#4540

Elements of a Weatherization Evaluation

by Miriam L. Goldberg

A weatherization evaluation can be described in four main stages: design, data collection, data analysis and reporting.

Design

Design, in a broad sense, means looking ahead to the form of results to be reported, selecting methods of analysis appropriate to provide those results, and setting out to collect data necessary to support that analysis. Thus, the first stage of a well-designed evaluation is to work backward from the anticipated end result, developing the plans and procedures needed to get there. Skimping on the design stage wastes money in the long run, with costly data collection efforts that fail to provide a basis for answering the questions of interest.

Step 1. Identify the questions to be answered.

The basic question of interest, "How much did this program save?" means different things to different interested people. What activities are included in "this program?" Savings are relative to what? Are savings to be measured in energy or dollars? If dollars, whose—occupants, owners, taxpayers, ratepayers, or stockholders?

Part of pinning down the questions is drawing boundaries around the study. Who and what is to be included in the study? Over what time period, geographic regions, types of buildings, types of occupants?

It's natural to try to answer a variety of related questions along the way, such as comparing different types of households or different service delivery systems. But the most effective design for answering one question may be very ineffective for others. Specify the question of primary interest, and the level of accuracy required, and design for that. Then consider what additional questions can also be answered with what accuracy at what additional cost.

Step 2. Develop a strategy to answer the questions.

This step typically leads back to Step 1, as the realities of what can be done within budget constraints requires a redefinition of the study objectives.

One important element of the design strategy is determining the type of "control group" to be used. Savings in "program" homes need to be compared to savings in untreated homes that are similar in all other respects. This control group provides the baseline of what would have happened without the program.

The next major design element is determining the sampling plan. Sampling means choosing a (usually small) subset for study from the "target population," (the total group of interest). The goal is to choose the sample in a way that allows valid conclusions, with calculable measures of uncertainty, to be drawn about the entire target population.

Design also includes the logistics of data collection and data quality control, as well as specification of the data analysis methods to be used once the data are collected.

Data Collection

The more accurate and complete the information collected for the database is, the more reliable the estimates from the study will be. Minimizing errors due to transcription, misunderstanding, information loss, and non-cooperation requires time and care, which in turn require money.

Data Collection Warnings

- *In general:* Retrieving existing data is not free; it takes cooperation and communication.
- *For surveys:* Meaningful results require high response rates. High response rates require substantial effort ($).
- *With utilities:* Billing records aren't stored for long; typically only 13 months are online. Every utility uses its own set of codes and definitions. Many require waivers for the release of billing records.

Data Analysis

The data analysis for an evaluation begins with basic savings estimates, such as those provided by PRISM, are only the beginning of an evaluation's data analysis. The collection of individual-house estimates must then be reduced to some summary measures describing the whole group, such as the mean or median. The "control-group adjustment," which gives the net program effect, can be based on the difference or ratio of means or medians, in absolute or percentage terms.

Which way to compute the overall program savings should be decided at the design stage, not after the data are in; choosing among alternate formulas after seeing the different results may call into question the credibility of the study.

Reporting Results

Even for bottom-liners, savings numbers are meaningful only in relation to a particular set of questions. Enough detail should be included in the report for readers to know which choices the evaluators made at each design step. The report should also include enough information so that someone with a different way of computing the bottom line can make alternative calculations. ■

Miriam Goldberg is a statistician in the Energy End Use Division at the U.S. Energy Information Administration. The opinions and conclusions expressed herein are solely those of the author and should not be construed as representing the opinions or policy of any agency of the United States Government.

MEASURING RESULTS

Inputs and Outputs

PRISM uses monthly energy consumption data from energy bills, along with daily average temperature data from a nearby weather station, to calculate a weather-normalized index of annual energy use, known as Normalized Annual Consumption (NAC)
amount of energy that would be cons
:r conditions. PRISM
umption data to dete
useful parameters:
stimate of the weath
ghting, cooking, ap
the marginal requi
F drop in outside te
rature τ (sometimes
re), the outside tempe
quires heating. Wha
umption analysis tool
emperature as a vari
rovides; PRISM calcul
rameters just mention
reliability of each. (Se
PRISM uses.)
es consumption for ea
-days constructed from
he same period. Linea
sumption vs. degree-da
α, β and τ), from which
= 365α +
= base level + spa
is the average number

days at the estimated reference temperature in an average year (based on several years of historical weather data).

