92.119

Prepared by:
Proctor Engineering Group, Ltd.
San Rafael, CA 94901
(415) 451-2480

Estimating Peak Reduction From Submetered Air Conditioned Data

Presented at:
1993 International Energy Program Evaluation Conference
Chicago, Illinois

Final Report
August 1993

Author:
John Proctor, P.E.

Creators of CheckMe!®

ESTIMATING PEAK REDUCTION FROM SUBMETERED DATA

John Proctor
Proctor Engineering Group
Corte Madera, California

Introduction

In 1991, Pacific Gas and Electric Company (PG&E) ran a pre-production test of the Appliance Doctor Project in Fresno, California. This test consisted of quality-assured distribution-duct sealing of residential central air conditioning systems (and repairs to a few units). This project was detailed in Reference 2 and showed a cooling savings of 21.6% for high-use customers and 9.2% for random customers (Ref. 6). The test included pre-treatment/post-treatment submetering of the air conditioners (ACs). Units were submetered from June 7 through July 10 (pre-treatment); Appliance Doctor repairs were made (treatment); and submetering continued from August 21 through September 25 (post-treatment). With this submetered data, the peak reduction due to these repairs was estimated.

The Problem

Valid estimation of electrical use of a residential AC on a peak day is intrinsically difficult. The evaluator is attempting to estimate an event that happens rarely (only once every 20 years, by some definitions). The event of interest is often outside the data set available and, when evaluating the effect of some treatment, the event may occur in neither the pre-treatment nor the post-treatment data set. The estimation of peak reduction, therefore, often depends on extrapolation beyond the values in the data set. When the event of interest occurs near the "edge" of the data, the "prediction interval" is wider than when it occurs near the mean of the data. As an evaluator extrapolates beyond the existing data, the convenient assumptions that were made in the mathematical model are likely to be less and less justified.

Peak electrical use is often driven by multiple variables. For residential ACs, for example, customer operation, outside temperature, relative humidity, time of day, day of week, solar incidence, sky cover, and prior temperatures are some of the variables that may affect the peak use. Utility planning requires an accurate estimate of peak.

Proctor Engineering Group investigated the assumptions and results of six analytical models applied to the Appliance Doctor data set: day-matching, a constrained regression model, a temperature-bin statistical model, hottest-days statistical aggregation, an hourly curve-fitting model, and a simplified regression model. This paper discusses the strengths and weaknesses of the models in estimating peak reduction from submeter data.

Data Set Description

For each AC, submetered data were recorded at 15 minute intervals. The data consisted of kWh readings for 60 residential ACs in three groups: 10 high-use customers that received Appliance Doctor (treated high-use); 13 random customers that received Appliance Doctor (treated random); and 37 random AC customers enrolled in PG&E's Appliance Metering Project in Fresno (comparison).

For the comparison of these models, only data from weekdays with maximum temperatures above 99°F were used. This consisted of five days in the pre-treatment timeframe and six days in the post-timeframe. Hours 15 (from 2 p.m. to 3 p.m.) and 19 (6 p.m. to 7 p.m.) were utilized because they represent system-peak and local residential distribution peak, respectively.

Characteristics of Residential AC Use

Residential AC usage varies for individual houses. Fixed thermostat settings are the exception rather than the rule (see Occupant Thermostat Control). Pooling all the homes in the group reduces the variation caused by occupant behavior. The pooled variable is the load that the utility sees from that group at each time interval. Both pooled and individual data sets were used in the analysis.

Residential AC use not only fluctuates for each customer, it also fluctuates with the time of day, independent of temperature. In this investigation, both temperature-pooled and hour-of-day data sets were used, leading to the conclusion that hour-of-day sensitive analysis provides a more valid estimate of peak use.

General Assumptions and Validity

The validity of each model is dependent on the accuracy of the following assumptions:

1. The experimental group is an unbiased sample of the population of interest.

2. The only significant difference between the comparison group and the experimental group is the treatment.

A comparison group as closely matched as possible should be employed in the analysis. This is particularly true because of the multiple factors influencing residential AC use.

