- If the respondent indicated that their residence was less than 250 square feet, they were assign a value of 200.
- If the survey response was 4001-5000 square feet, they were assigned a value of 4700.
- If the survey response was greater than 5000 square feet, they were assigned a value of 6000.

If *SQFT* was missing, the value was filled using the conditional means process described above.

The CDA requires a measure of residence surface area, *AREA*. Surface area was calculated using estimates of the relationship between surface area and square footage which were created using data collected for the Statewide Residential New Construction Energy Efficiency Baseline Study, Second-Year Report.[11] The relationship was estimated for single story, single family residences, multi story single family residences, and multi-family residences. Mobile homes were grouped with single story single family residences. The following equations list the relationship between surface area and *SQFT_A* for each residence type.

- For single story, single family residences and mobile homes

$$surface\ area = 5.9985 * SQFT\_A^{0.8528}$$

- For multi story, single family residences

$$surface\ area = 13.9694 * SQFT\_A^{0.7395}$$

- For multi-family residences

$$surface\ area = 0.5955 * SQFT\_A^{1.1034}$$

Household Income

The survey variable *INCOME* was transformed to the continuous variable *AVGINC* by using the mid-point of the survey range for all but the upper most value. For respondents who indicated that their household income was $150,000 or more, *AVGINC* was set to $175,000. If *INCOME* was missing, *AVGINC* was filled using the conditional means process described above.

Home Age

The survey variable *BUILTYR* was transformed into the year the home was built. Homes built between prior to 1940 were assign a built year of 1935, homes built between 1940 and 1949 were assigned 1945, those built between 1950 and 1959 were assigned 1955, and those built between 1960 and 1969 were assigned 1965. The year the home was built was used to create the variable *HOMEAGE* where *HOMEAGE* = 2003 - *BUILTYR*. If *BUILTYR* was missing, *HOMEAGE* was filled using the conditional means process described above. In order to facilitate comparison across housing ages, a new home variable was created using *HOMEAGE*. If *HOMEAGE* indicated that the home had been built between 1997 and 2003, the home was flagged as a new home where *NEWHOME* was equal to one, zero otherwise.

Number of Household Residents

The resident count variables (A16) were summed to create a count of the number of people in the household, *RESCNT*. For the CDA analysis, a new variable was created to represent the number of people in the household, NUMI. If RESCNT was non-missing, NUMI was set equal to RESCNT. If *RESCNT* was missing, NUMI was filled using the conditional means process described above. In the CDA, the number of people in the household is included as the log transformation of NUMI (labeled there as NHH: NHH = log (NUMI +1)).

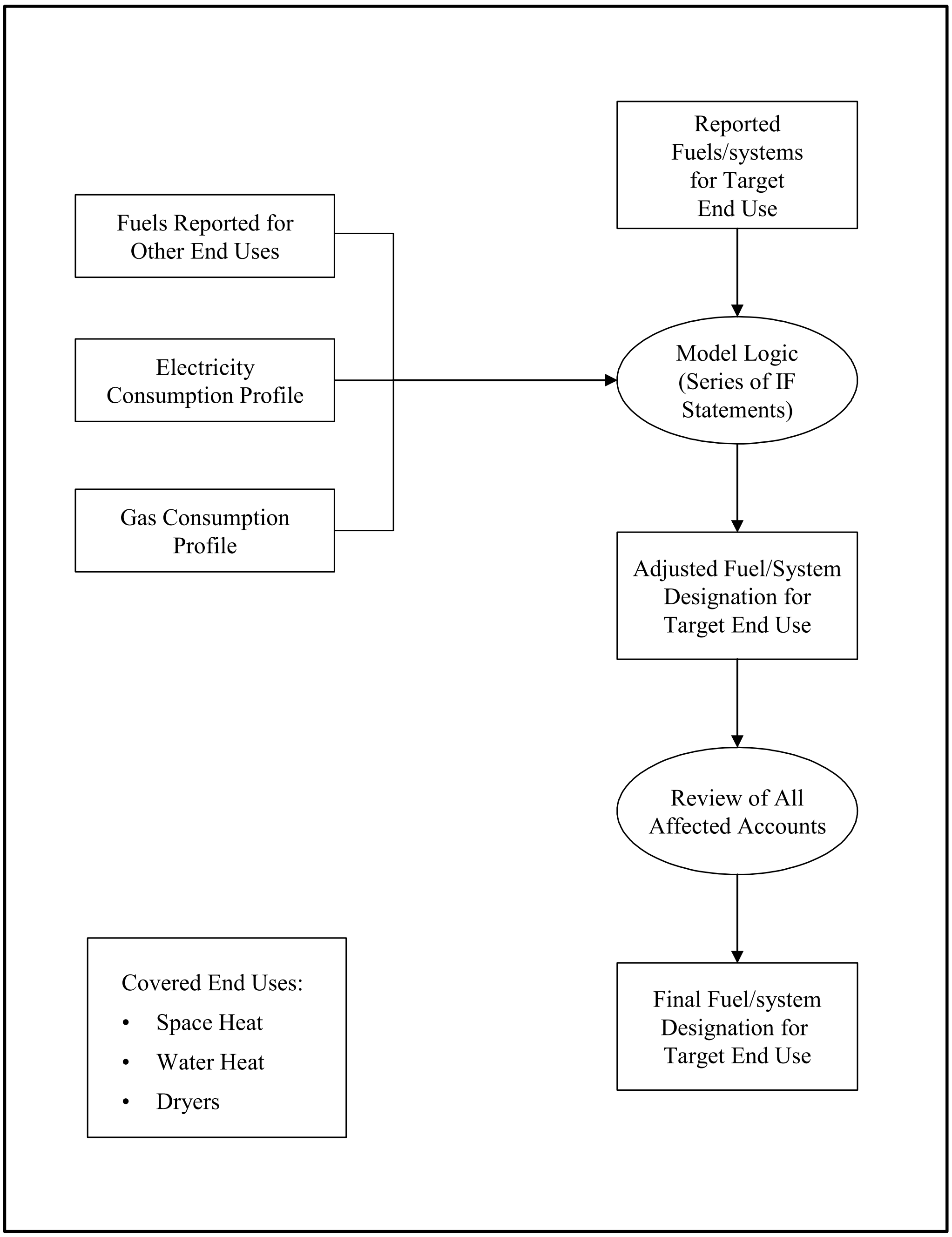
4.3.2 Creating New Fuel Variables

It is common knowledge that survey respondents often misreport fuels and system types. This type of misreporting can be troublesome for the process of disentangling end use consumption levels through the CDA modeling process. As a result, a considerable amount of care was taken to validate (and, where justified, to override) reported systems and fuels. As illustrated in Figure 2, and discussed in the data cleaning portion of this section, this process involved two steps:

- First, a series of logical overrides was developed. These overrides essentially checked the consistency of reported fuels/systems with other information and overrode responses when inconsistencies were found. For instance, if natural gas was reported as the heating fuel but no gas service was available, the household was assumed to have another heating fuel. If the electricity consumption profile exhibited enough seasonality to suggest (with a reasonable absence of ambiguity) the presence of electric space heating, the fuel type was set equal to electric; otherwise, it was left missing.

- Once the overrides were affected, all account information was printed out for all households affected by this process, and the correspondence between observed seasonality and imposed fuel types was double-checked.

Figure 4-2
Correcting Fuel Misreporting



Heating Fuel Variables

During the fuel cleaning process, the survey variable *PAYHEAT* and the summary variable representing the home's primary heating fuel (*PHTFUEL*), were updated through the creation of new variables *PAYHEAT2* and *PHTFUEL2*. The survey variables were maintained for comparison purposes. Table 4-10 lists the percentage of respondents who pay for their heat. Missing values are included in the survey response *PAYHEAT*. The fuel cleaned variable, *PAYHEAT2*, has all missing values plugged. If the rate tariff information indicated that the residence is on a master meter gas account, and the residence has gas heat, *PAYHEAT2* has been set to zero. If the survey response lists wood as their primary heating fuel, and they indicate that they do not have a system, *PAYHEAT2* was set equal to one.

Table 4-10
Heating Payment Question Cleaning

Do you pay to heat your home?	*PAYHEAT*	*PAYHEAT2*
Yes	90.9%	94.8%
No, it is part of my rent	2.7%	3.9%
No, do not have a system	3.2%	0.7%
Missing	3.2%	-
Master Meter Gas	-	0.6%

Table 4-11 lists the percentage of homes with alternative primary heating fuels, conditional on the residence paying for their heat (*PAYHEAT* was set equal to one or *PAYHEAT2* was set equal to one) or on the presence of a gas master meter (*PAYHEAT2* was set equal to zero). If the home did not pay for heat, the primary heating fuel is set to not applicable (99).[12] During the fuel cleaning process, approximately 2.5% of the sample was switched into natural gas, and 2% were switched out of electric. Following the fuel cleaning process no primary heating fuels are listed as missing.

The CDA contains a gas and an electric heating fuel indicator variable, *DGHEAT* and *DEHEAT*, respectively. *DGHEAT* was set equal to one if the residence pays for their heat (*PAYHEAT2* was set equal to one) and the primary heating fuel is natural gas (*PHTFUEL2* was set equal to one). *DEHEAT* was set equal to one if the residence pays for their heat and primary heating fuel is electric (*PHTFUEL2* was set equal to two).

Table 4-11
Primary Heating Fuel Cleaning

Primary Heating Fuel	*PHTFUEL*	*PHTFUEL2*
Natural Gas	73.2%	76.9%
Electric	13.3%	11.3%
Bottled Gas	4.0%	4.4%
Wood	2.9%	2.5%
Solar	0.03%	0.01%
Other	0.4%	0.3%
Missing	2.8%	-
Not Applicable	3.5%	4.6%

The CDA model also accounts for the presence of backup electric and gas heaters. If the household has a primary electric heater and a non-electric backup, *NONELEBK* is set to one, zero otherwise. If the household has a primary electric heater and an additional electric heater, *DAUXHT* is set to one, zero otherwise. If the household has a primary gas heater and a non-gas backup, *NONGBU* is set to one in the gas CDA model, zero otherwise. If the household has a primary natural gas heater and an additional natural gas heater, *DAUXHT* was set equal to one, zero otherwise.

Room heat and central heat consume different quantities of energy. To allow the CDA to estimate different consumption patterns, indicators of room heat were developed. If the residence had gas heat, and the primary heater was a floor or wall furnace, *GROOM* was set equal to one, zero otherwise. If the residence had electric heat, and the primary heater was a resistance heater, a through the wall heat pump, or a portable heater, *ROOM* was set equal to one, zero otherwise.

The CDA also contained variables for the average thermostat temperature, *HTTSET*, and an indicator variable to account for thermostat setback. *HTTSET* represents a weighted average of the residence's survey responses to B6, the average thermostat temperature for each time period during the heating season. If the household did not have thermostat or if they left B6 blank, means were assigned by fuel type and household type(*single-family, multi-family, and mobile home)*. *SETBK* is the heater thermostat setback variable. If the households nighttime heater setting (*HNITESET*) was lower than their average setting (*HTTSET*), *SETBK* was set to one, zero otherwise.

PHTFUEL3 was a variable that was added for the final dataset. It provides a combination of *PHTFUEL2* for all respondents who were included in the CDA modeling process. It adds in the previously established *PHTFUEL* variable for those who were not part of the CDA (primarily electrically master metered customers and dwellings with problematic or insufficient billing data). *PHTFUEL3* provides a single variable to report final heating fuel for all study participants.

Water Heating Fuel Variables

During the fuel cleaning process, the survey variable *PAYWH* and the summary variable representing the home's primary water heating fuel (*PWHFUEL*), were updated through the creation of new variables *PAYWH2* and *PWHFUEL2*. The survey variables were maintained for comparison purposes. Table 4-12 lists the percentage of respondents who pay for their water heating. Missing values are included in the survey response *PAYWH*. If the rate tariff information indicates that the residence is on a master meter gas account, and the residence has gas water heat, *PAYWH2* was set to zero. During the fuel cleaning process, *PAYWH2* increased by approximate 7.5 percentage points. This increase was fairly evenly split between those who initially stated that their water heater was included in their rent and those who did not respond to the *PAYWH* survey question.

Table 4-12
Water Heating Payment Question Cleaning

Do you pay for water heat?	*PAYWH*	*PAYWH2*
Yes	84.0%	91.5%
No, it is part of my rent	11.8%	7.6%
No, do not have a system	1.2%	0.2%
Missing	3.0%	-
Master Meter Gas	-	0.7%

Table 4-13 lists the percentage of homes with alternative primary water heating fuels, conditional on the residence paying for their water heat (*PAYWH* was set equal to one or *PAYWH2* was set equal to one) or on the presence of a gas master meter (*PAYWH2* was set equal to zero). If the home did not pay for water heat, the primary water heating fuel is set to not applicable (99).[13] The cleaned variable *PWHFUEL2* includes the plugged water heating variable for the CDA. During the fuel cleaning process, approximately 4.5% of the sample was switched into natural gas, 1% was switched out of electric and 3.4% was switched outing of missing. Following the fuel cleaning process there are no missing primary water heating fuels.

Table 4-13
Primary Water Heating Fuel Cleaning

Primary Water Heating Fuel	*PWHFUEL*	*PWHFUEL2*
Natural Gas	72.6%	78.1%
Electric	9.6%	8.8%
Propane	4.7%	5.1%
Solar	0.03%	0.01%
Other	0.1%	0.1%
Missing	3.4%	-
Not Applicable	9.6%	7.9%

The CDA contains a gas, an electric, and a solar water heating fuel indicator variable, *DGWH, DEWH, DGWHSOLAR*, and *DWHSOLAR* respectively. Note, a solar water heater with an electric backup system will have two indicator variables set equal to one, *DEWH* and *DWHSOLAR*. A solar water heater with a natural gas backup system will also have two indicator variables set equal to one, *DGWH* and *DGWHSOLAR*. In addition, if the residence indicated that they had more than one electric water heater, the indicator variable *ADDWHEL* was set equal to one, zero otherwise.

PWHFUEL3 was a variable that was added for the final dataset. It provides a combination of *PWHFUEL2* for all respondents who were included in the CDA modeling process. It adds in the previously established *PWHFUEL* variable for those who were not part of the CDA (primarily electrically master metered customers and dwellings with problematic or insufficient billing data). *PWHFUEL3* provides a single variable to report final heating fuel for all study participants.

Dryers

During the fuel cleaning process, the survey dryer variable *CDTYP*, was updated with the creation of two new variables *GDRY* and *EDRY*. The survey variables were maintained for comparison purposes. Table 4-14 lists the percentage of survey responses for alternative dryer types. The table also contains the filled values for *GDRY* and *EDRY*. *GDRY* and *EDRY* are simple indicator (0-one) variables. After cleaning, there was an increase in both gas and electric dryers. This increase comes primarily from households with missing values for *CDTYP*.

Table 4-14
Clothes Dryer Fuels

Type of Dryer	CDTYPE	GDRY	EDRY
No Clothes Dryer	8.8%	-	-
Natural Gas	37.8%	38.8%	-
Electric	31.2%	-	32.6%
Bottled Gas	1.7%	-	-
Missing	2.3%	-	-
Not Appropriate	18.2%	-	-
Not Gas	-	61.2%	-
Not Electric	-	-	67.5%

Central Air Conditioners

During the fuel cleaning process, the central air conditioning survey variable *PAYCOOL*, was updated through the creation of *DCAC*, a zero-one indicator variable indicating the presence of a central air conditioner. The survey variables were maintained for comparison purposes. Table 4-15 lists the percentage of survey responses for *PAYCOOL*. The table also contains the filled values for *DCAC*. After cleaning, there was a very small increase in the percentage of households with air conditioning.

Table 4-15
Central Air Conditioning Payment

Do you pay for central air?	PAYCOOL	DCAC
Yes	45.0%	45.5%
No, it is part of my rent	2.5%	-
No, do not have central air	46.6%	-
Missing	6.0%	-
No, part of rent or no central air	-	54.5%

The CDA also contains a variable for the average central air conditioner thermostat temperature, *TSETC*. *TSETC* represents a weighted average of the residence's survey responses to C5, the average thermostat temperature for each time period during the cooling season. If the household had central air, and they did not have thermostat or they did not respond to the thermostat question, temperature means were assigned *for single-family, multi-family, and mobile home households* .

4.3.3 Creating Indicator and Continuous Variables

Many of the variables needed in the CDA require the creation of indicator variables (values of one, two, or three) or continuous variables (often taken as the mean of a range variable).

Housing Indicator

The *RESIDENCE* variable was used to create the CDA housing variable *RESTYPE1* . *RESTYPE1* reduces the five choice *RESIDENCE* variable to three residence types: single family, multifamily and mobile home. After using the data cleaning algorithm discussed above, a very small percentage of the responses to *RESIDENCE* remained missing. Examination of the responses to these surveys indicated that they most resembled single family residences. These households were coded as single family for the *RESTYPE1* variable. Table 4-16 summarizes the percentage of survey household in each category.

Table 4-16
Cleaned and Filled responses to type of residence building

Description	Residence	Restype1
Single Family	65.3%	65.4%
Multifamily: Townhouse	8.4%	29.99
Multifamily: 2-4 unit apartment	7.6%	
Multifamily: 5+ unit apartment	16.0%	
Mobile Home	2.7%	2.7%
Missing	0.1%	-

Seasonal Home Indicator

RASS question A4 asked respondents if their current residence was a seasonal home. Seasonal homes may use substantially less energy than year-round residences. The responses to the seasonal questions A4 and A5 were combined to create a seasonal home indicator variable. If the respondent indicated that the home was a seasonal residence, vacation home, or a vacation rental home, and they indicated that they did not always occupy the home, the *SEASONAL* variable was set to one, zero otherwise.

Double Pane Windows and an Indicator for Homes in Colder Zones

Homes in colder climate zones are often expected to use more energy for heating. These same homes, hoWever, are frequently built with more insulation, double pane windows, and tighter window frames, reducing the impact of weather on energy usage. Prescriptive building requirements applying to CEUS weather zones 1, 161 and 162 suggest that new homes be built to withstand their colder climate. For the CDA, we created a Title 24 variable (T24) and set it to one if the home was located in zones 1, 161, or 162, zero otherwise. We also created an indicator of double pane windows. *DPWIN* was set equal to one if *WINDTYPE* was set equal to one (all or most double) or 2 (mixture of double and single), zero otherwise.

Fans

The CDA includes three types of fans; forced air fans, attic fans, and ceiling fans. Forced air fans are associated with central natural gas and central bottle gas furnaces. If *PHTFUEL2* was set equal to one or 3, and the heater is a central heater, than *DFFAN* was set equal to one, zero otherwise. Attic fans are used to cool the residence during the summer months. If the residence had an attic or a whole house fan, *DATTFAN* was set equal to one, zero otherwise. Ceiling fans were included in the electric CDA, in the miscellaneous term. *DCEILF* was set equal to one if the residence had at least one ceiling fan.

Room Air Conditioners

An indicator variable was created to indicate the presence of room air conditioners (*DRAC*) and a count variable was created to list the number of room air conditioners in the residence (*RACCNT*). The CDA also controls for room air conditioner usage. Using the responses to C8, room air conditioner usage, a usage variable *TSETUSE* was created. If the residence had a room air conditioner and C8 was missing, *TSETUSE* was filled using the mean by *RESTYPE1*.

