Final

2004 Smart Thermostat Program Impact Evaluation

Prepared for

San Diego Gas and Electric Company
San Diego, California

Prepared by

KEMA Inc.
Madison, Wisconsin

February 25, 2005

FINAL

2004 SMART THERMOSTAT PROGRAM IMPACT EVALUATION

Prepared for

San Diego Gas and Electric Company
San Diego, California

Prepared by

KEMA Inc.
Madison, Wisconsin

February 25, 2005

ma:project:wsdg0055:report:title

KEMA-XENERGY • 2001 W. Beltline Highway • Suite 200 • Madison, WI 53713 • (608) 277-9696

X EXECUTIVE SUMMARY

X.1 INTRODUCTION

X.1.1 Background

On March 27, 2001, the California Public Utility Commission (CPUC) issued Decision 01-03-073 (D.01-03-073) mandating San Diego Gas and Electric (SDG&E) to implement a pilot program designed to test the viability of a new approach to residential load control and demand responsiveness through the use of Internet technology and thermostats to affect residential air conditioning use. To meet this mandate, SDG&E implemented the Smart Thermostat Program beginning in the spring of 2002.

In the summer of 2002, the program was invoked once. Previous reports provided a process and impact evaluations of the 2002 program. The impact evaluation provided both estimates of impacts on the single 2002 re-set day and projections of savings under alternate conditions.

In the summer of 2003, the full program was not invoked but customers in the metering sample were re-set on the critical peak days of the Statewide Pricing Pilot (SPP). The impact evaluation report for that summer used these test re-sets to estimate impacts per unit for those days as well as projected savings under general conditions.

The present report provides the findings from an impact evaluation of the third summer of the program in 2004. As in 2003, customers in the metering sample were re-set on the critical peak days of the Statewide Pricing Pilot (SPP). This report provides estimates of impacts per unit for the re-set days as well as projected savings under alternate conditions. In addition, this report compares the results for the summer of 2004 with results from the 2003 program. The 2003 results have been revised slightly, so that the comparison is on a consistent basis with some improvements in the analytic approach developed this year.

X.1.2 Program Description

General Structure

The Smart Thermostat Program is designed to include approximately 5,000 residential customers representing an estimated 4 MW in peak demand reduction. Through the program, customers are provided the necessary technology installation and a small incentive for program participation. The equipment deployed allows SDG&E to remotely raise the cooling setpoints on participating customers' thermostats. Participating customers may over-ride the re-set, but forfeit a portion of their incentive each time they do so.

Conditions for Calling a Re-set Event

The program plan calls for the deployment of the Smart Thermostat system when the California Independent System Operator (ISO) calls for a Stage 2 Emergency Notice (Stage 2 Alert). This alert is based on statewide conditions and may occur at times when the weather in San Diego is mild.

As noted, the 2004 impact evaluation is based on a series of test re-sets. Customers in the metering sample were re-set during 12 SPP events. Savings are estimated for the program as a whole for these 12 events. The projected savings under alternate conditions are based on the same load data models used to estimate the impacts of the particular events.

There was also one full program event called on May 3, 2004—the day of a Stage 2 Alert. Savings are not directly estimated for this event because there was no comparison group left un-re-set. Instead, the May 3rd event is used as an example of the use of the saving projections provided in this report.

X.2 FINDINGS

X.2.1 Estimated Impacts for the Observed Re-set Events

Savings per unit enrolled in the program averaged over the re-set period ranged from a low of 0.10 kW to a high of 0.81 kW across the 12 events. The event average was statistically significantly different from zero (at 90 percent confidence) for 11 of the 12 events. Averaged across all 12 re-set periods, the program impact was 0.44 kW. The 90 percent confidence interval, reflecting variation across days as well as units, was from 0.03 to 0.81 kW. Thus, for the 2004 impact evaluation, the overall impact estimate was statistically different from zero.

The single best performance of the program occurred on September 9, 2004, when the per unit average savings reach 0.81 kW. If 5,000 units had received the re-set signal on that day, the estimated savings would have been 4.0 MW, with a 90 percent confidence interval of 2.3 to 5.8 MW. This estimate equals the *ex ante* estimate of 4 MW for 5,000 units. However, this is for the best, not average, day. Across all 12 2004 re-set periods, the savings from 5,000 units would have averaged an estimated 2.1 MW, with a 90 percent confidence interval of .1 to 4.1 MW. Four MW is within the confidence interval, but 4 MW of program savings remains an unlikely reality except on the most extreme days. These estimates are summarized in Table X-1.

Table X-1
Estimated Impacts

	Impact	Standard Error	90% Confidence Lower Bound	90% Confidence Upper Bound
Average of All Events				
per unit	0.37	0.24	-0.03	0.77
5000 units	1,834	1,219	-170	3,839
Average of 2004 Events				
per unit	0.43	0.24	0.03	0.83
5000 units	2,148	1,215	149	4,147
Average of 2003 Events				
per unit	0.30	0.23	-0.08	0.69
5000 units	1520	1,174	-410	3,451
Best Event - September 9th, 2004				
per unit	0.81	0.21	0.46	1.15
5000 units	4,031	1,051	2,302	5,760

X.2.2 What Fraction of Units Contribute to Savings

The fraction of participating units contributing to savings was found to be low in this analysis, as in previous years.

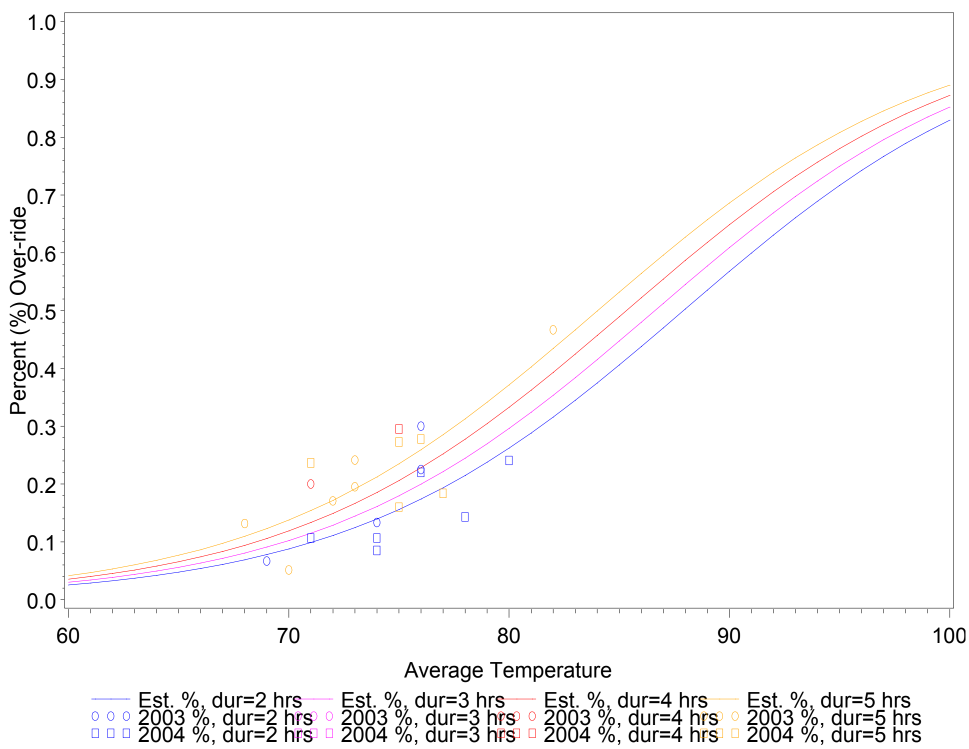
1. On average, just over 90 percent of the thermostats in the program appeared to operate correctly during each re-set event in 2004. The non-response rate (thermostats without confirmed signal receipt) increased for the 2004 analysis. The combined Smart Thermostat and SPP groups' non-response rate was 9 percent in 2004 compared to 6 percent in 2003. This difference was not found to be statistically significant.
2. In 2004, the over-ride rate continued to vary substantially across re-set events. The 2004 range, from a low of 9 percent to a high of 42 percent, was within the extremes of the 2003 rates. The average over-ride rate for the summer of 2004 was almost six percentage points higher than the average in 2003. This is consistent with an average temperature for the 12 2004 re-sets events that was 2°F higher than the 2003 average while average event duration and re-set amount were almost identical.
3. 2004 over-ride rates, when considered alone, showed variation primarily by event duration as opposed to average temperature. This was a change from the 2003 results where average temperature was identified as the strongest driver. A combined 2003/2004 model proved to be the most effective way to estimate over-ride percents as a function of average temperature and event duration. While the two years each primarily reflected the effect of only one of the explored variables, the combination provides a richer model that appears to provide a consistent and comprehensive structure across the years..
4. Eighteen percent of participating AC units were not used at all during the summer of 2003. For 2004, the percent decreased by only one percentage point to 17 percent.

While some of these units might be used during severe hot weather, they contribute no savings in the milder weather.

The average combined effect of non-response, over-ride, and non-use increased from 42 percent to 48 percent between 2003 and 2004. The increase was due to an increase in both non-response rates and over-ride rates. The result is that, on average, only slightly more than 50 percent of the participating units are "potential contributors" to impacts.

The relationship between outdoor temperature, event duration, and over-ride rate based on the combined 2003/2004 model is shown in Figure X-1.

Figure X-1
Predicted and Observed Over-ride Rates vs. Outdoor Temperature for Varying Event Duration



X.2.3 Projected Impacts for Future Events

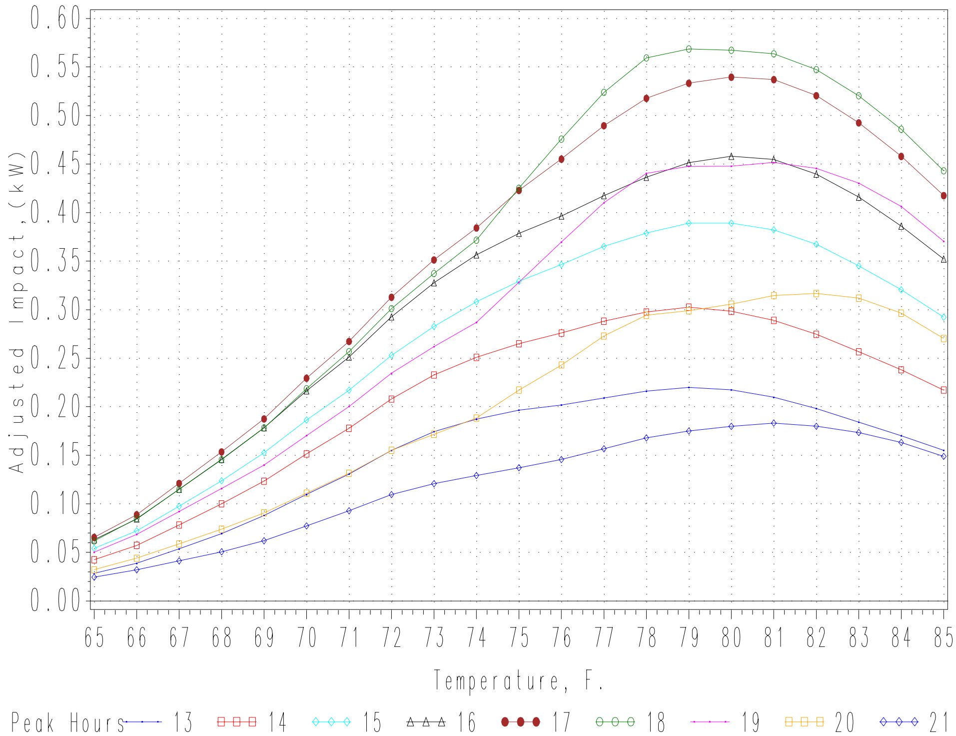
Impacts projected for a future re-set of 4°F of four hours in length are indicated in Table X-2 for the hour ending 6 PM. For 2004, this is the hour with the highest peak impacts for all ambient temperatures. Maximum savings occur at 79°F and 80°F. A 4°F re-set is estimated to yield 0.57 kW savings per thermostat at this temperature. The savings at noon is only 0.14 kW.

Table X-2
Projected Impacts per Participating AC Unit by Outside Temperature,
4°F Re-set of 4 Hours, Hour Ending 6 PM

Average Daily Temperature	Impact per Thermostat (kW)
70	0.22
71	0.26
72	0.30
73	0.34
74	0.37
75	0.43
76	0.48
77	0.52
78	0.56
79	0.57
80	0.57
81	0.56
82	0.55
83	0.52
84	0.49
85	0.44

Figure X-2 shows the projected savings per participating unit as a function of daily average temperature for an event duration of four hours for hours ending 1 PM to 9 PM. For the hours between noon and 7 PM, savings increase steadily into the high seventies at which point the over-ride rate takes over and the savings decrease.

Figure X-2
Projected Impacts per Participating AC Unit by Outside Temperature, 4°F Re-set of 4 Hours, Hour Ending 1 PM through 9 PM



X.2.4 *Projected Savings Applied to May 3rd Event*

A Stage 2 Alert on May 3, 2004, triggered deployment of the Smart Thermostat Program. We did not directly estimate the impact of this event by the planned methods comparing metered re-set and comparison groups because the full meter sample was re-set. We did, however, apply the projections derived for this evaluation to provide an indirect estimate of the event savings. Table X-3 shows the event savings estimated two different ways.

Table X-3
Application of Projects to May 3rd Re-Set Event

Projection Type	Non-Response Fraction (P_F)	AC Non-use Fraction (P_z)	Over-ride Fraction (P_{or})	Percent Not Contributing	Full Response Savings Estimate	Savings (Per Unit kW)	Full Program Savings (kW)
Actual Non-Response and Over-ride	5.4%	17.3%	30.2%	50.3%	1.28	0.64	3,053
Estimated Non-Response and Over-ride (Table X-2)	9.3%		35.3%	57.1%		0.55	2,640

In the first line, the non-response and over-ride fractions are the actual, program wide fractions for the May 3rd event. Combined using the formula in Section 2, they indicate a percent of non-contributors at almost exactly 50 percent. From Appendix C, the average full-response savings (i.e., without adjustment for nonresponse or over-ride) of hours 17 through 19 for a four-degree re-set at 82°F is 1.28 kW. Applying the percent non-contributors to this, the final savings estimate is .64 kW per unit. There were 4,796 confirmed units during the re-set event. The full program projected savings are just over 3 MW.

The second line of Table X-3 repeats the same process as the first line, but uses the estimated non-response and over-ride rates that are used in the savings projections shown in Table X-2 and Figure X-2 above. The estimated percent non-contributors is substantially higher than the actual rate. The May 3rd event was the first hot weather the San Diego area experienced in 2004, and this might explain the lower over-ride rate on that date relative to the estimates based on later events.

X.3 Implications of the Findings

The findings from this year's analysis are stronger than previous years' results. Eleven of the re-set events had statistically significant savings and the magnitude of those savings was higher than for 2003. Despite this, the 2004 analysis essentially confirms the finding from the previous two studies—that future performance of the program as a mechanism to respond to statewide emergencies is not fully reliable.

One factor is the limited used of air conditioning in the territory, with one-fifth of participating units never used over the summer. Another factor is that statewide emergency conditions do not necessarily coincide with hot weather in the San Diego area. As long as the emergency condition that triggers a re-set event is not tied to hot weather in San Diego, a high number of non-users is likely to be found during future re-sets. Finally, when the weather is hot, higher rates of over-ride are projected to occur. Thus, while the program is capable of savings of the desired 4 MW magnitude, it is unlikely to realize these savings in full on the day of a statewide emergency.

The 2004 results do provide one useful direction in this respect. While over-ride rates increase with average temperature, increased event duration also has an effect. This indicates that careful timing of re-set events is essential to focus maximal savings when they are most needed. Well-timed and shorter re-set events that end before a high percent of customers over-ride should provide the best participation rates.

1 INTRODUCTION

1.1 BACKGROUND

On March 27, 2001, the California Public Utility Commission (CPUC) issued Decision 01-03-073 (D.01-03-073) mandating San Diego Gas and Electric (SDG&E) to implement a pilot program designed to test the viability of a new approach to residential load control and demand responsiveness through the use of Internet technology and thermostats to affect residential air conditioning use. The Energy Division recommended a budget of $3.9 million per program year. To meet this mandate, SDG&E implemented the Smart Thermostat Program beginning in the spring of 2002. This report provides the findings from an impact evaluation of the second summer of this program, the summer of 2004.

1.2 PROGRAM DESCRIPTION

1.2.1 General Structure

The Smart Thermostat Program was designed to include approximately 5,000 residential customers representing an estimated 4 MW in peak demand reduction before 2002 year-end. In fact, the program enrollment reached 5,000 devices in November 2003. Through the program, customers are provided the necessary technology installation and a small incentive for program participation. The equipment deployed allows SDG&E control of the thermostat for emergency demand reduction, yet allows the customer the ability to over-ride the company signal remotely or directly at the thermostat.

The program's paging technology allows SDG&E to remotely raise the cooling setpoints on participating customers' thermostats. We refer to this action by SDG&E as a "re-set event." The effect of the higher setpoint is a reduction in the average demand of the air conditioners. This reduction is the desired demand impact.

1.2.2 Conditions for Calling a Re-set Event

The program plan calls for the deployment of the Smart Thermostat system when the California Independent System Operator (ISO) calls for a Stage 2 Emergency Notice (Stage 2 Alert). A Stage 2 Alert is issued when an Operating Reserve of less than 5 percent exists or is forecast to occur within the next two hours for the state. A Smart Thermostat Program re-set event is triggered by a Stage 2 Alert. This alert is based on statewide conditions and may occur at times when the weather in San Diego is mild.

When a Smart Thermostat re-set event is initiated, SDG&E will increase the setting of the thermostat in participants' homes for a period of two to six hours. The re-set may be extended or terminated as necessary. SDG&E has set a maximum of 20 re-sets per calendar year.

1.2.3 Incentives

The customer receives a state-of-the-art digital thermostat installed at no cost to the participant. In addition, the participant will receive up to $100 per year in incentives for the years 2002 through 2004. As noted, the participant may over-ride the increased setpoint of the re-set. However, each time the customer over-rides the re-set, the incentive will be reduced by $2. The incentive, less any reduction due to over-ride, will be paid each year.

1.2.4 Targeting

The targeting strategy for the program was prescribed by the CPUC in D.01-03-073, the decision mandating the program. The decision directed SDG&E to target the following three customer groups:

1. Residential customer whose average monthly electricity consumption is greater than average for their customer class, with the exact specified consumption level to be determined by SDG&E.
2. Residential customers residing in geographical areas in SDG&E's service territory known to have high electricity consumption due to climate.
3. Residential customers residing in known limited to moderate income areas.

Medical baseline customers are not permitted to participate due to the potential air conditioner needs of these customers.

SDG&E met criteria 1 and 2 by selecting customers from California Energy Commission (CEC) Climate Zone 10 who had average monthly summer consumption of 700 kWh or greater. Data from MIRACLE XIII, SDG&E's residential appliance saturation survey, were used to estimate the average consumption for those residing in SDG&E's Transitional Climate Zone with central air conditioning. The average monthly summer kWh consumption for SDG&E's Transitional Climate Zone residents with central air conditioners is 700 kWh. The Transitional Climate Zone was used as a proxy for CEC Climate Zone 10, since the MIRACLE survey data were collected for the SDG&E climate zones (Maritime, Coastal, and Transitional zones). Initially, residents in CEC Climate Zone 10 with average monthly summer consumption of 700 kWh or greater were selected. In an effort to increase participation, an additional mailing was conducted during October 2002 with a follow-up mailing to take place approximately one month later. Targeted customers for this mailing included those in CEC Climate Zone 10 with average monthly summer consumption of at least 600 kWh.

Criteria 3 was met by selecting customers under SDG&E's low-income rate class, the DR-LI rate, in CEC Climate Zone 10, whose average monthly summer consumption was 700 kWh or greater.

1.3 Impact Evaluation

SDG&E was required to evaluate this program effort, including both a process evaluation and a load impact evaluation component. The process evaluation was completed for program year 2002. The primary objectives of the process evaluation were to assess how efficiently and effectively SDG&E runs the program and to make suggestions for improvements. As part of that evaluation effort, survey data were collected from a sample of participants. These survey results shed some light on impact findings. An impact evaluation for the single re-set event in 2002 was also completed.

The load impact evaluation presented in this report provides estimates of the aggregate demand reduction and energy savings from summer 2004 re-sets. As with the 2003 evaluation, savings estimates were derived from 12 test re-sets of the metering sample. In these 12 cases, half of the meter sample was curtailed in conjunction with the Statewide Pricing Pilot (SPP). The 2004 test re-sets were all four degrees in magnitude and ranged from two to five hours in duration. Starting times ranges from 2 PM to 6 PM. The data from these test re-sets allows us to evaluate the program using the same basic methods as if the whole program were called. As in previous years, estimates are also provided for projected savings of future events as a function of the ambient temperature for the day and the length of the re-set event.

In addition to the 12 re-sets, there was one State 2 Alert on May 3, 2004, when the whole Smart Thermostat program was re-set. We apply our impact projections to conditions present on May 3rd to estimate the actual savings attained during this re-set event.

1.4 Organization of the Report

Section 2 describes the impact analysis methods, including the data sources and the analytic approach. The findings from the analysis are presented in Section 3. Section 4 reproduces results from the summer of 2003 using an approach consistent with the approach chosen for the 2004 analysis. Conclusions are summarized in Section 5. Plots of observed and estimated loads and impacts for each re-set day are given in Appendix A for the air conditioner load data analysis, and in Appendix B for the whole-house load data analysis. Tables of projected savings by temperature, time of day, and re-set amount are given in Appendices C and D.

2 METHODS

This section describes the various data used in the impact analysis as well as the methods by which demand impacts were estimated. The methods explained here include the original methodology proposed for the impact evaluation and used in the 2002 analysis. Also included are the modifications made necessary by the use of test re-sets as opposed to full program re-sets with a sample group control. These modifications were put in place for the 2003 analysis, the first year when test re-sets were used. With test re-sets continuing in 2004 the same approach is utilized again. Finally, within the general test re-set framework used in 2003, we have made improvements to some of the specific models used. These changes amount to refinements of some of the pieces of the analysis, but the overall structure remains consistent across the 2003 and 2004 evaluations. Nevertheless, the refinements introduced for 2004 would make it difficult to compare results across the two years. Thus, in Section 4 we reproduce results from the 2003 impact analysis using all the refinements from the 2004 analysis.

Section 2.1 discusses the data and how it was collected. Section 2.2 discusses the analytical approach to processing the data and estimating the demand impacts.

2.1 DATA SOURCES

There were three types of data collected for this study:

1. Interval metering data,
2. Weather data, and
3. Re-set operations data.

The data most necessary and difficult to collect were the interval metered energy consumption data from a sample of Smart Thermostat Program participants. A great effort was managed by SDG&E to gather that data. As a result of the 2002 evaluation, some limited data quality issues were identified. These issues were addressed at that time with future analysis in mind.

For the 2004 analysis, as with the 2003 analysis, SDG&E provided weather data from 10 local weather stations. This was a substantial improvement over the single weather station used for the 2002 analysis as modeling the dependency of air conditioning energy consumption on ambient temperature is a central part of the method employed in this impact analysis.

Silicon Energy, the implementation contractor responsible for the web-based control system, collected data on Smart Thermostat Program participants and on thermostat performance during re-set events. Those data were available directly from the Silicon Energy EEM Suite website.

2.1.1 Metered Data

Energy Consumption Data

Two streams of energy consumption data were collected at each study participant's premise:

1. Whole-premise and
2. Air conditioning (AC).

These streams were monitored on separate meters installed by SDG&E. Both meters recorded energy consumption accumulated over 15-minute intervals. All observations were recorded at quarter-hour intervals. SDG&E provided the energy consumption data sets at the end of the metering period. In addition to these data, SDG&E provided a meter installation survey data set. The survey data included information on nominal cooling capacity, estimated age of AC condenser, and AC type. The survey data also contained information necessary to collate the energy consumption data with the re-set event data, discussed below.

As the name suggests, whole-premise data included all loads at the premise including the AC condenser. Whole-premise data are valuable to the impact assessment of an AC demand reduction program because other loads may be affected by changes in the AC load. For example, greater use of ceiling, floor, or desk fans may accompany decreased cooling by the AC. Refrigerators will run more as less cooling allows the interior temperature to climb, and water heaters may run less. There may be an increased tendency among occupants to lessen internal heat gains, such as cooking, clothes drying, and lighting. These uncertain variables can have marked effects on the impact of an AC demand reduction program. Theoretically, the total impact at a premise is best viewed from the perspective of whole-house consumption.

Unfortunately, the variation of non-AC electrical loads at a premise can make it difficult to discern the impacts of AC demand reduction from whole-premise data alone. The fundamental dependency of AC use on ambient temperature may become more difficult to capture. For this reason, AC-only data were also collected.

The AC energy consumption data collected were taken from the circuit of the AC condenser, that part of the AC system located outdoors that dumps heat from the premise to the ambient environment. The condenser's load includes those of the refrigerant compressor motor, the cooling fan motor, condenser controls, and case or emollient heaters if present. The heaters are found generally in older condensers and serve to vaporize any liquid refrigerant that might enter the compressor. It seems that many run near continuously, perhaps even throughout the heating season.

The condenser is the largest but not the only load in an AC system. The system typically includes the same interior air distribution fan used by a forced-air furnace. The fan demand is approximately 150 Watts per nominal ton of AC capacity, or on the order of an additional 10 percent of condenser demand. Common air conditioner load control programs of the past involve controlling only the condensers with exterior control switches. This type of "cycling" control does not turn off the interior air distribution fan. By contrast, during re-set the Smart Thermostat is understood to turn off the interior air distribution fan just as it would under ordinary AC operation when the cooling setpoint is raised.

The interior air distribution fan is not on the same circuit as the condenser. In fact, it may be on a circuit with other non-AC loads. To collect data from both the condenser and the interior distribution fan alone thus might involve the time-consuming task of wiring sensors. For that reason, energy consumption data were collected from the condenser circuit alone and does not capture the impact of turning the interior fan off when the cooling setpoint is raised. This, then, is another reason to consider whole-premise data in a demand impact analysis.