PRISM's widest use is in measuring the effects of building retrofits, in which the NAC before and after the retrofit are compared, with the difference attributed to the retrofit. Generally it takes data for twelve consumption periods (from monthly bills), spanning a year before a retrofit, and twelve for the year after, to produce respectively NAC_{pre} and NAC_{post}, from which savings for that house are computed:

Savings = NAC – NAC

Error Statistics Mean?

ıdard error (se) of a model e
e to place in the estimate: t
less reliable the estimate is.
may be written as NAC ± se(N
idence that the *real* NAC is
rs of the PRISM estimate (an
lls within one standard err
in Table 1, there is a 95% p
etween 21,530 and 21,190 (21
that it is well determined. A
standard error of NAC, which
w percent of NAC. Specifically
of variation (CV) of NAC:

CV(NAC) = se(NAC)/

The example above has a well-det CV(NAC) of 0.01 or, in simpler but ec tive standard error of 1%.

The R^2 statistic provides a measure of the consumption data to the PRISM m month-to-month variability in the consum by PRISM (and thus by outdoor temper almost all of the variability in the above 0.998. The extent to which R^2 is less tha

NO POSTAGE NECESSARY IF MAILED IN THE UNITED STATES

BUSINESS REPLY CARD
FIRST CLASS MAIL PERMIT NO. 3244 BERKELEY, CA
POSTAGE WILL BE PAID BY ADDRESSEE

HOME ENERGY magazine
2124 Kittredge, Suite 95
Berkeley, CA 94704–9942

Table 1. Examples showing good and bad PRISM parameter results.*

	R^2	NAC (kWh/yr)	τ($\pm$se)** (°F)	α (kWh/day)	βH_0 (kWh/yr)
Example House with "Good" parameters	0.998	21,360	63($\pm$1)	35	8,540
[CV]***		[0.01]	[——]	[0.02]	[0.03]
Example House with "Bad" parameters	0.347	4,510	57($\pm$23)	11	436
[CV]		[0.05]	[——]	[0.09]	[0.68]

* "Good" house is HER 4; "Bad" is HER 39 from the 64-house study.

** For τ, the standard error (se) of the parameter is shown because the relative standard error of τ has no meaning.

*** The coefficient of variation of parameter X, CV(X), is computed as se(X)/X. (For example, in the first case, CV(NAC) = se(NAC)/NAC = 170/21,360 = 0.01.)

Nevertheless the corresponding NAC is well determined; its standard error is only 5%.

The "magic" of PRISM is that NAC can be well determined even if the individual parameters are not. This phenomenon occurs for complex statistical reasons,[4] but the practical outcome is that reliable savings estimates are available from PRISM for a far larger fraction of houses than the raw data might suggest. For this reason, criteria to determine the reliability of results for individual houses need to take advantage of the available PRISM error diagnostics. The box, "The PRISM Reliability 'Sieve,'" contains a description of a procedure for determining the reliability of NAC. This is based primarily on two statistics: the relative standard error of NAC, which we want to be as low as possible, and the R^2 value of the PRISM fit, which we want to be as close as possible to 1.0.

This set of reliability criteria was developed for a study of 64 electrically heated houses in New Jersey. The homeowners (pre-selected to include a range of consumption profiles) were the subjects of a pilot study of energy conservation feedback in which they were given a three-page report on their annual energy use for the most recent two years. Both raw and weather-adjusted consumption (NAC) were provided in addition to the difference in energy use between the two years.[5] The procedure in the box was used to determine whether the NAC estimates were sufficiently reliable to report. This set of 64 houses serves as our case study for this article.

How Do I Decide If I've Got Reliable Results?