Water Heater Usage

Energy usage for water heaters depends largely on the other systems in the residence, the number of individuals in the household, and the differential between the inflow water temperature and the desired temperature. The CDA model accounts for the other systems in the water heater usage analysis. *DWASHU* and *CWASHU* account for the water heater energy usage to run the dishwasher and the clothes washer. Both of these variables are usage variables, created respectively using the survey responses to F5, presence of a dishwasher, and E4, laundry load usage by load temperature. If households with the specified system did not answer the frequency of use questions, *DWASHU* and *CWASHU* were filled using the mean by *RESTYPE1*.

In addition, the number of baths and showers taken in a typical day impacts the hot water heaters energy usage. *WHTSHWRS* is a count on the total number of baths and showers taken in the home on the typical day. If survey respondent did not respond to questions D6, number of baths and showers, *WHTSHWRS* was filled with the mean by *RESTYPE1*.

To account for the differential between the water inflow temperature and the desired water temperature, an inflow temperature was constructed. The constructed inflow temperature was a weighted moving average of the outside temperature during the previous two months. The water heater temperature differential, *WHTEMP_DIFF*, was created as the difference between the residences typical water heater setting (*WHTEMP*) and the constructed inflow temperature. If the household did not respond to *WHTEMP*, the WHTEMP variable was filled with the mean by *RESTYPE1* .

Kitchen Appliances

The kitchen appliances included in the CDA were ovens and ranges, microwave ovens, and dishwashers. If the residence indicated that they had an electric range or oven *DERGOV* was set equal to 1, zero otherwise. If the range or oven was natural gas, *DGRGOV* was set equal to one. If the residence had a microwave oven, *DMWV* was set to one, zero otherwise. The presence of a dishwasher was captured by the indicator variable DDW.

Laundry

The CDA included gas and electric dryers and electric clothes washers. If the household had a natural gas dryer or an electric dryer, *DGDRY* or *DEDRY* were set to one, respectively. The weekly usage of the dryer was captured by the variables *GDRYU* or *EDRYU*. If the survey response to *DRYLDS* was missing, and the residence had a dryer, *GDRYU* or *EDRYU* was filled using the mean by *RESTYPE1*.

The indicator variable for electric clothes washers was set to one if the residence indicated that they use laundry equipment in their home (*LNDRYEQP*) and they chose either a top loading or a front loading washer (*CWTYP*). The variable indicating the presence of a clothes washer was *DCW*.

Outdoor Lighting

The CDA estimated the energy usage of outdoor lighting using information on both the presence of outdoor lights and the types of fixtures and bulbs. Survey question L2 asked the resident the number and type of fixtures on the outside of their home (Exterior Fixtures). The responses to this question were expanded to continuous variables and then summed to determine the total number of exterior fixtures (*OLTFIX*). If the sum of exterior fixtures was greater than zero, *DOLT* was set equal to one, zero otherwise. If the respondent left the entire series of questions on exterior fixtures missing, they were assumed to have no outdoor lighting.

Outdoor fixtures often employ bulbs that use less energy than incandescent bulbs and outdoor lighting controls are also common. Variables were created to represent the proportion of outdoor fixtures containing compact fluorescents (*ONOCFL*), the proportion on timers (*OPROPTIM*), and the proportion on motion detector or dusk to dawn sensors (*OPROPSENS*).

Televisions

The energy usage of televisions is a function of the number of televisions, the size of the televisions, and the total hours of usage. The RASS Survey questioned respondents about the number of large screen televisions, the number of standard size televisions, and the total number of hours of television usage per day. If the household had either a conventional screen TV or a big screen TV, DTV was set equal to one, zero otherwise. The number of hours of usage was *TVHRS*.

Big screen and conventional televisions use different quantities of electricity for a set number of hours of usage. The CDA accounts for the differential electricity usage with the variable *TVKW*. *TVKW* was set to 0.1 kWh per hour for conventional televisions and 0.25 per hour for big screen televisions. If the individual had both big and conventional television, the usage numbers were multiplied by the proportion of TVs of that type.

Personal Computers and Home Offices

The dramatic growth in the number of personal computers and the proportion of the population working from home, led to the inclusion of these end-uses in the CDA. If the respondent used a personal computer in their home, *DPC* was set equal to one, zero otherwise. Both the hours of usage and the number of PCs were included. PCs are often not turned off and are commonly connected to modems for Internet usage. *PCNUM* indicated the number of PCs and *PCHRS* represented the total number of hours the PCs were turned on each day.

The home office survey questions were asked separately from the PC questions. The number of PCs represented the total number in the residence. The home office energy usage represented the additional energy associated with running an office from home. *DHMOFF* was set equal to one if the individual indicated that someone in the home operated a business or worked from home. *HMOFFHRS* indicated the numbers of hours a week someone works out of the home.

Pools

If the respondent had a pool at their home and they pay for its energy use, *DPLPMP* was set equal to one. Only individuals living in single family residences were allowed to have pools. All other pools listed in the survey were assume to be pools located in common areas, and were disallowed in the CDA.[14] The number of hours per day used to filter the pool was captured by the variable *PLFILT*. This variable differs between summer months (May-October) and winter months (November-April). The pool size variable (*PLSIZE*) was set to 18,000 gallons for small pools, 30,000 for medium sized pools, and 42,000 for large pools.

If the pool was heated with electric heat, *EPLHT* was set to one, zero otherwise. If the pool was heated with natural gas, *DGPLHT* was set to one. The gas CDA also analyzed the impact of the frequency of pool heating, *GPLHTFREQ*. This variable was allowed to differ between summer and winter months. *PLCOV* indicates that the household used a pool cover. A pool cover may reduce the heating needs due to an increase in pool temperature or it may indicate a pool that was used more frequently, leading to an increase in heating needs.

Spas and Hot Tubs

If the respondent had a spa or hot tub at their home and they paid for its energy use, *DSPA* was set equal to one. Only individuals living in single family residences, town houses, or mobile homes were allowed to have spas and/or hot tubs. If the spa was heated with electricity or solar with electric backup, *DEHTSPA* was set to one. Spas heated by solar with electric backup also received an additional indicator variable, *SPASOLAR*. If the spa was heated with natural gas or solar with natural gas backup, *DGHTSPA* was set to one. Spas heated by solar with natural gas backup received an additional indicator variable, *SPAGSOLAR*.

The frequency of spa filtering (*SPAFREQ*) and electric (*SPAEHTFREQ*) and natural gas (*SPAGHTFREQ*) heating were allowed to differ between summer and winter months. The spa size variable was based on the number of people the spa holds. If the spa was small *SPASIZE* was set to 2, medium spas were set for 5 people, and large size spas for 8. *SPCOV* was set to one if the spa had an insulated cover.

4.3.4 Energy Usage for Refrigerators and Freezers

CDA models have difficulty accurately estimating end-uses with near 100% saturation. To improve the accuracy of refrigerator and freezer UECs, engineering estimates of refrigerator and freezer energy usage for each household were calculated.

Engineering Estimates of Refrigerator Energy Usage

The information collected about the refrigerators in the home was relatively extensive. The survey asked the age of the refrigerator, the door style (i.e., single-door, top freezer - bottom refrigerator, top refrigerator - bottom freezer, or side-by-side), whether the refrigerator was Frost Free or Manual Defrost, the size of the refrigerator (i.e., Mini (< 13 cu ft), Small (13 to 16 cu ft), Medium (17 to 19 cu ft), Large (20 to 23 cu ft), or Very Large (> 23 cu ft)), and whether the refrigerator had a through-the-door ice and water dispenser.

To formulate an engineering estimate of refrigerator energy usage, all of the refrigerator survey questions had to have non-missing responses. If the household did not respond to all of the refrigerator characteristic questions (G2), these variables were filled in order to facilitate engineering modeling. The following is the list of algorithms used to fill the refrigerator characteristics.

- If the survey did not contain information on the door style, the most common door style for the *RESTYPE1* , *RFNUM* combination was assigned.

- If the survey did not contain information on through-the-door ice, this was assigned based on door style. If the refrigerator was a side-by-side unit, *RFOTH* was set to one (ice), otherwise *RFOTH* was set to zero (no ice).

- If the survey did not contain information on the age of the refrigerator, it was filled with the mean age by *RESTYPE1* , *RFSTY*, and *RFNUM*.

- If the survey did not contain information on the size of the refrigerator, it was filled with the mean size by *RESTYPE1* , *RFSTY*, and *RFNUM*.

- If the survey did not contain information on the defrost style, it was assumed that the refrigerator was an automatic defrost unit unless the age of the refrigerator was greater than 15 years. Additionally, we assumed that all side-by-side refrigerators were automatic defrost units.

The Association of Home Appliance Manufacturers (AHAM) website (www.aham.org) contains historic refrigerator usage data by size and type of unit. These data were compiled to estimate annual usage, controlling for door style, adjusted volume, defrost, and automatic icemakers. Using the parameters calculated from the AHAM data, and the respondent's refrigerator characteristics, base engineering estimates of refrigerator energy usage were calculated for first and second refrigerators. The base engineering estimates were then calibrated using two adjustment factors. The first adjustment factor was determined by AHAM. It accounts for improvements in energy usage per cubic foot through time. This factor allows us to explicitly adjust energy usage for the age of the refrigerator beyond the age range available in the larger AHAM dataset. The second adjustment factor helps to calibrate the engineering estimate of usage to differences in AHAM published data and our simulation model. The resulting engineering estimates of refrigerator energy usage were *REFUSAGE1* and *REFUSAGE2*.

Engineering Estimates of Freezer Usage

The information collected on freezers was also extensive. The survey asked the age of the freezer, whether the freezer was Frost Free or Manual Defrost, and the size of the freezer (i.e., Small (< 13 cu ft), Medium (13 to 16 cu ft), or Large (> 16 cu ft).

To formulate an engineering estimate of freezer energy usage, all of the freezer characteristics had to have non-missing responses. If the household did not respond to all of the freezer characteristic questions (H2), these variables were filled. The following is the list of algorithms used to fill the freezer characteristics.

- If the respondent did not provide information on the freezer's style, the most prevalent style was assigned by *RESTYPE1* and *FZNUM*.
- If the respondent did not provide information on the freezer's age, the mean was assigned by *RESTYPE1*, *FZNUM*, and door style.
- If the respondent did not provide information on the freezer's size, the mean was assigned by *RESTYPE1*, *FZNUM*, and door style.

The AHAM freezer data were compiled to estimate annual usage, controlling for volume and defrost style. Using the parameters calculated from the AHAM data, and the respondent's freezer characteristics, base engineering estimates of freezer energy usage were calculated for first freezers. The base engineering estimates were then calibrated using two adjustment factors which were similar to the refrigerator adjustment factors. This process was used to calculate *FZUSAGE*.

5: DATA ANALYSIS METHODOLOGY

5.1 Introduction

This section describes the statistical model used to estimate unit energy consumption (UEC) values for specific residential end uses. UECs were developed using a statistical technique called Conditional Demand Analysis (CDA). [15] The CDA approach essentially makes use of the variation in appliance holdings and whole-house energy consumption across the study sample to econometrically disaggregate billed consumption into end use consumption values. Section 5.2 provides a general overview of the conditional demand framework. Sections 5.3 and 5.4 describe the derivation of the specific CDA model specifications used to characterize electricity and gas consumption for this project.

5.2 Overview of Conditional Demand Analysis

5.2.1 Graphical Overview

Figure 5-1 provides an overview of a basic conditional demand model. The underlying spirit of the approach is that a household's energy consumption is directly related to the stock of appliances present in the dwelling and the energy consumption levels associated with these appliances (unit energy consumption levels, or UECs). Unit consumption, in turn, is related to specific features of these appliances, dwelling characteristics, and the household's utilization patterns.

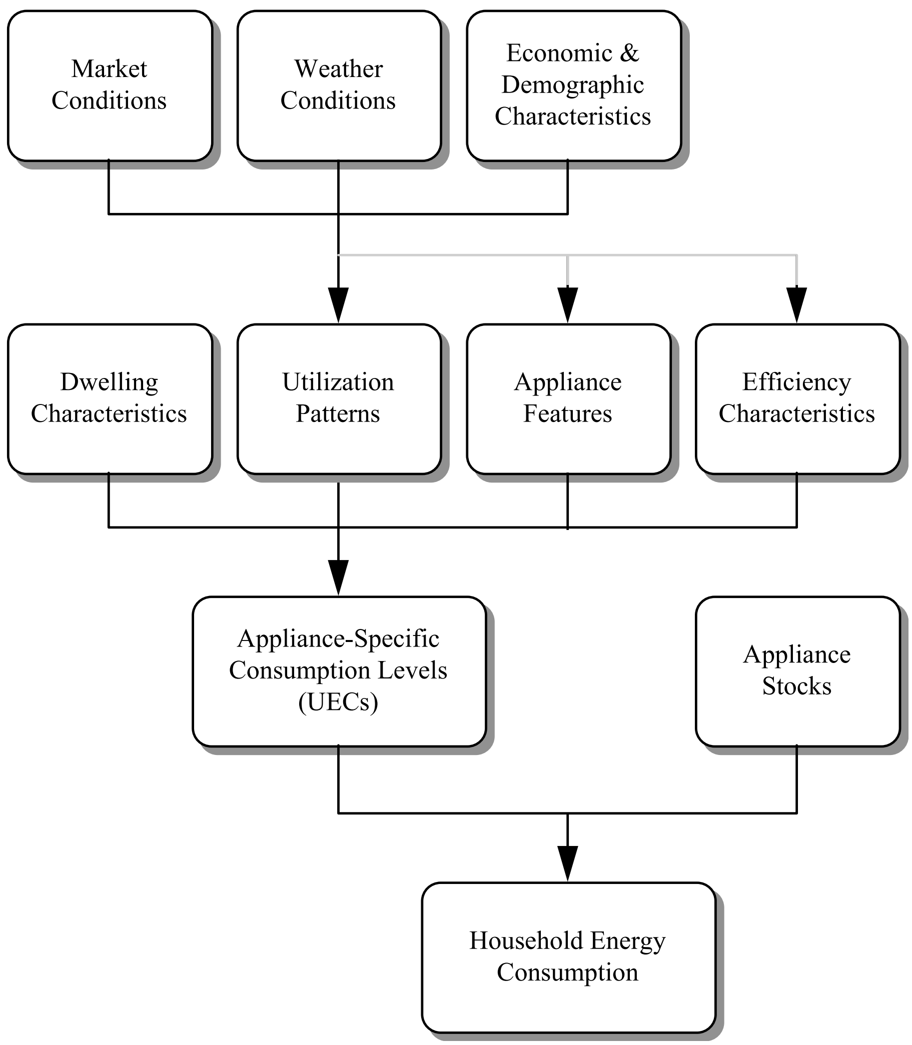
- Appliance stocks are typically represented in the CDA approach by a series of binary (0,1) or cardinal (1,2,3,..) indicators, generally defined for fairly specific appliance types. Binary indicators are used to indicate whether or not a particular system (e.g., central air conditioning) is present, whereas cardinal variables are used to represent appliances (say, TVs or refrigerators) where multiple units may be present.

- Appliance features include general characteristics like sizing (e.g., water heater capacities, air conditioner tonnage, etc.), as well as efficiencies and auxiliary equipment (e.g., intermittent ignition devices). Some direct information on these features can be available for the analysis. However, their roles can be recognized also indirectly by including variables that are expected to be correlated with the features (e.g. age of the structure, household size, etc.).

- Dwelling characteristics, which are most pertinent to space conditioning uses, can include surface area, insulation values, roofing materials, window areas, and other characteristics of the thermal shell.

- Utilization patterns include user-driven values such as cover thermostat settings on water heaters, pools/spas and space conditioning systems, as well as a variety of behavioral patterns relating to the use of other appliances. These utilization patterns are partially captured by surveys. When this information is unavailable, their effect can be incorporated indirectly into the model through the inclusion of market, weather, economic, and demographic variables likely to affect them.

Figure 5-1
Basic Overview of the Conditional Demand Model



5.2.2 A General Algebraic Specification of the Model

The basic conditional demand model can be represented in general algebraic form as:

(1) $$HEC_{ht} = \sum_{a=1}^{A} UEC_{hat} \cdot S_{hat}$$

where the following variable definitions are used:

HEC_{ht} = energy consumption by household h in period t

UEC_{hat} = energy consumption through household h's appliance a in period t

S_{hat} = a binary indicator of household h's ownership of appliance a in period t.

Both HEC_{ht} and S_{hat} are observable. Information on appliance stocks (S_{hat}) is obtained through the survey database. This information is collected only once for each sampled household, so the time subscript (t) is dropped in the remainder of the discussion. Data on whole-house energy consumption (HEC_{ht}) is extracted from billing records in monthly, bimonthly, seasonal, or annual form. Values of end-use energy consumption (UEC_{hat}), however, are directly observable only for those sites that have been metered. Nonetheless, the CDA methodology allows the analyst to infer values of UEC_{hat} from the other information used to develop the model. This feature is the most direct benefit associated with the use of CDA.

The disaggregation of whole-house consumption is accomplished through the statistical association of consumption with the presence of appliances. To illustrate this, consider a very simple specification in which the UECs are treated as constants across households and over time. This type of model can be written as:

(2) $$HEC_{ht} = \alpha_0 + \alpha_1 S_{h1} + \alpha_2 S_{h2} + \ ... \ + \alpha_n S_{hn} + e_{ht}$$

where e_{ht} is an error term (the nature of which is discussed more fully below) and the terms are the UECs. Using standard regression analysis, the analyst can estimate the α_n's. These estimates are based essentially on the tendency for household consumption to vary as appliance holdings vary. If homes with electric water heating tend to consume more energy than homes without this appliance, all other things equal, this tendency will be captured by the estimate of the coefficient on the water heating appliance variable. Each coefficient is interpreted as the increment in consumption due to the presence of the appliance in question, given the holdings of other appliances.

Of course, the above form of a conditional demand model is highly stylized, in that it treats UECs as constants across households. In fact, end-use consumption depends upon a variety of factors, as suggested earlier. This relationship can be formalized as:

(3) $$\mathrm{UEC_{hat}} = f_{\mathrm{a}}(\mathrm{AF_{ha}}, \mathrm{STRUC_h},\ \mathrm{EFF_{ha}},\ \mathrm{UP_{hat}}, \mathrm{e_{hat}})$$

where: AF_{ha} = features of household h's appliance stock

$STRUC_h$ = pertinent structural features

EFF_{ha} = factors relating to shell and equipment efficiencies

UP_{hat} = utilization patterns relating to appliance a

e_{hat} = a random error term for the end use.

As noted earlier, it seems reasonable to recognize the effect of weather conditions (call these WC_{ht}), market conditions (MC_{ht}), and the household's economic and demographic characteristics (EDC_h) on utilization patterns. So the model can be written as:

(4) $$\mathrm{UP_{hat}} = g_{\mathrm{a}}(\mathrm{WC_{ht}}, \mathrm{MC_{ht}}, \mathrm{EDC_h})$$

Explicit assumptions concerning the dependence of appliance features and structural characteristics on other variables could also be used, but assume for simplicity that data on these variables are available and that their values are taken as given for the purposes of the analysis.