Conventions in this Paper

A confidence interval states the expected range of the mean. For peak prediction, the interval of interest is the expected range of the extreme case. In estimating the change in peak, we wish to determine the "confidence interval" of the difference between the pre- and post-treatment extremes. In this paper, the listed confidence interval is the standard expected range of the mean difference.

The load is listed as the "fraction of maximum connected load." Hourly average kW has been divided by the maximum hourly average kW recorded by the submeter on that unit. This procedure produces normalized results.

The change in load between pre- to post-treatment is listed as a "plus" if the load increased and a "minus" if the load decreased.

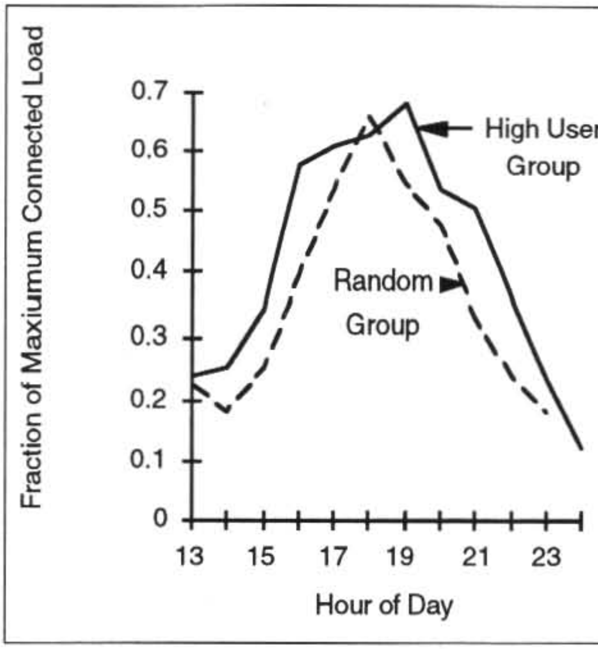

Figure 1. Load Profile on Pre-treatment Day

Model #1: Day-Matching

Description

Day-matching is defined as comparing data from pre-treatment days and post-treatment days that exhibited a high degree of similarity. Similarity is judged by: Weekday versus Weekend/Holiday, Maximum Ambient Temperature, Average Temperature, and Daily Temperature profile. The two closest matched pre/post days in the data set are pre-treatment (Monday, June 10, 1991) and post-treatment (Tuesday, September 3, 1991). The peak temperature on both days was 103°F.

Details

For these two matched days, use is pooled for each group by hour. The result is an AC load profile for each group. The pre-treatment load profiles for the high-user and random groups are shown in Figure 1.

The process is replicated for the post-treatment period, and the change in electrical load due to the treatment is estimated as:

$$\Delta L_{1A} = (L_{1Apost} - L_{1Apre}) - (L_{3Apost} - L_{3Apre})$$

where:

ΔL_{1A} = average net change in load for group 1 in hour A

L_{1Apost} = average load for group 1 in hour A on post-treatment day

L_{1Apre} = average load for group 1 in hour A on pre-treatment day

L_{3Apost} = average load for group 3 (comparison group) in hour A on post-treatment day

L_{3Apre} = average load for group 3 (comparison group) in hour A on pre-treatment day

Validity and Assumptions

The validity of this estimation model is dependent on how representative the matched days are of the peak day. In the end, the conclusions will be based on an assumption about the use at peak. These might be that the use at peak is:

1. The same as the use on the matched days.

2. Proportional to the use on the matched days [such as Peak Electrical Load = (1+x) • (Matched Day Load)].

3. Some other relationship to the matched days (such as the change in load at peak is equal to the change in load between the matched days).

Assumption #1 is unlikely to be valid because there are ample data to show that demand increases as the ambient temperature rises (unless the matched days have the same peak temperature as is predicted for the peak day). Assumptions #2 and #3 require some additional information to establish the relationship. Such an analysis is likely to develop into one of the other models.