Sample Design

The energy consumption data were collected from the same random sample of 100 premises of program participants that were selected early in the first year of the program. At that time, premises were limited to those with no more than two thermostats. The sample was divided randomly into two groups of approximately equal numbers of premises. The grouping was intended to allow one-half of the sample to serve as a comparison group for the other, for each re-set event. Thus, for each re-set, one group would be re-set while the other group continued to operate their AC as usual. With multiple re-set events, this would permit each group to be re-set in about half the events and to act as the comparison group for the other group in the other half of the event.

Table 2-1 describes the original sample in terms of numbers of premises, thermostats, and AC metered for each group. The table divides premises into categories by count of thermostats on the premise and numbers of AC metered. Each group had a two-thermostat premise where only one AC was metered. Otherwise, all AC were metered at all premises.

Table 2-1
Original 2002 Distribution of Premises, Thermostats, and Metered AC by Group in Sample

	Sample Group A			Sample Group B		
Premise Category	Premise Count	Thermostat Count	Count of Metered AC	Premise Count	Thermostat Count	Count of Metered AC
One AC, one metered	45	45	45	42	42	42
Two AC, one metered	1	2	1	1	2	1
Two AC, both metered	5	10	10	6	12	12
Total	51	57	56	49	56	55

The re-set and comparison groups differed by no greater than a count of one between premise, thermostat, and metered AC categories. The two groups likewise were very similar in terms of nominal cooling capacity. Sample Group A had a combined capacity of 214.5 tons, while sample Group B had a combined capacity of 202.5 tons. Average sizes were 3.8 and 3.7 tons per unit, respectively.

Subsequent program years have seen some attrition in the metering sample. Two meters have been removed altogether. The greater loss has been in actual program participation, that is, in receiving the initial remote thermostat re-sets. For the 2004 analysis, ex-participants with valid metering data remain in the analysis despite not actively participating in the program.

2.1.2 Weather Data

SDG&E provided observations of hour-ending average drybulb temperature and relative humidity for the period from November 2003 through September 2004 from 10 weather stations in the SDG&E service territory. SDG&E provided a list of program premises indicating the most appropriate weather station for this analysis. Seven of the 10 weather stations are used to describe the weather conditions for the 2004 sample of 94 premises. Table 2-3 shows the distribution of premises across the seven weather stations as well as the monthly mean temperature.

Table 2-2
Sample Group Distribution Across Weather Stations
with Summer Monthly Mean Temperature

Weather Station ID	Sample Group		Monthly Mean Temperature				
	A	B	June	July	August	September	October
S01	1		68	72	70	72	65
S02	19	22	66	71	71	73	64
S04	2		66	69	68	69	62
S05	14	16	68	74	74	74	66
S06		1	65	70	69	71	63
S08	8	6	66	73	72	71	61
S09	3	2	64	69	68	71	63

Multiple weather stations data, available for the 2003 and 2004 analyses, was a substantial improvement over the single weather station used for the 2002 impact analysis. The combination of the ocean and mountainous terrain has the potential to cause highly variable weather conditions across the SDG&E service territory. These weather data should better represent the varied ambient conditions faced by the sample of program participants. Figures 2-1 through 2-5 show the average day temperature for the seven weather stations represented in the sample for May and the four summer months. The variability across weather stations is clearly evident in these plots. Vertical lines indicate the days on which thermostat test re-sets took place. The May 3rd full program re-set is not marked, but coincides with the dramatic spike evident in the May plot. It should be noted that, despite the May 3rd temperature peak, re-set events do not necessarily coincide with peak temperatures. This is a visual reminder that San Diego area temperatures are not driving the Statewide Pricing Pilot curtailment events.

Figure 2-1

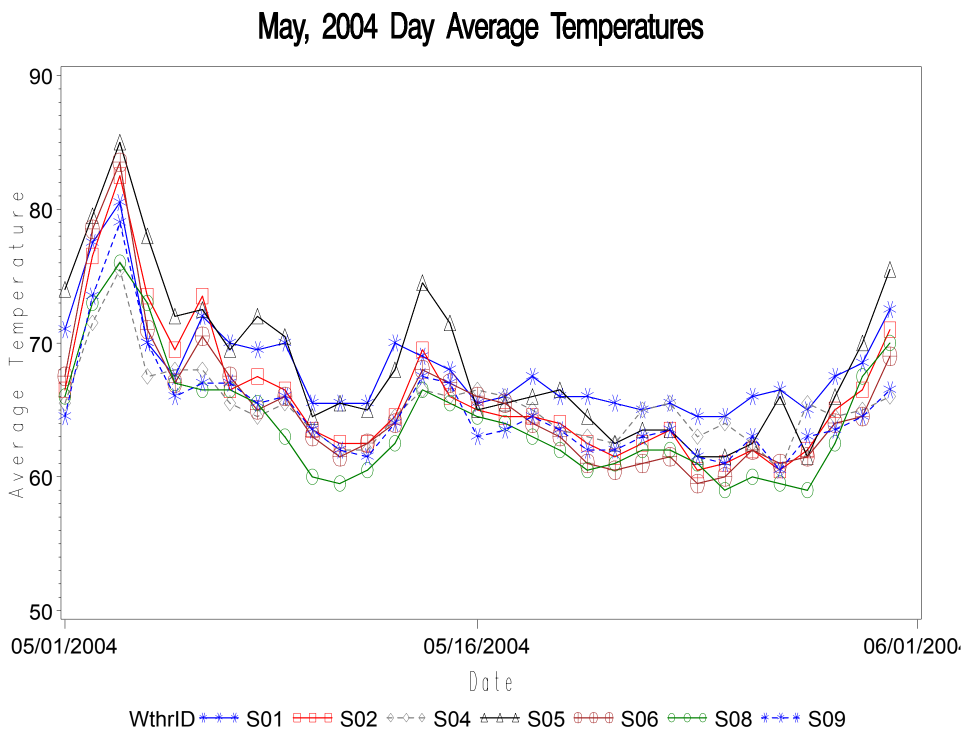


Figure 2-2

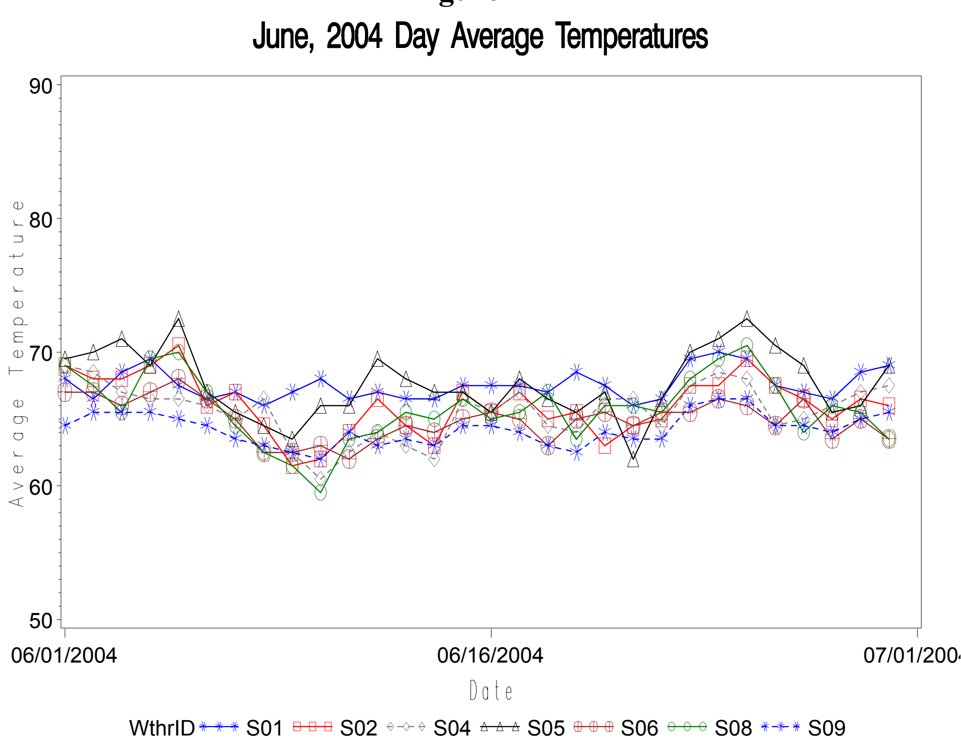


Figure 2-3

July, 2004 Day Average Temperatures

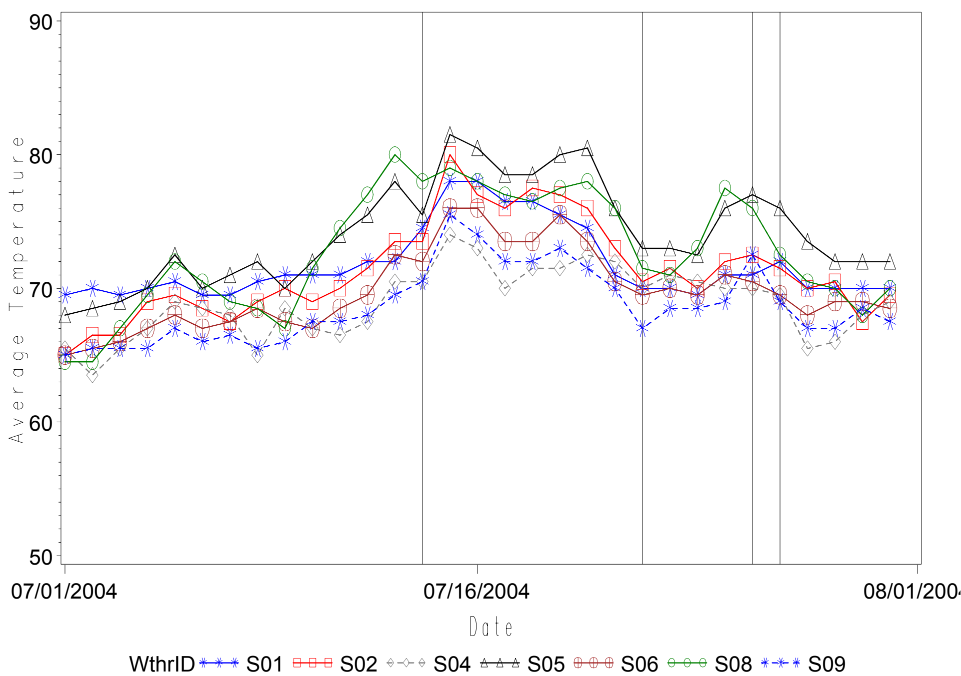


Figure 2-4

August, 2004 Day Average Temperatures

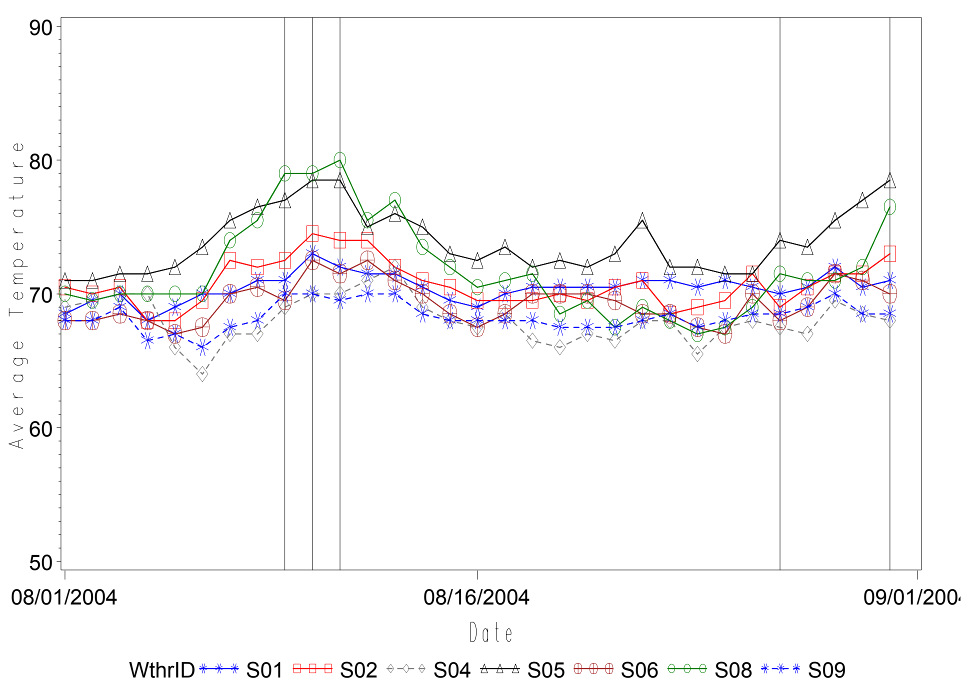
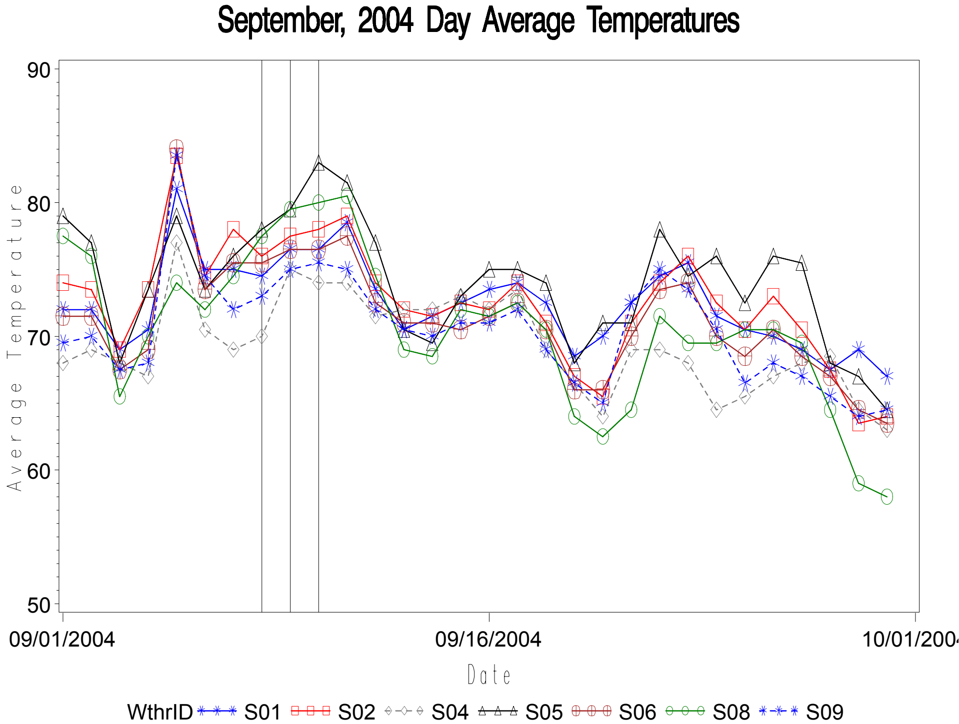


Figure 2-5



2.1.3 Event and Customer History Reports from Silicon Energy

The Silicon Energy EEM Suite website (rem.siliconenergy.com/siliconenergy/rem/asp/event_summary_setup.asp) allowed ready access to, and downloading of, data on customer participation in the summer's re-set events. These data included an observation for each thermostat that had been included in each re-set. Each observation identified the sample group to which the thermostat belonged as well as customer name and account number information. Additional fields described the start time and planned duration of the re-set event, the amount in degrees Fahrenheit of the thermostatic cooling setback, and time stamps of thermostat acknowledgement of re-set and of over-ride as appropriate. It was these last two time stamps that identified "non-responder" thermostats that did not appear to receive the re-set signal, and over-ride thermostats where the thermostat was manually lowered after being raised by the re-set signal.

2.2 METHODS

This section describes the methods by which the collected data were examined to estimate demand impacts. The same general approach was employed as for the original 2002 impact analysis. This approach focuses the savings calculations on the subset of participants who are potential contributors to savings. Then, data on all participants, or population data, make it possible to determine what percent of the whole program fit this definition. This two-part approach takes advantage of this population data to provide estimates with more precision.

The 12 test re-set events that took place in 2003 and 2004 provided more data to analyze than 2002 because they do not rely on Stage 2 Emergencies. However, the fact that only a subset of Smart Thermostat customers participated necessitated a change from the 2002 approach that utilized operations data from the full program population. Because the 2002 approach still represents the ideal given a full program implementation, we discuss it first and then discuss the changes necessary to modify the approach to the program data available.

The analysis has three main parts.

1. The fraction of units potentially contributing to savings for each event is determined.
2. The impacts for each re-set period are calculated from analysis of the load data for potential contributors, then adjusted for the fraction not contributing.
3. The impacts for a range of conditions are projected based on the same load models used for the analysis of the actual re-set days, and adjusted for the same fraction of non-contributors.

These steps are described below.

2.3 Potential Contributors

Not all AC units in the program provide savings during a re-set event. This analysis determines the average savings per unit in two parts. First, the average savings per unit is determined for the subset of units classified as "potential contributors" to savings. Savings for the remaining units are zero. The overall average savings across all units is then calculated by multiplying the average savings for potential contributors by the fraction of units in this category. Thus, for example, if only one-quarter of the units in the program are determined to be potential contributors to savings, the unit savings estimated for that quarter of the program, the potential contributors, are multiplied by one-quarter to get the savings per unit across all units in the program.

An alternative approach to accounting for units that do not contribute to savings would be simply to calculate savings directly over all units, both contributors and non-contributors, in the metered samples. With this more direct approach, however, the fraction of zero contributors in each metered group is random. This random variation in the proportion of zero contributors in each group adds to the variance of the estimated savings.

The two-part approach used for the 2002 evaluation provided a way to take advantage of available population data to calculate a more accurate estimate of the overall program savings. The accuracy was higher because we did not have to estimate the percentages of some kinds of non-contributors. We knew the actual percentage from the full population of participants in the event participation data. There is no estimation involved, thus no variance. Using this technique, the impact estimate for the whole group, including zero contributors, can be estimated with the variance of only a subset of participants.

Because the re-set events that occurred during the summers of 2003 and 2004 were SPP events rather than Smart Thermostat events, the whole population of the Smart Thermostat Program was not paged. This lack of population data required adjustments to the 2002 approach. Clearly we could not take advantage of zero-variance percentages, but the analysis followed a similar approach and took advantage of the general 2002 approach where possible.

There were still three reasons a unit might not provide demand savings during a re-set period.

1. The unit failed to receive the re-set signal.
2. The unit received the re-set signal, but the customer over-rides the re-set.
3. The unit was not in use at the time the re-set signal was sent, therefore had no reduction to provide.

If the full Smart Thermostat Program had been called, data on the fraction of units that did not receive signals and the fraction that over-rode would have been available from the Silicon Energy website for the full participant population for each re-set event. Instead, we had data for all Smart Thermostat participants who joined the SPP program in addition to those participants in our sample groups. We leveraged this additional data to lower the variance of the estimate of the percentage of non-responders, those with signal failure. The fraction of participants that over-rode was estimated using only the re-set sample group. As was always the case, whether or not an AC unit was in use on a particular day was determined only from the metering data for the combined sample groups.

2.3.1 Signal Failure Fraction

Signal receipt itself is not directly observed. What is known for all participating units is whether they returned a signal to the system head end, acknowledging receipt of the re-set signal. We used the percent of units that did not send an acknowledgement as an upper bound on the percent that did not receive a signal. If the signal transmission in each direction is such that virtually any unit that successfully received a re-set signal would successfully return an acknowledgement, this percent of non-responders is very close to the percent that didn't receive a signal and is not an overstatement.

On the other hand, if signal failure randomly affects a fraction of units essentially symmetrically and independently in each direction, the fraction non-responding overstates the fraction not receiving a signal. In this case, we can assume that half the non-responders did not receive a signal and half received a signal, but the response signal failed. Thus, we would treat one-half the observed fraction of non-responders as a lower bound on the percent not receiving the re-set signal.

For the 2002 analysis, the percent of non-responders across the whole population was known, with zero variance, from the participation data. In the summer of 2004, only the designated

Smart Thermostat sample group and the SPP participants got the re-set signal. Unlike 2002, we had to estimate the percent of non-responders. We could have used the sample group alone to estimate this percent. However, the full set of participants, including the SPP participants, provided more data for the estimation, lowering the variance of the estimate. The fundamental question was whether there were systematic differences between the curtailed sample group and the SPP participants with respect to re-set signal non-response. As SPP participants were originally Smart Thermostat Program participants, there appeared to be no difference from a hardware perspective. As the two groups were re-set as part of the same event by Silicon Energies, there appeared to be no differences with respect to the source of the re-set signal. We therefore assumed that the full set of participants were representative of the larger Smart Thermostat Program population and could be used to estimate the non-responder percentage.

2.3.2 Over-ride Fraction

The number of switches over-ridden was recorded directly in the event participation data. However, only those switches that received a signal can over-ride. Thus, we considered the over-ride fraction as a fraction of those that received the signal. Once again, for the 2002 analysis, this percentage was known directly from the full program population data.

For the 2004 program we had only the sample group and SPP participant data with which to determine the over-ride percent. As with the non-responder percent, we had to estimate the over-ride percent since we did not know it outright. Unlike the non-responder situation, it was not clearly reasonable to use the additional SPP data to improve the estimate. Unlike the non-response percent, which is a function of an automated communications process, the over-ride percent is a function of, among other things, the incentive structure of the program. The Smart Thermostat Program and SPP have different incentive structures. Thus, SPP over-ride data cannot be considered representative of the Smart Thermostat Program.

There are three remaining options for accounting for premises that over-ride the re-set.

1. **Separate over-ride percent adjustment.** Use the load data analysis to calculate average savings per unit for non-over-riders (potential contributors) only. Calculate the percent of over-riders from the sample for each re-set event. Adjust the non-over-rider savings per unit by the fraction of non-over-riders to provide the savings per unit across all potential contributors. This method is the same as the 2002 method, except that the percent of over-riders is calculated from the sample rather than from the full population.
2. **Over-riders included in the load data analysis.** Use the load data analysis to calculate average savings per unit across all responding users, regardless of whether they over-ride or not. No separate adjustment is needed for over-riders.
3. **Over-riders directly included in average savings per unit, but with savings set to zero.** Use the load data analysis to calculate the savings for each non-over-rider unit. Set over-rider unit savings to zero. Calculate the average savings per unit across the whole pool of responding users, non-over-riders, and over-riders combined. This method is similar to Method 1. It would give the same result as Method 1 if only re-set participants were included in the analysis without the “difference of differences” calculation. This

method also provides a basis for calculating the standard error of the savings per potential contributor.

As noted, Method 1 is most similar to the 2002 analysis. However, the reason for separating the over-ride percent from the average savings for non-over-riders in the 2002 analysis was that the over-rider percent could be determined without error. Using this information rather than relying on the sample percent over-ride reduced the variance of the overall savings estimate. Since this reduction is not possible for 2004, the separation offers no advantage in terms of variance reduction. However, isolation of the fraction of over-riders is useful in terms of understanding the program response as well as providing direct comparability to the previous results. Moreover, for the calculation of projected savings under general conditions, it is necessary to calculate the fraction of potential contributors accounting explicitly for the over-ride rate.

On the other hand, the variance calculation described below for the event-specific analysis requires that the over-riders and non-over-riders be combined. Method 3 provides nearly the same estimate as Method 1, but allows direct calculation of the standard error of the resulting estimate. For the event-specific analysis, therefore, we use Method 3.

Method 2 could be viewed as providing the most complete estimate, since it recognizes that over-rides do not take place instantly, but affect the savings differently over the duration of the re-set period. However, including the load data from over-riders in the analysis serves to increase the variance of the estimate. In addition, this approach is inconsistent with that used for the 2002 program and that which would be used if the full program were implemented. In the interests of consistency and variance reduction, we use Methods 1 and 3 for the primary analysis and presentation. We focus on Method 1, which corresponds to the original 2002 approach. However, we express the nearly equivalent results via Method 3 for the variance calculations and certain explanations.

If we used Method 2, including estimated over-ride impacts from the load data analysis, rather than Method 3, setting over-ride impacts to zero, we would likely get somewhat higher impacts in the early intervals of the re-set period and lower impacts in later intervals. In the early intervals, units that will over-ride but have not yet done so contribute positive savings rather than zero as in Method 3. In later intervals, the over-riders are likely to have some "pay-back" for the foregone cooling in the earlier intervals resulting in increased usage or negative savings rather than zero as in Method 3. However, as noted, the trade-off for this finer-grained look at the effect of over-rides would be increased overall variance.

2.3.3 Fraction Zero Use

Units that are never used during weekdays over the entire summer do not contribute to savings from this program at any time. We determine the fraction of zero users based on analysis of the metered air conditioning data. This fraction is determined from the full usable metering sample, not just those in the re-set group on the particular day a re-set occurred. The full sample is the largest group for which we can estimate this population characteristic. For 2004, this group includes both the curtail groups and the ex-participants. The "summer non-zero users" are those units that were used on a weekday at some time over the summer. Included in this group may be some units that had zero use on a particular re-set day. We do not attempt to estimate a zero use fraction separately by re-set event. The effects of zero use by a subset of those who are at least sometimes non-zero users are included in the average impacts estimated for the non-zero use group.

2.3.4 Potential Contributors and Non-contributors

Estimating Percent of Non-contributors

For the 2002 analysis we estimated the fraction of units that were complete non-contributors to savings as:

$$p_{NC} = p_F + (1-p_F)(p_{OR} + p_z),$$

where

p_{NC} = fraction of units that are non-contributors,

p_F = fraction of units that had signal failure,

p_{OR} = fraction of units that over-rode, out of those that did not have signal failure, and

p_z = fraction of units with zero weekday AC usage all summer.

That is, all units with signal failure (p_F) are non-contributors. Of the remaining units ($1-p_F$), those that cannot contribute to savings are those that over-ride (p_{OR}) and those that were never used (p_z). These proportions are additive because they are essentially mutually exclusive. Whether a unit has zero use is assumed to be independent of whether or not the signal was received.

For the 2003 and 2004 analyses, this calculation of the fraction of non-contributors is used in the calculation of projected savings. However, the over-ride and non-responding fractions are taken from the samples, as described above, rather than being provided for the entire population from the operating system.

For the 2004 event-specific analysis, we still consider non-responders (F), zero summer users (z), and over-riders (OR) all to be non-contributors to savings. Those who have non-zero summer use and respond to the signal we call ***responding users.*** Responding users can either be potential contributors or over-riders. While the 2002 over-ride rate was determined from the Silicon Energy population data, the 2004 over-ride rate p_{OR} is determined from the same pool of responding users providing the savings per unit via the load data analysis. For this reason, we break up the non-contributor fraction somewhat differently.