Substituting (4) into (3) yields:

(5) $$\mathrm{UEC_{hat}} = F_{\mathrm{a}}(\mathrm{AF_{ha}}, \mathrm{STRUC_h},\ \mathrm{EFF_{ha}}, \mathrm{WC_{ht}}, \mathrm{MC_{ht}}, \mathrm{EDC_h}, \mathrm{e_{hat}})$$

where F_a is on estimates form of f_a. And, finally, substituting (5) into (1) provides the final general specification:

(6) $$\mathrm{HEC_{ht}} = \sum_{\mathrm{a}=1}^{\mathrm{A}} F_{\mathrm{a}}(\mathrm{AF_{ha}}, \mathrm{STRUC_h},\ \mathrm{EFF_{ha}}, \mathrm{WC_{ht}}, \mathrm{MC_{ht}}, \mathrm{EDC_h}, \mathrm{e_{hat}}) \cdot \mathrm{S_{ha}}$$

5.2.3 Constructing Estimates of UECs

Once the Conditional Demand Model is estimated statistically, it can be used to infer unit energy consumption for individual households as well as designated subpopulations. Generally, these UEC values are defined for given reference values of time-dependent variables like weather and market conditions (call these $WCREF_h$ and $MCREF_h$, respectively). Using the notation developed above, the UEC equation can be written as:

$$(7) \quad UEC_{ha} = G_a(AF_{ha}, STRUC_h, WCREF_h, MCREF_h, EDC_h, e_{hat}), \text{ for } S_h = 1$$

where G_a is an estimated form of F_a. As an example, suppose that the arguments of F_a include square footage ($SQFT_h$), heating degree-days (HDD_{ht}), and the marginal electricity price (MP_{ht}). Suppose also that the form of the function is linear. Then the UEC for this end use would be equal to an estimated form of F_a, which might look like:

$$(8) \quad UEC_{ha} = \alpha_0 + \alpha_1 SQFT_h + \alpha_2 HDD_t + \alpha_2 MP_{ht} \qquad \text{for } S_h = 1$$

where the α_h's are estimated coefficients.

The standard practice for evaluating UECs involves using the average price level over the period of estimation, although another price could be inserted. It is also common to use normal weather conditions to derive UECs. When this practice is followed, the resultant estimates are called weather-normalized UECs.

Deriving UECs on a household-by-household basis (by substituting each household's value of the variables in the UEC equation) can be a useful intermediate step, however, estimates are generally derived for households as a whole or for selected classes of households (single-family residences only, residents of a particular operating region, or geo-demographic segments). The derivation of group UECs (call a group mean UEC_a) is straightforward. Perhaps the most common approach to this calculation is to average household-level UEC estimates over the relevant group. Allowing for the use of sampling weights on individual households (w_h), we would have:

$$(9) \quad UEC_a = \frac{(\sum_h w_h UEC_{ha})}{\sum_h w_h}$$

where the summation is over the set of households with the appliance in question (i.e., for whom S_{ha} = 1). Once household-specific UECs are calculated, these values are summarized very easily for household groupings using standard subroutines in various statistical programs.

5.3 Specification of the Electric CDA Model

5.3.1 Overview of End Uses

This section derives the California RASS CDA model for electricity consumption. The model is used to disaggregate whole-house electricity consumption into 25 end uses:

- Primary space heating
- Secondary space heating
- Central air conditioning
- Room air conditioning
- Evaporative coolers
- Forced air fans
- 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
- Waterbed heaters
- Well pumps
- Miscellaneous

A considerable attempt was also made to develop UECs for indoor lighting as part of this project. However, the lack of variation in the presence of indoor lighting across homes, coupled with the lack of detailed indoor lighting inventories, made it impossible to use the CDA for this approach. In order for CDA to isolate consumption associated with a specific end use, one of two conditions must be present: the presence of this end use varies across homes (that is, some homes have it and some do not), or the availability of detailed end use inventories allow the construction of engineering priors for end use consumption. Obviously, all homes in the sample have indoor lighting, so the first condition was not met. Moreover, as is common for mail surveys, the RASS survey did not yield detailed enough information on indoor lighting to support the development of reliable engineering estimates.

In the remainder of this section, we derive the end use elements of the electric CDA model.

5.3.2 Electric Model Derivation

Primary Electric Space Heating Model

The electric space heating UEC model is based on a fundamental balance equation:

$$(10) \qquad EHEATUSE_{ht} = \frac{[HEATLOSS_{ht} - BUHT_{ht}]}{EFFH_h}$$

where primary space heating usage ($EHEATUSE_{ht}$) is assumed to be equal to net heat loss ($HEATLOSS_{ht}$), less the heat loss replaced by non-electric secondary heating systems ($BUHT_{ht}$), converted by a system efficiency ($EFFH_h$). The net heat loss from a structure can be written as:

$$(11) \qquad HEATLOSS_{ht} = SURFLOSS_{ht} - SOLGAIN_{ht} - INTGAIN_{ht}$$

where $SURFLOSS_{ht}$ reflects losses through envelope surfaces and includes wall, floor, roof, chimney, and infiltration losses; $SOLGAIN_{ht}$ is solar gain through all surfaces during potential heating periods, and $INTGAIN_{ht}$ reflects internal gains during these periods.

Total surface losses can be determined from the familiar relation:

$$(12) \qquad SURFLOSS_{ht} = \alpha_1 \, U_h \, AREA_h \, TDIFF_{ht}$$

where U_h is the overall conductivity of the shell, $AREA_h$ is the total surface area, and $TDIFF_{ht}$ is the differential between inside and outside temperature levels, cumulated over all hours of the period for which the differential is positive.

Solar gain during potential heating periods is assumed to be related to surface area, minutes of sunlight ($MINOFLIGHT_{ht}$), and a variable indicating that the month is a winter month ($WINTER_t$):

(13) $$SOLGAIN_{ht} = \alpha_2 AREA_h MINSOFLIGHT_{ht} WINTER_t$$

Internal gain during the winter months is assumed to be proportional to surface area of the home:

(14) $$INTGAIN_{ht} = \alpha_3 AREA_h WINTER_t$$

Shell surface area is modeled as a function of square footage where ß is an elasticity of surface area with respect to square footage:

(15) $$AREA_h = \alpha_4 SQFT_h^{\beta}$$

This relationship was estimated using on-site data from the Residential New Construction Survey, an effort being conducted by Itron. The area equation was estimated separately for each residence type.

Shell conductivity is assumed to be related to the presence of double-pane glass ($DPWIN_h$), a binary variable reflecting the location of the home in a Title 24 Standards Climate Zone with stringent insulation requirements ($T24_h$), and a binary variable indicating that the home is a multi-family dwelling (MF_h):

(16) $$U_h = \alpha_5 + \alpha_6 T24_h + \alpha_7 DPWIN_h + \alpha_8 MF_h$$

The Title 24 variable was used to reflect differences in the expected shell integrity between zones with stringent standards and those with base standards. It was initially intended to use variables representing the presence of ceiling and wall insulation, but these variables did not perform well at all in the early process of model estimation. This is unsurprising, insofar as mail survey data on insulation tend to be unreliable. The multi-family variable is included to account for the influence of adiabatic walls in multi-family structures

The temperature differential is affected by both behavioral and weather factors. It can be written as:

(17) $$TDIFF_{ht} \equiv \sum \left[TDES_{ht} - T_{ht} \right] \qquad for\ TDES_{ht} \geq T_{ht}$$

where $TDES_{ht}$ is the desired internal temperature and T_{ht} is the outside temperature, and where the summation is across all hours of the period in question. The following specification was used to capture the influence of both outdoor temperatures and thermostat set points on effective temperature differentials:

$$(18)\quad TDIFF_{ht} = HDD_{ht}(\alpha_9 + \alpha_{10} INC_h + \alpha_{11} ROOM_h + \alpha_{12}\ SETBK_h + \alpha_{13} HTTSET_h + \alpha_{14} SEASONAL_h)$$

This specification is based on the notion that heating degree-days (HDD_{ht}), a variable that captures variations in outdoor temperatures but which is based on a single reference temperature for all homes, is a good proxy for the general temperature differential. However, the relation also recognizes the existence of household-specific variations in desired temperatures, as represented by reported daytime thermostat settings ($HTTSET_h$), nighttime setbacks ($SETBK_h$), the presence of room heating ($ROOM_h$), and a binary variable indicating that the home is a seasonal home ($SEASONAL_h$). Income (INC_h) is included in this expression to account for its influence on actual differences in operation across homes.

The contribution of non-electric secondary space heating can be modeled simply as:

$$(19)\quad BUHT_{ht} = \alpha_{15} NONELEBK_h HDD_{ht} AREA_h$$

where $NONELEBK_h$ is a binary variable indicating the presence of non-electric backup heat.

The variation in system efficiencies is accommodated through the recognition of specific electric heating technologies. In particular, we use the following specification in the model:

$$(20)\quad EFFH_h = CONV_h + \alpha_{16} HP_h$$

where $CONV_h$ and HP_h are binary variables indicating that the system is conventional or a heat pump, respectively. Note that conventional systems are assigned an efficiency of 1.0, while heat pumps are assumed to have relative heating efficiencies of α_{16} . For the purposes of our analysis, the value of the heat pump efficiency was taken to be 2.0. In our subsequent discussion of the model, note that these efficiency values are embedded in the other variables of the model (i.e., all heating cross-products are divided by 2.0 for heat pumps).

Combining equations (10) through (20), we obtain the full specification of the electric space heating model. This specification is presented below. Note that some cross-product terms have been dropped as a consequence of their poor performance in the subsequent estimation process.

(21)

$$EHEATUSE_{ht} = \left[\begin{array}{l} \alpha_1 HDD_{ht} AREA_h + \alpha_2 HDD_{ht} AREA_h T24_h \\ + \alpha_3 HDD_{ht} AREA_h DPWIN_h \\ + \alpha_4 HDD_{ht} AREA_h MF_h + \alpha_5 HDD_{ht} AREA_h AVGINC_h \\ + \alpha_6 HDD_{ht} AREA_h DPWIN_h INC \\ + \alpha_7 HDD_{ht} AREA_h MF_h INC_h \\ + \alpha_8 HDD_{ht} AREA_h ROOM_h \\ + \alpha_9 HDD_{ht} AREA_h DPWIN_h ROOM_h \\ + \alpha_{10} HDD_{ht} AREA_h MF_h ROOM_h \\ + \alpha_{11} HDD_{ht} AREA_h SETBK_h \\ + \alpha_{12} HDD_{ht} AREA_h DPWIN_h SETBK_h \\ + \alpha_{13} HDD_{ht} AREA_h MF_h SETBK_h \\ + \alpha_{14} HDD_{ht} AREA_h HTTSET_h \\ + \alpha_{15} HDD_{ht} AREA_h DPWIN_h HTTSET_h \\ + \alpha_{16} HDD_{ht} AREA_h MF_h HTTSET_h + \alpha_{17} AREA_h WINTER_{ht} \\ + \alpha_{18} AREA_h MINSUN_{ht} WINTER_{ht} \\ + \alpha_{19} HDD_{ht} SEASONAL_h \\ + \alpha_{20} NONEBU_h HDD_{ht} AREA \end{array} \right] / EFFH_h$$

Note that all parameters in (21) are redefined relative to the values specified in the derivation of the model. We do this to conserve on parameter names.

Secondary Electric Space Heating

A simple specification is included in the model for auxiliary electric space heating ($EAUXHTUSE_{ht}$):

(22)

$$EAUXHTUSE_{ht} = \alpha_{21} HDD_{ht} + \alpha_{22} HDD_{ht} AREA_h + \alpha_{23} HDD_{ht} AREA_h MF_h + \alpha_{24} HDD_{ht} AREA_h ADDFREQ_h$$

where $ADDFREQ_h$ is a variable representing the frequency with which auxiliary heating is used.

Central Air Conditioning

Central and room air conditioning are modeled separately. For homes with central air conditioning systems, cooling energy usage ($CACUSE_{ht}$) is assumed to be determined by a balance equation of the form:

(23) $$CACUSE_{ht} = \frac{HEATGAIN_{ht} - AUXCOOL}{EFFC_h}$$

where $HEATGAIN_{ht}$ represents both internal gains and heat gain through the structure, $AUXCOOL_{ht}$ reflects the use of auxiliary cooling (identified here as evaporativde cooling) and where $EFFC_h$ represents the efficiency of the system. Total heat gain is specified as:

(24) $$HEATGAIN_{ht} = SURFGAIN_{ht} + SOLGAINC_{ht} + INTGAINC_{ht}$$

where $SURFGAIN_{ht}$ is a measure of total convective heat gain through structural surfaces, $SOLGAINC_{ht}$ indicates total solar radiant gain during potential cooling periods, and $INTGAINC_{ht}$ is total internal gain during these periods. The total convective gain can be written as:

(25) $$SURFGAIN_{ht} = \beta_{25} U_h AREA_h TDIFFC_{ht}$$

where $TDIFFC_{ht}$ is the differential between the outside temperatures (Th_t) and the desired indoor temperatures ($TDES_{ht}$), cumulated over hours when the differential is positive. That is:

(26) $$\left[TDIFFC_{ht} \equiv \sum \left[T_{ht} - TDES_{ht}\right]\right] \quad \text{for } T_{ht} \geq TDES_{ht}$$

Solar gain during potential cooling periods are assumed to be related to minutes of sunlight, shell area and as an indicator of summertime ($SUMMER_t$):

(27) $$SOLGAIN_{ht} = \beta_{26} AREA_h MINSUN_{ht} SUMMER_t$$

Internal gain during the summer is assumed to be proportional to surface area.

(28) $$INTGAIN_{ht} = \beta_{27} AREA_h SUMMER_t$$

Shell surface area is modeled as a function of square footage where ß is an elasticity of surface area with respect to square footage:

(29) $$AREA_h = \beta_{28} SQFT_h^{\beta}$$

Shell conductivity is assumed to be related to the presence of ceiling and wall insulation, and indicated by the Title 24 stringency variable defined earlier, double-pane glass ($DPWIN_h$) and a binary variable indicating that the home is a multi-family dwelling (MF_h):

(30) $$U_h = \beta_{29} + \beta_{30} T24_h + \beta_{31} DPWIN_h + \beta_{32} MF_h$$

The latter variable is included to account for the influence of adiabatic walls in multi-family structures.

The cooling temperature differential is assumed to be a function of cooling degree-days (CDD_{ht}), which is defined with a common reference temperature of 65°F, income, and the cooling system thermostat setting ($TSETC_h$). The inclusion of the thermostat setting and the income term is designed to capture differences in cooling system operation across homes.

(31) $$TDIFFC_{ht} = \left(\beta_{33} + \beta_{34} INC_h + \beta_{35} TSETC_h\right) CDD_{ht}$$

Auxiliary cooling is specified as a function of the presence of evaporative cooling (DSWAMP), cooling degree-days, and the summer variable:

(32)
$$AUXCOOL_{ht} = \beta_{36} AREA_h CDD_{ht} DSWAMP_h + \beta_{37} AREA_h DSWAMP_h SUMMER_t$$

Inferences regarding air conditioner efficiency must be made similar to those for space heating where specific information is not available. Efficiency is assumed to be related to system vintage, with newer units assumed to be more efficient in general than older units. In order to reflect this assumption, a set of incremental terms involving a binary variable representing new homes (homes six years old or newer) was added to the model.

Combining equations 23 through 32 and adding the new home terms yields the full central air conditioning model. This specification is presented below (note again that we reuse parameter subscripts for economy):

$$CACUSE_{ht} = \left[\begin{array}{l} \beta_{25} CDD_{ht} AREA_h + \beta_{26} CDD_{ht} AREA_h T24_h \\ + \beta_{27} CDD_{ht} AREA_h DPWIN_h \\ + \beta_{28} CDD_{ht} AREA_h MF_h + \beta_{29} CDD_{ht} AREA_h INC_h \\ + \beta_{30} CDD_{ht} AREA_h DPWIN_h INC \\ + \beta_{31} CDD_{ht} AREA_h MF_h INC_h \\ + \beta_{32} CDD_{ht} AREA_h TSET_h \\ + \beta_{33} cDD_{ht} AREA_h DPWIN_h TSET_h \\ + \beta_{34} CDD_{ht} AREA_h MF_h TSET_h + \beta_{35} AREA_h SUMMER_t \\ + \beta_{36} AREA_h MINSOFLIGHT_{ht} SUMMER_t \\ + \beta_{37} CDD_{ht} AREA_h NEWHOME_h \\ + \beta_{38} CDD_{ht} AREA_h NEWHOME_h INC_h \\ + \beta_{39} CDD_{ht} AREA_h TSETC_h NEWHOME_h \\ + \beta_{40} AREA_h DSWAMP_h SUMMER_t \\ + \beta_{41} AREA_h CDD_{ht} DSWAMP_h \end{array}\right] \quad (33)$$

Room Air Conditioning

A similar albeit more parsimonious specification will be used for room air conditioning ($RACUSE_{ht}$), except that a term will be used to reflect the number of room air conditioning units ($RACCNT_h$). This stems from the assumption that total usage depends on the number of room air conditioners.

$$RACUSE_{ht} = \left[\begin{array}{l} \beta_{42} CDD_{ht} AREA_h \\ + \beta_{43} CDD_{ht} AREA_h DPWIN_h \\ + \beta_{44} CDD_{ht} AREA_h MF_h \\ + \beta_{45} CDD_{ht} AREA_h AVGINC_h \\ + \beta_{46} CDD_{ht} AREA_h DPWIN_h INC_h \\ + \beta_{47} CDD_{ht} AREA_h MF_h INC_h \\ + \beta_{48} CDD_{ht} AREA_h TSETUSE_h \\ + \beta_{49} CDD_{ht} SQFT_h^{\beta} MF_h TSET_h \\ + \beta_{50} CDD_{ht} AREA_h RACCNT_h \\ + \beta_{51} CDD_{ht} AREA_h DSWAMP_h \end{array} \right] \quad (34)$$

Evaporative Coolers

Usage by evaporative coolers ($EVAPCUSE_{ht}$) is assumed to depend upon cooling degree-days and surface area:

(35) $$EVAPCUSE_{ht} = \beta_{52} CDD_{ht} + \beta_{53} CDD_{ht} AREA_h$$

Forced Air Heating-Related Ventilation

Furnace fan usage ($FFANUSE_{ht}$) is assumed to be related to the presence of central forced air heating and to heating requirements. In order to economize on the number of parameters to be estimated, a simplified model for furnace fan usage is employed.

(36) $$FFANUSE_{ht} = \gamma_0 HDD_{ht} SQFT_h$$

Electric Water Heating

The spirit of the electric water heating equation is captured by the expression:

(37) $$EWHEATUSE_{ht} = \frac{[WHLOSS_{ht} + VUSE_{ht}]}{EFFWH_h}$$

where $EWHEATUSE_{ht}$ is total electricity consumption for water heating, $WHLOSS_{ht}$ reflects heat losses associated with standby losses from the heating unit, $VUSE_{ht}$ represents heat losses tied to water usage, and $EFFWH_h$ reflects the efficiency of the unit. Given the lack of survey information on unit efficiency, we assume that efficiency is constant across homes (except that the presence of solar assist is considered below).

For simplicity, we link standing tank losses to the number of household members (a proxy for tank size), tank temperature($WHTEMP_h$), residence type (as indicated by the binary multi-family indicator, MF_h), and the presence of more than one tank ($ADDWHEL_h$):

(38) $$WHLOSS_{ht} = \delta_0 + \delta_1 \log(NUMI_h + 1) + \delta_2 \log(NUMI_h + 1) MF_h + \delta_3 ADDWHEL_h + \delta_5 DWHSOLAR_h$$

This reflects the assumption that primary tank sizes are related to household size. The logarithmic functional form used for this relationship has been developed over a large number of CDA studies, and seems to best fit the data on water heating usage. The loss equation also reflects the likelihood that total piping lengths will be lower in multi-family dwellings than in single family structures, thus leading to lower losses. The last term in this expression is used to represent the replacement of heat loss through the presence of a solar system, where $DWHSOLAR_h$ indicates the presence of solar assist.

We assume that monthly usage-related fuel consumption depends upon the household size as well as the number of dishwasher loads ($DWASHU_h$), washing machine loads ($WMACHU_h$), showers ($TOTAL\_SHTSHWRS_h$) reported by the households in question, and the temperature differential between the tank temperature and the inlet temperature ($WHTEMP\_DIFF_{ht}$).