Results

The results of this analysis method for this study is shown in Figure 2 and compared to other models in Table 1. This analysis shows an estimated post-repair increase in load in hour 15. It also shows a load reduction in hour 19 for both high-use and random treated homes.

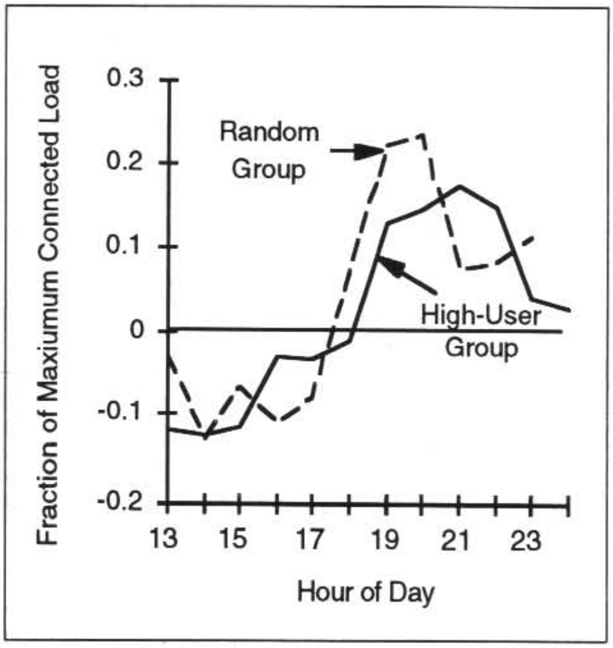

Figure 2. Load Reduction on 103°F Day

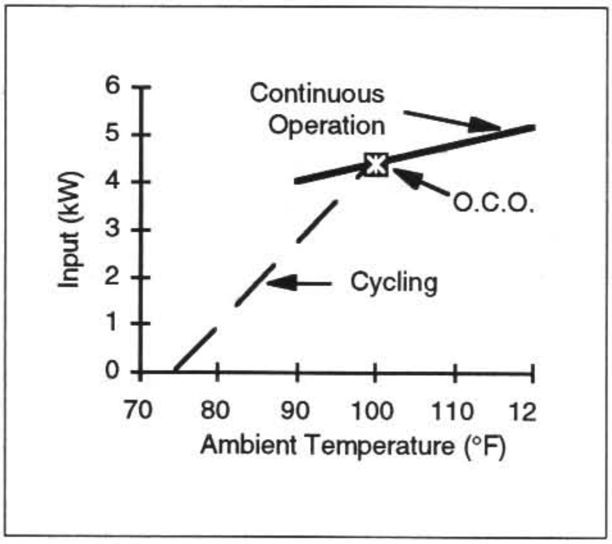

Figure 3. Model 2

The assumption that the peak use is the same as matched day use is violated by data recorded on July 3, 1991. The peak temperature on that day was 109°F and the use was 18% to 28% (of maximum connected load) higher than that recorded on the matched days. When other day matches are used (including days matched by comparison group usage data), the results are variable. Consequently, there is low confidence in the results of this analysis. Models #4, #5, and #6 are essentially extensions of Model #1 that attempt to improve and determine the prediction interval of the estimate.

Table 1. Change in Peak (Fraction of Maximum Connected Load)

	Model 1	Model 4	Model 5	Model 6
High Use Hr 19	-0.129	-0.213	-0.097	-0.233
95% interval	NC	±.133	NC	±.114
Random Hr 19	-0.223	-0.125	-0.086	0.059
95% interval	NC	±.151	NC	±.0730
High Use Hr 15	0.115	-0.080	0.084	0.120
95% interval	NC	±.151	NC	±.180
Random Hr 15	0.066	-0.050	0.014	0.093
95% interval	NC	±.131	NC	±.183

NC = Not calculated

Model #2: Linear Regression

Description

The linear regression model assumes that peak use can be modelled as though each AC is in one of three modes (off, cycling, or running continuously), and that prior ambient temperature is the predominant determinant of mode and Watt draw. This model most closely matches an AC controlled by a constant thermostat setting. There is no distinction between different hours of the day, and individual units are not pooled together.