First, we calculate the proportion of responding users as those who do not have a signal failure and do not have zero summer usage. These are assumed to be independent so that the combined probability is multiplicative:

$$p_{RU} = (1-p_F)(1-p_z).$$

We then determine the over-riders as a fraction of the responding users. The non-over-riders are the potential contributors. Thus,

$$p_C = (1-p_F)(1-p_z)(1-p_{OR}).$$

The non-contributors are everyone else. That is

$$\begin{aligned} p_{NC} &= 1-p_C \\ &= 1-[(1-p_F)(1-p_z)(1-p_{OR})] \\ &= 1-(1-p_F)(1-p_z-p_{OR} + p_z p_{OR}) \\ &= 1-[(1-p_F)-(1-p_F)(p_z+p_{OR})+(1-p_F)p_z p_{OR}] \\ &= p_F +(1-p_F)(p_z+p_{OR})+(1-p_F)p_z p_{OR} . \end{aligned}$$

This expression differs from that given above by the last term,

$$(1-p_F)p_z p_{OR} .$$

Strictly speaking, this term should be removed. It would be appropriate if zero summer use and over-ride were independent, whereas in fact they are mutually exclusive. However, the product is small, and the use of the expression for the proportion of contributors p_C simplifies the analysis.

Standard Error Calculation

In the 2002 analysis, only p_z entered the percent non-contributor equation as an estimated percent with an associated variance. The standard error was calculated as

$$\mathrm{SE}(p_{NC})= (1-p_F)\mathrm{SE}(p_z) = (1-p_F)(p_z*(1-p_z)/n_z)^{1/2},$$

where n_z is the number of premises in the combined samples or, alternatively, the denominator of the fraction that provides the percentage of non-users, p_z.

For the 2004 event-specific analysis, we develop an estimate of the standard error of the savings per unit across all responding users, and adjust this estimate by the proportion of responding users. Thus, we need an estimate of the standard error of this proportion.

This standard error is calculated as

$$\mathrm{SE}(p_{RU}) \quad = \mathrm{SE}[(1-p_F)(1-p_z)]$$

$$\simeq [(1-p_F)^2 \mathrm{SE}^2(p_z)+(1-p_z)^2 \mathrm{SE}^2(p_F)]^{1/2}.$$

The standard errors of the proportions are calculated using standard formulas for proportions from a simple random sample.

Similarly, the standard error of the contributor fraction is calculated as

$$\begin{aligned}\mathrm{SE}(p_{NC}) &= \mathrm{SE}(p_C)\\ &= [(1-p_F)^2(1-p_z)^2\mathrm{SE}^2(p_{OR})+(1-p_z)^2(1-p_{OR})^2\mathrm{SE}^2(p_F)\\ &\qquad +(1-p_F)^2(1-p_{OR})^2\mathrm{SE}^2(p_z)]^{1/2}.\end{aligned}$$

2.4 Impact Estimates on Re-set Day

2.4.1 Overview

To estimate the demand impact of a re-set event, it is necessary to have an estimate of the demand that would have been present without the re-set event. If the two groups into which our sample was divided were completely identical then we could simply use the comparison group as our estimate of what the load would have been. Taking the difference of the two groups' mean load would provide a good estimate of demand impact. Of course, in reality it is impossible to select two identical sample groups. Alternating curtailments between the sample groups ought to control for some of the differences, but that in turn implies conditions are the same across curtail days, and we know this was not the case.

Another approach to estimating the demand impact involves using a regression-based estimate of the load on the re-set days. This approach provides an alternative estimate of impacts but this assumes that re-set day consumption can be fully explained by the model.

For this and the two previous two Smart Thermostat analyses we have used a method that combines these two approaches to estimating demand impact. By combining the two approaches, we can overcome the weaknesses of each approach when used alone. The regression-based model controls for differences across the two sample groups and across re-set days. At the same time, the use of a comparison group controls for re-set day conditions not addressed by the regression-based estimates.

2.4.2 Load Model

The basic weather normalization model estimates load as a function of drybulb temperature, specifically, average daily heating or cooling degree-days. Using hour-specific dummy variables, the intercept and both degree-day measures enter into the model on an hour-specific basis. This means that each of the 24 hourly load measures for each day are regressed against an hour-specific intercept term and degree day term. The resulting parameter estimates, though based on only a single daily temperature measure, provide an hourly estimate of load as a function of weather.

Degree-days are calculated as the degrees above or below a base temperature. The ideal cooling base temperature is the minimum ambient temperature at which AC use begins and below which there tends to be no AC load. The heating base temperature is the maximum ambient temperature above which there tends to be no heating-related load. Base temperatures vary across premises because the inhabitants have different inside temperature preferences and houses are varied in their physical properties that relate to inside temperature. Our model estimates the same model across a wide range of cooling and heating degree-day bases and chooses the combination with the greatest explanatory power.

Eqn. 2-1 shows the model in equation form. It was fit separately for each premise to the AC or whole house consumption data. Hourly AC and whole house loads are calculated by summing the 15-minute interval data to the hour. The optimal combination of cooling and heating base temperatures was then chosen on the basis of the maximum R-square statistic.

$$L_{jdh} = \alpha_{jh} + \beta_{Hjh} H_d\left(\tau_{Hj}\right) + \beta_{Cjh}\, C_d\left(\tau_{Cj}\right) + \varepsilon_{jdh} \qquad \textbf{Eqn. 2-1}$$

where

L_{jdh}	=	sum of 15-minute interval AC consumption at hour h of day d for premise j;
$H_d(\tau_{Hj})$	=	heating degree-days at the heating base temperature τ_{Hj} for premise j, on day d, based on daily average temperature;
$C_d(\tau_{Cj})$	=	cooling degree-days at the cooling base temperature τ_{Cj} for premise j, on day d, based on daily average temperature;
ε_{jdh}	=	regression residual;
α_{jh}, β_{Hjh}, β_{Cjh}	=	coefficients determined by the regression; and
τ_{Hj} τ_{cj}	=	base temperatures determined by choice of the optimal regression.

The degree-day variables are calculated as

$$C_d(\tau_{Cj}) = \max((\mathrm{T_d} - \tau_{Cj}),0)$$

$$H_d(\tau_{Hj}) = \max((\tau_{Hj} - T_d),0),$$

where T_d is the "daily average temperature," calculated as the mean of the daily minimum and maximum for day d. Because of thermal lags in the house, this form of daily average tends to be a better predictor of heating and cooling loads than the current hourly temperature, or an average for particular hours of the day.

An alternative approach considered was to use lagged hourly temperature variables in the cooling model. This approach can be effective. However, hourly lag effects get confounded with time-of-day effects so that it may be difficult to obtain meaningful hourly coefficients if lag terms are also included. Using coefficients that do not vary by hour doesn't allow behavioral effects to be captured. The hourly coefficients β_{jh} account both for different behavior by time of day and also for the effects of thermal lags within the day.

For the 2004 analysis we did include a day lag variable. In addition to 24 hourly cooling parameters based on cooling degree days ($C_d(\tau_{Cj})$), we estimated 24 hourly cooling parameters based on a lagged temperature variable. The lagged cooling variable was a geometric combination of the previous three days' temperature. It was calculated as

$$\tau_{Lj} = \sum_{i=1}^{3} T_{d-i} e^{-i/3} \Big/ \sum_{i=1}^{3} e^{-i/3} .$$

Lagged cooling degree days are then calculated as with the other degree day calculations,

$$L_d(\tau_{Lj}) = \max((T_d - \tau_{Lj}),0).$$

The resulting load model, shown in Equation 2-2 is identical to equation 2-1 in structure except that it includes the additional 24 lagged cooling parameters.

$$L_{jdh} = \alpha_{jh} + \beta_{Hjh} H_d\left(\tau_{Hj}\right) + \beta_{Cjh} C_d\left(\tau_{Cj}\right) + \beta_{Ljh} L_d\left(\tau_{Lj}\right) + \varepsilon_{jdh} . \qquad \textbf{Eqn. 2-2}$$

This model explicitly accounts for thermal effects across days while still providing good estimates of behavior during the day.

Using regression coefficients from this fitted equation as indicated in Eqn. 2-3 by the overscript '^', and cooling, lag cooling and heating degree-days [$C_d(\tau_{Cj})$, $L_d(\tau_{Lj})$, and $H_d(\tau_{Hj})$] for day d of the re-set event, the estimated load (without re-set) L_{jdh}, was calculated for each premise, day, and hour using Eqn. 2-3.

$$\hat{L}_{jdh} = \hat{\alpha}_{jh} + \hat{\beta}_{Hjh} H_d\left(\hat{\tau}_{Hj}\right) + \hat{\beta}_{Cjh} C_d\left(\hat{\tau}_{Cj}\right) + \hat{\beta}_{Ljh} L_d\left(\hat{\tau}_{Lj}\right). \qquad \textbf{Eqn. 2-3}$$

A second adjustment made to the weather normalization model applied only to the AC-only model. The basic weather normalization model estimates a base load as well as heating and cooling parameters. For the AC-only model we would expect this base load to be zero unless the AC unit is actually a heat pump or, as mentioned above, there is some ongoing, low-level load used by the condenser. In instances where the basic weather normalization model produced base load parameters that were, in aggregate, negative, we constricted the weather normalization model to all-zero base load parameters. This restricted weather normalization model is identical to basic weather normalization model shown in Equation 2-2 except that it lacks the a_{jh} parameters.

2.4.3 Load Model Error Correction

Any load model will have some estimation error. The particular model used in this analysis is relatively simple using just the time of day, the daily average temperature, and now a lagged temperature variable. Effects of humidity, sunshine, and wind are not explicitly modeled.

Because of some of these physical factors, a portion of the modeling error for a given day and hour will be similar across AC units. The model may simply not have enough hot day data to estimate usage on the hottest days. Alternatively, if the day is very breezy, usage might tend to be lower than the temperature model would indicate. Further, even with a more sophisticated physical model there may be behavioral changes related to events in the news or holiday schedules that would be similar across homes.

The use of the comparison group provides a basis for correcting these systematic modeling errors. We take the average modeling error for the comparison group as an estimate of the likely average modeling error for the re-set group.

First we have to calculate the unadjusted impact estimate for the re-set group

$$S_{Rh} = \frac{1}{n_R} \sum_{j \in R} \left(\hat{L}_{jh} - L_{jh} \right),$$

where

$\hat{L}_{jh}$ is the weather normalized estimate of hourly load,

L_{jh} is actual hourly load,

S_{Rh} is the unadjusted load impact estimate of the re-set group, and

n_R is the number of units in the re-set group.

The model estimates for each premise tell us what would have happened without the re-set. The differences from the observed load for each premise are the estimated savings. The premise-level impacts are averaged over the group to get the mean unadjusted load impact estimate per unit. This estimate is still “unadjusted” because the quality of the weather-normalized estimate is unknown.

The model estimate does not need to be perfect, only consistent across the two sample groups. With this assumption, the average modeling error, or what we are considering the error adjustment, is, in fact, the same calculation for the comparison group.

$$S_{Ch} = \frac{1}{n_C} \sum_{j \in C} \left(\hat{L}_{jh} - L_{jh} \right),$$

where

S_{Ch} is the unadjusted load impact estimate of the comparison group, and
n_C is the number of units in the comparison group.

If the weather normalization model were perfect, the model estimate for each comparison premise would be identical to the observed load. The mean "impact" across the comparison group, $\hat{L}_{ch}$, would equal zero and there would be no adjustment necessary. However, we do not expect the model to be perfect. We use the comparison group average error to estimate the average error for the re-set group. Thus, the comparison group average modeling error indicates if the model tends to be high or low and by how much. The adjustment is made by taking the difference of these two differences.

$$S_h = S_{Rh} - S_{Ch}.$$

If the model, on average, over-estimates the comparison group's actual load for a particular interval, then it will also give too much impact credit to the re-set group. In this case, the error adjustment will be positive and will be subtracted from the inflated re-set group estimate. If the model is low, a negative error adjustment is removed (a double negative) so the original re-set impact estimate is increased.

This "difference of differences" approach combines the model estimation and comparison group approaches to determining "what would have been." The above explanation implies the model estimate is the primary step, with the comparison group serving to adjust the model-based result. The method can just as easily be explained the other way around and this may be more intuitive for some. With this approach, the difference between the observed loads of the comparison and re-set groups is the primary impact estimate. The weather normalized load estimates are only compared with each other to determine if any systematic influences are affecting the two groups differently. These two approaches are mathematically identical.

2.4.4 Savings Estimates by Time Interval

The load model is estimated on an hourly basis, and the savings equations above indicate estimates for each hour. However, the load data were available on a quarter-hour basis. Kilowatt-hour savings for each quarter-hour interval were calculated analogously to the hourly equations indicated above. For the quarter-hourly estimates, the load in each time increment was estimated using the load model coefficients for the hour that included that increment.

Savings were also calculated for the average of the entire re-set period. The re-set periods are all listed as starting and ending on the hour. Apparently, though, the actual start and end times are slightly offset so as not to have too extreme a system affect when the group is returned to full cooling. Despite this we only consider intervals within the specific curtail period.

For the overall re-set period savings, each AC unit's average observed load during the re-set period was calculated across all increments in the period. Each unit's estimated load was similarly averaged across all re-set period time increments. The difference of differences calculation was then applied to these re-set period averages to obtain the re-set period average kW savings.

2.4.5 Final Impact Estimate

The difference of differences method gives the savings per unit among potential contributors. When we include over-riders (with impacts set to zero) in the re-set group average, the resulting difference of differences gives the savings per unit across all responding users for each time interval in the re-set period. Multiplying by the fraction of participants that are responding users gives the savings per unit across all participants.

Thus, the final impact S_{Th} for each interval h is given by

$$S_{Th} = p_{RU} S_h \ ,$$

where S_h is the average impact per responding user, as defined above.

2.4.6 Standard Error of the Impacts

The standard error of the impact estimate is calculated from the separate standard errors of the proportion of responding users p_{RU} and the savings per unit S_h for this group. This responding user unit savings is calculated by the difference of differences method.

The corresponding standard error is calculated for each interval by first calculating the standard error of each group's difference between observed and modeled load. This standard error is simply the standard deviation of individual units' modeling errors, divided by the square root of the number in the group. The standard error of the difference of differences impact estimate, $\mathrm{SE}(S_h)$, is the square root of the sum of squared standard errors for the re-set and comparison groups.

The standard error of the final estimate S_{Th} is then

$$\mathrm{SE}(S_{Th}) = [p_{RU}^2 \mathrm{SE}^2(S_h) + S_h^2 \mathrm{SE}^2(p_{RU})]^{1/2},$$

where the calculation of the standard error of the proportion was given above.

2.4.7 Assessing Comparability of the Comparison Group

The savings estimation approach assumes that the modeling error for the comparison group is a good indicator of the likely modeling error for the re-set group if no re-set had occurred. Thus, an important step prior to applying this method was to assess whether the two groups were in fact similar.

Premises were selected at random for the metering sample and were randomly assigned to Group A or B. Thus, there was no *a priori* reason the groups should have been different. However, random effects could result in observable differences at the outset that would suggest a need for some kind of adjustment.

A particular concern was that the sizes of the air conditioning units in the two samples might be different. In this case, the comparison group error might be a good indicator of the re-set group error, but a scaling factor might need to be applied to the comparison group error to adjust for the size difference. Our original plan was to calculate savings after normalizing the two groups' observed and estimated loads by dividing by their respective average air conditioner capacity, in tons.

The 2002 analysis decided the two groups had practically the same distribution of AC unit size, and this normalization was not necessary. That analysis compared the two groups in terms of the mean, median, minimum, maximum, and standard deviation of tons, both for the full sample and for the smaller sample used in different stages of the analysis. In terms of these distribution statistics, the two groups were very similar to one another and were similar also across the different subsets used in the analysis. This comparison is repeated for the groups used in the 2004 analysis.

An additional check is also repeated. It plots the average re-set group model error against the average comparison-group model error for warm weekday afternoons excluding the re-set day. This plot is presented in Section 3.

This comparison showed a strong relationship between the two groups' errors. The comparison also showed a similar standard deviation of error between the two groups indicating no scale difference. A regression of re-set average error on comparison to group average error had an intercept very close to zero indicating no systematic shift between the two. These comparisons support the use of the comparison group without scale adjustment.

Even with very comparable groups, normalization by capacity could be considered as a variance reduction technique. Ratio estimation, such as calculating savings per ton rather than mean savings per unit, can often be effective in reducing the variance of impact estimates. However, for this method to be effective in variance reduction, it is necessary to have the normalization variable known for the entire population. In this study, capacity data were collected for the metering sample to allow for scaling between the re-set and comparison groups if necessary, but were not available for the general population of participating AC units. Thus, once it was determined that scale adjustment was not required between the two groups, no normalization by capacity was used in calculating the savings estimate.

2.4.8 Whole-premise Analysis

For each re-set event, the same analysis method was applied to the whole-premise data as the AC data. The same units identified as potential contributors by the end-use analysis were included in the whole-premise analysis.

The results presented in Section 3 show that, as in 2002 and 2003, the whole premise analysis was less reliable in terms of the standard errors of the resulting estimates than was the AC analysis. We therefore continue to rely on the AC results for the impacts.

2.5 PROJECTED IMPACT ESTIMATES FOR GENERAL CONDITIONS

This section describes the methods by which demand impacts were estimated under general conditions. A general condition is defined simply by a daily average temperature an hour of the day. The methods for general conditions used the same load models as described above, but essentially applied a theoretical model of equivalent temperature differences to describe the effect of re-set. The initial projection estimate of demand impacts is analogous to the estimate of potential contributor impacts. These projections can be adjusted taking into consideration actual or estimated non-response, over-ride and zero usage percents.

2.5.1 Model AC Loads at Different Temperatures

The load models described above to estimate load *without* re-set were used here to estimate load both with and without re-set. Loads *with* re-set were estimated using the daily average temperature less the thermostat setback. This in effect lowers the average daily temperature and thereby decreases the cooling load. That is, the effect of setting the thermostat forward by δ degrees is essentially the same as the effect of dropping the ambient temperature by δ degrees. The magnitude of the thermostat setback, in degrees Fahrenheit, thus was a critical determinant of the load with re-set.

The lagged temperature variable must also be provided for the projection model. One could simulate a variety of situations depending on the choice of the lag variable. Finally, The demand impact estimate for a single premise is calculated as the difference between the two loads, with and without re-set. Program projections are the average impacts across all potential contributors. For the projections this group includes everyone except the AC non-users. The adjustments to this estimate provide projections for the whole program.

2.5.2 Accounting for Non-contributors

Impact projections provided by the weather normalization load models are for potential contributors. As with the impact estimates above, these potential contributor estimates must be adjusted to account for non-responders, over-riders, and zero AC users. The non-response and zero AC user percents for the purposes of projection are effectively constant. We use the average across the 12 re-set events for the non-response percent. The zero AC usage percent is already a single percent as used for the impact adjustments. The over-ride percent is more challenging for the projections because over-ride is clearly a function of temperature, re-set duration, time of re-set, and the re-set amount.

Modeling the Over-ride Rate

To account for the varying over-ride rates with re-set event conditions, we modeled the over-ride rate using a logistic regression. The over-ride rate p_{OR} is transformed to the log odds ratio

$$f(p_{OR}) = \ln(p_{OR}/(1-p_{OR})).$$

The log odds ratio was then modeled as a linear function of temperature, duration, and re-set amount. The predicted log odds was then transformed back to the predicted over-ride proportion as

$$p_{OR} = e^f/(1+e^f).$$

The transformation ensures that the predicted over-ride rates from the fitted model all fall in the range from 0 to 100 percent. With the resulting estimates of over-ride percent, we can adjust impact projections in a way that is consistent with the desired combination of re-set event conditions.

2.5.3 Comparison with Actual Re-set Events

The impact estimates developed for the individual re-set events were compared to projections estimated using this more general approach, with the corresponding average temperatures and re-set amounts. If the projections were consistently high or low, a suitable calibration could be devised for the projection model.

3

FINDINGS

This section describes the findings of the analysis of the metered consumption data and the re-set event data for Summer 2004. We first describe the data screening used to determine which meters had usable data for the analysis. We then present the results of the analysis steps described in Section 2:

- Estimation of the Fraction Noncontributing
- Impacts for the Re-set Event
- Projected Impacts for General Conditions.

As discussed in Section 2, the 2004 analysis both follows the same methodological approach as the 2003 analysis but also makes refinements of certain pieces of the analysis.

3.1 UNITS USED IN THE ANALYSIS

3.1.1 Identifying Participants Still in Program

One hundred sites were originally chosen for the Smart Thermostat Program sample. As of the start of the third summer of the program (July 1, 2004), 20 of the original participants were no longer active in the program. The thermostats at these sites were no longer contacted for programs re-sets or tests. For the majority of these sites, however, the interval data meters were still providing interval data to the program. Facing a loss of 20 percent of the original sample, with the prospect of further attrition for future years of the program, we decided to continue to include ex-participant sites in the analysis. Because they are no longer curtailed, ex-participants must be included with the comparison group for all re-set events. We refer to ex-participants as Group C. Group C is made up of ex-members of Groups A and B, but in the analysis, it will function as a separate group joining Group A or B depending on which is the comparison group.

Table 3-1 describes the distribution of sites, thermostats, and AC units for ex-participants and those remaining in the program.

Table 3-1
Sites Included in the 2004 Impact Report, Distribution of Sites, Thermostats and Metered AC by Curtail Group

		Sample Group A			Sample Group B			Total		
Participant Status	Premise Category	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermost at Count	Count of Metered AC
Left Program before July 1st, 2004 (Group C)	One AC, One metered	8	8	8	9	9	9	17	17	17
	Two Ac, both metered	3	6	6				3	6	6
Participated in at least on 2003 re-set event	One AC, One metered	37	37	37	33	33	33	70	70	70
	Two Ac, one metered	1	2	1	1	2	1	2	4	2
	Two Ac, both metered	2	4	4	6	12	12	8	16	16
	Total	51	57	56	49	56	55	100	113	111

3.1.2 Identifying Meters with Good Data

Of the original 100 participants, 94 sites provided acceptable 15-minute interval energy consumption data for our analysis. The six exceptions were sites with missing or suspicious interval data. Table 3-1 shows the overall breakout for sites, thermostats, and metered AC units.

Table 3-2
Premises, Thermostat, and Metered AC for Removed Sites

Premise Category	Site Count	Thermostat Count	Count of Metered AC
All original sites	100	113	111
Bad or insufficient data	6	7	7
Sites for Analysis	94	106	104

The complete interval data for 96 sites was received from SDG&E. It included 12 months of data from November 2003 through October 2004. The interval data was not sent where two customers moved and metering equipment was removed. In addition, two other sites had either AC or WH data missing for the majority of the year. Of these four sites, three were ex-participants and one was still enrolled in the program.

We conducted two further checks of the meter data. AC and WH usage data were compared by interval. AC usage from the AC meter should never be larger than the WH usage from the WH meter. For the 2003 analysis, seven sites were identified with this problem. Because the problems could not be resolved prior to the completion of the 2003 analysis, these sites were not included. This year, all but two of the questionable sites were cleared for inclusion in the analysis. These two sites are from the active participants group. They have failed this test all three years of the program.

Table 3-3 separates ex-participants from current participants (vertically) as well as breaking out the curtail groups.

Table 3-3
Premises, Thermostat, and Metered AC for Removed Sites, by Groups

	Sample Group A			Sample Group B			Total		
Premise Category	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermost at Count	Count of Metered AC
Ex-Participants Used in Comparison Group	11	14	14	9	9	9	20	23	23
Insufficient Data	2	2	2	1	1	1	3	3	3
Participants for Group C	9	12	12	8	8	8	17	20	20
All Current Participants	40	43	42	40	47	46	80	90	88
Insufficient Data	1	2	2				1	2	2
Bad Data	1	1	1	1	1	1	2	2	2
Participants for Curtail Groups A and B	38	40	39	39	46	45	77	86	84

As a final test of the interval data, we flagged zero whole-premises usage. Valid zero readings are unlikely for a whole-premises meter unless there is a system outage. There were intervals with zero usage but no sites had widespread zero WH usage and no zero usage intervals took place on curtail days. Assuming that zero WH usage did not reflect normal usage, these intervals were not included in the weather normalizing load model portion of the analysis.

In total, six sites were removed from the analysis. Three were ex-participants and three were still active participants. This left 17 ex-participants that could be included in the analysis as a third group, Group C, that was never re-set and always joined the comparison group. The two curtail groups had a similar number of sites, 38 and 39, but Group B had six of the seven multiple AC units in the sample. This should not affect results as we normalized both WH and AC results. Each group includes a single site with two thermostats but AC metering on only one AC unit. Thus, WH results were normalized by the number of thermostats and AC results by the number of metered AC units.

3.1.3 Units Included in Each Analysis Component

As described in Section 2, the AC units were classified as either “non-contributors” or “potential contributors” for each re-set day. Sites that had zero usage on all summer weekdays were non-contributors for the whole scope of the analysis. Sites that did not receive the re-set signal for a particular event, or who over-rode the re-set signal once it was received, were considered non-contributors for that event. Thus, for any particular re-set event, potential contributors were those with successful signal receipt, no over-ride, and non-zero usage during summer weekdays. Load data analysis was used to determine the savings per unit for potential contributors. This unit savings was then adjusted by the estimated population percent of potential contributors to obtain the average savings over all units, including the non-contributors.

Identifying AC Non-users

AC non-users were identified by the absence of AC data indicating more than minimal AC use during weekday afternoons. AC use was defined as a quarter-hourly consumption observation greater than 0.025 kWh to allow for the possibility of continuously running case or emollient heaters in the condenser. Minimal AC use then was defined as having less than one percent of quarter-hourly observations between 10 AM and 10 PM on weekdays between May 1 and October 1 showing AC use.

Since only one AC energy consumption meter was used at any one site, two-thermostat sites considered non-users necessarily showed no AC use from either thermostat. If they showed AC use, it could not be discerned whether one thermostat might have been a non-user. It is also recognized that metering errors could result in the appearance of no AC use at any hour.

Table 3-3 lists the counts of sites, thermostats and metered AC units for AC non-users in the remaining sample members. Group C is not broken out to avoid confusion. In fact, only the overall, full sample counts are relevant to the further analysis. Groups A and B are compared to show that their incidence of non-users is similar. This is not changed by the removal of Group C.