(39) $$VUSE_h = \delta_6 + \delta_7 DWASHU_h + \delta_8 WMACHU_h + \delta_9 WHTSHWRS_h + \delta_{10} \log(NUMI_h + 1) + \delta_{11} WHTEMP\_DIFF_{ht}$$

Substituting (38) and (39) into equation (37), we obtain our basic water heating relation:

(40)
$$EWHEATUSE_{ht} = (\delta_0 + \delta_6) + (\delta_1 + \delta_{10})\log(NUMI_h + 1) + \delta_2 \log(NUMI_h + 1)MF_h + \delta_3 ADDWHEL_h + \delta_4 WHTEMP\_DIFF_{ht} + \delta_5 DWHSOLAR_{h15} + \delta_7 DWASHU_h + \delta_8 WMACHU_h + \delta_9 WHTSHWRS_h + \delta_{10} \log(NUMI_h + 1) + \delta_{11} WHTEMP\_DIFF_{ht}$$

Primary Refrigerators

For primary refrigerator usage ($REF1USE_h$), we use the simple relation:

(41)
$$REF1USE_{ht} = \lambda_1 REFUSAGE1_h$$

where REFUSAGE1 is an engineering estimate of usage based on unit size and efficiency. Insofar as the survey did not provide information on efficiency, this characteristic was inferred from the unit type and age, coupled with AHAM shipments data on refrigerator efficiency by type and age. The algorithm used for this calculation was discussed in Section 4.3.4.

Secondary Refrigerators

Second refrigerator usage ($REF2USE_h$) is specified as:

(42)
$$REF2USE_{ht} = \left(\lambda_2 + \lambda_3 MF_h + \lambda_4 SUMMER_t\right) REFUSAGE2_h$$

where $REFUSAGE2_h$ is an engineering estimate of usage based on reported number, size, age and type of second and subsequent refrigerators, coupled with AHAM shipments data, and $SUMMER_t$ is an indicator that the period is a summer month. (See Section 4.3.4.)

Freezers

Freezer consumption ($FREEZUSE_h$) is modeled in terms of an engineering estimate of usage ($FRZRUSAGE_h$). This estimate is based on the number, type(s), size(s) and age(s) of the freezers owned by the household. Like refrigeration above, AHAM shipments data was used to obtain the engineering estimates for freezer usage. (See Section 4.3.4.) The freezer equation is a simple adjustment function given by:

(43) $$FREEZUSE_{ht} = \lambda_5 FRZRUSAGE_h$$

Ranges and Ovens

Energy consumption through kitchen ranges and ovens ($RNGEOVNUSE_h$) will be specified as:

(44) $$RNGEOVNUSE_h = \mu_1 + \mu_8 \log(NUMI_h + 1) + \mu_3 \log(NUMI_h + 1) INC_h - \mu_4 \log(NUMI_h + 1) MICRO_h$$

where $MICRO_h$ reflects the presence of a microwave oven. The negative sign on the microwave variable indicates our expectation that these units act as substitutes for ranges in at least some activities. The sign on real income is theoretically indeterminate but is probably negative. It is unlikely that increases in income cause increased range usage. Instead, higher income households may tend to use the range less because of a higher propensity to eat away from home and a lower likelihood of being home during lunch time.

Microwave Ovens

The impact of microwave ovens on range/oven consumption was addressed above. Of course, microwaves consume power and must be included in the electricity equation. We incorporate microwave consumption ($MICWAVUSE_{ht}$) as a function of household size:

(45) $$MICWAVUSE_{ht} = \mu_5 \log(NUMI_h + 1)$$

Dishwashers

Dishwashers affect energy consumption both directly and indirectly. The indirect impacts operate through water heating requirements and have been treated above. The direct effects entail the use of electricity for operation of the units (motor loads and, in the case of some dishwashers, electric water heater boosters). We assume the following simple relation:

(46) $$DWASHUSE_{ht} = \mu_6 + \mu_7 \log(NUMI_h + 1)$$

where $DWASHUSE_{ht}$ indicates direct consumption through dishwashers.

Clothes washers

Washing machines also affect energy consumption both directly and indirectly. The indirect impacts operate through water heating requirements and have been treated above. The direct effects entail the use of electricity for operation of the units. We assume the following simple relation:

(47) $$WMASHUSE_{ht} = \mu_8 + \mu_9 \log(NUMI_h + 1)$$

where $WMACHUSE_{ht}$ indicates direct consumption through washing machines (motor usage).

Electric Dryers

Energy consumption by clothes dryers ($EDRYERUSE_{ht}$) will be assumed to be related to household size and reported dryer loads ($EDRYU_h$)

(48) $$EDRYERUSE_{ht} = \mu_{10} + \mu_{11} EDRYU_h + \mu_{12} \log(NUMI_h + 1)$$

Outdoor Lighting

The outdoor lighting model will explain outdoor lighting use ($OLTUSE_{ht}$) in terms of the total number of fixtures ($OLTFIX_h$) and the usage per fixture ($OLTUT_{ht}$). Usage per fixture is assumed to be a function of the proportion of CFLs ($OPROPCFL_h$), the proportion of fixtures using HID lamps ($OPROPHID_h$), the proportion of fixtures on motion sensors ($OPROPSENS_h$), the proportion on dusk-to-dawn sensors or timers ($OPROPTIM_h$), and the number of hours of darkness in the month in question ($HRDK_{ht}$). The outdoor lighting equation is thus:

(49) $$\begin{aligned} OLTUSE_{ht} = (\ &\eta_0 + \eta_1 OPROPCFL_h + \eta_2 OPROPHID_h + \eta_3 OPROPSENS_h \\ &+ \eta_4 OPROPTIM_h + \eta_5 HRDK_{ht}\ ) OLFIX_h \end{aligned}$$

Proportions of CFLs, fixtures on dimmers, and fixtures on sensors were derived from the numbers of fixtures of these types and the total number of fixtures.

Televisions

Electricity consumption through televisions ($TVUSE_{ht}$) is assumed to be related to the total connected load for televisions ($TVKW_h$) as well as hours of use ($TVHRS_h$):

(50) $$TVUSE_h = \sigma_0 TVKW_h + \sigma_1 TVHRS_h TVKW_h$$

Connected loads were derived from the numbers of TVs (standard and big-screen) and prior estimates of the connected load per unit for these types of units.[16]

Home Office Equipment

Electricity use associated with home offices ($EHMOFFUSE_h$) will be modeled very simply as:

(51) $$EHMOFFUSE_{ht} = \sigma_2 + \sigma_3 HMOFFHRS_h$$

where $HMOFFHRS_h$ is hours of use of the home office.

Personal Computers

Electricity use for personal computers ($PCUSE_{ht}$) is assumed to be related to number of personal computers ($PCNUM_h$) and the total number of hours of use per day ($PCHRS_h$):

(52) $$PCUSE_{ht} = \sigma_4 + \sigma_5 PCNUM_h + \sigma_6 PCHRS_h PCNUM_h$$

Swimming Pool Pumps

Energy consumption through the operation of swimming pool pumps ($PLPUMPUSE_{ht}$) is assumed to be linked to household size, temperatures and the season in question.

(53) $$PLPUMPUSE_{ht} = \sigma_7 + \sigma_8 PLFILT_{ht} + \sigma_9 PLFILT_{ht} PLSIZE_h$$

where $PLSIZE_h$ is pool size and $PLFILT_h$ is hours of use of filters (which is assumed to vary across seasons).

Spa Pumps

Electricity consumption through spa pumps ($SPAPUMPUSE_{ht}$) is assumed to depend upon spa size ($SPASIZE_h$) and frequency of use of the spa filter ($SPAFFREQ_h$):

(54) $$SPAPUMPUSE_{ht} = \sigma_{10} + \sigma_{11} SPAFFREQ_h + \sigma_{12} SPAFFREQ_h SPASIZE_h$$

Electric Spa Heat

Electric spa heating usage ($SPAHTUSE_{ht}$) is assumed to be determined by spa size, the frequency with which the spa is heated in the season in question ($SPAHTEFREQ_h$), the presence of an insulated cover ($SPACOV_h$), and the presence of solar assist ($SPASOLAR_h$):

(55) $$SPAHTUSE_{ht} = \sigma_{13} + \sigma_{14} SPAHTEFREQ_{\mathrm{h}} + \sigma_{15} SPAHTEFREQ_{ht} SPASIZE_h + \sigma_{16} SPACOV_h SPASIZE_h + \sigma_{17} SPASOLAR_h$$

Waterbed Heaters

Consumption of electricity for heated waterbeds ($WBEDHTUSE_h$) is assumed to be proportional to the number of waterbeds ($WBEDHTN_h$):

(56) $$WBEDHTUSE_h = \sigma_{18} WBEDHTN_h$$

Well Pumps

Well pump usage ($WELLPUSE_{ht}$) is assumed to be associated with the household size:

(57) $$WELLPUSE_{ht} = \sigma_{19} + \sigma_{20}(NUMI_h + 1)$$

Miscellaneous

A variety of other electric appliances may be owned by households (fans, mixers, etc.). To account for consumption through these other specified and unspecified uses ($MISCUSE_{ht}$), we use the formulation:

(58)

$$
\begin{aligned}
MISCUSE_{ht} &= \omega_0 + \omega_1 AVGINC_h + \omega_2 SQFT_h + \omega_3 \log(NUMI_h + 1) \\
&+ \omega_4 NEWHOME_h + \omega_5 MF_h + \omega_6 SEASONAL_h + \omega_7 EPLHT + \omega_8 DCEILF_h \\
&+ \omega_9 DATTFAN_h CDD_{ht} + \omega_{10} DATTFAN_h AREA_h CDD_{ht}
\end{aligned}
$$

Where $SEASONAL_h$ reflects that the home is a seasonal home, $EPLHT_h$ indicates the presence of electric pool heat, $DATTFAN_h$ indicates the presence of an attic fan, and $DCEILF_h$ reflects the presence of ceiling fans.

Summary of Electric Model

The electric model is derived by summing the above usage specifications, each multiplied times a binary variable representing the presence of the electric end use in question:

(59)

$$
\begin{aligned}
& ELECUSE_{ht} = EHEATUSE_{ht} DEHEAT_h + EAUXHTUSE_{ht} DEAUXHT_h \\
& + CACUSE_{ht} DCAC_h + RACUSE_{ht} DRAC_h + FFANUSE_{ht} DFFAN_h \\
& + EWHEATUSE_{ht} EWHFRAC_t DEWH_h \\
& + REF1USE_{ht} DREF1_h + REF2USE_{ht} DREF2_h \\
& + FREEZUSE_{ht} DFRZ_h + RNGEOVNUSE_{ht} DERNGOV_h \\
& + MICWAVUSE_{ht} MICWVFRAC_t DMWV_h + DWASHUSE_{ht} DWFRAC_t DDW_h \\
& + EDRYERUSE_{ht} DRYFRAC_t DEDRY_h + OLTUSE_{ht} DOLT_h \\
& + TVUSE_{ht} DTV_h + EHMOFFUSE_{ht} DHMOFF_h + PCUSE_{ht} DPC_h \\
& + PLPUMPUSE_{ht} DPLPMP_h + EPLHEATUSE_{ht} DEPLHT_h + SPAPUMPUSE_{ht} DSPA_h \\
& + ESPAHTUSE_{ht} DESPAHT_h + WBEDHTUSE_{ht} DEWB_h + WELLPUSE_{ht} DWELLP_h \\
& + MISCUSE_{ht}
\end{aligned}
$$

where the variables beginning with the prefix D are binary indicators of the presence of the electric end use. and the variables with the suffix FRAC are relative usage variables defined for specific end uses on the basis of monthly shapes developed in previous studies. For some end uses these binary variables are further interacted with monthly fractions for the end use in question based on prior load research. Such fractions are used for water heat, microwave ovens, dryers, dishwashers, and clothes washers. The application of these fractions helps the model to distinguish seasonal patterns across end uses.

5.3.3 Estimated Electricity Model

The electric model was estimated using RASS survey data, billing records covering the period January 2002 through August 2003, and weather data for the same period. The model was estimated with least squares regression analysis, with a correction for autocorrelation (correlation of the error term across time). Early tests indicated a fairly high level of first order autocorrelation in the residuals, so a standard generalized least squares technique was used to transform the data as a means of correcting this problem.[17]

Electric model estimated coefficients and their respective standard errors are presented in Table5-1. The overall fit of the model was reasonably good, with an adjusted coefficient of determination (R-squared) of 0.49. The coefficients for first refrigerators, forced air fans, microwaves, and clothes washers were restricted during the estimation process. The coefficients for first refrigerators and microwaves are difficult to estimate due to the near one hundred percent saturation. The coefficients on forced air fans and clothes washers were restricted due to problems with multi-collinearity. The restricted parameter estimates are labeled with an (R) in Table 5-1.

Table 5-1
Electric Model

Variable	Parameter	SE	T-Value
Intercept	0.0443	1.45576	0.03
(1/EFFH)*DHEAT*HDD*AREA	0.000033	0.00005506	0.6
(1/EFFH)*DEHEAT*HDD*AREA*DPWIN	-0.00008386	0.00006554	-1.28
(1/EFFH)*DEHEAT*HDD*AREA*MF	-0.00112	0.00008599	-13.06
(1/EFFH)*DEHEAT*HDD*AREA*INC	-2.90E-10	1.07E-10	-2.71
(1/EFFH)*DEHEAT*HDD*AREA*INC*DPWIN	1.77E-10	1.26E-10	1.41
(1/EFFH)*DEHEAT*HDD*AREA*INC*MF	2.01E-11	1.74E-10	0.12
(1/EFFH)*DEHEAT*HDD*AREA*ROOM	-0.00003423	0.00001057	-3.24
(1/EFFH)*DEHEAT*HDD*AREA*ROOM*DPWIN	0.00002347	0.00001292	1.82
(1/EFFH)*DEHEAT*HDD*AREA*ROOM*MF	0.00015439	0.00001706	9.05
(1/EFFH)*DEHEAT*HDD*AREA*SETBK	-0.00000748	0.00001155	-0.65
(1/EFFH)*DEHEAT*HDD*AREA*SETBK*DPWIN	-0.00001515	0.00001361	-1.11
(1/EFFH)*DEHEAT*HDD*AREA*SETBK*MF	0.00005879	0.00001799	3.27
(1/EFFH)*DEHEAT*HDD*AREA*HTTSET	0.0000035	8.74E-07	4
(1/EFFH)*DEHEAT*HDD*AREA*HTTSET*DPWIN	-1.64E-07	0.00000105	-0.16
(1/EFFH)*DEHEAT*HDD*ARE*HTTSET*MF	0.00001861	0.0000014	13.26
(1/EFFH)*DEHEAT*HDD*AREA*NONELEBK	0.00004832	0.0000063	7.67
(1/EFFH)*DEHEAT*AREA*WINTER	0.18559	0.00632	29.37
(1/EFFH)*DEHEAT*AREA*WINTER*MINSOFLIGHT	-0.00025469	0.00000891	-28.6
(1/EFFH)*DEHEAT*AREA*HDD*T24	-0.00004063	0.0000074	-5.49
(1/EFFH)*DEHEAT*HDD*SEASONAL	-0.15854	0.02977	-5.33
DAUXHT*HDD	0.01261	0.01127	1.12
DAUXHT *HDD*AREA	0.00003403	0.00000332	10.24
DAUXHT *HDD*AREA*MF	-0.00001016	0.00000677	-1.5
DAUXHT *HDD*AREA*ADDFREQ	0.00000178	1.59E-07	11.19
DCAC*CDD*AREA	0.00149	0.00003898	38.09
DCAC*CDD*AREA*NEWHOME	0.0000485	0.00005925	0.82
DCAC*CDD*AREA*DPWIN	-0.0001195	0.00004688	-2.55
DCAC*CDD*AREA*MF	0.00105	0.00008713	12.1
DCAC*CDD*AREA*INC	9.42E-11	4.38E-11	2.15
DCAC*CDD*AREA*INC*NEWHOME	-1.68E-10	5.82E-11	-2.9
DCAC*CDD*AREA*INC*DPWIN	1.25E-10	4.98E-11	2.5
DCAC*CDD*AREA*INC*MF	-2.11E-09	8.78E-11	-24.01
DCAC*CDD*AREA*TSETC	-0.00001516	4.93E-07	-30.75
DCAC*CDD*AREA*TESTC*NEWHOME	-2.14E-07	7.53E-07	-0.28
DCAC*CDD*AREA*TSETC*DPWIN	9.03E-07	5.93E-07	1.52
DCAC*CDD*AREA*TSETC*MF	-0.00001014	0.00000111	-9.13
DCAC*AREA*MINSOFLIGHT*SUMMER	0.00010001	0.0000034	29.38
DCAC*AREA*DSWAMP*SUMMER	0.01272	0.00198	6.43
DCAC*CDD*DSWAMP*AREA	-0.00016875	0.00000612	-27.55
DCAC*AREA*SUMMER	-0.07495	0.00275	-27.21
DRAC*CDD*AREA	0.00005146	0.00000754	6.82
DRAC*CDD*AREA*DPWIN	-0.00001868	0.00000473	-3.95
DRAC*CDD*AREA*MF	0.00001129	0.00001076	1.05
DRAC*CDD*AREA*INC	-5.83E-10	5.72E-11	-10.2
DRAC*CDD*AREA*TSETC	0.00001805	0.00000141	12.82
DRAC*CDD*AREA*RACCNT	0.00001597	0.00000457	3.49

Table 5-1
Electric Model (cont'd)

Variable	Parameter	SE	T-Value
DRAC*CDD*DSWAMP*AREA	-0.00008934	0.00000589	-15.16
DSWAMP*AREA*CDD	0.00006345	0.00000767	8.27
DSWAMP*CDD	0.19156	0.01999	9.58
DFFAN*HDD*AREA (R)	0.000023	0	Infty
DRF1*REFUSAGE1 (R)	0.0833	0	Infty
DRF2*REFUSAGE2	0.1366	0.00202	67.69
DRF2*SUMMER*REFUSAGE2	-0.00404	0.00156	-2.58
DRF2*REFUSAGE2*MF	-0.053	0.00586	-9.04
DFRZR*FZUSAGE	0.12464	0.00219	56.79
DEWH*FACTAWH*DWASHU	28.89343	1.02908	28.08
DEWH*FACTAWH*CWASHU	9.98225	0.68911	14.49
DEWH*FACTAWH*WHTSHWRS	18.4293	1.86502	9.88
DEWH*FACTAWH*DWHSOLAR	-127.56103	11.68353	-10.92
DEWH*ADDWHEL*FACTAWH	15.96034	3.89104	4.1
DEWH*FACTAWH* Log(NUMI+1)	42.08176	7.24915	5.81
DEWH*FACTAWH* Log(NUMI+1)*MF	-73.10609	3.82932	-19.09
DEWH*FACTAWH*WHTEMP_DIFF	0.03581	0.00603	5.94
DEWH*FACTAWH	73.0256	7.01039	10.42
DERNGOV* Log(NUMI+1)	37.1557	5.11421	7.27
DERNGOV* Log(NUMI+1)*INC	0.00005195	0.0000188	2.76
DERNGOV* Log(NUMI+1)*MICRO	-5.78601	3.77348	-1.53
DERNGOV	-22.0967	4.0174	-5.5
DMWV *FACTAMI* Log(NUMI+1) (R)	8.33	0	Infty
DDW* Log(NUMI+1)*FACTADW	9.89775	2.98564	3.32
DDW*FACTADW	-6.41515	3.81725	-1.68
DCW*FACTACW* Log(NUMI+1) (R)	37.09798	3.17859	11.67
DCW*FACTACW (R)	-40.09798	3.17859	-12.62
DEDRY*FACTADR*EDRYU	16.78199	0.46556	36.05
DEDRY*FACTADR* Log(NUMI+1)	5.5022	3.53861	1.55
DEDRY*FACTADR	-27.02423	4.17348	-6.48
DOLT*OLTFIX*ONOCFL	-5.65594	0.57041	-9.92
DOLT*OLTFIX*OPROPHID	5.26879	1.19711	4.4
DOLT*OLTFIX*OPROPSENS	-4.17967	0.68911	-6.07
DOLT*OLTFIX*OPROPTIM	11.10408	0.47871	23.2
DOLT*OLTFIX*HRDK	2.11248	0.06226	33.93
DOLT*OLTFIX	-20.00278	0.75837	-26.38
DTV*TVKW*TVHRS	36.48776	0.96943	37.64
DTV*TVKW	99.84392	6.58883	15.15
DHMOFF*HMOFFHRS	0.80713	0.09919	8.14
DHMOFF	-0.712	2.05713	-0.35
DPC*PCNUM	16.48716	1.3221	12.47
DPC*PCNUM*PCHRS1	1.68823	0.0487	34.66
DPC	6.52058	2.04486	3.19
DPLPMP*PLFILT	-17.9017	1.64402	-10.89