In estimating the peak use of residential ACs, there are *physical constraints* that set the upper boundary of electrical use. The maximum power draw from an individual unit is limited by the design and condition of the machine. The maximum draw for any one AC is determined primarily by outside temperature, inside temperature, and humidity. The upper limit constraint is utilized in this model.

Details

A number of useful features are provided by this model. First, it will not overpredict the use of ACs that are running continuously. Second, it easily explains different modes of operation and their effects on peak.

The design of the simplified model is detailed in Refs. 5 and 6. For each house, a linear regression correlating maximum AC input to hourly outdoor temperature data is calculated. This is the continuous operation line in Figure 3. Similarly, another regression line describing AC use when the unit is cycling is computed (the cycling line). The intersection of these two lines is the onset of continuous operation (OCO) and indicates the temperature above which the AC (if running) is modeled to run continuously.

The model is calculated for each house in both pre- and post-treatment periods, and the change in load due to the treatment is estimated as:

$$DL_{n(103)} = L_{n(103)post} - L_{n(103)pre}$$

where:

$DL_{n(103)}$ = raw change in load for house n at 103°F ambient

$L_{n(103)pre}$ = load for house n at 103°F ambient before treatment

if ambient temperature (103) > OCOnpre,
$= D_{on103pre} * (C_{ncontpre} + s_{ncontpre} * 103)$;
if ambient temperature (103) < OCOnpre,
$= D_{on103pre} * (C_{ncycpre} + s_{ncycpre} * 103)$

OCO_{npre} = onset of continuous operation for house n before treatment

$D_{on103pre}$ = probability that unit in house n before treatment will be on at 103°F

$C_{ncontpre}$ = intercept of the continuous operation line for house n before treatment

$s_{ncontpre}$ = slope of the continuous operation line for house n before treatment

$C_{ncycpre}$ = intercept of the cycling operation line for house n before treatment

s_{ncycpr} = slope of the cycling operation line for house n before treatment

and,

$L_{n(103)post}$ = load for house n at 103°F ambient after treatment (calculated in the same manner as $L_{n(103)pre}$)

The average net change in load due to the treatment is estimated as:

$$\Delta L_{1(103)} = \text{average}\,(DL_{1(103)}) - \text{Average}\,(DL_{3(103)})$$

where,

$\Delta L_{1(103)}$ = average net change in load for group 1 at 103°F ambient

Avg $(DL_{1(103)})$ = average raw change in load for group1 at 103°F ambient

Avg $(DL_{3(103)})$ = average raw change in load for group 3 (comparison) at 103°F ambient

Validity and Assumptions

The validity of this method in estimating the peak reduction is dependent on how well the model matches the peak-use patterns of each individual AC. The following assumptions are made:

1. The probability that the unit will be running increases linearly at higher temperatures (and cannot exceed 1).
2. Any differences in electrical load that are caused by variables other than outside temperature are cancelled by the use of the comparison group.
3. The cycling regression and probability of the unit running capture thermostat adjustments at high temperature.

The accuracy of Assumption #1 can easily be tested at high temperatures below the peak temperature. If it proves to be inaccurate at those temperatures an alternate assumption can be substituted. Occupant attitudinal variables may prove helpful in modelling occupant control behavior at high temperatures. While Assumption #2 may be accurate, the inclusion of all hours of the day in the same analysis increases the variability. This can be improved by separate analysis of each hour of the day. Assumption #3 is problematic. As discussed below, a constant thermostat setting is not the predominant control pattern. However, the model does not assume that a constant thermostat setting exists; it only models the cycling use as a linear function. Start-up and shut-down cycles can be determined from submeter data. With these data, the effects of these types of cycle can be estimated.