Table 3-4
Premises, Thermostat, and AC Meter Count of AC Non-users

	Sample Group A			Sample Group B			Total		
Premise Category	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermost at Count	Count of Metered AC
All sites for analysis	47	52	51	47	54	53	94	106	104
AC Non-Users	9	10	10	8	9	8	17	19	18
Potential Impact Contributors	38	42	41	39	45	45	77	87	86

Non-responding Thermostats

Non-responding thermostats were identified as non-responders on the Silicon Energy EEM Suite website. The non-responders were identified by event reports available from that website (sdgerem.siliconenergy.com/siliconenergy/rem/asp/event_summary_setup.asp). Non-responder thermostats had neither an acknowledgement time stamp nor an over-ride time stamp in the event report.

As discussed in Section 2, for some non-responders, the thermostat may in fact have received the signal and raised the cooling setpoint successfully but failed to send an acknowledgement reply to the system head end. Thus, the percentage of thermostats reported as non-responders could be viewed as an upper bound on the signal failure rate. On the other hand, there could also be cases where the signal was received but the re-set did not occur. Recognizing these potential sources of over- and under-statement, we treat the percent not responding to the re-set signal as the percent that were not re-set.

Over-riding Thermostats

Over-ride thermostats also were identified by event reports available from the Silicon Energy EEM Suite website. Over-ride time stamps were available in those reports. They were believed to indicate the time of receipt of the over-ride acknowledgement message. Thus, there could be some delay between the time the occupant changed the setpoint and the reported over-ride time. The possible range of delay times is believed to exceed 15 minutes.

3.2 Fractions Potentially Contributing and Not Contributing to Savings

The method employed in the 2002 analysis utilized population data where possible to provide more accurate estimates of per unit load impact. Because of the effective changes in program delivery, data reflecting the whole population were not available for the 2003 analysis or this, the 2004 analysis. We will follow the method employed for the 2003 analysis. This method followed the same basic method as the 2002 analysis with some necessary changes. Whereas for the 2002 analysis only the non-zero AC use portion of the non-participant fraction had to be estimated, for the 2003 and 2004 analyses, all three parts (non-response, over-ride, and non-zero AC use) must be estimated.

3.2.1 AC Non-users

For the 2004 analysis, as with both of the previous analyses, the percentage of AC non-users must be estimated from the final set of sites with good data. Table 3-3 above indicates that 18 units were categorized as non-users out of the total 104 metered units still in the analysis. For every re-set day, then, the fraction of AC non-users is 18/104.

3.2.2 Non-responding Thermostats

The non-response fraction must be calculated in the same manner as the 2003 analysis. This differs from the original method employed in 2002. If the full Smart Thermostat Program was re-set then there would be a confirmation page returned from each affected thermostat in the whole program population. This took place on the single re-set day in the summer of 2002. Two hundred thirty-two participants did not confirm receipt of the re-set. The fraction of non-responding participants, 232/2,259, was known with certainty. It entered into the impact estimate with zero variance.

The 2003 and 2004 program activities did not include the full Smart Thermostat population, only the sample groups and other Smart Thermostat participants who had opted to participate in the SPP. As a result, there was no full population for which the non-response fraction could be known with certainty. To determine an impact estimate that reflected the whole Smart Thermostat Program population, we instead had to estimate the fraction of non-responders. Estimating this fraction will add variance to the ultimate impact estimates. On the other hand, the larger the population used for the estimate, the smaller the additional variance.

One option for reducing the variance was to include the SPP population in the estimate. The increase of the number of units by 131 in addition to the individual Smart Thermostat sample groups (at 33 and 38 units) decreased the variance of this estimate by a factor of roughly $\sqrt{3}$. If the two populations are not believed to have systematic differences, then use of the large group is clearly preferable. As re-set confirmation is a purely mechanical issue, there is little reason to suspect systematic differences on thermostat response. Table 3-5 compares the re-set percentages for the groups.

Table 3-5
Comparison of Re-set Rates Between the Smart Thermostat Sample Groups and the SPP Program Participants

Re-set Date	Start Time	Sample Group	Smart Thermostat Sample AC			SPP AC Counts		
			Confirmed	No Response	No Response Percent	Confirmed	No Response	No Response Percent
7/14/2004	14:00:00	B	43	4	9%	123	8	6%
7/22/2004	14:00:00	A	42	1	2%	117	14	11%
7/26/2004	15:00:00	B	43	4	9%	103	28	21%
7/26/2004	18:00:00	A	42	1	2%	124	7	5%
8/9/2004	14:00:00	B	40	7	15%	104	27	21%
8/10/2004	14:00:00	A	39	4	9%	107	24	18%
8/11/2004	16:00:00	B	39	8	17%	126	5	4%
8/27/2004	16:00:00	A	42	1	2%	122	9	7%
8/31/2004	14:00:00	B	42	3	7%	125	6	5%
9/8/2004	16:00:00	A	40	3	7%	122	9	7%
9/9/2004	16:00:00	B	42	3	7%	120	11	8%
9/10/2004	16:00:00	A	41	2	5%	123	8	6%

The average non-response rate for the Smart Thermostat and SPP participants across the 12 re-set events for 2004 was 9 percent. This was substantially higher than the average of 6 percent for 2003. With testing, this difference proved to not be statistically different from zero. This indicates that, despite the seemingly large absolute difference, these two averages could come from the same distribution.

3.2.3 Over-ride Thermostats

The over-ride stamp always indicates that the setpoint has been reduced from the re-set signal level. A thermostat that was set to a higher setpoint than that set by the re-set signal, or an AC unit that was turned off, would not be registered as over-riding. Thus, over-riding thermostats always reduce the total savings.

As with the non-response percentage discussed above, because the full Smart Thermostat population did not participate in re-set events, over-ride percents must be estimated. Once again, taking advantage of the combined SPP/Smart Thermostat group numbers could have lowered the variance on this estimate. Unfortunately, the choice to over-ride is much more complex than the mechanical possibility of non-response. In addition, the Smart Thermostat Program sample and the SPP participants are responding to different programs. Table 3-6 compares the over-ride rates for the two groups. This comparison appears to show quite different rates of over-ride.

Table 3-6
Comparison of Over-ride Rates Between the Smart Thermostat Sample Groups and the SPP Program Participants

			Smart Thermostat Sample AC			SPP AC Counts		
Re-set Date	Start Time	Sample Group	Confirmed	Over-ride	Over-ride Percent	Confirmed	Over-ride	Over-ride Percent
7/14/2004	14:00:00	B	43	18	42%	123	22	18%
7/22/2004	14:00:00	A	42	13	31%	117	14	12%
7/26/2004	15:00:00	B	43	4	9%	103	5	5%
7/26/2004	18:00:00	A	42	5	12%	124	14	11%
8/9/2004	14:00:00	B	40	15	38%	104	14	13%
8/10/2004	14:00:00	A	39	15	38%	107	25	23%
8/11/2004	16:00:00	B	39	11	28%	126	11	9%
8/27/2004	16:00:00	A	42	5	12%	122	8	7%
8/31/2004	14:00:00	B	42	8	19%	125	25	20%
9/8/2004	16:00:00	A	40	9	23%	122	37	30%
9/9/2004	16:00:00	B	42	7	17%	120	19	16%
9/10/2004	16:00:00	A	41	13	32%	123	26	21%

For determining an over-ride fraction we will not make use of the additional participants in the SPP. The fraction will be estimated from the Smart Thermostat sample alone.

For purposes of this analysis, we use the percent that over-rode at any point during each event. Since over-rides increase as the re-set event continues, this over-ride percent is the maximum that occurred over the event—that is, the percent that had over-ridden as of the end of the event. Figure 3-1 shows the percent overriding as a function of time from the start of each re-set event.

Figure 3-1
Over-ride Percent as Function of Time During Event

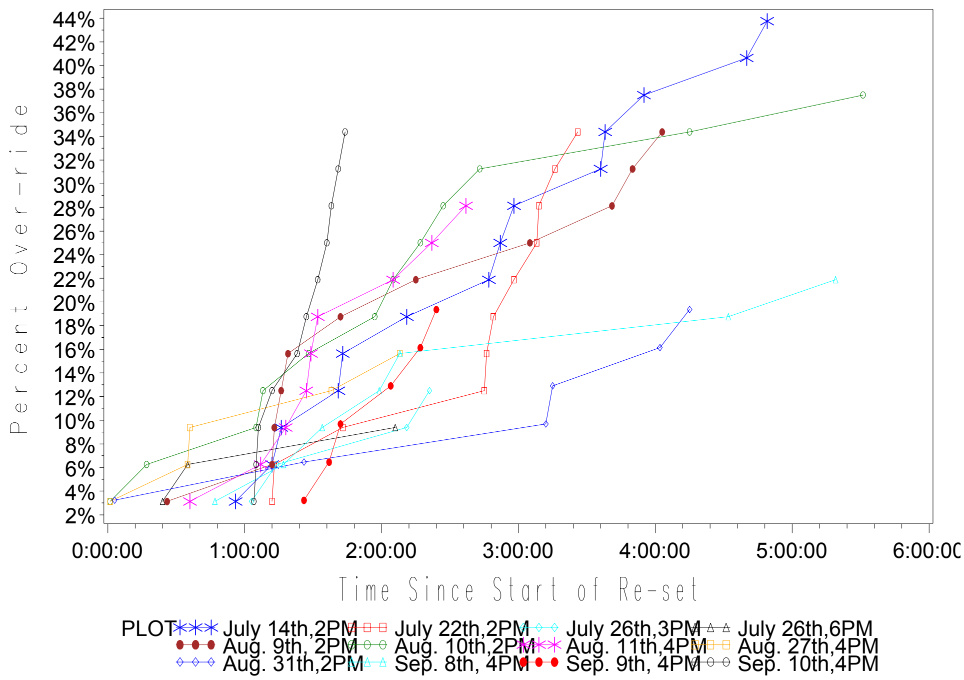


Table 3-7 shows the percent of over-riders with the factors most likely to be correlated with over-riding. For the 2003 analysis, there was a clear correlation between over-ride rate and average temperature. For the 2004 analysis the correlation between over-ride rate and average temperature was difficult to discern while there appeared to be a correlation between over-ride rate and re-set duration.

Table 3-7
Percent Over-ride Compared to Average Temperature, Re-set Amount Event Duration and Time

Date	Sample group	Start Time	End Time	Hours Duration	Degrees Setback	Average Te	Percent Over-ride
7/14/04	B	14:00	18:00	4	4	75	42%
7/22/04	A	14:00	19:00	5	4	71	31%
7/26/04	B	15:00	17:00	2	4	74	9%
7/26/04	A	18:00	20:00	2	4	74	12%
8/9/04	B	14:00	19:00	5	4	75	38%
8/10/04	A	14:00	19:00	5	4	76	38%
8/11/04	B	16:00	18:00	2	4	76	28%
8/27/04	A	16:00	18:00	2	4	71	12%
8/31/04	B	14:00	19:00	5	4	75	19%
9/8/04	A	16:00	21:00	5	4	77	23%
9/9/04	B	16:00	18:00	2	4	78	17%
9/10/04	A	16:00	18:00	2	4	80	32%

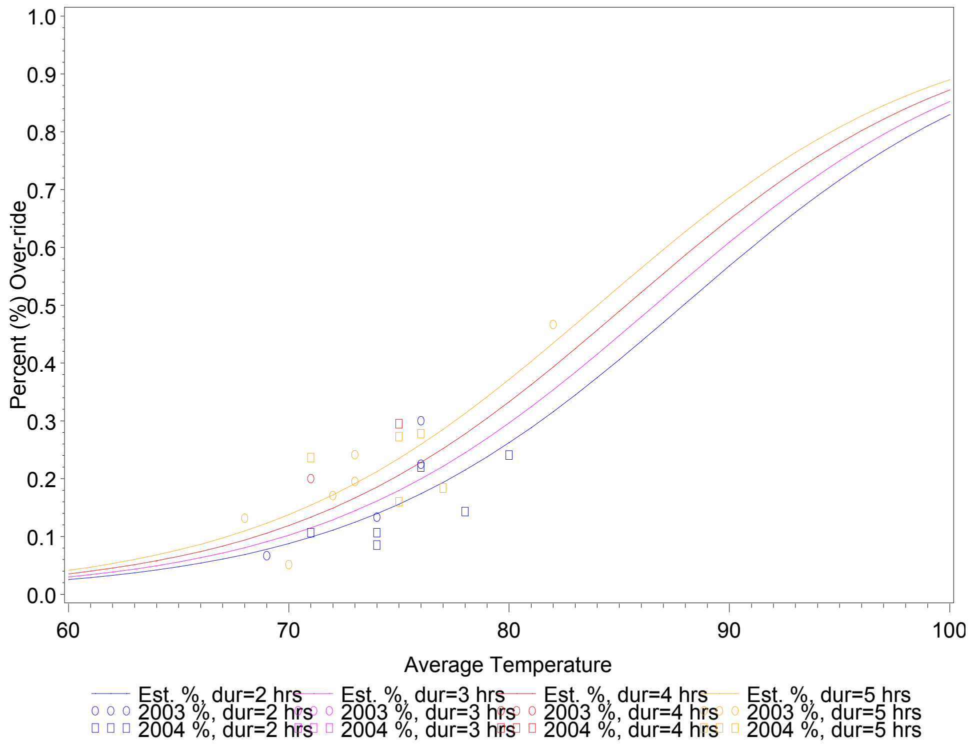
Modeling the Over-ride Rate

We explored whether the over-ride rate should be modeled separately by year using a single model. Comparing the separate year model results, there were substantial differences. Unlike the 2003 analysis where temperature was found to be the dominant driver of the over-ride rate, in 2004, re-set duration was the driver. The duration coefficient was significantly different from zero both when included alone in the model and in combination with temperature. Temperature, both alone and in combination with duration, was never significantly different from zero. In 2004, re-set amount was constant across all re-set days so could not be included in the model.

Based on 2004 evidence alone, it appeared unnecessary to include temperature in the over-ride rate model. The 2003 evidence, however, indicated that temperature alone should be included. Since the purpose of this model is for use in savings projections, ideally it would work across different years with different drivers. We tested a model with both average temperature and event duration and also using data from both years. Despite the apparent differences across the years, a dummy variable for year was not statistically significant. As a result, for the 2004 (and new 2003) projections, we use an over-ride model that includes data from both years.

Figure 3-2 shows the fitted regression line and the observed data. The four lines represent the four possible event durations, the longer the duration the higher the percent over-ride. The color coding representing the four different durations carries through to the plots of actual over-ride rates.

Figure 3-2
Predicted and Observed Over-ride Rates vs. Outdoor Temperature and Event Duration



3.2.4 Percent Not Contributing

Table 3-8 summarizes the fraction not contributing to impacts for the 12 re-set events in the summer of 2004. As discussed, the non-responder fractions reflect the percent non-response of the combined Smart Thermostat/SPP group for each date. The AC non-use percent is constant across the whole summer and reflects the faction of units still remaining in the analysis with zero AC usage. The over-ride fraction is calculated from the re-set sample group alone. The combined percent not contributing ranged from 32 to 63 percent across the 12 re-set days in 2004. This can be compared to a range of 26 to 68 percent not contributing for the 12 2003 events and 40 percent not contributing for the single 2002 re-set event. The average percent not contributing across the 12 2004 re-set events was 48 percent, 6 percent higher than for 2003. Both the non-response fraction and the over-ride fraction are higher. The higher non-response fraction accounts for approximately two thirds of the increase.

Table 3-8
Percent Non-Responding User and Percent Not Contributing During 2004 Re-set Events

Re-set Date	Non-Response Fraction: No Response	Non-Response Fraction: ST/SPP participants	Non-Response Fraction: Fraction (P_F)	AC Non-use Fraction (P_z)	Over-ride Fraction: Over-riders	Over-ride Fraction: ST Sample Participants	Over-ride Fraction: Fraction (P_{or})	Percent Non-Responding User: PF +(1-PF)(Pz)	Percent Not Contributing: P_F +(1-P_F)(P_z+P_{or})
07/14/04	12	178	7%	17%	18	43	42%	23%	62%
07/22/04	15	174	9%	17%	13	42	31%	24%	53%
07/26/04	32	178	18%	17%	4	43	9%	32%	40%
07/26/04	8	174	5%	17%	5	42	12%	21%	32%
08/09/04	34	178	19%	17%	15	40	38%	33%	63%
08/10/04	28	174	16%	17%	15	39	38%	31%	63%
08/11/04	13	178	7%	17%	11	39	28%	23%	49%
08/27/04	10	174	6%	17%	5	42	12%	22%	33%
08/31/04	9	176	5%	17%	8	42	19%	22%	40%
09/08/04	12	174	7%	17%	9	40	23%	23%	44%
09/09/04	14	176	8%	17%	7	42	17%	24%	39%
09/10/04	10	174	6%	17%	13	41	32%	22%	52%

3.3 Validation of Load Models and Comparison Group

3.3.1 Re-set and Comparison Group Characteristics

As described in Section 2, it is necessary to compare the size distributions of the different curtail groups. The primary reason was to determine if there was a need to scale the savings by capacity and the appropriate magnitude of the scaling. The review also would reveal anomalous units. The 2002 and 2003 analyses concluded there was no need for scaling with their respective sample groups. The comparison was repeated for 2004 taking into account both the updated curtail groups and the ex-participant group (C) which was always included with the comparison group.

Table 3-9 shows the distribution of AC unit capacity for each analysis group. For the two curtail groups, the results were similar to those for the 2003 analysis. In addition, the ex-participants group was also quite similar in make-up though slightly lower in capacity than either curtail group. The smaller Group C had the affect of lessening differences between the curtail and comparison groups when Group B was curtailed and increasing the capacity difference when Group A was curtailed. We tested to make sure that this disparity did not negatively affect the final results. Removing Group C from the analysis had almost no effect on the impact estimates but did increase the variance of the estimate.

Table 3-9
Distribution of AC Unit Capacity (tons) by Data Scope

Data Scope	Group	Sites	AC Units	Mean	Median	Min	Max	Standard Deviation
Original Sample	A	51	57	3.8	4.0	2.0	6.0	0.9
	B	49	56	3.7	3.5	2.0	6.0	1.0
	Total	100	113	3.7	4.0	2.0	6.0	0.9
All Units With Good Data for the 2004Analysis	A	38	40	3.9	4.0	2.0	6.0	0.8
	B	39	46	3.6	3.5	2.0	6.0	1.0
	C	17	20	3.4	3.0	2.0	5.0	0.9
	Total	94	106	3.7	4.0	2.0	6.0	0.9
All Units with Good Data and Non-Zero Usage	A	32	34	3.9	4.0	2.0	6.0	0.9
	B	32	38	3.7	3.5	2.0	6.0	1.0
	C	13	15	3.6	3.5	2.0	5.0	0.9
	Total	77	87	3.8	4.0	2.0	6.0	1.0

3.3.2 Observed and Modeled Loads

Another type of method validation was examination of the quality of the load model fits for both the re-set and comparison groups. We considered both the AC end-use data and the whole-house data.

Table 3-10 summarizes key regression diagnostics for the end-use and whole-house model fits. The table indicates that the whole-house fits were generally better than the AC fits. The mean WH R^2 statistic was substantially higher though the t-statistics for the WH peak hour cooling slopes were not as high as the AC t-statistics.

Table 3-10
Mean Regression Diagnostics for End-use and Whole-house Load Model Fits

Regression Statistic		AC Data	Whole-Premise Data
Median R-Squared		0.50	0.79
Median Cooling Slope t-statistic	Hour		
	12	3.32	2.86
	13	5.68	4.02
	14	7.56	5.21
	15	10.23	6.55
	16	11.54	7.15
	17	11.84	7.33
	18	11.56	7.26

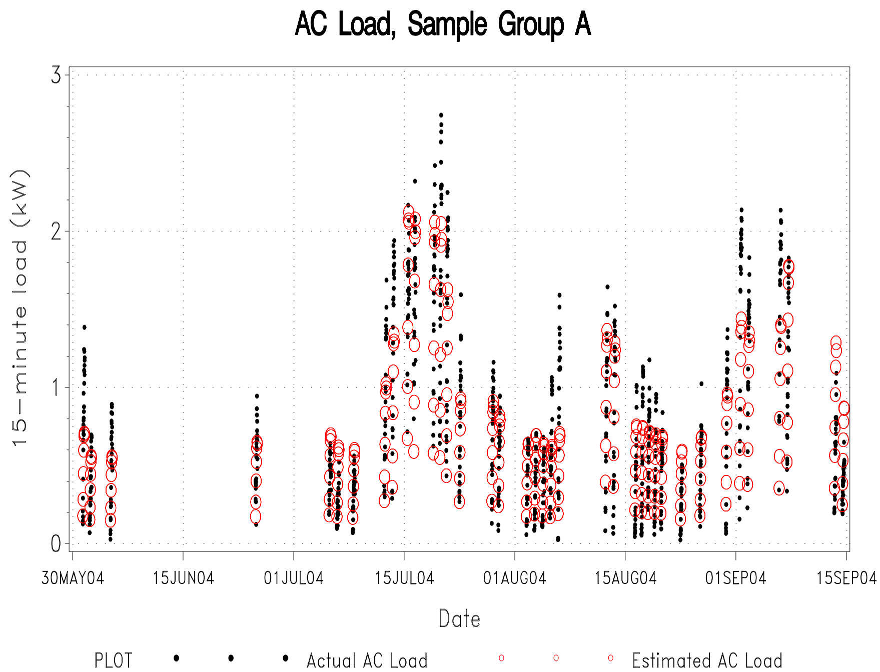
These comparisons are somewhat deceiving, because the data in the two models are different. As discussed in Chapter 2, the weather normalization model will include heating, cooling, and base parameters as determined by the choice of optimal model configuration. Heating parameters are only rarely included in the AC weather regressions. This happened primarily when a heat pump was the source of cooling. Thus, the WH models are explaining a wider range temperature-related of variation, which includes both heating and cooling. The AC models are explaining a narrower range of variation, including cooling only.

The peak hour cooling t-statistics measure the strength of the cooling degree day- related usage estimates. Unlike the AC load, not all of the whole-premise load is weather sensitive. The increased non-weather related variability would be expected to decreased the accuracy of the whole-premise cooling parameters, as is seen.

Figures 3-3 and 3-4 show observed and modeled AC loads for the re-set and comparison groups, respectively. The plotted data are limited to intervals from weekdays with an average temperature of 68°F or higher between the hours 11 PM and 6 PM. The data shown are for the 64 sites classed as "potential contributors" as well as the 13 ex-participant sites. The black dots represent actual 15-minute load levels while the red-circles, because model estimation is done hourly, are hourly load levels. Average temperature may be above 68°F for only a subset of sites. Only days for which the majority of sites are at or above 68°F are included in the plot.

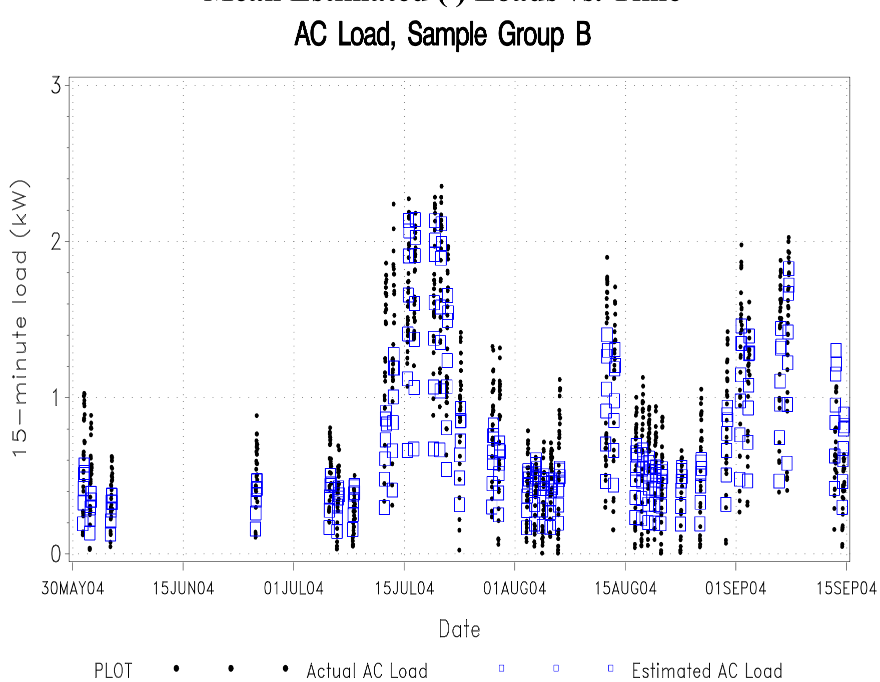
The weather normalization model does not provide particularly good estimates of the highest loads on the hottest days. The weather conditions in the San Diego area are particularly difficult to model. There is variation across the region with respect to temperature, humidity, and wind speed. Furthermore, it just isn't hot enough for consistent AC use. As described in Section 2, we attempted to account for the increasing use of air conditioners after multiple days of hot weather by adding a lag term to the model. The remaining underestimation in peak conditions indicates that there are further factors driving the peak, such as humidity, that remain unaccounted for in the model. As illustrated further below, this type of systematic modeling error across customers is a reason for the use of the comparison group in the analysis. The comparison group correction appears to take care of the residual unaccounted for peak loads.

Figure 3-3
Group A Warm Weekday 15-minute Mean Observed (•) and Mean Estimated (◦) Loads vs. Time



Points plotted are average values up to 41 “potentially contributing” AC units in Group A.