Table 5-1
Electric Model (cont'd)

Variable	Parameter	SE	T-Value
DPLPMP*PLFILT*PLSIZE	0.00116	0.00005773	20.06
DPLPMP	177.43949	2.84182	62.44
DSPA*SPAFREQ	1.8575	0.61018	3.04
DSPA*SPAFREQ*SPASIZE	0.6434	0.11184	5.75
DEHTSPA*SPAEHTFREQ	4.11848	0.55963	7.36
DEHTSPA*SPAEHTFREQ*SPASIZE	-0.19491	0.11672	-1.67
DEHTSPA*SPASIZE*SPCOV	7.22828	0.80349	9
DEHTSPA*SPASOLAR	6.29138	17.02186	0.37
DWB*WBEDHTN	59.92947	3.1606	18.96
DWELLP* Log(NUMI+1)	55.41209	6.98169	7.94
DWELLP	0.64884	9.02897	0.07
INC	0.00030879	0.00002009	15.37
SQFT	0.04769	0.00105	45.45
Log(NUMI+1)	43.11824	3.05322	14.12
NEWHOME	-42.01492	2.42332	-17.34
MF	-8.54592	1.64028	-5.21
SEASONAL	-142.36973	4.49941	-31.64
DCEILF	19.19172	1.19237	16.1
DATTFAN*CDD	0.35164	0.02095	16.79
DATTFAN*CDD*AREA	-0.00007051	0.00000574	-12.28
EPLHT	88.18653	13.11469	6.72

In general, the estimated coefficients take on the expected signs, and most are highly significant. There are a few issues to point out with respect to these coefficients:

- First, due to the high level of interaction of the explanatory variables, the influence of some variables is dependent on the values of the others. For instance, in the electric space heating equation, the influence of AREA depends upon HDD, INC, MF, and the other variables with which AREA is interacted. Accordingly, the signs of the individual coefficients should be interpreted carefully. They relate only to the incremental effect of the term in question, not the overall effect of any of its components.

- Second, a few coefficients may appear to have inappropriate signs, but do not. In the electric range equation, for instance, the incremental income term takes on a negative sign. We have gotten this result in every conditional demand analysis we have done, and it probably indicates the effect of income on the propensity to eat out rather than cooking at home.

- Third, it should be understood that the sign of some coefficients may reflect the influence of confounding conditions associated with the term. For instance, the coefficient of the spa cover in the electric spa heating equation is positive and significant. We do not suggest that, all things given, spa covers cause more heating energy to be used; rather, it is likely that households with spa covers probably use their spas more frequently in spite of the presence of the cover because of more frequent use. While we have included a spa use frequency variable to control for this factor, reported usage may not be a very good indicator of actual usage.

- Fourth, the coefficients of the outdoor lighting equation probably deserve some comment. As expected, usage is positively related to hours of darkness and negatively related to both the proportion of CFLs and the proportion on motion sensors. It is probably also reasonable that the influence of timers is positive, given that the installation of a timer probably indicates a greater preference of the use of outside lighting. The positive influence of the proportion of HIDs may also make sense. While HID lighting may be more efficient that incandescent, the presence of HIDs may indicate considerably higher lumens and, in spite of better efficiencies, greater total wattages than in homes without any HID lighting.

5.4 Specification of the Natural Gas CDA Model

5.4.1 Overview of Gas End Uses

This subsection derives the CDA model for natural gas consumption. The model is used to disaggregate whole-house natural gas consumption into eight end uses:

- Primary space heating
- Secondary space heating
- Water heating
- Ranges and ovens
- Dryers
- Swimming pool heat
- Spa heat
- Miscellaneous

End-use specifications are derived in the remainder of this subsection.

5.4.2 Derivation of the Gas Model

Primary Gas Space Heating

The gas space heating UEC model is based on a fundamental balance equation:

(60) $$GHEATUSE_{ht} = \frac{[HEATLOSS_{ht}\text{-}BUHT_{ht}]}{EFFH_h}$$

where primary gas space heating usage ($GHEATUSE_{ht}$) is assumed to be equal to net heat loss ($HEATLOSS_{ht}$), less the heat loss replaced by non-gas secondary heating systems ($BUHT_{ht}$), converted by a system efficiency ($EFFH_h$). The net heat loss from a structure can be written as:

(61) $$HEATLOSS_{ht} = SURFLOSS_{ht} - SOLGAIN_{ht} - INTGAIN_{ht}$$

where $SURFLOSS_{ht}$ reflects losses through envelope surfaces and includes wall, floor, roof, chimney, and infiltration losses; $SOLGAIN_{ht}$ is solar gain through all surfaces during potential heating periods, and $INTGAIN_{ht}$ reflects internal gains during these periods.

Total surface losses can be determined from the familiar relation:

(62) $$SURFLOSS_{ht} = \theta_1\ U_h\ AREA_h\ TDIFF_{ht}$$

where U_h is the overall conductivity of the shell, $AREA_h$ is the total surface area, and $TDIFF_{ht}$ is the differential between inside and outside temperature levels, cumulated over all hours of the period for which the differential is positive.

Solar gain during potential heating periods is assumed to be related to surface area, minutes of sunlight ($MINOFLIGHT_{ht}$), and a variable indicating that the month is a winter month ($WINTER_t$):

(63) $$SOLGAIN_{ht} = \theta_2 AREA_h MINSOFLIGHT_{ht} WINTER_t$$

where, as noted in the derivation of the electric model, surface area is modeled by residence type as a function of square footage. Internal gain during the winter months is assumed to be proportional to surface area of the home:

(64) $$INTGAIN_{ht} = \theta_3 AREA_h WINTER_t$$

Shell conductivity is assumed to be related to the presence of double-pane glass ($DPWIN_h$), a binary variable representing a new home (homes six years old or newer), a binary variable reflecting the location of the home in a Title 24 Standards Climate Zone with stringent insulation requirements ($T24_h$), and binary variables indicating that the home is a multi-family dwelling (MF_h) or a mobile home (MH_h):

(65) $$U_h = \theta_5 + \theta_6 NEWHOME + \theta_7 T24_h + \theta_8 DPWIN_h + \theta_9 MF_h + \theta_{10} MH_h$$

The rationale for the Title 24 variable was discussed earlier with reference to electric heating. Note that the gas model includes two residence type variables rather than one. This is the case because the higher saturation of gas space heating allows a more extensive specification.

The temperature differential is affected by both behavioral and weather factors. It can be written as:

(66) $$TDIFF_{ht} \equiv \sum \left[TDES_{ht} - T_{ht}\right] \qquad for\ TDES_{ht} \geq T_{ht}$$

where $TDES_{ht}$ is the desired internal temperature and T_{ht} is the outside temperature, and where the summation is across all hours of the period in question. The following specification was used to capture the influence of both outdoor temperatures and thermostat set points on effective temperature differentials:

(67) $$TDIFF_{ht} = HDD_{ht}(\theta_{11} + \theta_{12} INC_h + \theta_{13} ROOM_h + \theta_{14} SETBK_h + \theta_{15} HTTSET_h + \theta_{16} SEASONAL_h)$$

This specification is based on the rationale presented above for electric space heating.

The contribution of non-gas secondary space heating ($NGBUHT_{ht}$) can be modeled simply as:

(68) $$NGBUHT_{ht} = \theta_{17} NONGBU_h HDD_{ht} AREA_h$$

where $NONGBU_h$ is a binary variable indicating the presence of non-gas backup heat.

System efficiencies are represented indirectly in terms of the age of the system ($GHTAGE_h$):

(69) $$1/EFF = \theta_{18} + \theta_{19} GHTAGE_h$$

The full gas space heating specification is derived from the combination of equations (60) through (69), although some cross-product terms are omitted to conserve on degrees of freedom. The model is presented below. Note that the parameters have been redefined to simplify the presentation.

$$GHEATUSE_{ht} = \begin{bmatrix} \theta_1 HDD_{ht} AREA_h + \theta_2 HDD_{ht} NEWHOME + \theta_3 HDD_{ht} AREA_h DPWIN_h \\ + \theta_4 HDD_{ht} AREA_h T24_h + \theta_5 HDD_{ht} AREA_h MF_h + \theta_6 HDD_{ht} AREA_h MH_h \\ + \theta_7 HDD_{ht} AREA_h NEWHOME_h + \theta_8 HDD_{ht} AREA_h AVGINC_h \\ + \theta_9 HDD_{ht} AREA_h DPWIN_h INC \quad + \theta_{10} HDD_{ht} AREA_h MF_h INC_h \\ + \theta_{11} HDD_{ht} AREA_h MH_h INC_h + \theta_{12} HDD_{ht} AREA_h NEWHOME_h INC_h \\ + \theta_{13} HDD_{ht} AREA_h ROOM_h + \theta_{14} HDD_{ht} AREA_h DPWIN_h ROOM_h \\ + \theta_{15} HDD_{ht} AREA_h MF_h ROOM_h + \theta_{16} HDD_{ht} AREA_h SETBK_h \\ + \theta_{17} HDD_{ht} AREA_h DPWIN_h SETBK_h + \theta_{18} HDD_{ht} AREA_h MF_h SETBK_h \\ + \theta_{19} HDD_{ht} AREA_h TSET_h + \theta_{20} HDD_{ht} AREA_h DPWIN_h TSET_h \\ + \theta_{21} HDD_{ht} AREA_h MF_h TSET_h \\ + \theta_{22} AREA_h WINTER_t + \theta_{23} AREA_h WINTER_t T24_h \\ + \theta_{24} AREA_h MINSOFLIGHT_{ht} WINTER_t \\ + \theta_{25} NONGBU_h HDD_{ht} AREA_h + \theta_{26} HDD_{ht} SEASONAL_h \\ + \theta_{27} HDD_{ht} GHTAGE_h + \theta_{28} HDD_{ht} AREA_h GHTAGE_h \\ + \theta_{29} HDD_{ht} AREA_h INC_h GHTAGE_h \\ + \theta_{30} HDD_{ht} AREA_h ROOM_h GHTAGE_h \\ + \theta_{31} HDD_{ht} AREA_h SETBK_h GHTAGE_h \\ + \theta_{32} HDD_{ht} AREA_h TSET_h GHTAGE_h \end{bmatrix}_h$$

Secondary Gas Space Heating

A simple specification will be included in the model for auxiliary gas space heating ($GAUXHTUSE_{ht}$):

(70)

$$GAUXHTUSE_{ht} = \theta_{33} HDD_{ht} + \theta_{34} HDD_{ht} AREA_{h} + \theta_{35} HDD_{ht} AREA_{h} MF_{h}$$

Gas Water Heating

The spirit of the gas water heating equation is captured by the expression:

(71) $$GWHEATUSE_{ht} = WHLOSS_{ht} + VUSE_{ht}$$

where $GWHEATUSE_{ht}$ is total gas consumption for water heating, $WHLOSS_{ht}$ reflects heat losses associated with standby losses from the heating unit, $VUSE_{ht}$ represents heat losses tied to water usage. Given the improved efficiency of newer homes, we assume that efficiency is higher in newer homes than older homes.

For simplicity, we link standing tank losses to the number of household members, residence age, the difference between tank temperatures and inlet temperatures, and the presence of gas solar assist ($GWHGSOLAR_{h}$):

(72)
$$\begin{aligned} WHLOSS_{ht} = {} & \rho_0 + \rho_1 \log(NUMI_h + 1) + \rho_2 WHTEMP\_DIFF_{ht} \\ & + \rho_3 DWHGSOLAR_h + \rho_4 NEWHOME \end{aligned}$$

The rationale for this specification is the same as for electric water heat, except that the multi-family incremental term was not found to be necessary for the model.

We assume that monthly usage-related fuel consumption depends upon the household size as well as the number of dishwasher loads, washing machine loads, the number of showers taken by the household, the temperature differential, and a variable representing that the home is a seasonal home:

(73)

$$\begin{aligned} VUSE_h = {} & \rho_5 + \rho_6 \log(NUMI_h + 1) SEASONAL_h + \rho_7 DWASHU_h + \rho_8 WMACHU_h + \rho_9 WHTSHWRS_h \\ & + \rho_{10} \log(NUMI_h + 1) + \rho_{11} WHTEMP\_DIFF_{ht} \end{aligned}$$

Substituting (72) and (73) into equation (71), we obtain:

(74)

$$GWHEATUSE_{ht} = (\rho_0 + \rho_5) + (\rho_1 + \rho_{10})\log(NUMI_h + 1) + (\rho_2 + \rho_{11})WHTEMP\_DIFF_{ht} + \rho_3 DWHGSOLAR_h + \rho_6 \log(NUMI_h + 1) SEASONAL_h + \rho_7 DWASHU_h + \rho_7 WMACHU_h + \rho_9 WHTSHWRS_h + \rho_4 NEWHOME_h$$

Ranges and Ovens

Gas consumption through kitchen ranges and ovens ($GRNGEOVNUSE_h$) will be specified as:

(75)
$$GRNGEOVNUSE_h = \pi_1 + \pi_2 \log(NUMI_h + 1) + \pi_3 \log(NUMI_h + 1) INC_h + \pi_4 \log(NUMI_h + 1) MICRO_h$$

The rationale for this specification mirrors that of the electric cooking equation.

Gas Dryers

Gas consumption by clothes dryers ($GDRYERUSE_{ht}$) will be assumed to be related to household size and reported dryer loads ($EDRYU_h$)

(76)
$$GDRYERUSE_{ht} = \pi_5 + \pi_6 EDRYU_h + \pi_7 \log(NUMI_h + 1)$$

Gas Pool Heat

Gas pool heating usage ($GPLHEATUSE_{ht}$) is assumed to be related to pool size and $PLHTFREQ_{ht}$, an indicator of the frequency of pool heating (which varies by summer and winter).

(77)

$$GPLHEATUSE_{ht} = (\beta_{22} + \beta_{23} PLHEATFREQ_h + \beta_{24} PLHEATFREQ_{ht} PLSIZE_h) OCC_{ht}$$

Gas Spa Heat

Gas spa heating usage ($GSPAHTUSE_{ht}$) is assumed to be determined by spa size, the frequency with which the spa is heated in the season in question ($SPAHTFREQ_h$), and the presence of an insulated cover ($SPACOV_h$).

(78)
$$GSPAHTUSE_{ht} = \pi_{12} + \pi_{13} SPAHTFREQ_h + \pi_{14} SPAHTFREQ_{ht} SPASIZE_h + \pi_{15} SPACOV_h SPASIZE_h$$

Miscellaneous

Gas miscellaneous usage ($GMISC_h$) is limited to two pieces of identified equipment: medical equipment ($DGMED_h$) and barbeques ($DGBBQ_h$):

(79)
$$GMISC_h = \pi_{16} DGMED_h + \pi_{17} DGBBQ_h$$

Summary of Gas Model

The gas model is derived by summing the above usage specifications, each multiplied times a binary variable representing the presence of the electric end use in question:

(80)

$$\begin{aligned} GASUSE_{ht} &= GHEATUSE_{ht} DGHEAT_h + GAUXHTUSE_{ht} DGAUXHT_h \\ &+ GWHEATUSE_{ht} GWHFRAC_t DGWH_h + GRNGEOVNUSE_{ht} RNGFRAC_t DGRNGOV_h \\ &+ GDRYERUSE_{ht} DRYFRAC_t DGDRY_h + GPLHEATUSE_{ht} DGPLHT_h \\ &+ GSPAHTUSE_{ht} DGSPAHT_h + GMISC_h \end{aligned}$$

where the variables beginning with the prefix D are binary indicators of the presence of the gas end use, and the variables with the suffix FRAC are relative usage variables defined for specific end uses on the basis of monthly shapes developed in previous studies.

5.4.3 Estimated Gas Model

The natural gas model was estimated with data on individually metered gas customers. The requisite data included billing records, survey data, and weather data. The overall fit of the natural gas model was quite good, with an adjusted coefficient of determination of 0.70. Natural gas model estimated coefficients and their respective standard errors are presented in Table 5-2. Again, a few comments with respect to these estimates.

- First, almost all coefficients take on the expected signs, and most are significant.

- Second, the coefficient on pool covers takes on the wrong sign and is significant. This probably indicates that homes with pool covers use more pool heating energy than others in spite of the conservation effect of the covers, due to higher preferences for more frequent pool use and perhaps bigger pools. Again, we have included size and frequency of use variables to control for these factors, but there may be significant reporting errors in these variables.

- The presence of solar assist appears to positively influence gas spa heat usage, a result that again probably reflects the result of this variable acting as a proxy for frequency of use.