A constant thermostat setpoint was evident in only 26% of the houses in this sample (Ref. 6) Other investigators have shown similar thermostat control patterns in room ACs (Refs. 3 and 4). Lutzenhiser (Ref. 4) found only 29% of the window ACs in his study were controlled by a constant thermostat setting. An investigation in California by Berkeley Solar Group (Ref. 1) yielded similar results— 33% of the occupants using central ACs reported constant setpoint.

Results

This is a very useful model in explaining the modes of central AC use. One of its analytical strengths is that it provides justification for extrapolation to temperatures hotter than those observed. The strength of this model is lost when all hours are pooled. In that form, the results are not usable due to large standard errors and low R^2. This model should be further developed in an hour-by-hour analysis.

Model #3: Temperature Bin

Description

The third model eliminates assumption of linearity from Model #2. For each AC, this model combines all the load data for temperature bins 5°F wide. This model makes no distinction between different hours of the day and does not pool the individual units.

Details

For each house, all the data were binned by ambient temperature. Each bin contained data from all hours in which the ambient temperature was within 2°F. For example, the 95°F bin contained all data for hours when the temperature was between 93°F and 97°F. A mean and 95% confidence interval was then calculated for each bin and plotted, as shown in Figure 4.

This method showed wide confidence intervals at high temperatures. These confidence intervals were further widened by the inclusion of data from differing hours of the day. Since AC use is so dependent on customer control behavior, use at 100°F at 3 p.m. can be substantially different from use at 100°F at 7 p.m. Because of this time-of-day dependence, many of the graphs showed little change with outside temperature (see Figure 4).

Validity and Assumptions

Extrapolation to peak temperature is difficult with this model. Each individual AC/home/occupant combination has its own characteristic signature, which is not easily standardized by this method. This model also depends on the assumption that the use of a comparison group will cancel effects other than temperature. The inclusion of all hours of the day in the analysis increases the variability.

Results

This model drove home the point that each hour of the day must be analyzed individually.

Model # 4: Hottest-Days Aggregation

Description

In the "hottest-days statistical aggregation," load data for each hour from the hottest days are aggregated in a single bin. This is an hour-differentiated, pooled high-temperature, single-bin version of Model #3. It is also a robust version of Model #1.

Details

For each group, the load was pooled for each hour of the hottest weekdays (100°F or above). The treatment group data were paired with the comparison group data for the same day and the difference (D_A) was computed. The change in electrical load due to the treatment is estimated as the difference between the mean D_{Apre} and the mean D_{Apost}.

Validity and Assumptions

This model attains its maximum validity when each hour of the day is utilized in its own analysis and when the temperatures in the pre- and post-periods are similar to each other. Nevertheless, there is no structural reason to assume that the difference between the means is constant at higher temperatures.

Results

This analysis is the only one that predicts peak reductions for both groups for both time periods (see Table 1).