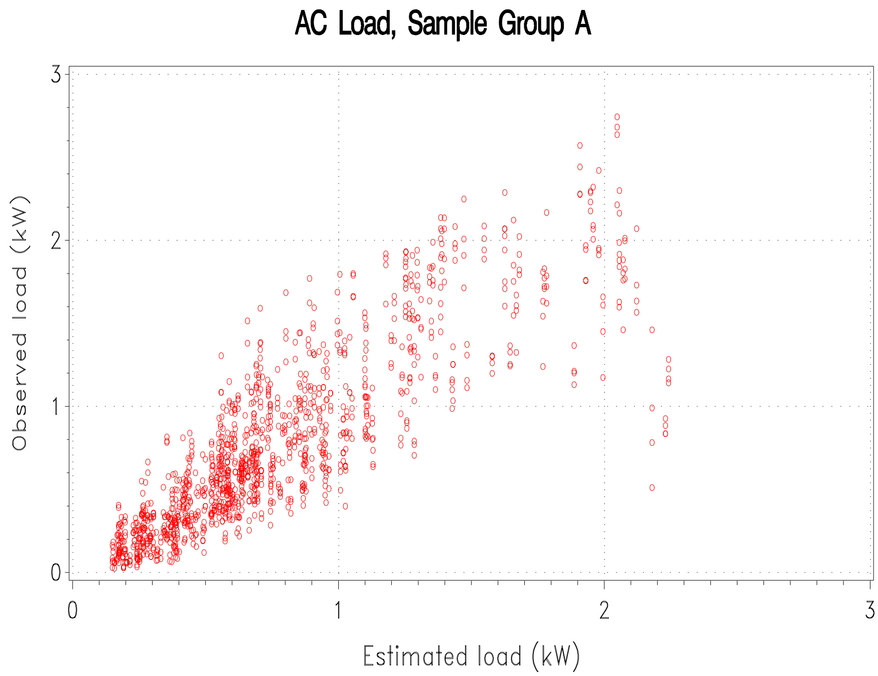
Figure 3-4
Group B Warm Weekday 15-minute Mean Observed (•) and Mean Estimated (◦) Loads vs. Time



Points plotted are average values of up to 45 “potentially contributing” AC units in Group B.

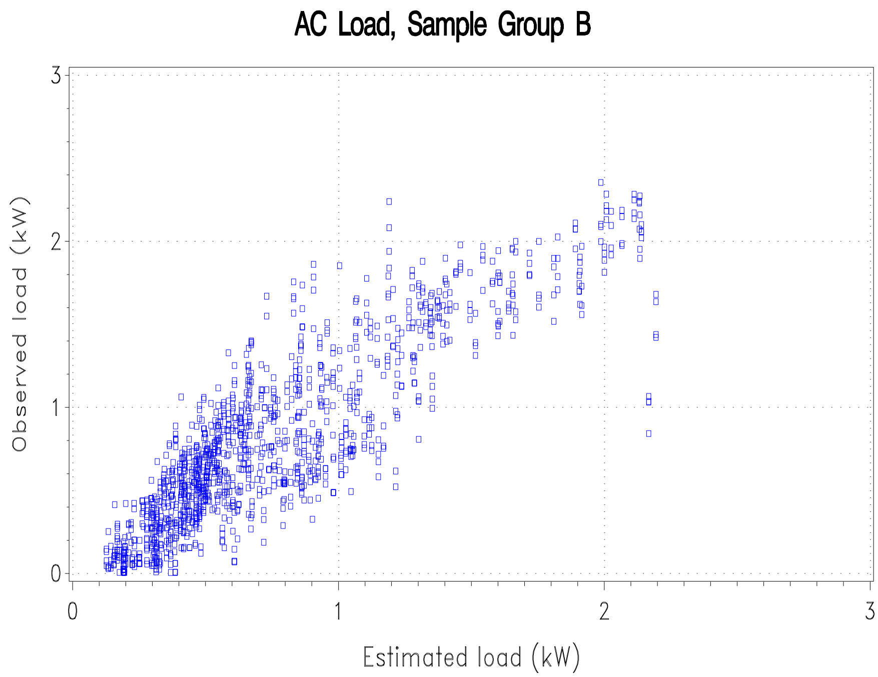
The same data shown in Figures 3-3 and 3-4 are plotted in Figures 3-5 and 3-6. These charts show observed versus modeled hourly mean loads. Both charts show a fairly uniform linear relationship along a 1:1 ratio of observed to estimated load. This is a good indicator of model fit. Still there is a fair amount of estimation error given that each point is an average error over 41 and 45 AC units.

Figure 3-5
Group A Warm Weekday Observed vs. Modeled 15-minute Mean Loads



Points plotted are average values up to 41 “potentially contributing” AC units in Group A.

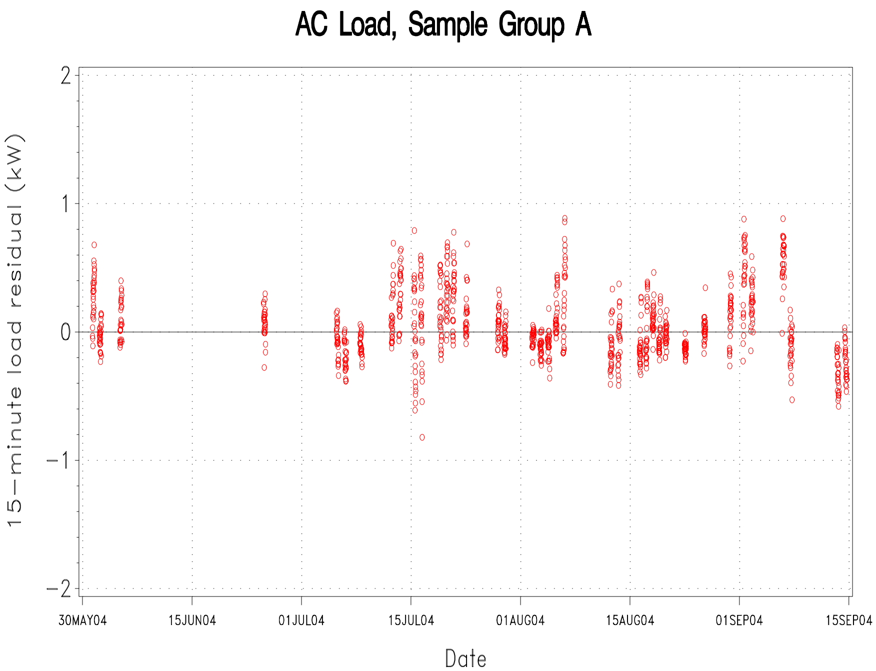
Figure 3-6
Group B Warm Weekday Observed vs. Modeled 15-minute Mean Loads



Points plotted are average values of up to 45 “potentially contributing” AC units in Group B.

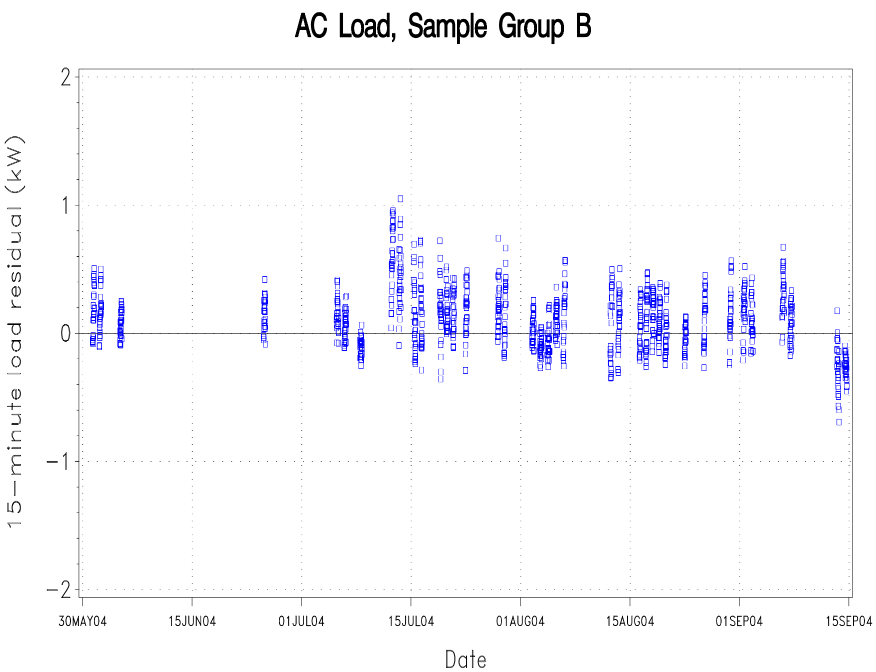
Figures 3-7 and 3-8 show the mean residuals, or errors, of the model estimates of hourly mean load for warm days from May 30 to September 15 for the re-set and comparison groups, respectively. As this is only the warm days from the period we would not necessarily expect the mean of the residuals to be zero. The larger magnitude errors are concentrated among a few hours and not scattered across all hours. The patterns of errors over time are very similar between the re-set and comparison groups. This relationship is consistent with the conjecture that particular weather conditions for those days with larger errors create systematic modeling errors across sites. The similarity of the error pattern also shows that errors of the comparison group can be a good indicator of the error of the re-set group error for a given day and hour.

Figure 3-7
Group A Warm Weekday 15-minute Mean Load Residual vs. Time



Points plotted are average values up to 41 “potentially contributing” AC units in Group A.

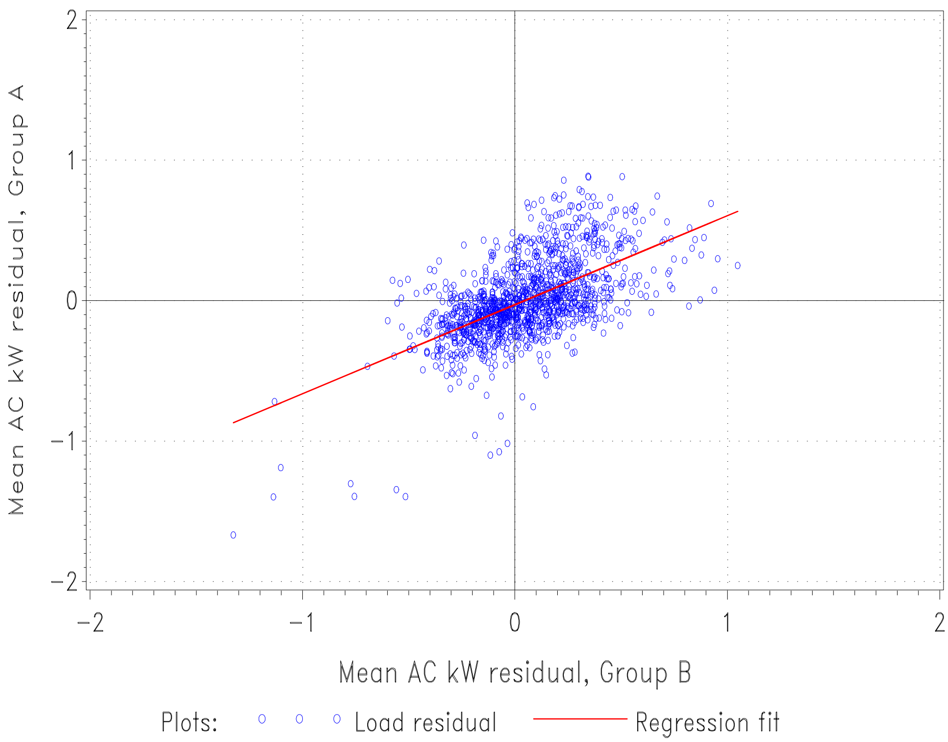
Figure 3-8
Group B Warm Weekday 15-minute Mean Load Residual vs. Time



Points plotted are average values of up to 45 “potentially contributing” AC units in Group B.

The difference of difference method requires not just that the two groups be similar in actual load, but also that the modeling error for the comparison group be a good indicator of the modeling error for the other. Figure 3-9 shows a plot of the comparison group's hourly mean residuals against the re-set group's residuals. Also shown is the regression line, which gave an R^2 of 0.40.

Figure 3-9
Curtail Groups A vs. B 15-minute
AC Mean Load Residuals and Regression

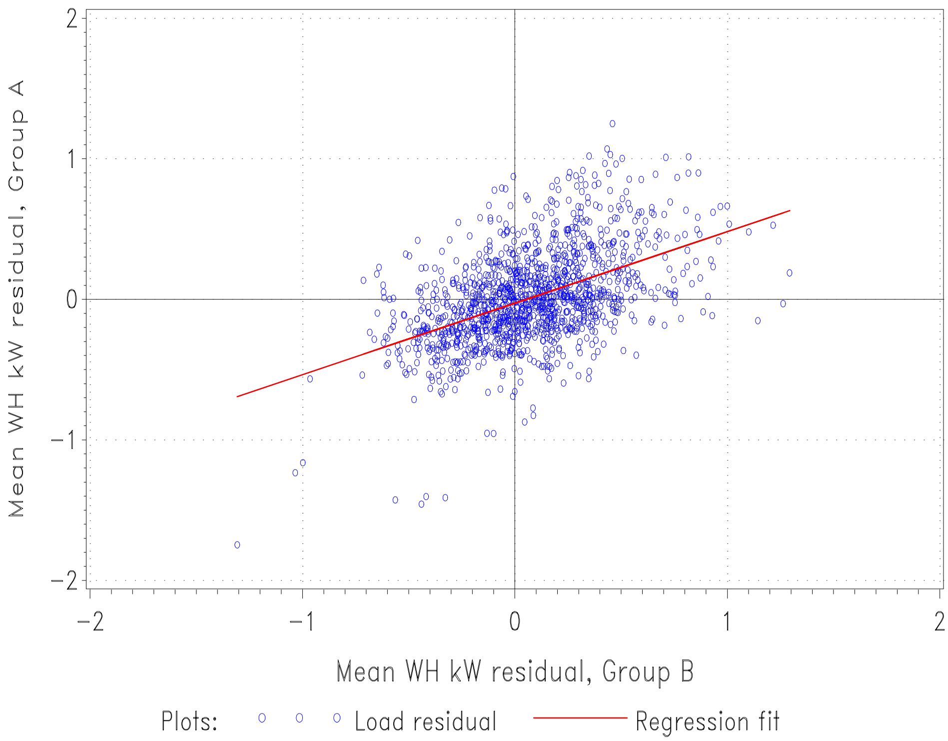


Points plotted are average values over 33 Group A and 38 Group B "potentially contributing" AC units.

The figure shows a strong relationship between comparison group and re-set group modeling error, with the regression line passing very close to the center point (0,0). The plot also indicates that the scale of the errors is similar so that no scaling adjustment is required when using the comparison group to estimate the re-set group error. Thus, the difference of differences method appears to be well founded for the end-use AC data.

Figure 3-10 shows a similar plot for the whole-house data. The general pattern of correspondence between the two groups in the whole-house analysis is similar to that for the AC analysis. However, the range of variation is greater, as indicated before. Thus, the difference of differences method is appropriate using the whole-house data, but the accuracy of the estimates will be worse than with the AC data.

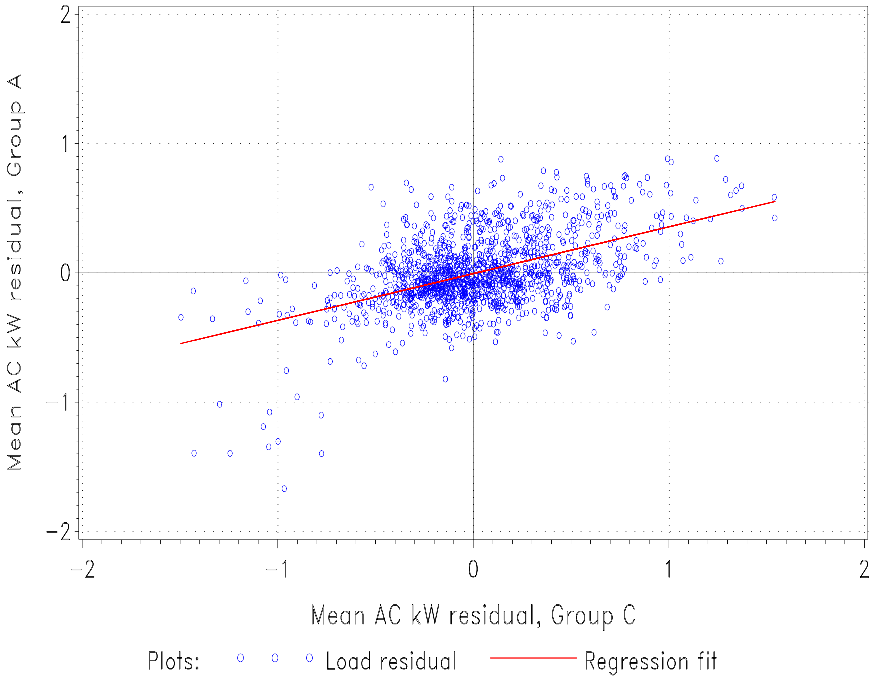
Figure 3-10
Curtail Groups A vs. B Modeling Error
Whole-House Data



Points plotted are average values over 34 Group A and 38 Group B "potentially contributing" AC units.

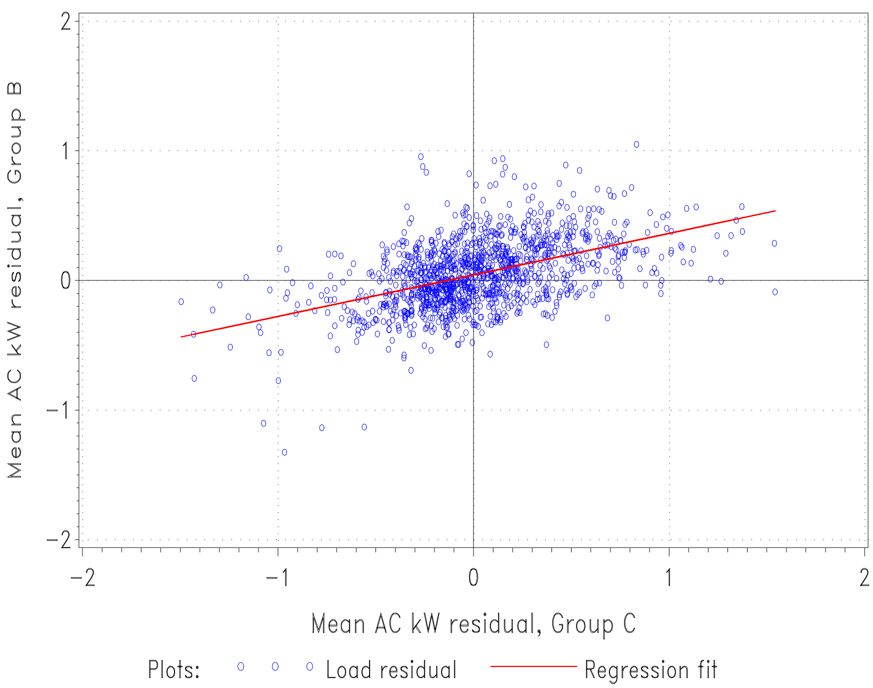
The previous plots compare curtail Groups A and B. Curtail Group C was not included because it is, relatively speaking, smaller, and as it joins each of the other curtail groups in turn, in aggregate its effect should be minimal. The plot comparing two groups' residuals is the one plot that is important to reproduce including curtail Group C. We want to make sure that Group C does not undermine the relationship between curtail Groups A and B.

Figure 3-11
Curtail Groups A vs. C 15-minute
AC Mean Load Residuals and Regression



Points plotted are average values over 33 Group A and 15 Group C "potentially contributing" AC units.

Figure 3-12
Curtail Groups B vs. C 15-minute
AC Mean Load Residuals and Regression



Points plotted are average values over 38 Group B and 15 Group C "potentially contributing" AC units.

These plots are very similar to each other. They exhibit the same strong relationship between the two groups' modeling error and regression lines show no systematic shift away from the origin. While slightly different from the A vs. B plots shown earlier, the important issue for curtail Group C is whether it would have a differential effect by being added to one curtail group as opposed to the other. These plots indicate curtail Group C has a similar relationship to the two primary curtail groups.

3.4 ESTIMATED IMPACTS OF THE RE-SET EVENT

There were 12 re-set events during the summer of 2004. Table 3-11 gives an overview of the times, degrees re-set, and sample group for each re-set day.

Table 3-11
Re-set Event Times, Degrees Re-set, and Sample Group

Date of Re-set	Sample Group Re-set	Start time	End time	Average Temperature	Degrees Setback
7/14/04	B	14:00	18:00	4	75
7/22/04	A	14:00	19:00	4	71
7/26/04	B	15:00	17:00	4	74
7/26/04	A	18:00	20:00	4	74
8/9/04	B	14:00	19:00	4	75
8/10/04	A	14:00	19:00	4	76
8/11/04	B	16:00	18:00	4	76
8/27/04	A	16:00	18:00	4	71
8/31/04	B	14:00	19:00	4	75
9/8/04	A	16:00	21:00	4	77
9/9/04	B	16:00	18:00	4	78
9/10/04	A	16:00	18:00	4	80

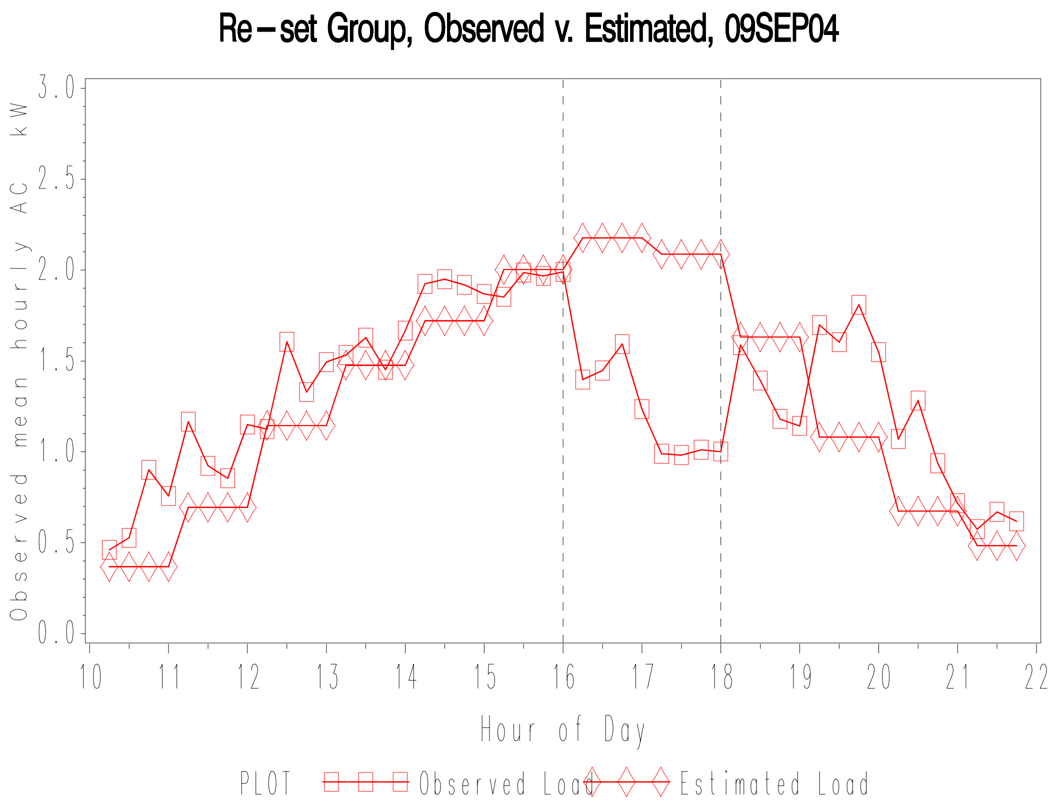
The following section displays the methodology used in this analysis in visual plots. The re-set event that occurred on September 9th is used as an example. Similar plots for all remaining re-set events can be found in Appendix A. Corresponding plots for the whole-house analysis are in Appendix B. These plots show the impact of all responding users as opposed to potential contributors. As discussed in Section 2, including over-ride sites directly in the estimate of impacts makes sense as there is no decrease in variance to be gained by including them with the non-participating fraction.

One further anomaly should be noted. The September 10th re-set event may have experienced technical difficulties. Unlike any other event, three separate pages were sent, with two of them subsequently terminated. If the records are taken literally, there were two overlapping re-set periods for curtail Group A. Furthermore, for this day alone, many of the acknowledgement timestamps are days, even weeks, after September 10th. On the other hand, some participants were clearly re-set because they over-rode their re-set. This confusion is particularly unfortunate because this was the hottest re-set day of the summer. We continue to treat September 10th as a re-set event but the results for this day have to be considered suspect.

3.4.1 Modeled and Observed Load on a Re-set Day for Responding Users

Group B was re-set on September 9th between 4 PM and 6 PM or hours 17 and 18. Figure 3-13 shows the re-set group's observed load compared to the estimated load.

Figure 3-13
Observed (—) and Estimated (--) AC Loads on Re-set Day vs. Time

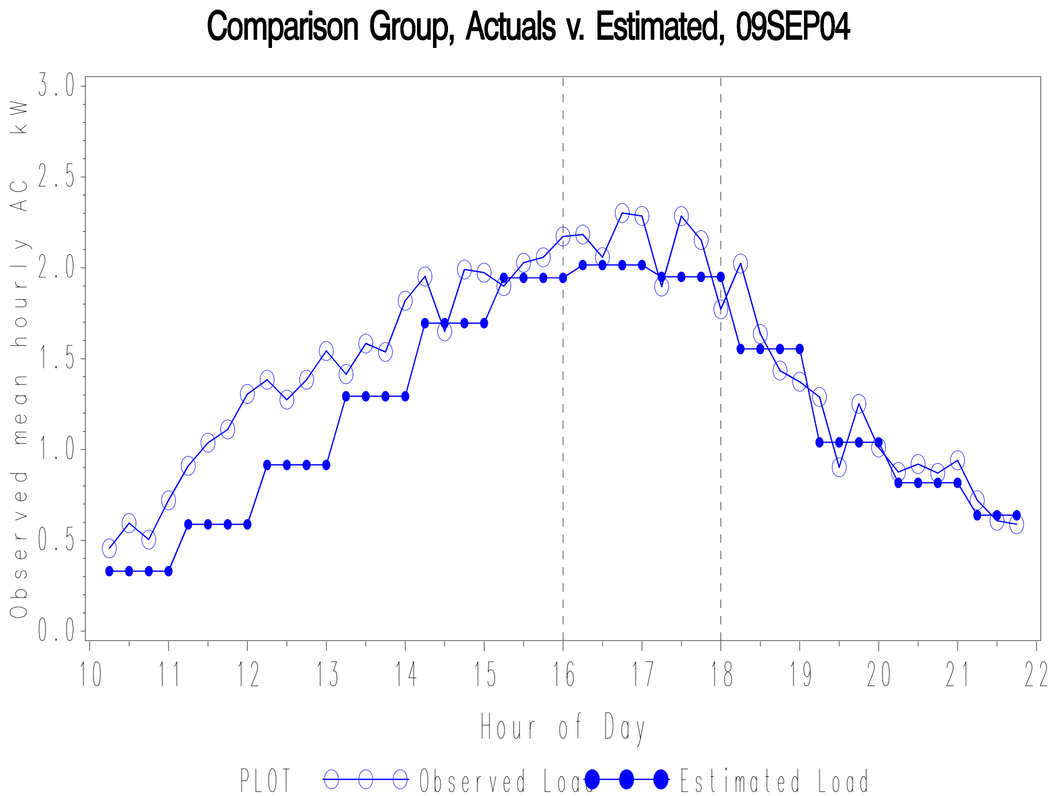


Points plotted are average values over re-set group comparison group "potentially contributing" AC units.