Table 5-2
Gas CDA Coefficients

Label	Estimate	Error	t Value
DGHEAT*AREA*WINTER*T24	0.000238	0.000287	0.83
DGHEAT* HDD* AREA *T24	-1.6E-05	4.85E-07	-32.31
DGHEAT*HDD* AREA	-2.68E-07	1.06E-07	-2.53
DGHEAT*HDD* AREA *NEWHOME	-6.7E-06	1.13E-06	-5.91
DGHEAT*HDD* AREA *GHTAGE	-1.9E-06	7.08E-08	-26.33
DGHEAT*HDD* AREA *DPWIN	-2.5E-06	1.38E-06	-1.81
DGHEAT*HDD* AREA *MF	-4E-05	3.63E-06	-11.12
DGHEAT*HDD* AREA *INC	4.73E-11	4.08E-12	11.57
DGHEAT*HDD* AREA *INC*NEWHOME	6.42E-12	5.85E-12	1.1
DGHEAT*HDD* AREA *INC*GHTAGE	-6.31E-13	1.88E-13	-3.36
DGHEAT*HDD* AREA *INC*DPWIN	-1.97E-11	3.57E-12	-5.52
DGHEAT*HDD* AREA *INC*MF	-1.11E-11	7.92E-12	-1.4
DGHEAT*HDD* AREA *GROOM	2.26E-06	6.72E-07	3.35
DGHEAT*HDD* AREA *GROOM*GHTAGE	-3.13E-07	2.80E-08	-11.18
DGHEAT*HDD* AREA *GROOM*DPWIN	4.56E-06	5.82E-07	7.85
DGHEAT*HDD* AREA *GROOM*MF	2.27E-06	1.02E-06	2.23
DGHEAT*HDD* AREA *SETBK	-5.18E-07	4.35E-07	-1.19
DGHEAT*HDD* AREA *SETBK*GHTAGE	-1.32E-07	2.00E-08	-6.63
DGHEAT*HDD* AREA *SETBK*DPWIN	1.73E-06	3.80E-07	4.56
DGHEAT*HDD* AREA *SETBK*MF	4.95E-06	7.88E-07	6.29
DGHEAT*HDD* AREA *HTTSET	5.36E-07	8.69E-09	61.64
DGHEAT*HDD* AREA *HTTSET*GHTAGE	3.04E-08	1.13E-09	26.89
DGHEAT*HDD* AREA *HTTSET*DPWIN	-6.13E-08	2.25E-08	-2.72
DGHEAT*HDD* AREA *HTTSET*MF	5.96E-07	5.98E-08	9.97
DGHEAT*HDD* AREA *NONGBU	-1.7E-06	1.73E-07	-9.81
DGHEAT* AREA *WINTER	0.01694	0.000186	91.11
DGHEAT* AREA *WINTER*MINSOFLIGHT	-2.3E-05	2.65E-07	-86.05
DGHEAT*HDD*GHTAGE	-0.00847	0.00311	-2.73
DGHEAT*HDD*NEWHOME	0.00104	4.94E-05	21.03
DGHEAT*HDD*SEASONAL	-0.00771	0.00298	-2.59
DGHEAT*HDD*AREA *MH	5.23E-06	1.4E-06	3.72
DGHEAT*HDD*AREA*INC*MH	-4.42E-11	3.49E-11	-1.27
DNGAUXHT*HDD	0.65463	0.03224	20.3
DNGAUXHT*HDD* AREA	0.45847	0.01962	23.37
DNGAUXHT*HDD* AREA *MF	-2.67182	1.43665	-1.86
DGWH*FACTAWH* Log(NUMI+1)	-3.13922	0.25027	-12.54
DGWH*FACTAWH*DWASHU	-9.0196	0.64293	-14.03
DGWH*FACTAWH*CWASHU	13.98212	0.67417	20.74
DGWH*FACTAWH*DWHGSOLAR	0.00966	0.00778	1.24
DGWH*FACTAWH*LOG(NUMI+1)*NEWHOME	0.21075	0.05043	4.18
DGWH*FACTAWH* Log(NUMI+1)*SEASONAL	6.31861	0.45331	13.94
DGWH*FACTAWH	-3.1E-06	1.41E-06	-2.21
DGWH*FACTAWH*WHTEMP_DIFF	-1.23934	0.3011	-4.12
DGWH*FACTAWH*TOTAL_SHTSHWRS	-3.18378	0.41413	-7.69
DGRNGOV* Log(NUMI+1)	0.000238	0.000287	0.83
DGRNGOV* Log(NUMI+1)*INC	-1.6E-05	4.85E-07	-32.31
DGRNGOV* Log(NUMI+1)*MICRO	-2.68E-07	1.06E-07	-2.53

Table 5-2
Gas CDA Coefficients (cont'd.)

Label	Estimate	Error	t Value
DGRNGOV	-6.7E-06	1.13E-06	-5.91
DGDRY*FACTADR*GDRYU	0.6391	0.04373	14.62
DGDRY*FACTADR* Log(NUMI+1)	0.50575	0.35162	1.44
DGDRY*FACTADR	-1.53717	0.42913	-3.58
DGPLHT	-1.30781	1.78322	-0.73
DGPLHT*GPLHTFREQ	2.76838	0.06357	43.55
DGPLHT*PLSIZE	0.00046	6.2E-05	7.42
DGPLHT*PLSIZE*DPLCOV	0.000234	3.17E-05	7.39
DGHTSPA	3.5606	0.4036	8.82
DGHTSPA*SPAGHTFREQ	0.81287	0.12965	6.27
DGHTSPA*SPAGHTFREQ*SPASIZE	0.00161	0.02307	0.07
DGHTSPA*SPASIZE*SPCOV	-0.12805	0.10758	-1.19
DGHTSPA*SPAGSOLAR	1.64078	1.04384	1.57
DGMED	27.02511	5.89721	4.58
DGBBQ	2.22319	0.23987	9.27

ENDNOTES

[1] The population of concern is comprised of households in PG&E, SCE, SDG&E, and LADWP electric service territories.

[2] The population of concern is comprised of master-metered units in PG&E, SCE, SDG&E, and LADWP electric service territories.

[3] The 30% success rate for gathering phone numbers is similar to that achieved by RoperASW and KEMA-XENERGY in California. "Do Not Call" lists and other customer privacy considerations have reduced the ability to match up phone numbers to customers in a given sample.

[4] "Roughly" because the special handling of some Zip Codes and customers and the reassignment of cases from telephone interviewers to in-person staff resulted in a total base that could only be approximated.

[5] Energy Commission provided results for LADWP from FERC filing were 7,345 GWh serving 1,378,725 customers for a total energy use per customer of 5,327 kWh. The RASS sample frame yielded 3,581 GWh serving 895,199 customers for a total energy use per customer of 4,064 kWh.

[6] The fall in the percentage missing pre and post cleaning is due to the fall in sample observations which results from the elimination of invalid surveys.

[7] SCE, SCG, and SDG&E provided customer identification variables. The RASS survey was in the field during the spring of 2003. For residences with a change in customers, the final customer was retained in an attempt to correctly match survey information with billing data. DWP and PG&E did not provide a customer identifier on their billing databases.

[8] PG&E and SCG provided bill start and end dates. For SCE, SDG&E, and LADWP only end dates are identified. If the start date of the billing cycle was not provided, the start date was calculated as the end date minus the number of billing days in the billing cycle.

[9] LADWP collects billing data on a bi-monthly basis. The calendarization routine follows the same basic steps for monthly or bi-monthly data. LADWP’s daily dataset simply spreads over a two month period, instead of a one month period. The longer billing period, and the averaging that occurs during the calendarization process, is likely to decrease the estimated impact of degree days on consumption within LADWP’s service territory.

[10] While it is not possible to precisely determine the number of accounts with missing PG&E gas bills, Itron identified 56 customers who stated that they had PG&E gas and their electric bills did not appear to substantiate electric space heat. Upon examination, PG&E found that 37 of these customers had

PG&E gas accounts. While the number of missing gas bills is likely to exceed 37 customers, the relative number of customers with missing gas bills is believed to be very small.

[11] Regional Economic Research, Inc. 2002. *Statewide Residential New Construction Energy Efficiency Baseline Study, Second-Year Report.* Prepared for Pacific Gas & Electric. San Diego, CA, Sept (2002).

[12] Prior to the fuel cleaning and plugging process, homes could state that they did not pay for their heat (*PAYHEAT* = 2, 3, or 97) and still list a system type. This system information is included in *PHTFUEL*.

[13] Prior to the fuel cleaning and plugging process, homes could state that they did not pay for their water heat (*PAYWH* = 2, 3, or 97) and still list a system type. This system information is included in *PWHFUEL*.

[14] Individuals could have answered yes, I have a pool and I pay for its energy use, when pools were located in common areas. Home Owners' association fees often include a set amount for the expense of heating and filtering common area pools. To help reduce this possibility, we restricted the analysis of pools to single family homes.

[15] For a more thorough description of the CDA process, refer to: "The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector" The Rand Journal of Economics, Spring 1980.

[16] Standard TVs were assigned connected loads of 100 W, while big screen units were assigned connected loads of 250 W each.

[17] The correction entails multiplying the lagged value of each variable by an autocorrelation coefficient, then subtracting the resulting product from the current value of the variable in question. The transformation is applied to both the dependent variable (whole-house consumption) as well as to all of the regressors (right-hand model variables).

CALIFORNIA STATEWIDE RESIDENTIAL APPLIANCE SATURATION STUDY

VOLUME 2, STUDY RESULTS FINAL REPORT

Consultant Report

Prepared for:
California Energy Commission

Prepared by:
KEMA-XENERGY
Itron
RoperASW

June 2004
300-00-004

Prepared by:
KEMA-XENERGY
Itron
RoperASW

Contract No. 300-00-004

Prepared for:
California Energy Commission

Contract Manager: ***Glen Sharp***

Sponsors:
Pacific Gas and Electric (PG&E)
San Diego Gas and Electric (SDG&E)
Southern California Edison (SCE)
Southern California Gas Company (SoCalGas)
Los Angeles Department of Water and Power (LADWP)

1: RASS RESULTS INTRODUCTION

For the first time in California, the large Investor Owned Utilities (IOUs) pooled resources and performed a RASS and Unit Energy Consumption (UEC) Study as a team. The project was administered by the California Energy Commission and sponsored by Pacific Gas and Electric (PG&E), San Diego Gas and Electric (SDG&E), Southern California Edison (SCE), Southern California Gas Company (SoCalGas), and Los Angeles Department of Water and Power (LADWP). KEMA-XENERGY was the prime consultant. Itron provided data cleaning and performed the Conditional Demand Analysis. RoperASW fielded the non-response follow-up.

The RASS effort has resulted in a research product that provides both statewide and utility-specific results. The study was designed to allow comparison of results across utility service territories, climate zones and other variables of interest (i.e. dwelling type, dwelling vintage, and income). The study includes results for 21,920 residential customers that are weighted to the population represented by the sponsoring utilities. The saturation results capture both individual and master metered dwellings. This rich set of customer data includes information on all appliances, equipment, and general usage habits. The study also includes a detailed conditional demand analysis that calculates unit energy consumption (UEC) values for all individually metered customers.

The study was initiated in late 2002 and the sampling plans and survey implementation occurred throughout 2003. The data was collected using a two stage direct mail survey targeted to a representative sample of California residential customers. The survey requested customers to provide details on their energy equipment and behaviors. A non-response follow-up survey was implemented at the end of the double mailing phase to a sub-sample of non-respondents. The non-response follow-up included telephone and in-person interviews in an effort to minimize non-response bias by using alternative surveying techniques.

The results from the RASS study were used to develop a CDA model. This analytical method uses a combination of customer energy use with the responses from the customer survey to model end uses and develop unit energy consumption results for those end uses. The results of the CDA are included in summary form along with the general study results in this executive summary and are provided in further detail in the methodology section of the report.

The study also includes onsite metering for a sample of 180 RASS participants. The onsite metering sample was designed to over-sample air conditioning use, with the meters gathering both a whole-house and central air conditioning usage at each dwelling. The onsite meters are in the field at the time of publication and the final results from that portion of the project will be delivered as whole house and air conditioning load shapes after the 2004 cooling season has ended.

Because of the need to serve a wide range of users, the study was designed to produce multiple products:

- A high level summary of key findings;
- Detailed saturation tables for all appliances and equipment holdings;
- Detailed UEC tables for 25 electric and 8 gas end uses;
- Whole house and air conditioning load shapes; and
- An Internet-based tool providing customized data filtering and viewing.

The concept of using a statewide survey instrument provided the CEC and other parties with a consistent set of questions and study results to use for statewide planning and cross utility comparisons. In addition, the sample includes sufficient data enabling utility specific analyses. The project required a cooperative effort among the sponsors as they came together to create a unified research plan, program materials, and implementation strategy. The sponsors all shared project costs and final results. Each utility provided the data necessary to create a unified sampling plan. Each utility also provided customer specific information for customers who were selected for the sample. In order to insure individual customer anonymity, the study participants were assigned a generic identification number that includes details about sampling their strata. Respondent zip codes are the only other information that is generally available in the final study database as to the customer's location. In addition to the "non-confidential" data, each utility received a "confidential" dataset of results for their service territory with customer identification information as provided by the utility initially. This key allows the utility to match up the RASS data with their own account information.

This report is split into two volumes because of the size. Volume One describes the study design and implementation methods while Volume Two details the results in the form of UEC banners and saturation banners.

Volume Two includes a detailed description of the CDA models and results followed by a series of cross tabulations depicting the results to all survey questions. The cross tabs are a series of tables that present weighted final results for the particular group in the set. All cross tab counts have been divided by 1,000 to save space on the page. As an example, the total study population is shown as 10,347 which is 10.347 million customers. Each banner contains answers to all of the questions on the survey (as well as some final plugged/cleaned values as noted in the survey documentation).

Banners consist of three types for the population at large and four for each utility as follows:

- **Banner 1** is by Education, Primary Language, Ethnicity, and Income.
- **Banner 2** is by Square Footage, Type of Occupant, Primary Heating Fuel, Type of Air Conditioning, and Water Heating Fuel.
- **Banner 3** is by Dwelling Type, Ownership, House Occupancy (full or part-time), Dwelling Age, and Gas Utility.
- **Climate** banners are by a combination of CEC Forecast climate zone and a condensed dwelling type (Single Family, Multi Family, and Mobile Home).

2: CONDITIONAL DEMAND ANALYSIS RESULTS

2.1 Introduction

This section presents the results of the CDA analysis. Estimated UECs were derived from the electric and gas models using the estimated model coefficients and the relevant values of the regressors. The final UECs were computed in four steps, as described below.

- First, the UECs were computed at the household level using the estimated UEC equations, the binary appliance variables representing the presence of each end use, and actual 2002 weather for each weather station.
-
- Second, these estimated UEC values were calibrated to benchmark into utility information on overall usage per residential customer for all customers in 2002. This calibration process entailed the multiplication of all UECs for a specific service area by the ratio of two variables: (a) average usage per residential customer (for the overall service area population); and (b) model-based predictions of whole-house consumption for 2002 for the population of customers. Table 2-1 presents these calibration ratios. It should be noted that the application of these factors recognizes that average usage in the sample may differ from population average usage. Such differences may occur as a result of survey response bias, some of which may still be present in spite of considerable efforts to minimize it as part of the overall survey protocol. As shown, the ratios are reasonably close to 1.0, suggesting that response bias is fairly low.

Table 2-1
Calibration Ratios

	Type of Energy	
Utility Service Area	Electricity	Natural Gas
PG&E	1.020	0.935
SCE	1.044	-
SDG&E	0.958	0.924
SoCalGas	-	1.121
LADWP	1.089	-

- Third, the slightly larger SoCalGas calibration ratio, 1.121, may be due to the design of the survey. The reasons for this are as follows.

- ▪The RASS survey was designed as an electric end-use survey. The strata and the resulting sample weights were based on electric service territory and electricity usage, while SoCalGas's weights were set equal to the household's electric weight. This is a simplified weighting method, but it may not accurately weight up to SoCalGas' population.

- ▪The possible survey population for RASS households does not include all groups of homes served by SoCalGas. SoCalGas provides gas to customers from several different municipal electric utilities. The RASS survey population does not include households from small electric municipal service territories.

- ▪Natural gas billing data matched to the RASS survey does not include all homes who believe that their gas company is SoCalGas. During the billing matching process, a concerted effort was made to match survey participants in SoCalGas's territory with their gas bills, however a small percentage of the bills for SoCalGas households may have been missed.

- ▪SoCalGas's calibration ratio may be higher due to their prevalence in LADWP's service territory. The sample provided by LADWP has been found to have some gaps in coverage and there were some apparent problems with either the billing data or the strata definitions.

- Fourth, in a step applying only to weather-sensitive end uses, the UECs were weather normalized through the solution of the model using normal weather conditions for each of the weather stations used in the analysis. These weather-normalized UECs were calibrated using the same factors as derived in the second step. Table 2-2 lists the actual and weather normalized annual heating and cooling degrees for the four electric utilities.

Table 2-2
Actual and Normalized Weather

• Utility	• Actual Weather		• Normalized Weather	
•	• Heating Degree Days	• Cooling Degree Days	• Heating Degree Days	• Cooling Degree Days
• PG&E	• 2697	• 777	• 2421	• 735
• SDG&E	• 1457	• 433	• 1182	• 592
• SCE	• 1609	• 1164	• 1431	• 1261
• LADWP	• 1475	• 946	• 1235	• 1110

- Fifth, as a means of providing summary values of the UEC estimates, household values were used, along with the relevant case weights, to compute weighted averages for various customer segments. While the database of household-level UECs provided to the utilities as a project deliverable can be used to develop UECs for any customer segment, we confine our attention here to the following segmentation variables: residence type (single family homes, town homes, 2-4 unit apartments, 5+ unit apartments, and mobile homes); new home versus existing homes; utility service area; and, for weather-sensitive end uses only, CEC forecasting climate zones along with residence type.

In what follows, we discuss the estimated electric and gas UECs by customer segment. When analyzing these UECs, special care must be taken to account for the end use's saturation and the size of the segment. The estimated UEC for end uses with very low saturations, and/or in segments with very small populations, may not accurately represent the actual energy usage for the end use. We recommend that caution be used when examining UECs from end uses that are the result of fewer than thirty observations and that extreme care be employed if fewer than ten observations were used to calculate the segment's end use UEC. Finally, due to the statistical properties of Conditional Demand Analysis (especially the relative ease of disentangling weather-sensitive end-use consumption levels), the number of observations needed to accurately determine a segment's end use UEC, will be larger for non-weather sensitive end-uses than for space conditioning and weather sensitive end-uses.

2.1.1. Estimated Electric UECs

Estimated calibrated and weather-normalized electric UECs, segment frequencies, and the associated saturations are presented in Tables 2-3 through 2-18. Table 2-3 provides UECs, segment frequencies, and the associated saturations by residence type. Tables 2-4 and 2-5 provide estimates by structural vintage (dwelling age). Table 2-6 presents estimates by service area. Tables 2-7 and 2-8 provide estimates of weather-sensitive end uses by CEC Forecasting Climate Zone. Finally, Tables 2-9 through 2-18 present UECs, segment frequencies, and saturations of space conditioning end uses by CEC Forecasting Climate Zone and residence type.

These UEC estimates are discussed briefly below, with special emphasis on values that may differ appreciably from values used by the CEC and/or the utilities in prior work. The discussion is organized by end use groupings, and focuses on the results by residence type. Prior to Table 2-4 is a discussion of the estimates by structural vintage. As discussed above, care must be used when interpreting the UEC estimates for smaller segments with low saturations.

Space Heating

UECs were developed for both conventional (resistance) electric space heating and heat pump space heating. As shown in Table 2-3, weather-normalized conventional space heating varies from 584 kWh for 2-4 unit apartments to just under 1,500 kWh per year for single family homes. Heat pump space heating (note that this excludes the cooling side of heat pump usage) ranges from 315 kWh for 2-4 unit apartments to 1,077 kWh for single family residences. It should be noted here that comparisons of the conventional and heat pump heating UECs should not be used to infer savings associated with replacing resistance heating with heat pump heating, insofar as the characteristics of the households with these two types of systems differ considerably. Homes with heat pumps tend to be newer and larger than those of their counterparts with conventional electric heating.

The UECs for CEC Forecast Climate Zone 7, presented in Tables 2-7 and 2-9 through 2-18, help to demonstrate the care that needs to be used when end use UECs are divided into population segments. The segment frequencies presented in Table 2-7 shows that the RASS survey included 384 households from Forecast Climate Zone 7. The weighted saturations in Table 2-7 for this zone indicate that only one percent of these households had conventional electric heating. While the UEC for this end use appears to be reasonable, it was calculated using household characteristics from five or fewer households. Tables 2-9 through 2-18 further cut Zone 7 into residence types. The conventional electric heating UECs for Zone 7 by residence type were calculated using, at most, two to three households for each residence type. If these few households differ substantially from the norm for their residence type, these UECs will be unreliable estimates of the energy consumption for this population segment.

Ventilation

UECs were derived for furnace fans. Furnace fan usage (defined as the forced air fan energy used to distribute gas space heating, but not electric space heating and not central air conditioning) varies from 51 kWh for apartment units in buildings with 5+ units to 162 kWh for single family homes.