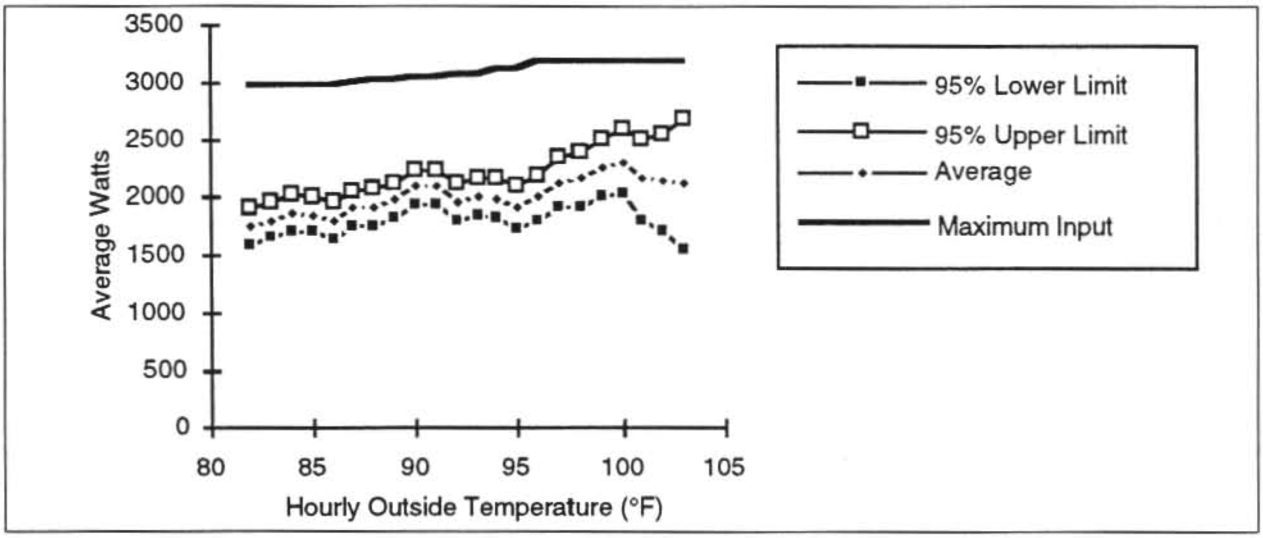

Figure 4. Post-Treatment Bin Analysis—House 12

921149

Model #5: Hourly Curve-Fitting

Description

This model consists of a piece-wise linear regression for every hour of the day and pools the individual ACs into groups. In this model, two variables (the percent of the units on during the hour and the duty cycle of the units on during the hour) and their product (coincidence) are assessed. These variables are regressed against the maximum ambient temperature for the day. Separate regression coefficients are calculated for higher temperatures (100°F to 115°F) and lower temperatures (88°F to 99°F). Since the temperatures of interest are the higher ones, only the results at those temperatures are discussed here.

Details

The first step in this analysis is to separate coincidence (also called diversity) into two components. The first component (%on) is simply the number of units on at any time in the hour divided by the number of units in the group. For example, if 8 out of 10 units ran during an hour, the %on would be 80%. This component captures information on ACs that are switched off at high temperatures. Using only ACs that were on sometime during the hour, the second component (Yon), average duty cycle as a percent of the hour is calculated. For example, if all the units that were on at any time during the hour were running continuously, the Yon would equal 100%. The relationship between these variables and the diversified load in that hour is:

$$L_{1A} = \text{load for group 1 in hour A}$$

$$= \text{average}(CL_{1A}) \bullet \%_{on1A} \bullet Y_{on1A}$$

where,

CL_{1A} = average connected load for group 1 in hour.

The change in electrical load due to the treatment is estimated as:

$$\Delta L1 = (L_{1Apost} - L_{1Apre}) - (L_{3Apost} - L_{3Apre})$$

where each of the loads is calculated from the high temperature regression, such as:

$$L_{1Apre} = I_{1Apre} + S_{1Apre} \bullet T_p$$

where,

I_{1Apre} = high temperature pre-treatment regression intercept for group 1 in hour

S_{1Apre} = high temperature pre-treatment regression slope for group 1 in hour A

T_p = maximum temperature for the peak day

Validity and Assumptions

This model is a pooled hourly, hot days regression model. It is a further development of Model #4, in that a regression against peak temperature is applied to the high temperature bin. The validity of this model is dependent on two assumptions:

1. The percentage of units that will be running increases linearly at higher temperatures (and cannot exceed 1).
2. The average duty cycle of the ACs in the group increases linearly with outside temperature.

This model attempts to overcome the shortcomings of Model #4 by determining the slope of the data in the high temperature bin. The accuracy of this method depends on the temperature range of the available data. A wide range of data and the presence of temperatures near the peak temperature will improve the validity of this estimate substantially.

Results

The results of this model for 103°F are displayed in Table 1. This model and the day-matching model both produced changes in peak that were of the same sign for both the high-use and random groups—peak reduction in hour 19 and peak increase in hour 15.

Model #6: Simplified Multivariate Linear Regression

Description

This model deals with every hour of the day. It pools the individual ACs into groups, pairs treatment, and comparison group averages by day, and utilizes the post period as a "dummy" variable in the regression.