The re-set group's observed load diverged dramatically from the estimated load at the 4 PM start time. For this event, it took almost an hour for maximum reductions to take place. During the re-set period, units will start to cycle again because the internal temperature of the thermostat has reached the new setpoint. This effect takes over, though not dramatically, in the second hour of the re-set. After 6 PM, the re-set group's load jumps above the expected load as those AC units cycle on full time to compensate for lost cooling. The period after 6 PM, when observed load is higher than expected load, is the "payback" period discussed earlier. The difference between observed and estimated load for the re-set group is the unadjusted estimate of the impact for this re-set period. Note that prior to the re-set period, observed load appears somewhat higher than estimate load.

Figure 3-14 shows the comparison group observed load compared to the estimated load. The difference between observed and estimated load for the comparison group provides the adjustment for the impact estimate shown in Figure 3-12.

Figure 3-14
Observed (—) and Estimated (--) AC Loads on Re-set Day vs. Time



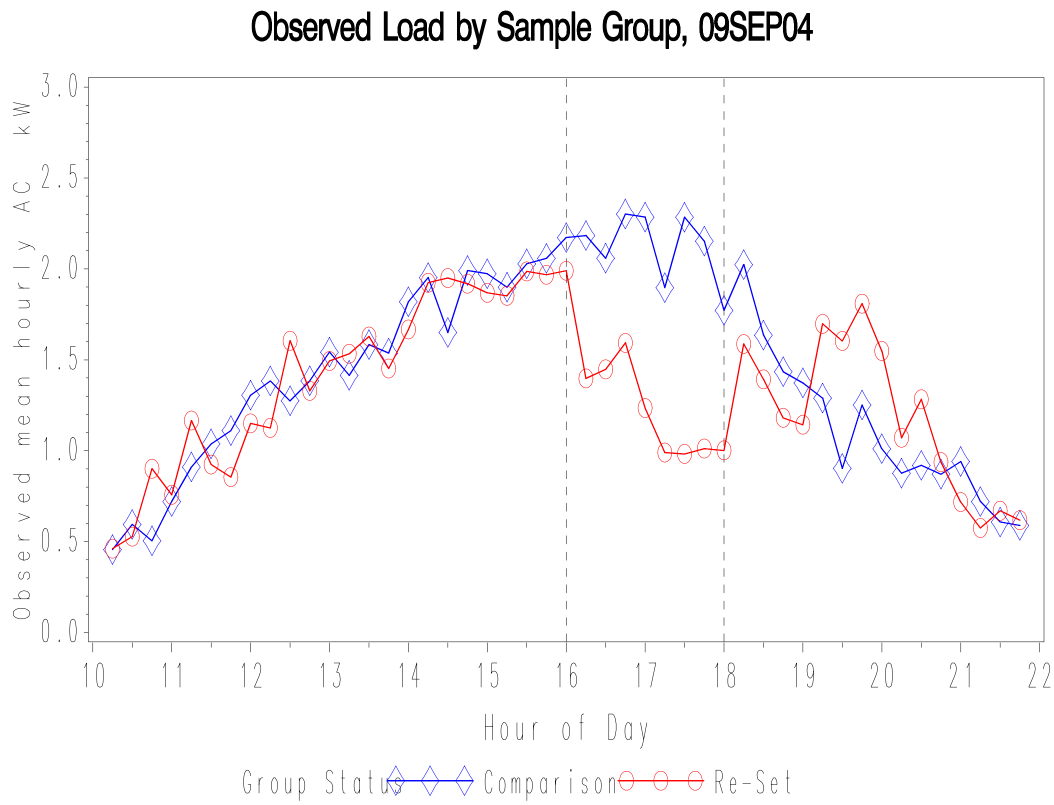
Points plotted are average values over comparison group "potentially contributing" AC units.

The comparison group gives us an idea of the effectiveness of our load modeling on that particular day. Figure 3-14 shows that the estimated load was generally below the observed load for the comparison group. This indicates that the estimated load for the re-set group in Figure 3-11 was too low by the same amount. Increasing the re-set group estimated load in Figure 3-11 by the difference in Figure 3-14 gives an estimate of load adjusted for the specific conditions of that day.

The difference of differences approach achieves just this. Subtracting observed from estimated load in Figure 3-14 results in a negative error adjustment. When this is subtracted from the unadjusted impact estimate in Figure 3-13 to get the difference of differences result, the double negative makes for a net increase to the impact estimate.

As indicated above we can conceive of the "difference of differences" method starting with either difference. Above we started with an unadjusted impact estimate for the re-set group derived from the estimated load and then adjusted it with the error from the comparison group. Alternatively, we can start with the difference between the two sample groups and adjust that impact estimate with the difference in the model-estimated loads. Figure 3-15 shows the re-set and comparison group loads on the re-set day.

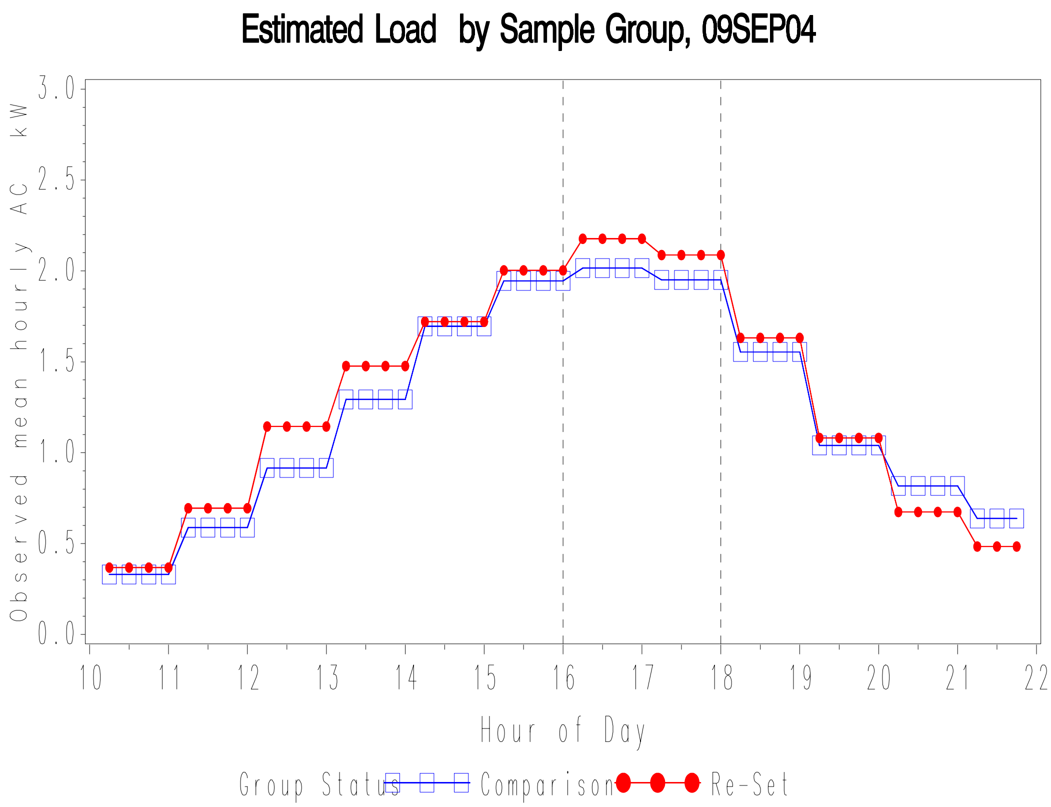
Figure 3-15
Observed 15-min Average AC Loads on Re-set Day vs. Time



Points plotted are average values over re-set and comparison group "potentially contributing" AC units.

This plot needs to be adjusted by the difference of the estimated loads for the two sample groups. Figure 3-16 compares the two load estimates.

Figure 3-16
Estimated 15-min Average AC Loads on Re-set Day vs. Time



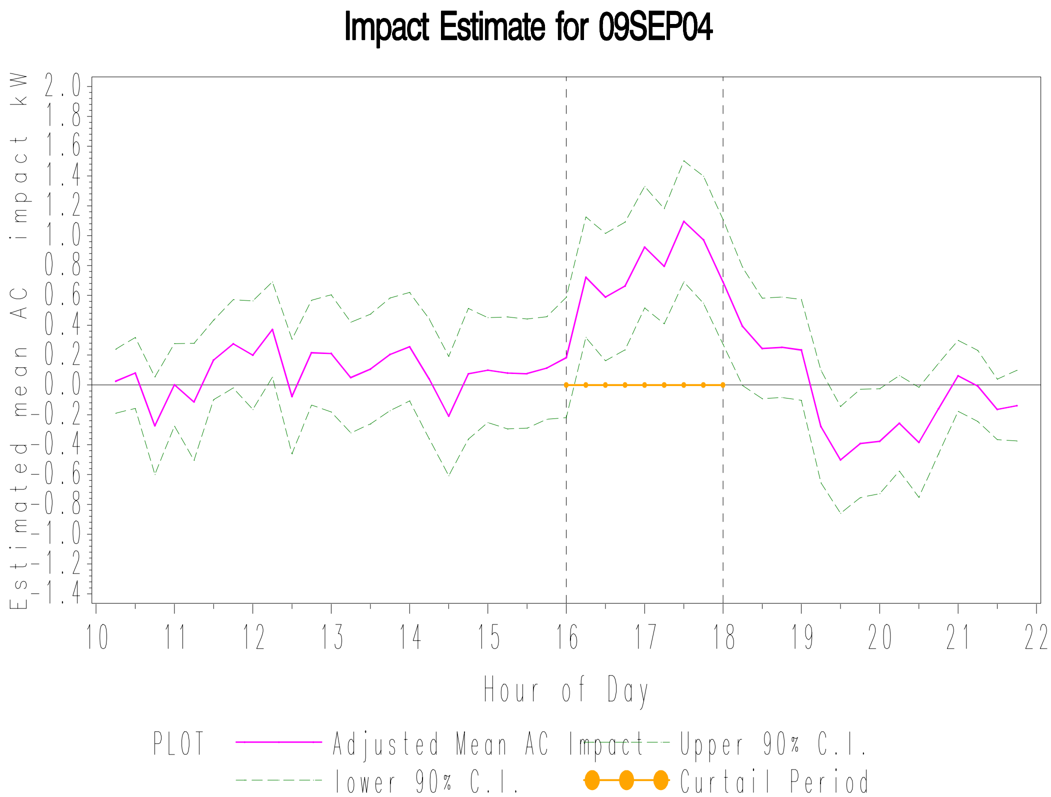
The re-set group estimated load is higher than the comparison group through the re-set period. This indicates that in Figure 3-15, all weather related effects being equal, either the re-set group load should have been lower by this amount to put the two groups on the same terms or, alternatively, the comparison group load increased by this amount. Either way, the original impact estimate is increased by the error adjustment. Viewed this way, the error adjustment is correcting for the tendency, noted earlier, for Group B, the re-set group on this day, to have higher load than Group A.

3.4.2 Savings Estimates

Air Conditioning Unit Impacts

The adjusted impact estimates, derived through the difference in differences approach, reflect the per-unit impact of all responding user units. These are units that have non-zero consumption for at least some part of the summer and received a re-set signal. It is still necessary to adjust this result so that it reflects a per-unit impact for all units in the program. This result is adjusted by multiplying the savings per responding user by the fraction these responding users represent of the whole. The percent of non-responding users for each day was given in Table 3-8 above. The remainder is the percent responding users. The final adjusted impact estimates for the September 9th re-set event are displayed in Figure 3-17.

Figure 3-17
Estimated Impacts on Re-set Day vs. Time



The 90 percent confidence interval lines indicate the level of statistical confidence in the estimate. If the confidence interval includes zero, then the estimate cannot be considered statistically different from zero at a 90 percent confidence level. Every 15-minute interval in the re-set period for September 9th is statistically significant.

These 15-minute interval results can be presented in aggregate form for the whole re-set event. That is, the average kW savings across all intervals in the re-set period is determined. Table 3-12 presents the results for all 12 re-set events.

Table 3-12
AC Impacts and Standard Errors for 12 Re-set Days

Re-set Event Date	Start time	End time	Sample Group Re-set	Average Temperature	Degrees Setback	Group A AC Count	Group B AC Count	Mean Impact for Responding AC Users	Percent Responding AC Users	Mean Impact per AC Unit	Standard Error
7/14/2004	2:00 PM	6:00 PM	B	75	4	48	34	0.73	77%	0.56	0.13
7/22/2004	2:00 PM	7:00 PM	A	71	4	32	53	0.44	76%	0.33	0.11
7/26/2004	3:00 PM	5:00 PM	B	74	4	48	34	0.72	68%	0.49	0.14
7/26/2004	6:00 PM	8:00 PM	A	74	4	32	53	0.40	79%	0.32	0.18
8/9/2004	2:00 PM	7:00 PM	B	75	4	48	31	0.82	67%	0.55	0.11
8/10/2004	2:00 PM	7:00 PM	A	76	4	31	53	0.37	69%	0.26	0.12
8/11/2004	4:00 PM	6:00 PM	B	76	4	48	30	0.84	77%	0.64	0.18
8/27/2004	4:00 PM	6:00 PM	A	71	4	32	53	0.40	78%	0.32	0.10
8/31/2004	2:00 PM	7:00 PM	B	75	4	50	33	0.60	78%	0.47	0.15
9/8/2004	4:00 PM	9:00 PM	A	77	4	31	53	0.40	77%	0.31	0.13
9/9/2004	4:00 PM	6:00 PM	B	78	4	50	33	1.06	76%	0.81	0.21
9/10/2004	4:00 PM	6:00 PM	A	80	4	31	53	0.13	78%	0.10	0.23

As indicated by the confidence intervals in Figure 3-15, the average kW impact for the September 9th re-set event is, in fact, statistically significant at the 90 percent level. For the 2004 analysis, only the suspect September 10th re-set event had a mean impact that was not significantly different from zero. The impact for this day is highlighted.

In these AC impact results, the savings associated with Group B are consistently higher than curtail Group A. This is despite that fact that the methods employed here are specifically employed to account for systematic differences across the samples. As was stated throughout this section, there is no evidence that points to problems with group make-ups. The addition of the ex-participants' dedicated comparison cohort was an obvious suspect but removing this group from the analysis did not solve the problem and, as expected, increased the error of the estimate. The difference in capacities, likewise, does not appear to be a sufficient explanation, especially as the pattern is hardly present in 2003. Finally, it is possible that, for no discernable reason, Group B just behaved differently than Group A in 2004.

Given the evidence, there is no clear explanation for the disparity in impacts. Ultimately, we fall back on the simplest reason for having alternating control groups. Because the two groups were re-set in an alternating fashion across the summer's re-set events, the overall mean impact results will not reflect any bias due to differences between the two curtail groups.

Table 3-13 provides the impact results with confidence intervals for both a single unit and for 5,000 units, the target size of the Smart Thermostat Program.

Table 3-13
AC Impacts with Confidence Intervals, Per Unit and for 5,000 Units

Date	Mean kW Per Unit				kW for 5000 Units		
	Impact	Standard Error	90% Confidence Lower Bound	90% Confidence Upper Bound	Impact	90% Confidence Lower Bound	90% Confidence Upper Bound
7/14/04	0.56	0.13	0.34	0.78	2,809	1,698	3,920
7/22/04	0.33	0.11	0.15	0.51	1,657	758	2,555
7/26/04	0.49	0.14	0.26	0.72	2,444	1,311	3,576
7/26/04	0.32	0.18	0.01	0.62	1,596	69	3,123
8/9/04	0.55	0.11	0.36	0.73	2,731	1,824	3,637
8/10/04	0.26	0.12	0.06	0.46	1,284	290	2,279
8/11/04	0.64	0.18	0.34	0.94	3,210	1,699	4,721
8/27/04	0.32	0.10	0.16	0.47	1,578	783	2,374
8/31/04	0.47	0.15	0.23	0.72	2,372	1,146	3,598
9/8/04	0.31	0.13	0.10	0.52	1,549	485	2,612
9/9/04	0.81	0.21	0.46	1.16	4,031	2,278	5,785
9/10/04	0.10	0.23	-0.28	0.49	517	-1,403	2,438

Whole-premises Impacts

The energy savings based on the whole-premises metering data are presented in Table 3-14.

Table 3-14
Whole-premises Impacts and Standard Errors for 12 Re-set Days

Re-set Event Date	Start time	End time	Sample Group Re-set	Average Temperature	Degrees Setback	Group A AC Count	Group B AC Count	Mean Impact for Responding AC Users	Percent Responding AC Users	Mean Impact per AC Unit	Standard Error
7/14/2004	2:00 PM	6:00 PM	B	75	4	49	34	0.89	77%	0.68	0.15
7/22/2004	2:00 PM	7:00 PM	A	71	4	33	53	0.64	75%	0.48	0.13
7/26/2004	3:00 PM	5:00 PM	B	74	4	49	34	0.79	67%	0.53	0.16
7/26/2004	6:00 PM	8:00 PM	A	74	4	33	53	0.41	78%	0.32	0.22
8/9/2004	2:00 PM	7:00 PM	B	75	4	49	31	0.98	66%	0.65	0.13
8/10/2004	2:00 PM	7:00 PM	A	76	4	32	53	0.50	69%	0.34	0.15
8/11/2004	4:00 PM	6:00 PM	B	76	4	49	30	1.06	76%	0.81	0.21
8/27/2004	4:00 PM	6:00 PM	A	71	4	33	53	0.28	77%	0.22	0.14
8/31/2004	2:00 PM	7:00 PM	B	75	4	51	33	0.76	78%	0.59	0.18
9/8/2004	4:00 PM	9:00 PM	A	77	4	32	53	0.37	76%	0.28	0.15
9/9/2004	4:00 PM	6:00 PM	B	78	4	51	33	1.20	76%	0.90	0.26
9/10/2004	4:00 PM	6:00 PM	A	80	4	32	53	-0.12	77%	-0.09	0.25

The whole-premises data impact results are consistent with the AC data results. In all but two instances, the results are higher than the end-use savings estimates. This could reflect the savings related to decreased use of the interior air distribution fan that is not included in the air conditioning metering data. In general, savings from the whole-premises data is 5 percent greater than the end-use data results.

As with the two previous analyses, the greater variation in the whole-premises data makes the final savings estimates less reliable. Only 9 of the 12 re-set days have statistically significant impact as opposed to the 11 days with the AC data. For this reason, we rely on the AC results as the primary impact estimates.

Table 3-15 provides the whole-premises impact results with confidence intervals for both a single premise and for 5,000 sites.

Table 3-15
Whole-premises Impacts with Confidence Intervals, Per Unit and for 5,000 Sites

	Mean kW Per Unit				kW for 5000 Units		
Date	Impact	Standard Error	90% Confidence Lower Bound	90% Confidence Upper Bound	Impact	90% Confidence Lower Bound	90% Confidence Upper Bound
7/14/04	0.68	0.15	0.43	0.94	3,411	2,127	4,694
7/22/04	0.48	0.13	0.26	0.70	2,409	1,297	3,521
7/26/04	0.53	0.16	0.27	0.80	2,674	1,339	4,008
7/26/04	0.32	0.22	-0.05	0.69	1,597	-270	3,463
8/9/04	0.65	0.13	0.43	0.87	3,249	2,155	4,344
8/10/04	0.34	0.15	0.09	0.59	1,705	446	2,964
8/11/04	0.81	0.21	0.47	1.15	4,051	2,327	5,775
8/27/04	0.22	0.14	-0.02	0.46	1,099	-87	2,285
8/31/04	0.59	0.18	0.29	0.89	2,948	1,447	4,448
9/8/04	0.28	0.15	0.03	0.53	1,408	155	2,660
9/9/04	0.90	0.26	0.47	1.34	4,519	2,353	6,686
9/10/04	-0.09	0.25	-0.52	0.33	-467	-2,587	1,653

3.5 Projected Impacts By Temperature and Re-set Amount

Projected impacts at various outside temperatures and re-set amounts were estimated from the same load models developed in the analysis of the specific re-set event, as described in Section 2. For each unit with good data and non-zero summer use, the unit's load model was used to calculate the load for each hour of the day at a given daily average temperature. The same model was used also to calculate the hourly loads assuming an increase in the thermostat setpoint. This increase was represented in the model as an increase in the unit's cooling reference temperature. Because a single cooling reference setpoint was estimated for both the day-of and lag coefficients, the setpoint is increased for both. The difference in the model's estimate of load with and without the setpoint change was the estimated savings at that outside temperature and re-set amount for each hour.

The addition of the lagged temperature variable for the 2004 analysis added a complication to this procedure. For any day-of average temperature the lag temperature representing the three previous days will vary depending on whether we are projecting savings for the first day of a hot spell or the fourth. Over the 24 re-set events in 2003 and 2004, the median lag temperature was .8ºF lower than average temperature of the re-set day. The first and third quartile differences are 2.2ºF lower and .3ºF higher, respectively, showing a range of just over 1ºF in each direction.

The bulk of the projections provided are based on the median scenario, a lag temperature that was .8ºF lower than the day-of average temperature. We also provide an indication of the range of outcomes given the range of possible lag temperatures. This could allow a program manager to fine tune projections to reflection weather for the previous days.

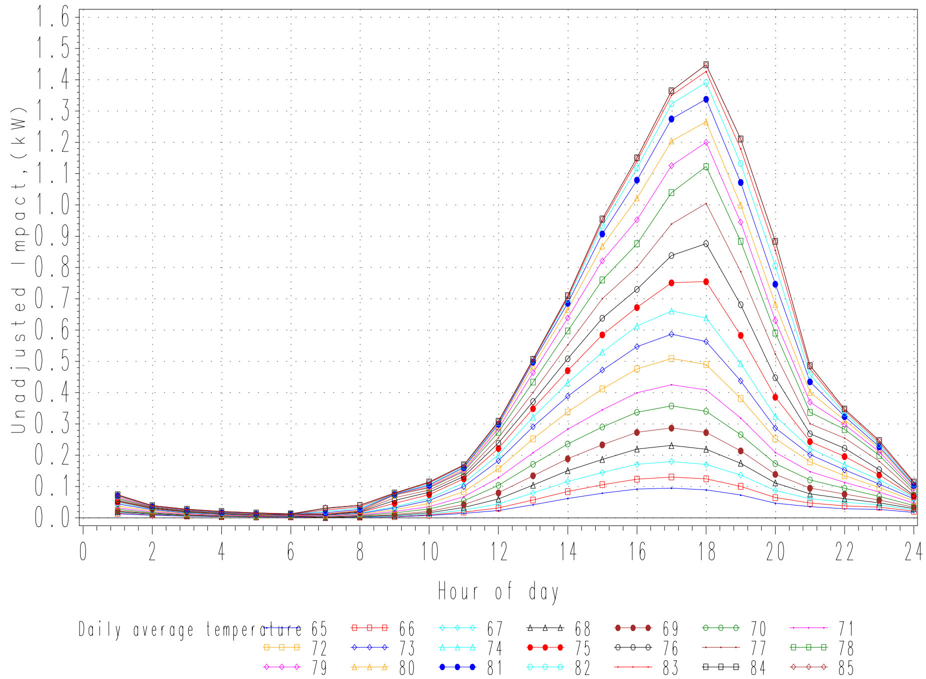
Once individual site savings estimates were produced, they were averaged across all units in the sample for which the model could be estimated. For this projection analysis, the assignment of units to re-set or comparison group was not relevant.

The unadjusted savings estimates apply to the universe of potential contributors. Adjusting these results by the fraction of potential contributors provides estimates of savings given expected rates of zero use, non-response, and over-ride.

The results are plotted by time of day in Figure 3-18 for a 4°F re-set, and various daily average outside temperatures. These are the impacts per potential contributor, without adjustment for signal failure, over-rides, or zero summer use. That is, Figure 3-18 shows the unadjusted projected impacts. Projected impacts, without adjustment for non-contributors, are tabulated in Appendix C for each combination of re-set amount and average outside temperature.

The figure shows that unadjusted savings are low at low outside temperatures, where air conditioning use is low and higher at higher outside temperatures. Savings are also low in the early morning and overnight. Savings per unit are greater at higher outside temperatures because a larger fraction of AC units are on. At lower temperatures, many of the units have zero estimated load and zero savings.

Figure 3-18
Unadjusted Projected Impacts by Hour, Average per AC Unit
(For 4°F Re-set)



For outside temperatures above 84°F, there is no additional increase in the unadjusted projected savings. This leveling off occurs once the outside temperature exceeds the point where all the units are projected to be on, based on the individual load model fits. The load models assume a linear relationship between load and outside temperature above each unit's reference temperature. Thus, a 4°F shift in reference temperature has the same affect on load for all outside temperatures above this reference point.

Figures 3-19a and 3-19b show the projected impacts adjusted by an estimate of the potential contributors percentage from the 12 re-set events during the summer of 2004. The first plot shows the adjusted impacts on the same scale as the unadjusted impacts displayed in Figure 3-16. The second plot expands the scale so that the patterns can be seen more easily.

For the adjusted projected impacts shown in Figures 3-19a and 3-19b, average signal failure and zero use percentages are applied with an estimated override percent. Override is estimated as a function of temperature and event duration as discussed in Section 3.2.3. Projected impacts, adjusted for non-contributors, are tabulated in Appendix D for each combination of re-set amount and average outside temperature.