Air Conditioning

Three air conditioning UECs were developed: central air conditioning, room air conditioning, and evaporative coolers. Central air UECs range from to just over 700 kWh for town homes to just over 1,400 kWh per year for single family dwellings. In general, these values are lower than expected based on prior research relating to residential energy usage in California. However, our confidence in these results is high, given the relative ease of isolating air conditioning usage through statistical analyses like CDA. Remembering that the billing data reflecting summer usage were from 2002, one could hypothesize that the lower than expected central air conditioning UECs reflects the influence of the energy crisis of 2001 and the programs promoting conservation(e.g., the 20/20 program) in the summer of 2002. Room air conditioning UECs are also fairly low relative to prior estimates, varying from105 kWh for multi-family units in buildings with 5+ units to 227 kWh for single family homes and mobile homes. Evaporative cooling UECs vary from 374 kWh to 688 kWh across residence types. Again, we caution the reader that comparisons of evaporative cooling UECs with air conditioning UECs should not be used to infer potential savings from replacing the latter with the former, since the characteristics of households with these appliances vary.

Water Heating

Conventional electric water heating UECs vary from 1,567 for multi-family units in large buildings to 3,079 kWh per year for single family homes. These results are generally consistent with the results of metering studies, but lower than often derived from engineering calculations. A considerable effort was made in cleaning data relating to water heating and trying different model specifications for this end use, so we are reasonably confident in the reasonableness of these results. Unsurprisingly, solar electric water heating UECs tend to be considerably lower than conventional electric UECs; however, nothing should be inferred from these estimates about the electric consumption impact of adding solar assist. Care should also be used when using the solar water heat UECs due to extremely low saturations.

Dishwashers, Clothes Washers and Dryers

Dishwasher UECs range from 47 kWh for mobile homes to 84 kWh for single family dwellings. Clothes washer UECs, which include only motor loads and not incremental water heating usage, vary from 11 kWh in mobile homes to 127 kWh in single family residences. While the single family and multi-family estimates are probably reasonable, the mobile home estimated may be anomalous. Electric dryer UECs range from 429 kWh to 713 kWh across residence types.

Refrigerators and Freezers

First refrigerator UECs fall in the range of 721 to 824 kWh for the represented residence types. Second refrigerator UECs (which reflect the usage of all secondary units) are higher than primary unit UECs for single family homes and mobile homes (as a result of both multiple units and lower efficiencies), but slightly lower for multi-family homes (where secondary units may be very small). Freezer UECs all fall in the range between 877 and 964 kWh.

Pools and Spas

The pool pump UEC for single family homes is 2,671 kWh. Spa pumps use from 180 kWh for mobile homes to 467 kWh for single family homes. Note that the RASS survey included 563 mobile homes, and that only three percent of these residences have a spa. The mobile home spa pump UEC is based on the characteristics of fewer than twenty households, and should not be taken too literally. Spa electric heating ranges from 694 for town homes to over 3,500 for mobile homes. Again, though, note that fewer than fifteen mobile homes have an electric spa. With a sample this small, the electric spa UECs should only be used with extreme caution.

Outdoor Lighting

Outdoor lighting UECs vary from 173 kWh to 284 kWh. We are unaware of other estimates of UECs for this end use, so it is difficult to assess the reasonableness of these estimates. However, the single family estimate would be consistent with three 75 Watt bulbs being used 3.5 hours per night, and this seems to be a reasonable order of magnitude.

Cooking

The Electric range/oven UEC varies from 191 to 301 kWh across residence types. Microwave oven UECs range from 113 kWh to 140 kWh.

Televisions

Television UECs (usage from all units, rather than usage per unit) fall into the range of 436 to 519 kWh. These estimates are slightly lower than estimates derived in some previous studies, but seem reasonable.

Personal Computers and Home Offices

Personal computer UECs vary from 458 to 591 kWh. Home office UECs are considerably smaller than this (ranging from 121 to 158 kWh), and they should be interpreted as the incremental effects of home office equipment over and above personal computers.

Water Beds

Water bed heating UECs are in the range of 732 to 840 kWh.

Well Pumps

Well pumps are estimated to use, on average, anywhere from 724 to 862 kWh. While we did allow multi-family homes to have well pumps, we are suspicious of the positive responses in these residence types and would suggest that the associated UECs not be considered particularly reliable.

Miscellaneous

Miscellaneous usage is estimated to fall in the range of 1,257 to 2,147 kWh per year. A major element of miscellaneous usage would be interior lighting; other elements would be various plug loads and ceiling and attic fan. Ceiling and attic fans have been included in the miscellaneous term due to unreliable survey responses and unrealistic UEC estimates for these end uses.

It was not possible to disentangle indoor lighting usage from the other loads covered by miscellaneous usage. However, we believe that a good estimate of indoor lighting usage is approximately sixty percent of the residence types miscellaneous usage. Given this assumption, indoor lighting usage would range from 754 kWh in 5+ unit apartment residences to 1288 kWh in single family homes.

Table 2-4 lists the electric UECs for new and old homes. Estimates show that older homes use an average of 5,846 kWh while newer homes use 7,035 kWh. Much of this increased energy usage is due to the increased size of newer homes. New dwellings are 42% larger than the average for the existing housing stock, 2,061 and 1,448 square foot, respectively. The increased size is due in part to the preponderance of single-family homes among the newer housing stock. 74% of newer homes are single family residences while only 58% of the existing housing stock are single family residence. Residents of newer homes also have a substantially higher average household income than residents of older homes, $87,402 verses $58,978. Finally, the saturation of central air conditioners, pools, and computers is much higher for newer homes than for the existing housing stock.

Table 2-3
Electric UECS Calibrated and Normalized, by Residence Type

	Single Family		Town Home		2-4 Unit Apt		5+ Unit Apt		Mobile Home	
	UEC	**Sat.**	**UEC**	**Sat.**	**UEC**	**Sat.**	**UEC**	**Sat.**	**UEC**	**Sat.**
All Household	7,105	13,824 homes	4,469	1,780 homes	3,877	1,608 homes	3,807	3,377 homes	5,662	563 homes
Conv. Eheat	1,494	0.04	724	0.06	584	0.15	658	0.23	1,150	0.10
HP Eheat	1,077	0.01	392	0.01	315	0.02	335	0.05	1,031	0.03
Aux Eheat	296	0.28	114	0.21	85	0.19	74	0.13	298	0.31
Furnace Fan	162	0.68	73	0.54	65	0.32	51	0.26	118	0.58
Central Air	1,423	0.46	713	0.41	1,019	0.28	749	0.32	1,143	0.39
Room Air	227	0.15	148	0.14	120	0.16	105	0.22	227	0.34
Evap Cooling	688	0.05	595	0.02	374	0.02	403	0.02	537	0.27
Water Heat	3,079	0.05	1,723	0.04	1,657	0.09	1,567	0.10	3,258	0.17
Solar Water Heater	1,708	0.00	407	0.00	.	0.00	32	0.00	.	0.00
Dryer	713	0.34	591	0.32	429	0.17	548	0.17	549	0.42
Clothes Washer	127	0.95	63	0.76	62	0.37	14	0.26	11	0.86
Dish Washer	84	0.70	63	0.61	66	0.38	59	0.48	47	0.55
First Refrigerator	824	1.00	769	1.00	722	1.00	721	1.00	809	1.00
Second Refrigerator	1,245	0.25	739	0.11	700	0.06	586	0.04	1,143	0.13
Freezer	937	0.24	877	0.09	964	0.07	908	0.04	951	0.30
Pool Pump	2,671	0.14	.	0.00	.	0.00	.	0.00	.	0.00
Spa	467	0.13	270	0.03	.	0.00	.	0.00	180	0.03
Outdoor Lighting	284	0.67	173	0.56	228	0.32	206	0.25	232	0.56
Range/Oven	301	0.41	240	0.44	191	0.41	207	0.49	208	0.27
TV	519	0.96	465	0.92	439	0.92	436	0.96	457	0.93
Spa Electric Heat	1,719	0.07	694	0.02	.	0.00	.	0.00	3,550	0.02
Microwave	140	0.97	125	0.92	125	0.91	122	0.92	113	0.96
Home Office	148	0.20	158	0.19	145	0.17	144	0.15	121	0.13
PC	578	0.75	591	0.68	521	0.54	532	0.59	458	0.45
Water Bed	840	0.02	748	0.02	732	0.00	757	0.01	773	0.03
Well Pump	862	0.05	842	0.01	911	0.01	816	0.01	724	0.18
Miscellaneous	2,147		1,532		1,339		1,257		1,462	

Table 2-4
Electric UEC by House Age

	New House		Old House	
	UEC	Saturation	UEC	Saturation
All Household	7,035	1,393 homes	5,846	19,760 homes
Conv. Eheat	1,167	0.05	861	0.09
HP Eheat	414	0.01	595	0.02
Aux Eheat	319	0.19	240	0.24
Furnace Fan	167	0.82	136	0.53
Central Air	1,411	0.77	1,215	0.39
Room Air	302	0.06	178	0.17
Evap Cooling	1,013	0.01	616	0.04
Water Heat	2,858	0.04	2,371	0.07
Solar Water Heater	.	0.00	1,345	0.00
Dryer	746	0.33	657	0.29
Clothes Washer	131	0.90	107	0.73
Dish Washer	84	0.92	76	0.60
First Refrigerator	763	1.00	791	1.00
Second Refrigerator	999	0.24	1,193	0.17
Freezer	861	0.19	940	0.18
Pool Pump	2,712	0.13	2,667	0.08
Spa	455	0.14	461	0.08
Outdoor Lighting	418	0.64	253	0.54
Range/Oven	316	0.41	260	0.42
TV	542	0.96	486	0.95
Spa Electric Heat	988	0.06	1,761	0.04
Microwave	137	0.98	133	0.95
Home Office	152	0.23	147	0.18
PC	580	0.84	564	0.68
Water Bed	762	0.03	823	0.01
Well Pump	858	0.04	849	0.04
Miscellaneous	1,820		1,833	

Table 2-5 lists the whole household electric UEC by utility and residence type. These calculations show that the statewide increase in electricity usage in newer homes is due to the increased usage in single family homes. All four of the electric utilities experienced an increase in electricity usage in newer single family homes. Three out of the four utilities, however, have a reduction in usage for newer multi-family homes (town homes, 2-4 unit apts, and 5+ unit apts) as compared to their existing multi-family housing stock.

Table 2-5
Electric Household UEC by House Age, Utility and Residence Type

	New House		Old House	
	Household UEC	Count	Household UEC	Count
All	7,035	1,393	5,846	19,760
All PG&E	7,013	689	6,215	8,576
SF PG&E	8,117	537	7,278	5,926
MF PG&E	3,451	145	4,134	2,395
All SDG&E	6,340	199	5,358	2,328
SF SDG&E	7,170	163	6,456	1,515
MF SDG&E	3,060	36	3,612	779
All SCE	7,659	468	6,018	7,511
SF SCE	8,203	354	7,082	4,895
MF SCE	4,430	104	4,089	2,370
All LADWP	3,219	37	4,084	1,345
SF LADWP	6,128	8	5,437	426
MF LADWP	2,888	28	3,598	909

Table 2-6
Electric UECs by Utility

	PG&E		SDG&E		SCE		LADWP	
	UEC	Saturation	UEC	Saturation	UEC	Saturation	UEC	Saturation
All Household	6,265	9,265 homes	5,445	2,527 homes	6,102	7,979 homes	4,071	1,382 homes
Conv. Eheat	1,113	0.10	581	0.13	734	0.06	542	0.09
HP Eheat	799	0.02	458	0.03	555	0.01	201	0.03
Aux Eheat	331	0.26	156	0.24	192	0.23	103	0.17
Furnace Fan	180	0.58	91	0.60	115	0.56	71	0.26
Central Air	1,108	0.39	644	0.35	1,494	0.48	1,075	0.29
Room Air	181	0.14	63	0.09	202	0.20	158	0.25
Evap Cooling	469	0.05	277	0.01	797	0.05	372	0.02
Water Heat	2,585	0.09	2,151	0.06	2,342	0.05	1,387	0.05
Solar Water Heater	1,193	0.00	1,501	0.01	1,508	0.00	.	0.00
Dryer	652	0.45	648	0.26	717	0.18	474	0.07
Clothes Washer	97	0.78	75	0.77	129	0.77	125	0.36
Dish Washer	77	0.67	69	0.71	80	0.60	73	0.27
First Refrigerator	788	1.00	780	1.00	801	1.00	754	1.00
Second Refrigerator	1,201	0.19	1,054	0.19	1,210	0.19	933	0.06
Freezer	928	0.23	841	0.17	983	0.15	880	0.05
Pool Pump	2,580	0.08	2,557	0.12	2,772	0.10	3,096	0.02
Spa	428	0.08	445	0.12	495	0.10	423	0.02
Outdoor Lighting	260	0.56	268	0.53	276	0.55	218	0.42
Range/Oven	268	0.61	241	0.49	271	0.27	200	0.17
TV	474	0.95	446	0.94	520	0.96	479	0.94
Spa Electric Heat	1,346	0.05	903	0.06	2,514	0.04	895	0.01
Microwave	131	0.95	119	0.96	139	0.96	140	0.89
Home Office	152	0.20	159	0.19	141	0.16	134	0.18
PC	602	0.72	614	0.78	515	0.66	516	0.55
Water Bed	787	0.02	925	0.01	818	0.02	848	0.00
Well Pump	829	0.08	831	0.01	952	0.02	890	0.01
Miscellaneous	1,852		1,750		1,912		1,495	

Table 2-7
Electric UEC for Weather Sensitive End Uses
in Forecast Zones 1-7

	Forecast 1		Forecast 2		Forecast 3		Forecast 4		Forecast 5		Forecast 7	
	UEC	Sat.	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All Hhold	7,519	780 homes	6,668	804 homes	7,052	1,676 homes	6,544	3,314 homes	4,971	2,691 homes	7,088	384 homes
Conv. Eheat	1,580	0.15	1,306	0.08	1,232	0.09	1,107	0.09	915	0.13	1,235	0.01
HP Eheat	1,225	0.03	664	0.04	1,148	0.02	605	0.01	572	0.02	953	0.00
Aux Eheat	464	0.26	394	0.22	285	0.26	338	0.25	310	0.28	434	0.29
Furnace Fan	274	0.43	170	0.63	152	0.57	180	0.68	189	0.48	194	0.70
Central Air	941	0.41	1,082	0.69	1,548	0.67	885	0.42	226	0.06	1,902	0.57
Room Air	106	0.18	176	0.24	326	0.25	94	0.12	20	0.04	247	0.12
Evap Cooling	313	0.11	375	0.05	618	0.12	320	0.03	46	0.00	606	0.26
Water Heat	2,668	0.35	2,361	0.10	3,010	0.13	2,592	0.05	1,913	0.06	2,979	0.07
Solar Water Heater	932	0.00	1,587	0.00	1,711	0.00	794	0.00	.	0.00	.	0.00

Table 2-8
Electric UECs for Weather Sensitive End Uses
in Forecast Zones 8 to 13

	Forecast 8		Forecast 9		Forecast 10		Forecast 11		Forecast 12		Forecast 13	
	UEC	Sat.	UEC	Sat	UEC	Sat	UEC	Sat.	UEC	Sat	UEC	Sat
All Hhold	5,417	3,175 homes	5,660	2,461 homes	7,529	1,959 homes	3,736	951 homes	4,849	431 homes	5,445	2,527 homes
Conv. Eheat	571	0.08	837	0.05	969	0.05	560	0.07	515	0.12	581	0.13
HP Eheat	445	0.01	495	0.01	769	0.01	177	0.03	254	0.03	458	0.03
Aux Eheat	153	0.19	170	0.30	220	0.21	95	0.20	134	0.12	156	0.24
Furnace Fan	94	0.52	112	0.45	127	0.73	64	0.19	77	0.43	91	0.60
Central Air	848	0.36	1,509	0.40	1,908	0.74	915	0.15	1,169	0.61	644	0.35
Room Air	126	0.15	215	0.26	262	0.21	153	0.19	164	0.39	63	0.09
Evap Cooling	286	0.01	772	0.03	934	0.12	369	0.02	379	0.02	277	0.01
Water Heat	1,955	0.06	2,392	0.03	2,800	0.05	1,399	0.03	1,377	0.10	2,151	0.06
Solar Water Heater	3,586	0.00	.	0.00	1,370	0.00	.	0.00	.	0.00	1,501	0.01

Table 2-9
Space Conditioning Electric UEC for Single Family Residences in Forecast Zones 1-7

Residence 1	Forecast 1		Forecast 2		Forecast 3		Forecast 4		Forecast 5		Forecast 7	
Single Family	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All HHold	8,633	607 homes	7,390	653 homes	8,139	1208 homes	7,415	2409 homes	6,047	1586 homes	7,454	288 homes
Conv. Eheat	1,770	0.14	1,856	0.03	1,553	0.08	1,671	0.03	1,639	0.03	1,473	0.01
HP Eheat	1,305	0.03	1,102	0.02	1,325	0.02	1,190	0.00	1,217	0.01	1,403	0.00
Aux Eheat	570	0.25	432	0.25	353	0.26	388	0.29	393	0.36	395	0.30
Central Air	1,003	0.44	1,213	0.72	1,749	0.70	1,053	0.43	278	0.09	1,985	0.56
Room Air	117	0.17	210	0.22	407	0.21	110	0.12	24	0.05	263	0.11

Table 2-10
Space Conditioning Electric UEC for Single Family Residences in Forecast Zones 8-13

Residence 1	Forecast 8		Forecast 9		Forecast 10		Forecast 11		Forecast 12		Forecast 13	
Single Family	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All HHold	6,499	1850 homes	6,760	1692 homes	8,351	1419 homes	4,918	295 homes	7,484	139 homes	6,536	1678 homes
Conv. Eheat	1,152	0.01	1,197	0.03	1,550	0.03	1,115	0.03	1,437	0.01	1,182	0.04
HP Eheat	1,107	0.00	1,246	0.00	893	0.01	.	0.00	513	0.00	751	0.02
Aux Eheat	186	0.25	211	0.34	250	0.21	161	0.26	241	0.22	184	0.30
Central Air	1,119	0.36	1,674	0.49	1,941	0.76	1,623	0.17	1,715	0.82	784	0.38
Room Air	178	0.12	246	0.27	296	0.18	333	0.16	227	0.15	84	0.07

Table 2-11
Space Conditioning Electric UEC for Town Homes in Forecast Zones 1-7

Residence 2	Forecast 1		Forecast 2		Forecast 3		Forecast 4		Forecast 5		Forecast 7	
Town Home	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All HHold	3,475	25 homes	4,507	40 homes	4,495	70 homes	4,963	304 homes	4,280	281 homes	5,326	16 homes
Conv. Eheat	.	0.00	1,530	0.05	485	0.12	1,074	0.05	842	0.06	.	0.00
HP Eheat	678	0.04	401	0.05	505	0.02	.	0.00	.	0.00	157	0.04
Aux Eheat	190	0.12	107	0.20	100	0.20	185	0.25	169	0.26	184	0.21
Central Air	659	0.65	623	0.43	912	0.51	500	0.59	143	0.07	957	0.80
Room Air	25	0.01	53	0.15	245	0.44	51	0.07	6	0.01	144	0.32

Table 2-12
Space Conditioning Electric UEC for Town Homes in Forecast Zones 8-13

Residence 2	Forecast 8		Forecast 9		Forecast 10		Forecast 11		Forecast 12		Forecast 13	
Town Home	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All HHold	4,723	353 homes	4,218	212 homes	5,673	109 homes	3,414	89 homes	4,650	51 homes	3,930	230 homes
Conv. Eheat	623	0.06	824	0.05	943	0.09	395	0.09	185	0.02	549	0.08
HP Eheat	360	0.02	345	0.01	367	0.01	.	0.00	668	0.02	278	0.01
Aux Eheat	77	0.16	55	0.22	106	0.21	71	0.14	111	0.15	53	0.28
Central Air	523	0.47	872	0.44	1,514	0.84	877	0.04	1,066	0.62	356	0.28
Room Air	78	0.19	239	0.20	160	0.29	292	0.09	142	0.43	50	0.07