Details

This model takes the general form:

$$D_A = c + (a \times POST) + (b \times P.TEMP)$$

where,

D_A = difference in pooled load between paired by date) treatment and comparison groups in hour A

POST = dummy variable (0/1) for post time period

P.TEMP = maximum temperature for the paired day

The regression coefficient *a* estimates the change in electrical load due to the treatment.

Validity and Assumptions

The validity of this model is dependent on two assumptions:

1. The change in peak (captured in the coefficient a) is a constant value (independent of the peak temperature).
2. The underlying structure of the data is linear and the important variables randomly vary in both the comparison group and the treatment group.

Sample size and comparison group selection are critical to the validity of this model. In order to obtain linearity in the data, two days had to be excluded from the analysis. July 5 (the day after the holiday) and June 11 were highly significant outliers. June 11 appeared to be an outlier due to the small sample size (two ACs that normally operated during the hours of analysis were off on that day).

Results

The estimates obtained with this model are displayed in Table 1. For hour 19, the range of temperatures in the high temperature data set, the load difference (D_A) was independent of the peak temperature. For hour 15, the peak temperature exhibited a moderate influence on D_A.

Summary, Conclusions, and Recommendations

Table 1 summarizes the estimated change in peak load in hour 13 (near system peak) and hour 19 (near the residential local area distribution peak). Based on the consistent results across models, it can be concluded that high-use customers in this locale will exhibit a lower load on the distribution system in the early evening when their ducts are sealed. We reserve judgment on the efficacy of reducing early afternoon peak from a duct-sealing program.

Accurate estimation of peak reduction from submetered data is difficult for small samples. Accurate estimation requires significant advanced planning. We strongly recommend that peak reduction on residential ACs not be inferred from measured kWh savings. For future evaluations and verifications, we recommend:

- *Select a strong comparison group*. The validity of all the models is highly dependent on a comparison group that does not differ significantly from the treatment group. It is not out of the question to utilize a flip/flop method for many treatments. Even duct leakage could be automatically changed on a fixed schedule with motor-operated dampers.
- *Continue development of Model #2*. This model is the only one with a physical justification for extrapolation to higher temperatures. It should be modified to an hour-by-hour analysis, the statistical package developed, and the results tested.
- *Analyze hour-by-hour*. Because of the large time-of-day effect on residential AC use, utilize an hour-by-hour analysis.
- *Increase sample size and sampling period*. Future samples should include at least 20 units, and the pre- and post-treatment periods should each be extended over an entire cooling season.
- *Explore weighted analysis*. Larger data sets offer the opportunity to use data that are weighted more heavily closer to the anticipated peak temperature.

References

(1) Berkeley Solar Group. "New Home End Use Analysis Report." Sacramento, CA: California Energy Commission, 1990.

(2) Jacobson, Richard, John Proctor, and Amy Polak. "PG&E Appliance Doctor Pre-Production Test," *Proceedings of the ACEEE 1992 Summer Study on Energy-Efficiency in Buildings*. Washington, DC: American Council for an Energy-Efficient Economy, 1992. Volume 5, pp 109-115.

(3) Kempton, Willett, Daniel Feuermann, and Arthur E. McGarity. "I Always Turn It on Super: User Decisions about When and How to Operate Room Air Conditioners," *Energy and Buildings*, Vol. 18, 1992, pp. 177-191.

(4) Lutzenhiser, Loren. "A Question of Control: Alternative Patterns of Room Air Conditioner Use," *Energy and Buildings*, Vol. 18, 1992, pp. 193-200.

(5) Proctor, John. "Pacific Gas and Electric Appliance Doctor Pilot Project: Final Report." San Francisco, CA: Pacific Gas and Electric Company, 1991.

(6) Proctor, John, and Ron Pernick. "Getting it Right the Second Time: Measured Savings and Peak Reduction from Duct and Appliance Repair," *Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings*. Washington, DC: American Council for an Energy-Efficient Economy, 1992. Volume 2, pp. 217-224