Figure 3-19a
Adjusted Projected Impacts by Hour, Average per AC Unit
(For 4°F Re-set, 4-Hour duration)

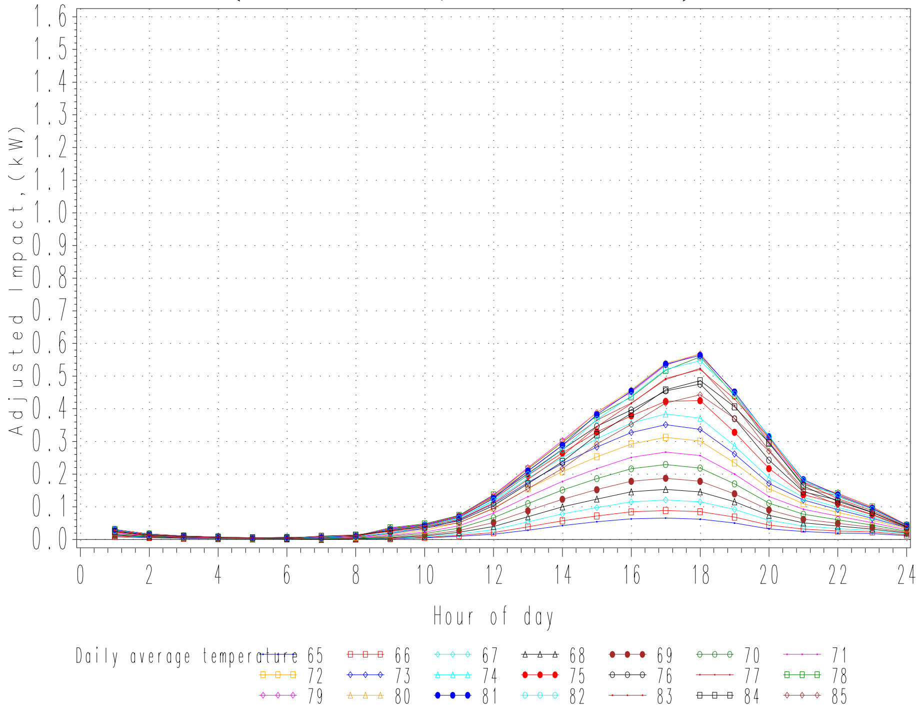
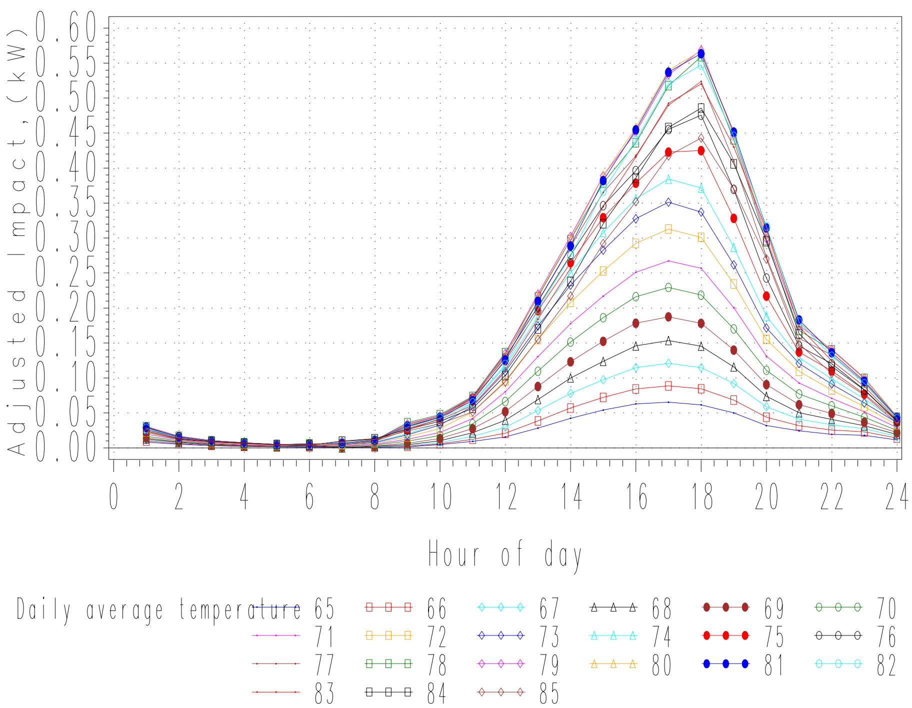


Figure 3-19b
Adjusted Projected Impacts by Hour, Average per AC Unit, Full Scale
(For 4°F Re-set, 4-Hour duration)



In Figures 3-19a and 3-19b the adjusted projected savings show no additional increase above 79°F. Savings is a function of temperature while over-ride rates are a function of both temperature and re-set duration. Both increase with increasing temperature, but at different rates. Figures 3-19a and 3-19b indicate that the effect of the increase in override percent is greater above 79°F than the increase in savings.

Figure 3-20, below, illustrates the relationship between savings and temperature for the peak hours. For the 2004 analysis, adjusted savings increases steadily into the high seventies at which point the over-ride rate takes over and the savings decrease.

Figure 3-20
Adjusted Projected Impacts by Temperature, Average per AC Unit
(For 4°F Re-set, 4-Hour duration)

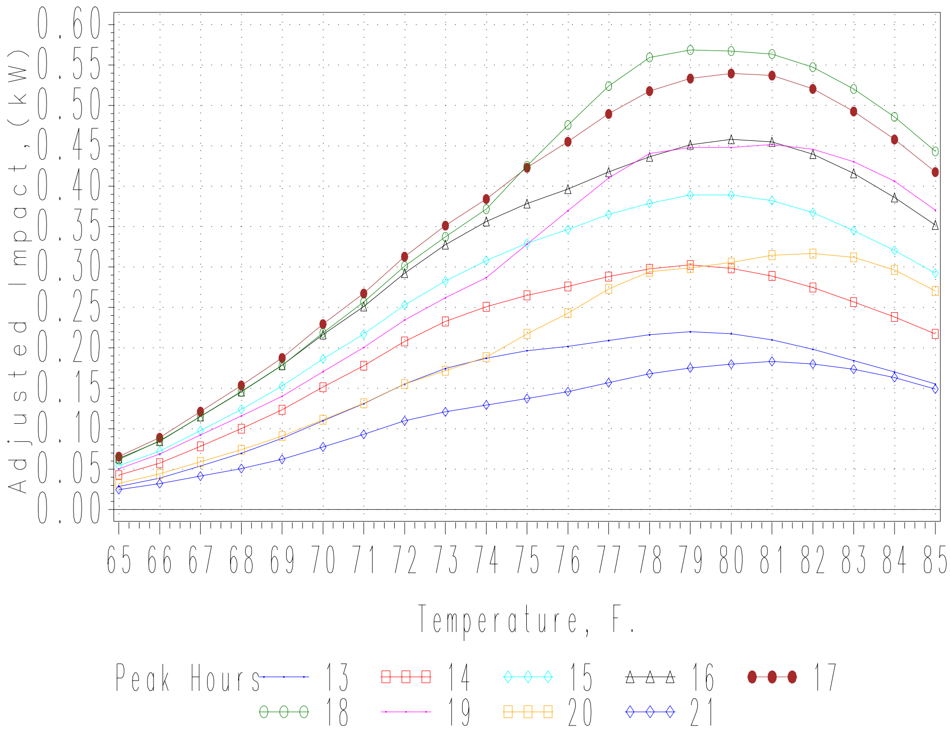


Figure 3-21 shows the range of hour 17 savings depending on the lag day temperature effect. The middle line is the same as hour 17 in Figure 20. This reflects the median relationship between day-of and lag day temperatures. The other two lines reflect the first and third quartile of the same relationship. If the previous days were slightly warmer than the day of the re-set, then savings would be higher.

Figure 3-21
Hour 17 Adjusted Projected Impacts by Temperature, Average per AC Unit with High and Low Lag Temperatures (For 4°F Re-set, 4-Hour duration)

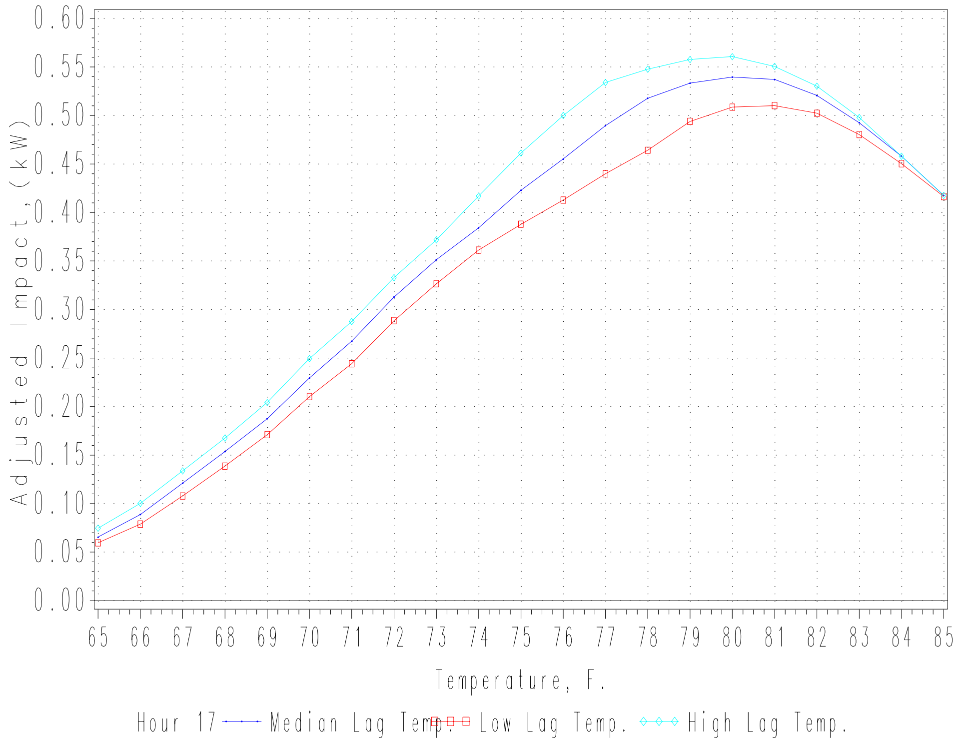


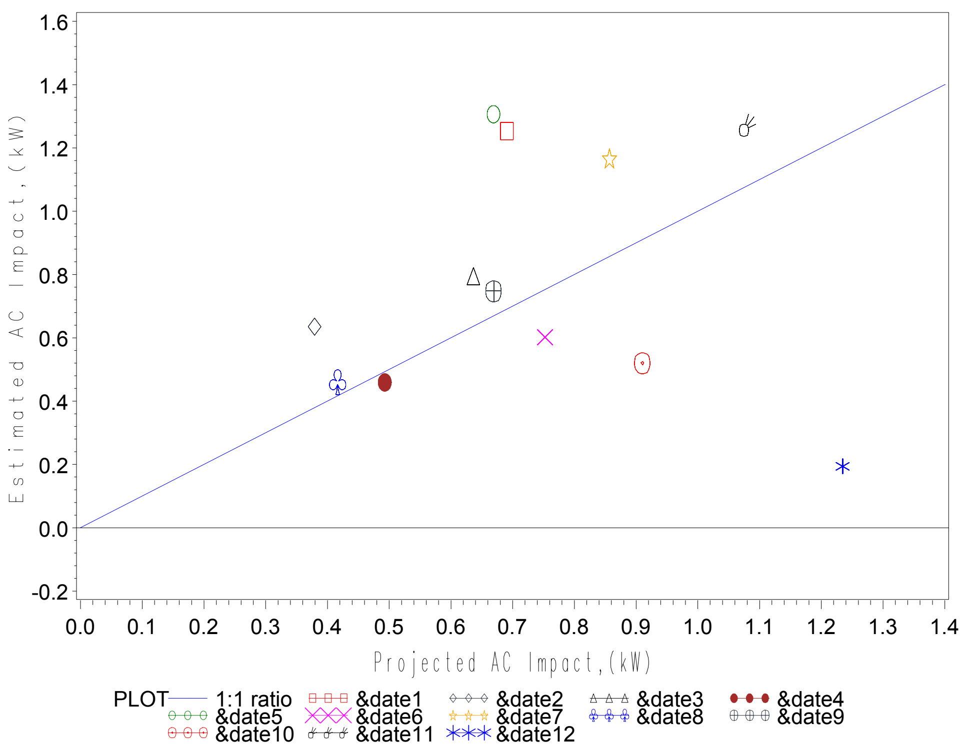
Figure 3-22, below, compares the unadjusted mean AC impact estimated earlier with the difference in differences method with the projected impacts. Projected impacts for each re-set are calculated as for Figure 3-18, but using the actual re-set amount for each event.

In this figure, the over-riding fraction has been backed out of the mean impact per responding user to provide the mean impact per potential contributor. We focus on this comparison, because it captures all the differences between the general and particular-day estimates. Immediately after an actual re-set event for the regular program, the fractions of over-riders and non-responders would be known directly from the program operation system so that the same adjustments would be applied for either method. The fraction of non-zero users assumed at this point would also be the same for either.

The diagonal blue line in the figure represents a 1:1 ratio. For the 2004 analysis, the plot indicates that our projections are on the low side for a number of events. This may result from the difficulty we had successfully modeling extreme temperatures. The projections do not have the comparison group to control for modeling errors of this sort. Also, the prediction model, based on the two curtail groups combined, tends to underestimate savings for the curtail Group B re-sets and slightly over estimate savings for curtail Group A re-sets. All but one of the re-set events above the 1:1 ratio line are curtail Group B re-sets.

Another potential reason for the disparity between the particular-day estimates and the corresponding projections for the general conditions is in how the particular day is mapped to the general condition. The general condition is defined by an ambient temperature and re-set amount. However, the sites in the study are modeled using seven different weather stations. For purposes of this comparison, we used the average temperature across all the non-zero users in the metering sample to define the general condition. Effectively then, the projection assigns the same temperature to all sites, whereas the particular-day estimates used the local temperature for each. This difference contributes to the variation seen in Figure 3-22.

Figure 3-22
Unadjusted Estimated vs. Projected AC Impact
Average Impact for Each Event



4 2003 ANALYSIS FINDINGS

This section provides updated results for the 2003 analysis. A combination of improved methods and resolution regarding the inclusion of additional data make direct comparison between the original 2003 results and the 2004 results difficult. This section reproduces all of the relevant 2003 analysis results consistent with the approach used in the 2004 analysis. Some of the charts examining load model results and providing examples of the difference of differences approach are not reproduced. Despite not reproducing all tables in Section 3 we keep the same numbers on tables and charts for ease of comparison.

4.1 CHANGES FOR THE 2004 ANALYSIS

Here is a list of the changes made for the 2004 analysis.

- Ex-participants (Group C) are no longer discarded but are included with the comparison group for each re-set event. This allows the inclusion of an additional six sites left out of the original 2003 analysis.
- Data issues for six sites were resolved clearing their inclusion in both the 2003 and 2004 analyses.
- Weather normalizing load model now includes a lagged average temperature variable as well as the option to force base load across all hours to zero.
- The over-ride model now includes both event duration and average temperature.
- Whole-premise impact results are now normalized to produce results on a per-unit basis as has always been the case with the AC results. This was motivated by the extreme disparity between curtail Groups A and B for the 2004 analysis with respect to the number of houses with multiple AC units.

4.2 UNITS USED IN THE ANALYSIS

4.2.1 Identifying Participants Still in Program

Table 4-1
Sites Included in the 2004 Impact Report, Distribution of Sites, Thermostats and Metered AC by Curtail Group

		Sample Group A			Sample Group B			Total		
Participant Status	Premise Category	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermost at Count	Count of Metered AC
Left Program before July 7th, 2003	One AC, One metered	5	5	5	3	3	3	8	8	8
Participated in at least on 2003 re-set event	One AC, One metered	40	40	40	39	39	39	79	79	79
	Two Ac, one metered	1	2	1	1	2	1	2	4	2
	Two Ac, both metered	5	10	10	6	12	12	11	22	22
	Total	51	57	56	49	56	55	100	113	111

Table 4-2
Premises, Thermostat, and Metered AC for Removed Sites

Premise Category	Site Count	Thermostat Count	Count of Metered AC
All original sites	100	113	111
Bad or insufficient data	6	7	7
Sites for Analysis	94	106	104

Table 4-3
Premises, Thermostat, and Metered AC for Removed Sites, By Group

	Sample Group A			Sample Group B			Total		
Premise Category	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermostat Count	Count of Metered AC
All Removed Participants As of summer 2003	5	5	5	3	3	3	8	8	8
Insufficient Data	1	1	1	1	1	1	2	2	2
Participants for Comparison Group	4	4	4	2	2	2	6	6	6
All 2003 Participants	46	52	51	46	53	52	92	105	103
Insufficient Data	2	3	3	0	0	0	2	3	3
Bad Data	1	1	1	1	1	1	2	2	2
Participants in analysis	43	48	47	45	52	51	88	100	98

Table 4-4
Site, Thermostat, and AC Meter Count of AC Non-users

	Sample Group A			Sample Group B			Total		
Premise Category	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermostat Count	Count of Metered AC	Site Count	Thermostat Count	Count of Metered AC
All sites for analysis	47	52	51	47	54	53	94	106	104
AC Non-Users	12	12	12	8	8	8	20	20	20
Potential Impact Contributors	35	40	39	39	46	45	74	86	84

4.2.2 Savings Estimates

Air Conditioning Unit Impacts

Table 4-12
AC Impacts and Standard Errors for 12 Re-set Days

Re-set Event Date	Start time	End time	Sample Group Re-set	Average Temperature	Degrees Setback	Group A AC Count	Group B AC Count	Mean Impact for Responding AC Users	Percent Responding AC Users	Mean Impact per AC Unit	Standard Error
7/17/2003	2:00 PM	3:45 PM	A	74	5	34	47	0.20	77%	0.16	0.19
7/28/2003	2:00 PM	7:00 PM	B	72	5	41	41	0.23	75%	0.17	0.11
8/8/2003	3:00 PM	5:00 PM	B	76	3	42	40	1.01	75%	0.76	0.19
8/15/2003	2:00 PM	7:00 PM	A	82	3	34	49	0.42	74%	0.31	0.18
8/27/2003	4:00 PM	6:00 PM	A	76	3	34	49	0.68	78%	0.53	0.14
9/3/2003	2:00 PM	7:00 PM	A	73	4	33	49	0.28	71%	0.20	0.08
9/12/2003	2:00 PM	7:00 PM	B	73	4	42	41	0.41	75%	0.31	0.09
9/22/2003	2:00 PM	6:00 PM	A	71	4	34	49	0.40	78%	0.31	0.10
9/29/2003	2:00 PM	7:00 PM	B	70	4	43	39	0.32	78%	0.25	0.07
10/9/2003	3:00 PM	5:00 PM	A	69	4	34	49	0.03	73%	0.02	0.07
10/14/2003	2:00 PM	7:00 PM	B	68	4	44	38	0.22	74%	0.16	0.08
10/20/2003	3:00 PM	5:00 PM	A	74	4	34	49	0.60	79%	0.48	0.18

Table 4-13
AC Impacts with Confidence Intervals, Per Unit and for 5,000 Units

	Mean kW Per Unit				kW for 5000 Units		
Date	Impact	Standard Error	90% Confidence Lower Bound	90% Confidence Upper Bound	Impact	90% Confidence Lower Bound	90% Confidence Upper Bound
7/17/03	0.16	0.19	-0.16	0.47	777	-793	2,346
7/28/03	0.17	0.11	-0.02	0.36	854	-92	1,800
8/8/03	0.76	0.19	0.44	1.07	3,781	2,224	5,339
8/15/03	0.31	0.18	0.01	0.61	1,548	30	3,065
8/27/03	0.53	0.14	0.29	0.77	2,639	1,437	3,840
9/3/03	0.20	0.08	0.06	0.34	996	303	1,690
9/12/03	0.31	0.09	0.16	0.46	1,554	801	2,307
9/22/03	0.31	0.10	0.13	0.48	1,540	670	2,410
9/29/03	0.25	0.07	0.13	0.37	1,252	642	1,863
10/9/03	0.02	0.07	-0.10	0.14	118	-478	714
10/14/03	0.16	0.08	0.04	0.29	806	178	1,434
10/20/03	0.48	0.18	0.18	0.78	2,379	881	3,877

Whole-premises Impacts

Table 4-14
Whole-premises Impacts and Standard Errors for 12 Re-set Days

Re-set Event Date	Start time	End time	Sample Group Re-set	Average Temperature	Degrees Setback	Group A AC Count	Group B AC Count	Mean Impact for Responding AC Users	Percent Responding AC Users	Mean Impact per AC Unit	Standard Error
7/17/2003	2:00 PM	3:45 PM	A	74	5	35	48	0.39	77%	0.30	0.20
7/28/2003	2:00 PM	7:00 PM	B	72	5	42	42	0.30	75%	0.22	0.14
8/8/2003	3:00 PM	5:00 PM	B	76	3	43	41	0.95	75%	0.71	0.21
8/15/2003	2:00 PM	7:00 PM	A	82	3	35	50	0.43	74%	0.32	0.20
8/27/2003	4:00 PM	6:00 PM	A	76	3	35	50	0.99	78%	0.77	0.17
9/3/2003	2:00 PM	7:00 PM	A	73	4	34	50	0.30	72%	0.21	0.12
9/12/2003	2:00 PM	7:00 PM	B	73	4	43	42	0.38	76%	0.28	0.11
9/22/2003	2:00 PM	6:00 PM	A	72	4	35	50	0.49	78%	0.38	0.14
9/29/2003	2:00 PM	7:00 PM	B	70	4	44	40	0.47	78%	0.37	0.10
10/9/2003	3:00 PM	5:00 PM	A	69	4	35	50	0.06	74%	0.05	0.09
10/14/2003	2:00 PM	7:00 PM	B	68	4	45	39	0.25	74%	0.18	0.09
10/20/2003	3:00 PM	5:00 PM	A	74	4	35	50	0.87	80%	0.69	0.22

Table 4-15
Whole-premises Impacts with Confidence Intervals, Per Unit and for 5,000 Sites

	Mean kW Per Unit				kW for 5000 Units		
Date	Impact	Standard Error	90% Confidence Lower Bound	90% Confidence Upper Bound	Impact	90% Confidence Lower Bound	90% Confidence Upper Bound
7/17/03	0.30	0.20	-0.04	0.64	1,503	-209	3,215
7/28/03	0.22	0.14	-0.01	0.46	1,123	-65	2,311
8/8/03	0.71	0.21	0.37	1.06	3,567	1,827	5,306
8/15/03	0.32	0.20	-0.01	0.65	1,591	-74	3,255
8/27/03	0.77	0.17	0.48	1.06	3,852	2,405	5,299
9/3/03	0.21	0.12	0.01	0.42	1,068	57	2,080
9/12/03	0.28	0.11	0.11	0.46	1,425	533	2,317
9/22/03	0.38	0.14	0.15	0.61	1,894	743	3,046
9/29/03	0.37	0.10	0.20	0.54	1,847	997	2,698
10/9/03	0.05	0.09	-0.11	0.20	235	-551	1,021
10/14/03	0.18	0.09	0.02	0.34	914	123	1,706
10/20/03	0.69	0.22	0.32	1.06	3,444	1,584	5,304

4.3 Projected Impacts By Temperature and Re-set Amount

Figure 4-18
Unadjusted Projected Impacts by Hour, Average per AC Unit
(For 4°F Re-set)

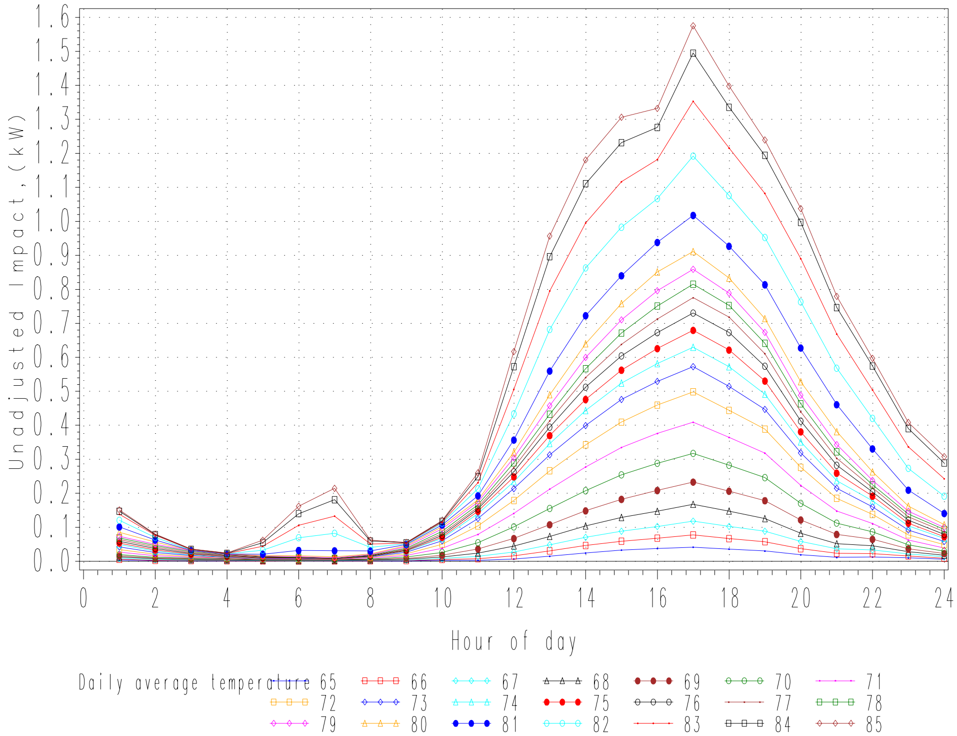


Figure 4-19a
Adjusted Projected Impacts by Hour, Average per AC Unit
(For 4°F Re-set, 4-Hour Duration)

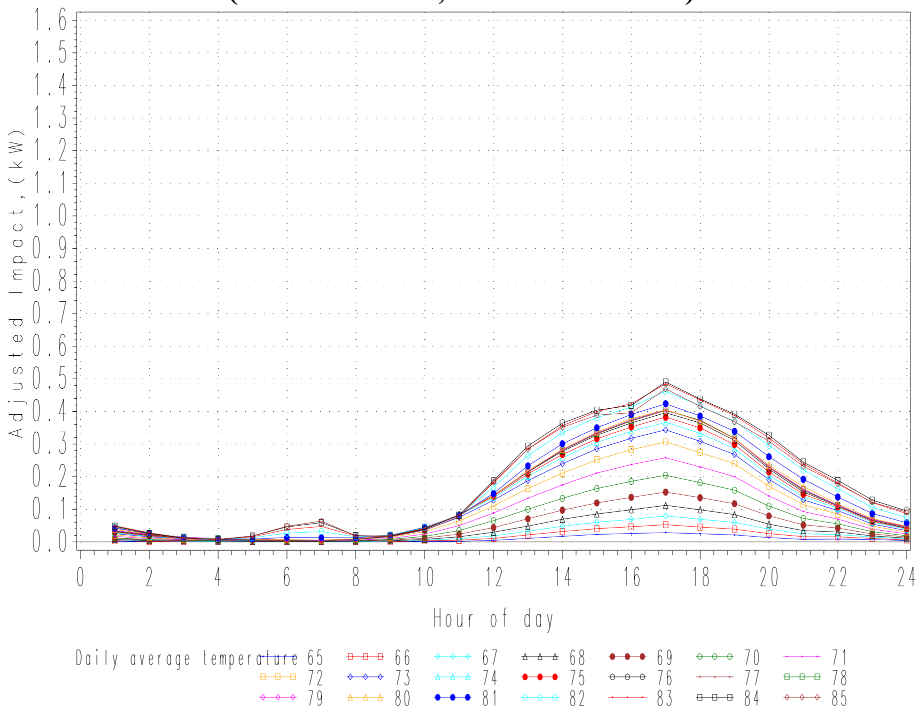


Figure 4-19b
Adjusted Projected Impacts by Hour, Average per AC Unit, Full Scale
(For 4°F Re-set, 4-Hour Duration)