Table 2-13
Space Conditioning Electric UEC for 2-4 Unit Apartments in Forecast Zones 1-7

Residence 3	Forecast 1		Forecast 2		Forecast 3		Forecast 4		Forecast 5		Forecast 7	
2-4 Unti Apt	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All HHold	2,945	45 homes	4,285	38 homes	3,928	98 homes	4,674	181 homes	3,622	278 homes	8,064	22 homes
Conv. Eheat	207	0.11	1,035	0.23	709	0.15	903	0.19	814	0.15	368	0.06
HP Eheat	.	0.00	206	0.17	798	0.01	597	0.06	445	0.00	.	0.00
Aux Eheat	87	0.38	97	0.15	73	0.16	90	0.17	135	0.27	122	0.09
Central Air	725	0.22	517	0.79	931	0.58	533	0.32	103	0.01	2,818	0.68
Room Air	79	0.03	55	0.20	176	0.30	61	0.09	14	0.03	247	0.17

Table 2-14
Space Conditioning Electric UEC for 2-4 Unit Apartments in Forecast Zones 8-13

Residence 3	Forecast 8		Forecast 9		Forecast 10		Forecast 11		Forecast 12		Forecast 13	
2-4 Unit Apt	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All HHold	4,122	294 homes	3,397	159 homes	5,221	107 homes	3,155	168 homes	3,852	40 homes	3,368	179 homes
Conv. Eheat	487	0.21	492	0.05	311	0.17	269	0.05	335	0.06	342	0.26
HP Eheat	358	0.02	132	0.03	555	0.00	131	0.03	49	0.00	222	0.02
Aux Eheat	57	0.11	48	0.18	91	0.28	57	0.19	83	0.09	57	0.23
Central Air	457	0.26	1,063	0.19	2,326	0.69	836	0.05	1,244	0.60	354	0.27
Room Air	107	0.13	172	0.21	165	0.22	77	0.15	131	0.40	35	0.14

Table 2-15
Space Conditioning Electric UEC for 5+ Unit Apartments in Forecast Zones 1-7

Residence 4	Forecast 1		Forecast 2		Forecast 3		Forecast 4		Forecast 5		Forecast 7	
5+ Apt	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All HHold	3,503	51 homes	4,116	63 homes	4,238	159 homes	4,195	381 homes	3,748	526 homes	4,317	20 homes
Conv. Eheat	951	0.22	1,005	0.26	657	0.07	933	0.33	811	0.36	1,295	0.04
HP Eheat	910	0.04	313	0.02	386	0.02	434	0.04	443	0.06	546	0.02
Aux Eheat	137	0.35	132	0.06	98	0.26	95	0.10	107	0.12	117	0.14
Central Air	662	0.33	551	0.45	1,084	0.76	381	0.31	93	0.05	1,349	0.80
Room Air	64	0.09	108	0.40	165	0.19	63	0.20	13	0.04	217	0.09

Table 2-16
Space Conditioning Electric UEC for 5+ Unit Apartments in Forecast Zones 8-13

Residence 4	Forecast 8		Forecast 9		Forecast 10		Forecast 11		Forecast 12		Forecast 13	
5+ A[t	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All HHold	3,758	652 homes	3,478	355 homes	4,682	175 homes	3,229	396 homes	4,419	193 homes	3,550	406 homes
Conv. Eheat	540	0.20	497	0.10	467	0.13	539	0.11	534	0.18	414	0.38
HP Eheat	391	0.04	372	0.02	295	0.02	187	0.06	242	0.05	262	0.07
Aux Eheat	65	0.11	51	0.24	96	0.12	49	0.17	67	0.09	61	0.07
Central Air	503	0.35	860	0.19	1,716	0.74	518	0.21	892	0.55	326	0.32
Room Air	86	0.23	121	0.28	164	0.27	84	0.24	169	0.46	47	0.15

Table 2-17
Space Conditioning Electric UEC for Mobile Homes in Forecast Zones 1-7

Residence 5	Forecast 1		Forecast 2		Forecast 3		Forecast 4		Forecast 5		Forecast 7	
Mobile Home	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All HHold	6,384	52 homes	8,174	10 homes	6,328	141 homes	6,998	39 homes	4,724	20 homes	5,074	38 homes
Conv. Eheat	1,525	0.29	3,092	0.10	972	0.12	1,642	0.06	1,118	0.12	.	0.00
HP Eheat	1,095	0.02	923	0.04	1,056	0.08	.	0.00	582	0.05	.	0.00
Aux Eheat	402	0.33	406	0.42	221	0.42	172	0.41	281	0.29	796	0.42
Central Air	627	0.11	805	0.90	1,167	0.45	511	0.13	185	0.30	1,009	0.40
Room Air	81	0.58	.	0.00	252	0.48	85	0.10	18	0.05	207	0.17

Table 2-18
Space Conditioning Electric UEC for Mobile Homes in Forecast Zones 8-13

Residence 5	Forecast 8		Forecast 9		Forecast 10		Forecast 11		Forecast 12		Forecast 13	
Mobile Home	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All HHold	3,539	26 homes	4,816	43 homes	5,604	149 homes	4,903	3 homes	4,320	8 homes	4,307	34 homes
Conv. Eheat	690	0.02	1,189	0.03	940	0.02	.	0.00	.	0.00	902	0.23
HP Eheat	.	0.00	1,495	0.00	786	0.01	.	0.00	.	0.00	.	0.00
Aux Eheat	125	0.09	169	0.24	221	0.24	143	0.36	274	0.26	177	0.12
Central Air	830	0.28	1,132	0.49	1,580	0.47	492	0.36	808	0.50	553	0.35
Room Air	284	0.07	86	0.13	299	0.45	.	0.00	248	0.25	87	0.03

2.1.2. Estimated Natural Gas UECs

Estimated calibrated and weather-normalized natural gas UECs and the associated saturations are presented in Tables 2-19 through 2-25. Table 2-19 provides UECs and the associated saturations by residence type. Tables 2-20 and 2-21 provide estimates by structural vintage (home age). Table 2-22 presents estimates by service area. Finally, Tables 2-23 through 2-25 provide estimates of weather-sensitive end uses by CEC Forecasting Climate Zone. These UEC estimates are discussed briefly below, with special emphasis on values that may differ appreciably from values used by the CEC and/or the utilities in prior work. The discussion is organized by end use groupings, and concentrates strictly on the results by residence type.

The presentation of the Gas UECs and saturations differs slightly from the electric results. The RASS survey was designed as an electric survey. A total of 5,034 of the 21,153 individually metered households in the survey are households for which we have with no natural gas billing data. Some of the 5,034 households may have gas service. Some of these households may receive their gas from smaller municipal gas utilities not included in the survey or the billing matching process may have been unable to obtain and match their gas bills to the electric RASS frame. Gas UEC were calculated for 17,382, leaving 3,771 households as all electric households.

Tables 2-19 through 2-25 present Gas UECs and saturations for all RASS household by residence type and for gas RASS households by residence type. The gas RASS households are limited to RASS households with natural gas billing data from one of the three gas utilities, PG&E, SDG&E, or SoCal Gas. Examination of the two sets of results, the All Homes and the Homes w/Gas, indicates that the end-use saturations differ substantially while the end use UECs remain relatively constant. The saturations are higher for homes with gas bills and lower if all homes are used to determine the saturation rates. To determine if a population segment is sufficiently large to produce reliable estimate of the end use UEC, analysts should use the Homes w/Gas saturations and population counts listed with the saturation data in the all household UEC row.

As stated above, the Homes w/Gas Data columns are restricted to homes with gas billing data from PG&E, SDG&E, and SoCalGas. Homes that receive their natural gas service from other, small providers, have estimated UECs, but the lack of gas billing data eliminated these households from the Homes w/Gas column. Unfortunately, due to the difficulty of identifying SoCal Gas customers, some SoCal Gas customers also lack gas billing data. Gas UECs have been estimated for these customers, but those without billing data are not included in the Homes w/Gas columns.

Space Heating

Primary gas Heating UECs vary quite a bit across residence types, from 95 therms per year for multi-family units in large apartment buildings to 243 therms for single family homes. Auxiliary gas space heat UECs range from 37 to 73 therms.

Water Heating

Conventional (non-solar) gas water heat UECs fall in the range of 183 to 206 therms. Solar assisted gas water hear UECs are only modestly below these values, but it should again be kept in mind that the homes with solar heating are very different from those with conventional gas water heat. While the gas solar water heating UECs appear reasonable, caution is called for when using them. Extremely low saturation rates may reduce their reliability.

Dryers

Gas dryers are estimated to use between 13 and 31 therms per year, depending on residence type.

Ranges/Ovens

Gas range/oven UECs are estimated to be between 28 and 46 therms.

Pool and Spa Heat

The overall average gas pool heating UEC is 222 therms, while the spa heating UECs vary from 81 to 114 therms per year. Given the samples sizes, 11,273 single family homes and 247 mobile homes with gas data, and the extremely low saturation rates, the single family spa heating UEC may be a more reliable estimate of the true gas spa usage.

Miscellaneous

Miscellaneous gas usage is estimated to be 1 or 2 therms per year.

Table 2-20 lists the Gas UECs for new and existing homes. While the data appear to indicate that newer homes use slightly more gas than older homes, the correct interpretation of the estimates is more complicated. Table 2-21 lists the whole household UEC for new and existing homes by utility and residence. These data show that newer homes in PG&E and SoCalGas use less gas than existing homes, while new single family homes in SDG&E use more gas than older homes. New homes in Table 2-20 appear to use more gas than older homes because most new homes are more predominantly single family homes and because SDG&E's new single family homes use substantially more gas than their older homes. These two characteristics of the data appear to dominate the statewide averages, hiding the fact that newer homes in PG&E's and SoCal Gas' service territories use less gas than older homes.

Table 2-19
Gas UECs and Saturations, by Residence Type, for all Households and for Homes w/Gas Account Data

	Single Family				Town Home				2-4 Unit Apt				5+ Unit Apt				Mobile Home			
	All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data	
	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.
All Household UEC	454	13,824 homes	508	11,273 homes	300	1,780 homes	326	1,496 homes	222	1,608 homes	284	1,195 homes	151	3,377 homes	232	1,908 homes	235	563 homes	433	247 homes
Primary Heat	242	0.87	243	0.98	114	0.85	114	0.95	101	0.66	102	0.85	92	0.49	95	0.75	216	0.53	209	0.99
Auxiliary Heat	71	0.00	73	0.00	40	0.00	44	0.00	37	0.00	41	0.00	36	0.00	37	0.00	72	0.00	72	0.00
Conv. Gas Water Heat	206	0.89	206	0.99	195	0.87	194	0.95	184	0.69	183	0.89	185	0.46	186	0.71	192	0.53	193	0.99
Solar Water Heat w/Gas Backup	162	0.00	160	0.00	170	0.00	170	0.00	138	0.00	148	0.00	117	0.00	114	0.00	.	0.00	.	0.00
Dryer	31	0.50	31	0.55	24	0.32	23	0.35	24	0.14	23	0.17	21	0.07	20	0.11	13	0.19	13	0.39
Range/Oven	46	0.66	46	0.73	39	0.57	38	0.60	40	0.58	42	0.73	38	0.48	37	0.69	26	0.49	28	0.90
Pool Heat	220	0.03	222	0.04	.	0.00	.	0.00	.	0.00	.	0.00	.	0.00	.	0.00	.	0.00	.	0.00
Spa Heat	81	0.06	81	0.07	90	0.01	89	0.01	.	0.00	.	0.00	.	0.00	.	0.00	114	0.01	114	0.03
Miscellaneous	2		2		3		1		1		1		0		0		1		2	

Table 2-20
Gas End Use UECs and Saturations by House Age for all Households and for Homes w/Gas Account Data

	New House				Old House			
	All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data	
	UEC	Saturation	UEC	Saturation	UEC	Saturation	UEC	Saturation
All Household	370	1,393 homes	434	1,107 homes	355	19,760 homes	431	16,119 homes
Primary Heat	198	0.83	199	0.96	201	0.76	202	0.93
Auxiliary Heat	82	0.00	82	0.00	61	0.00	65	0.00
Conv. Gas Water Heat	160	0.84	162	0.97	203	0.77	203	0.93
Solar Water Heat w/Gas Backup	142	0.00	142	0.00	152	0.00	157	0.00
Dryer	31	0.49	31	0.57	30	0.35	30	0.42
Range/Oven	42	0.77	42	0.89	43	0.60	43	0.70
Pool Heat	259	0.05	261	0.06	214	0.02	215	0.02
Spa Heat	85	0.08	84	0.10	80	0.04	81	0.04
Miscellaneous	4		5		1		2	

Table 2-21
Gas Household UECs by House Age By Utility and by Residence type

	New Homes		Older Homes	
	Gas Household UEC	Count	Gas Household UEC	Count
All Utilities	434	1,107	431	15,012
All PG&E	427	490	436	6,255
SF PG&E	489	391	516	4,431
MF PG&E	224	98	267	1750
All SDG&E	377	142	349	1,544
SF SDG&E	423	122	399	1,149
MF SDG&E	183	20	223	375
All SoCalGas	457	475	443	7,213
SF SoCalGas	508	362	529	4,818
MF SoCalGas	243	105	283	2,251

Table 2-22
Gas UECs and Saturations by Utility for all Households and for Homes w/Gas Account Data[1]

	PG&E				SDG&E				SCG				Other	
	All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data			
	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.	UEC	Sat.
All Household	343	8789 homes	436	6747 homes	279	2275 homes	351	1686 homes	407	8773 homes	443	7688 homes	179	1316 homes
Primary Heat	245	0.74	245	0.94	135	0.74	136	0.91	181	0.85	181	0.93	188	0.38
Auxiliary Heat	84	0.00	85	0.00	33	0.01	34	0.01	52	0.00	57	0.00	40	0.00
Conv. Gas Water Heat	183	0.74	183	0.94	180	0.76	181	0.96	219	0.85	219	0.93	197	0.42
Solar Water Heat w/Gas Backup	133	0.00	144	0.00	149	0.00	155	0.00	176	0.00	176	0.00	167	0.00
Dryer	25	0.22	25	0.28	23	0.43	23	0.54	33	0.48	33	0.53	26	0.21
Range/Oven	37	0.42	37	0.53	35	0.58	35	0.71	48	0.80	48	0.86	38	0.41
Pool Heat	224	0.02	225	0.02	218	0.03	217	0.04	218	0.03	222	0.03	206	0.00
Spa Heat	76	0.02	76	0.03	86	0.05	86	0.07	82	0.05	83	0.06	69	0.01
Miscellaneous	1		1		2		2		2		2		1	

[1]For households w/"other" gas utility providers, the California Statewide Energy Survey did not collect gas account data.

Table 2-23
Gas UECs for Forecast Zones 1-4

	Zone 1				Zone 2				Zone 3				Zone 4			
	All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data	
	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All Household	117	780 homes	420	208 homes	352	804 homes	435	603 homes	317	1,676 homes	416	1,121 homes	403	3,314 homes	467	2,844 homes
Primary Heat	226	0.27	215	0.97	253	0.74	252	0.91	220	0.73	218	0.97	262	0.82	261	0.96
Auxiliary Heat	138	0.01	138	0.03	88	0.00	88	0.00	78	0.01	80	0.01	72	0.00	73	0.00
Conv. Gas Water Heat	172	0.26	173	0.96	182	0.75	182	0.93	182	0.73	182	0.96	191	0.81	191	0.94
Solar Water Heat w/Gas Backup	.	0.00	.	0.00	84	0.00	84	0.00	141	0.00	141	0.00	121	0.00	150	0.00
Dryer	25	0.08	25	0.31	32	0.15	32	0.19	30	0.21	30	0.28	26	0.27	26	0.31
Range/Oven	30	0.18	30	0.67	39	0.47	39	0.59	40	0.36	40	0.47	40	0.40	40	0.46
Pool Heat	144	0.02	142	0.06	283	0.01	283	0.01	173	0.01	174	0.01	225	0.03	226	0.04
Spa Heat	70	0.02	68	0.08	77	0.01	77	0.02	75	0.02	73	0.03	84	0.03	84	0.04
Miscellaneous	1		2		1		1		1		1		1		1	

Table 2-24
Gas UECs for Forecast Zones 5-9

	Zone 5				Zone 7				Zone 8				Zone 9			
	All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data	
	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All Household	361	2,691 homes	425	2,344 homes	416	384 homes	521	186 homes	369	3,175 homes	437	2,607 homes	391	2,461 homes	447	1,808 homes
Primary Heat	244	0.77	244	0.91	275	0.77	253	0.99	163	0.80	164	0.95	167	0.83	173	0.94
Auxiliary Heat	71	0.00	71	0.00	90	0.00	90	0.00	30	0.00	32	0.00	41	0.00	62	0.00
Conv. Gas Water Heat	182	0.79	182	0.94	220	0.75	224	0.96	224	0.79	225	0.93	216	0.86	221	0.95
Solar Water Heat w/Gas Backup	161	0.00	161	0.00	.	0.00	.	0.00	172	0.00	171	0.00	175	0.00	181	0.00
Dryer	23	0.26	23	0.31	30	0.42	38	0.47	32	0.48	32	0.56	35	0.51	36	0.58
Range/Oven	35	0.56	35	0.67	34	0.61	41	0.76	46	0.71	46	0.82	50	0.82	51	0.90
Pool Heat	244	0.01	244	0.01	196	0.01	189	0.01	240	0.03	240	0.04	218	0.01	229	0.01
Spa Heat	71	0.02	71	0.02	58	0.04	70	0.02	76	0.06	76	0.07	74	0.04	78	0.04
Miscellaneous	0		1		2		2		3		3		2		2	

Table 2-25
Gas UECs for Forecast Zones 10-13

	Zone 10				Zone 11				Zone 12				Zone 13			
	All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data		All Homes		Homes w/Gas Data	
	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat	UEC	Sat
All Household	452	1,959 homes	528	1,401 homes	271	951 homes	316	759 homes	226	431 homes	329	319 homes	298	2,527 homes	368	1,919 homes
Primary Heat	232	0.85	237	0.98	110	0.66	113	0.78	138	0.55	143	0.81	142	0.76	143	0.92
Auxiliary Heat	77	0.00	112	0.00	36	0.00	37	0.00	34	0.00	34	0.00	33	0.01	34	0.01
Conv. Gas Water Heat	217	0.86	219	0.98	200	0.75	201	0.87	211	0.50	211	0.73	185	0.77	186	0.96
Solar Water Heat w/Gas Backup	.	0.00	.	0.00	188	0.00	188	0.00	162	0.00	148	0.00	149	0.00	155	0.00
Dryer	34	0.52	35	0.61	28	0.24	27	0.27	26	0.22	25	0.30	24	0.44	24	0.55
Range/Oven	50	0.76	50	0.86	45	0.86	44	0.96	46	0.69	47	0.94	36	0.59	36	0.71
Pool Heat	190	0.03	192	0.04	274	0.00	274	0.00	262	0.01	262	0.02	215	0.03	214	0.04
Spa Heat	89	0.06	88	0.08	112	0.01	112	0.01	77	0.01	77	0.02	88	0.06	88	0.08
Miscellaneous	2		3		0		0		2		2		2		3	