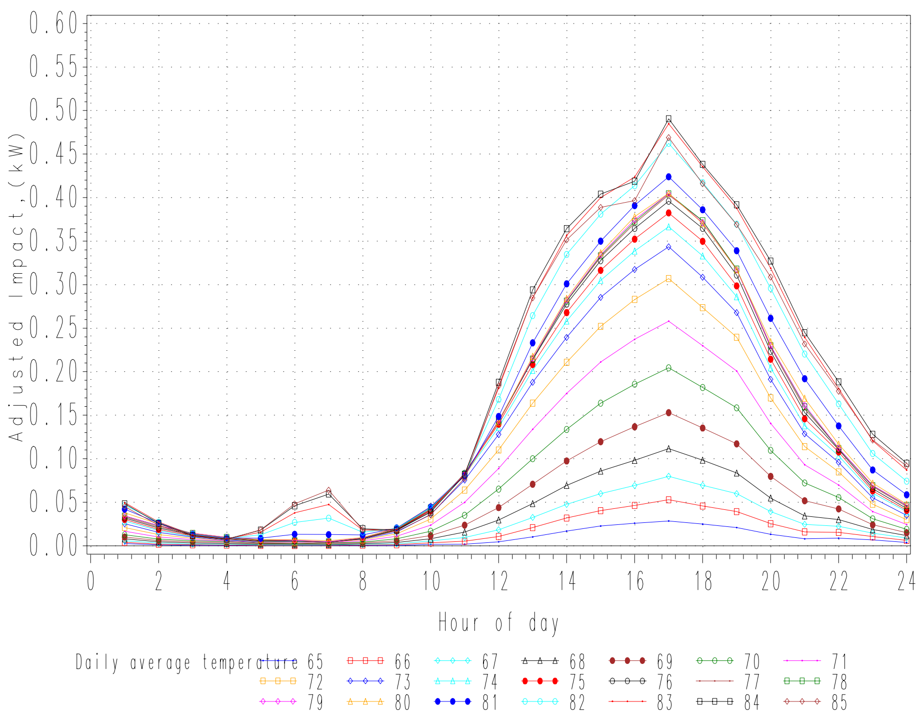


Figure 4-20
Adjusted Projected Impacts by Temperature, Average per AC Unit
(For 4°F Re-set, 4-Hour Duration)

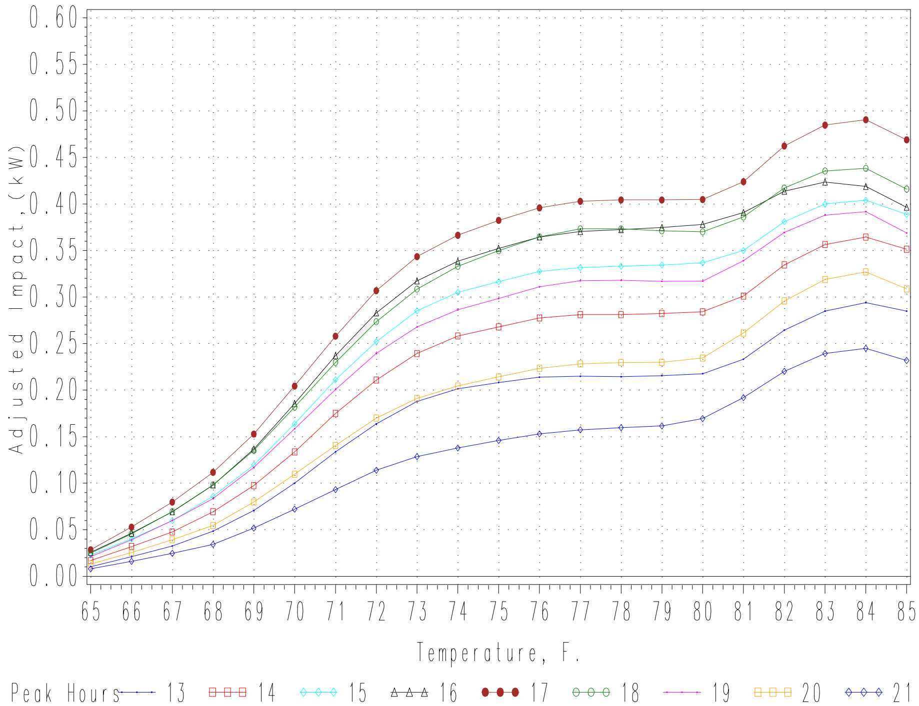


Figure 4-21
Hour 17 Adjusted Projected Impacts by Temperature, Average per AC Unit with High and Low Lag Temperatures (For 4°F Re-set, 4-Hour Duration)

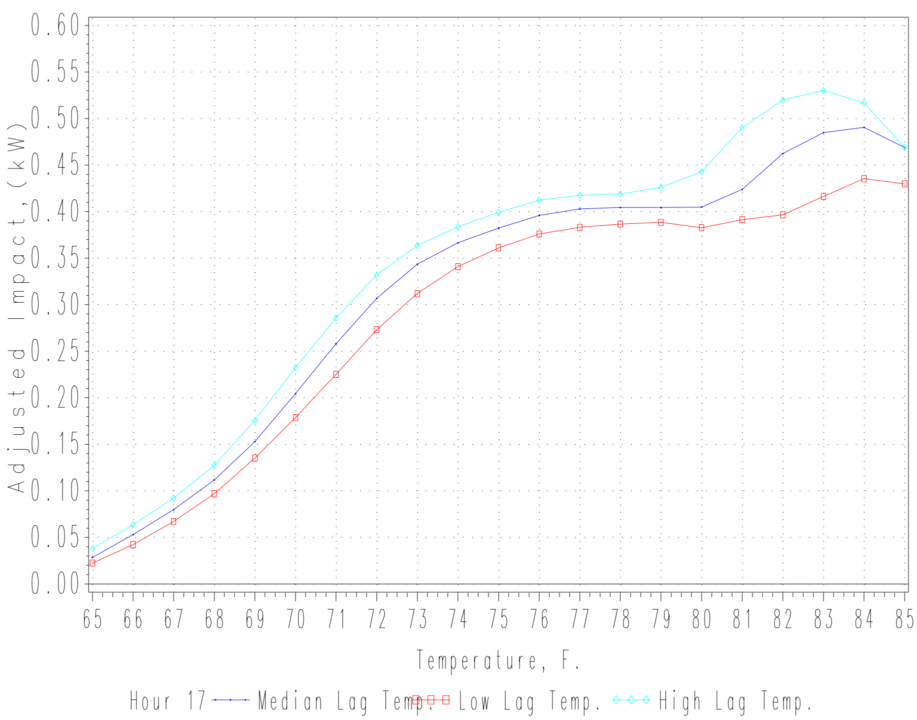
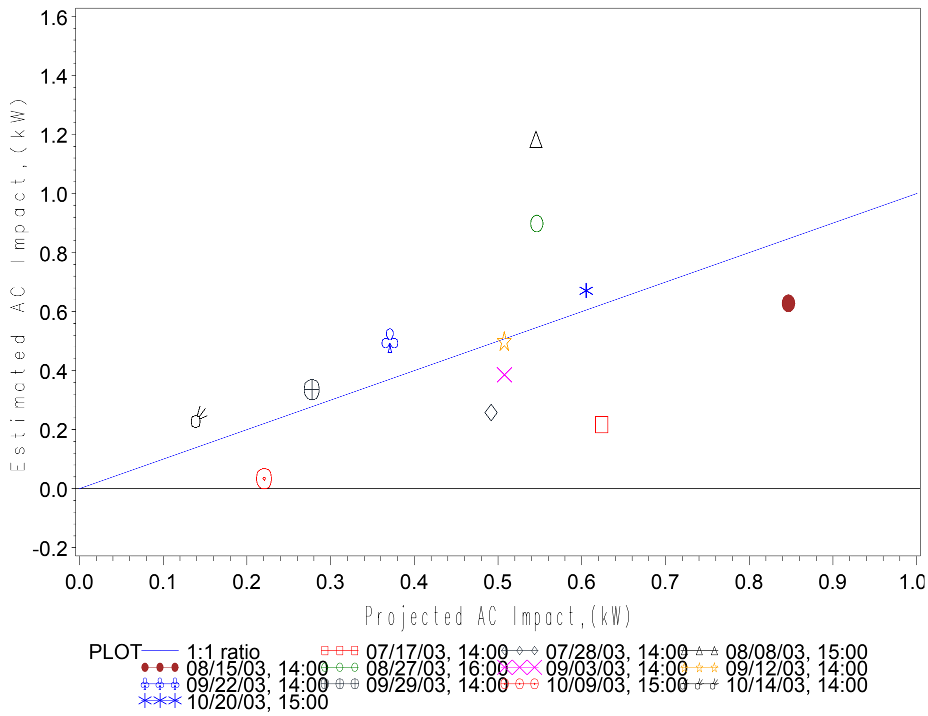


Figure 4-22
Unadjusted Estimated vs. Projected AC Impact
Average Impact for Each Event



5 CONCLUSIONS

In the summer of 2004, the Smart Thermostat Program did have one full program enactment. The full program was re-set on May 3rd during a Stage 2 emergency declared by the California ISO. The whole analysis metering group was re-set for this event so we cannot directly estimate the impact. Below, as an example of the use of the projections developed for this report, we use actual data from the May 3rd re-set to estimate the amount of savings produced on that day.

Following are some of the key findings from the 2004 study with comparisons to the 2003 results.

5.1 WHAT FRACTION CONTRIBUTE TO SAVINGS

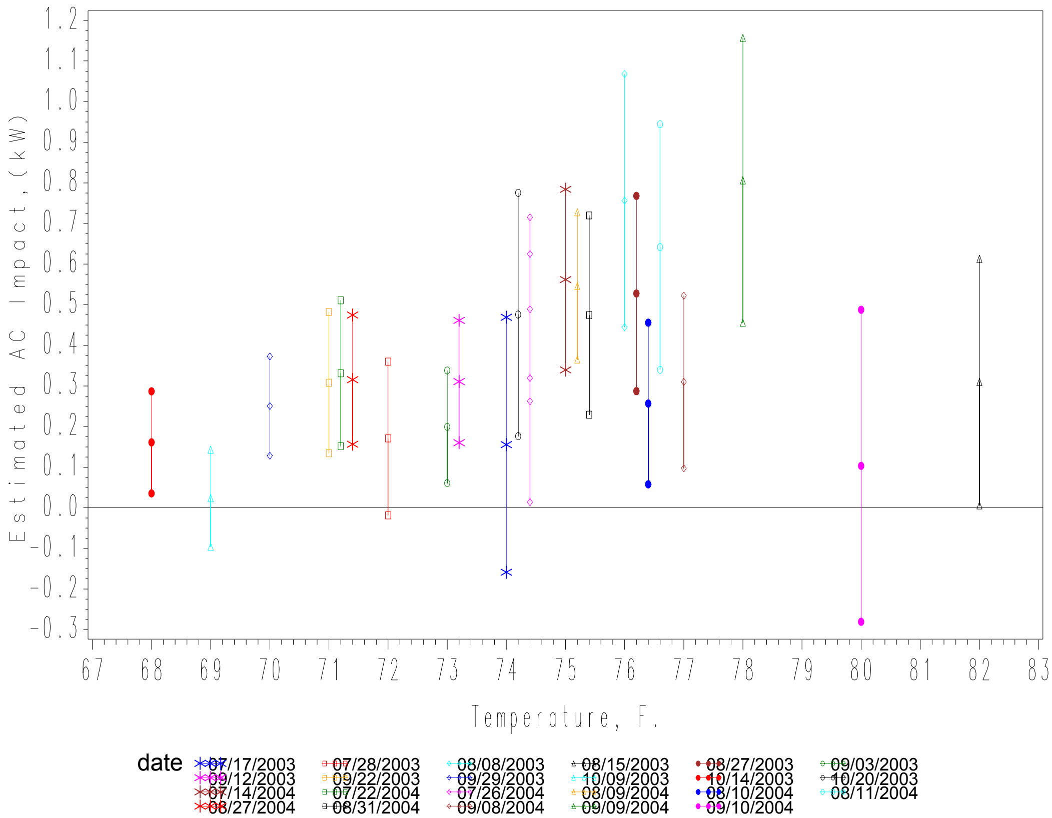
1. On average, just over 90 percent of the thermostats in the program appeared to operate correctly during each re-set event in 2004. The non-response rate increased for the 2004 analysis. The combined Smart Thermostat and SPP groups' non-response rate was 9 percent in 2004 compared to 6 percent in 2003. This difference was not found to be statistically significant.
2. In 2004, the over-ride rate continued to vary substantially across re-set events. The 2004 range, from a low of 9 percent to a high of 42 percent was within the extremes of the 2003 rates. The average over-ride rate for the summer of 2004 was almost six percentage points higher than the average in 2003. This is consistent with an average temperature for the twelve 2004 re-sets events that was 2°F higher than the 2003 average while average event duration and re-set amount were almost identical.
3. 2004 over-ride rates, when considered alone, showed variation primarily by event duration as opposed to average temperature. This was a change from the 2003 results where average temperature was identified as the strongest driver. A combined 2003/2004 model proved to be the most effective way to estimate over-ride percents as a function of average temperature and event duration. While the two years each primarily reflected the effect of only one of the explored variables, the combination provides a richer model that appears to provide a consistent and comprehensive structure across the years.
4. Eighteen percent of participating AC units were not used at all during the summer of 2003. For 2004, the percent decreased by only one percentage point to 17 percent. While some of these units might be used during severe hot weather, they contribute no savings in the milder weather.

The average combined effect of non-response, over-ride, and non-use increased from 42 percent to 48 percent between 2003 and 2004. The increase was due to an increase in both non-response rates and over-ride rates. The result is that, on average, only slightly more than 50 percent of the participating units are "potential contributors" to impacts.

5.2 SAVINGS FOR THE RE-SET DAYS

The 24 individual re-set event results across the 2003 and 2004 program years are summarized in Figure 5-1. The midpoint of each bar represents the impact estimate while the range of each bar indicates the 90 percent confidence intervals. The results are organized by average temperature. Where multiple event days have the same temperature, the results are offset slightly so they can be seen.

Figure 5-1
2003 Impact estimates by Date with Error Bars



From 2003 to 2004, savings per AC unit enrolled in the program, averaged across the 12 re-set events, increased from .30 kW to .44 kW. The higher 2004 average savings was also significantly different from zero. Eleven of the 12 individual events were statistically significant. These results are stronger than the 2003 results where only 9 of the 12 individual events were statistically significant and the overall average across participants and re-set days was also not statistically significant.

Averaged across all 24 test re-sets, the savings per AC unit are .37 kW. This result is, once again, not statistically different from zero. However, this result for the overall average does not mean that on a statistical basis we can't be sure if any savings occur at all. Clearly the re-set has an effect. We simply have relatively small average savings over the conditions under which the program has been invoked, small enough that we don't have an accurate estimate of the magnitude. These results are summarized in Table 5-1.

Table 5-1
Estimated kW Impacts

	Impact (kW)	Standard Error	90% Confidence Lower Bound	90% Confidence Upper Bound
Average of All Events				
per unit	0.37	0.24	-0.03	0.77
5000 units	1,834	1,219	-170	3,839
Average of 2004 Events				
per unit	0.43	0.24	0.03	0.83
5000 units	2,148	1,215	149	4,147
Average of 2003 Events				
per unit	0.30	0.23	-0.08	0.69
5000 units	1520	1,174	-410	3,451
Best Event - September 9th, 2004				
per unit	0.81	0.21	0.46	1.15
5000 units	4,031	1,051	2,302	5,760

The single best performance of the program occurred on September 9, 2004, when the per unit average savings reach .81 kW. If 5,000 units had received the re-set signal on that day, the estimated savings would have been 4.0 MW, with a 90 percent confidence interval of 2.3 to 5.8 MW. This estimate equals the *ex* ante estimate of 4 MW for 5,000 units. However, this is for the best, not average, day. Across all 12 2004 re-set periods, the savings from 5,000 units would have averaged an estimated 2.1 MW, with a 90 percent confidence interval of .1 to 4.1 MW. Four MW is within the confidence interval, but 4 MW of program savings remains an unlikely reality except on the most extreme days.

5.3 PROJECTED SAVINGS FROM FUTURE RE-SET EVENTS

Both unadjusted and adjusted savings estimates were projected for a variety of average temperature and event durations. The unadjusted projections are reported in Appendix C. The unadjusted projections only vary by temperature. The adjusted projections are reported in Appendix D. Because they are based in part on the estimation of over-ride percent, the adjusted projections vary by temperature and event duration.

Figures 5-2 and 5-3 summarize the adjusted projections for hours 17 and 18 for a four-hour event. Hour 17 was the hour of peak savings in 2003 while hour 18 was the peak hour in 2004.

Figure 5-2
2003 and 2004 Hour 17 Projected Adjusted Mean Impacts v. Average Temperature, 4-Hour Event Duration

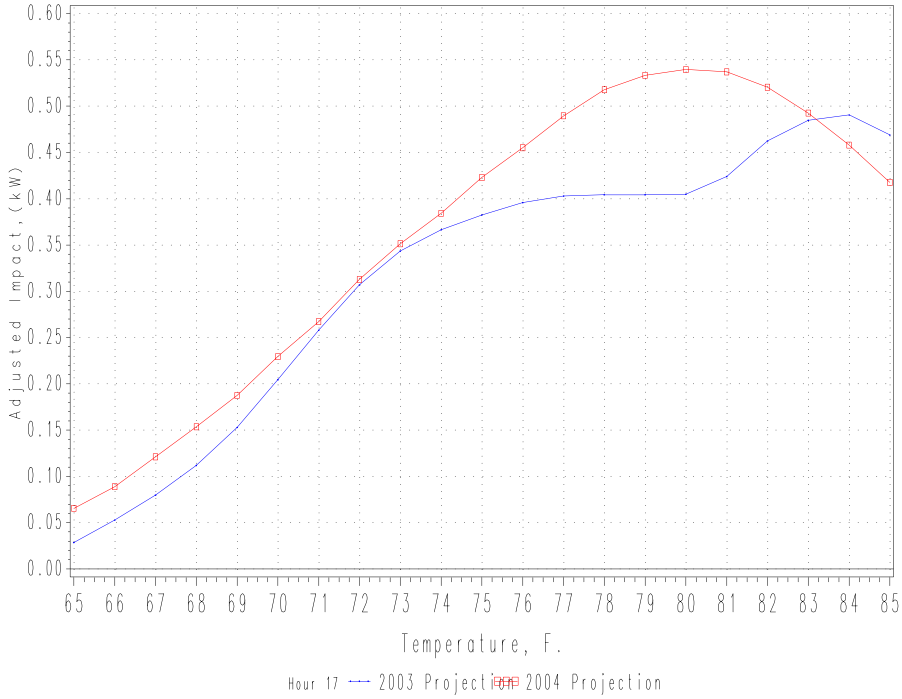
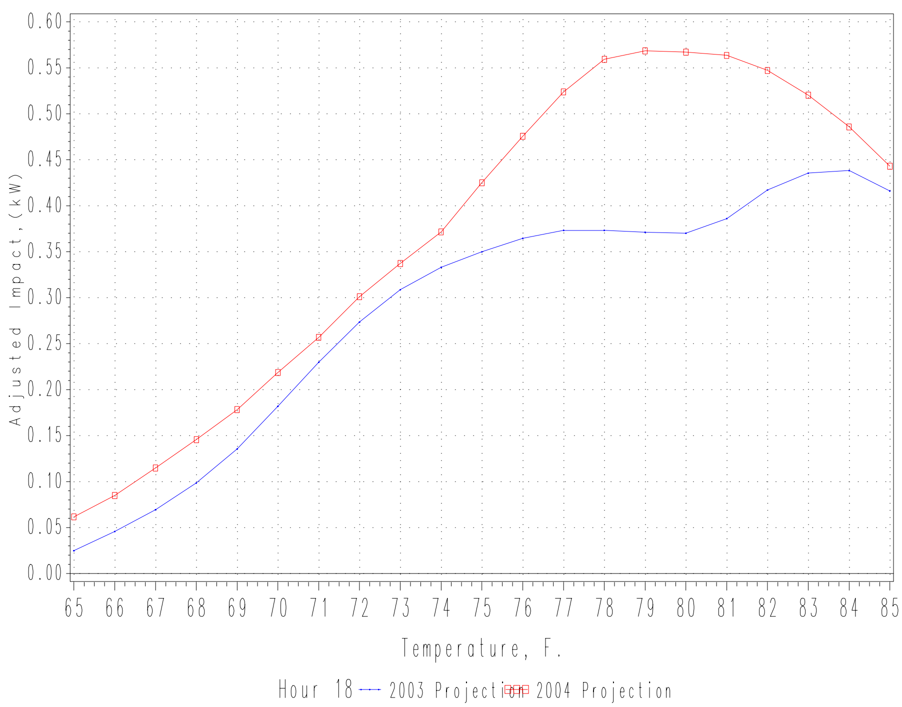


Figure 5-3
2003 and 2004 Hour 17 Projected Adjusted Mean Impacts v. Average Temperature, 4-Hour Event Duration



The 2004 impact projections provide a new look at expected savings for the Smart Thermostat Program. The impact projections from 2003 and 2004 are quite similar at lower temperatures and again at very high temperatures. However, in the critical range of the high seventies and low eighties the projections are quite different. These projections are produced using the same model for over-ride rate so the difference lies not in differential over-ride effects but different savings gained. Through the high seventies and low eighties, the 2003 projections show a slowing in the rate of savings as the temperature increases.

The comparison of the two projections raises questions about the 2003 results. The shape of the 2004 projections is more consistent with what we would expect in a large and diverse population. The unexpected 2003 shape, as produced by the projection model, reflects the specific cooling behavior of a just a handful of units with thermostat setpoints in the high seventies. This could be a true representation of the full population cooling behavior or the result of a projection model that is overly sensitive to a relatively small metering sample. It does appear the 2003 projections are on the whole slightly lower than the 2004 projections. It is not clear, however, that the substantially lower projections through the high seventies and low eighties are reliable. As a result, the 2004 adjusted projections are provided in Appendix D and should be used going forward.

5.3.1 Application of the Projections

The projected savings can be used in two ways. One is to provide an estimate of savings for a recently completed re-set event without requiring collection and analysis of metering data. In this case the over-ride and non-response rates are known. These rates, together with an estimate of the fraction of units with zero summer use, provide an estimate of the fraction of units in the program that were potential contributors in that event. This fraction is then applied to the unadjusted projected savings in Appendix C to produce the impact estimate for the particular event.

The first line of Table 5-2 shows this process. The non-response, AC non-use, and over-ride fractions are the actual, program wide fractions for the May 3rd event. Combined using the formula in Section 2, they indicate a percent of non-contributors at almost exactly 50 percent. From Appendix C, the average unadjusted savings of hours 17 through 19 for a four-degree re-set at 82°F is 1.28 kW. Applying the percent non-contributors to this, the final savings estimate is .64 kW per unit. There were 4,796 confirmed units during the re-set event. The full program projected savings are just over 3 MW.

Table 5-2
Application of Projects to May 3rd Re-Set Event

Projection Type	Non-Response Fraction (P_F)	AC Non-use Fraction (P_z)	Over-ride Fraction (P_{or})	Percent Not Contributing	Unadjusted Savings Estimate	Savings (Per Unit kW)	Full Program Savings (kW)
Unadjusted (Actual Non-Response and Over-ride)	5.4%	17.3%	30.2%	50.3%	1.28	0.64	3,053
Adjusted (Estimated Non-Response and Over-ride)	9.3%		35.3%	57.1%		0.55	2,640

The other use is to project savings in advance of a re-set event, either for general planning or to guide a specific operational decision. In this case, the over-ride and non-response rates must also be projected. The projection of the over-ride rate, in particular, adds additional uncertainty to the estimate. The adjusted projected savings shown in Appendix D include the adjustments for non-contributors based on the projected rates of over-ride, non-response, and zero summer usage.

The second line of Table 5-3 repeats the same process as the first line but using the estimated non-response and over-ride rates that are used in the adjusted savings projections. The estimated percent of non-contributors is substantially higher than the actual rate. The May 3rd event was the first hot weather the San Diego area experienced in 2004 and this might explain the lower over-ride rate on that date relative to the estimates based on later events.

5.4 FUTURE PROGRAM PERFORMANCE

The findings from this year's analysis are stronger than previous years' results. Eleven of the re-set events had statistically significant savings and the magnitude of those savings was higher than for 2003. Despite this, the 2004 analysis essentially confirms the finding from the previous two studies that future performance of the program as a mechanism to respond to statewide emergencies is not fully reliable.

One factor is the limited used of air conditioning in the territory, with one-fifth of participating units never used over the summer. Another factor is that statewide emergency conditions do not necessarily coincide with hot weather in the San Diego area. As long as the emergency condition that triggers a re-set event is not tied to hot weather in San Diego, a high number of non-users is likely to be found during future re-sets. Finally, when the weather is hot, higher rates of over-ride are projected to occur. Thus, while the program is capable of savings of the desired 4 MW magnitude, it is unlikely to realize these savings in full on the day of a statewide emergency.

The 2004 results do provide one useful direction in this respect. While over-ride rates increase with average temperature, increased event duration also has an effect. This indicates that careful timing of re-set events is essential to focus maximal savings when they are most needed. Well-timed and shorter re-set events that end before a high percent of customers over-ride should provide the best participation rates.