In simplified terms, the PRISM NAC estimate may be considered reliable if the relative standard error of NAC is 6% or less (i.e., CV(NAC) $\leq$ 0.06) and if the R^2 statistic is 0.7 or larger. As shown in the scatterplot of CV(NAC) vs. R^2 for our study sample of 64 electrically heated houses (Fig. 2), most of the houses cluster in the bottom right-hand corner of the plot, representing the 51 cases (80% of the sample of 64) meeting the criteria.[6] It is evident from the plot that high R^2 and low CV(NAC) go together. In some applications more stringent criteria may be desirable.[7] For example, a criterion of $R^2 \geq 0.9$ for

The PRISM Reliability 'Sieve'

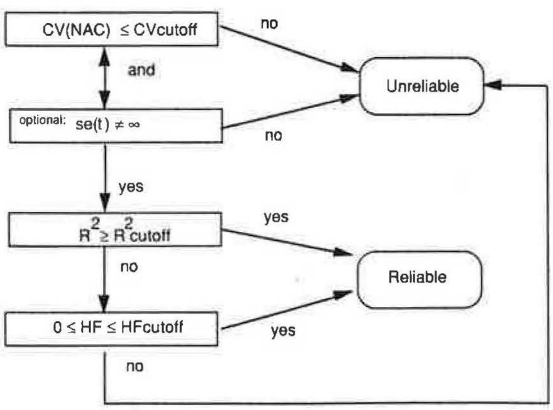
The procedure we have developed for testing the reliability of NAC uses several PRISM statistics. The most important two are the relative standard error of NAC, or CV(NAC), and the R^2 statistic, whose closeness to 1.0 provides a measure of goodness of fit of the PRISM model. Consideration of R^2 and CV(NAC) together provide useful reinforcement of reliability.

In addition, the standard error of the reference temperature, se(τ), is useful for flagging "problem" houses; a very large or infinite value of se(τ) (indicated as a "–9" in the PRISM output) often indicates a consumption data anomaly. Most important, when se(τ) is infinite (i.e., not estimable), se(NAC) cannot be correctly computed and thus the NAC value may be suspect.

A fourth statistic, the heating fraction of NAC, given as HF = $\beta H_0(\tau)$/NAC, is an indication of whether heating is too insignificant to allow accurate determination of the other parameters (especially τ). A house with a low value of HF can have a low R^2 (from a low value of the heating slope β) and in addition a very high se(τ), since τ is physically meaningless and thus indeterminable when there is no heating. Its total consumption index, NAC, can nevertheless be highly reliable. Therefore, when HF is very low, the "sieve" uses only CV(NAC) and ignores R^2.

Since applications of PRISM vary widely in their sample sizes and data quality, and particularly in the extent to which the scorekeeper can afford to trade sample size reduction for increased accuracy, the specific reliability sieve shown in the figure below is intended as a suggested starting point. The indicated cut-offs worked well in our feedback study of 64 houses: the cut-off of 0.7 for R^2 was chosen to maximize the number of interviewed participants to receive reasonably reliable feedback. A higher cut-off (e.g., R^2 of 0.9) is likely to be not too stringent in an evaluation of a large sample of treated houses, especially since (in our experience) median R^2s of 0.95 or higher are typical. The exclusion of houses with infinite se(τ) is shown as an optional mesh in the sieve, and probably is needed more for individual-house feedback than for group-level evaluation.

Reliability Sieve for NAC



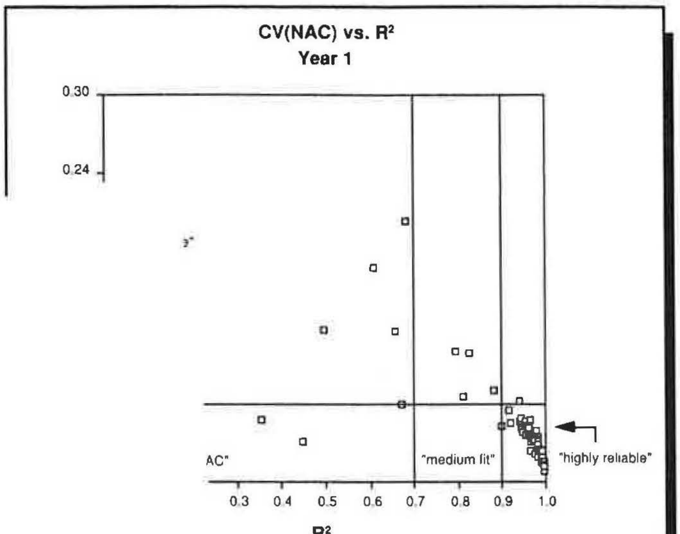
Sample cut-offs are:

CV(NAC) = 0.06

R^2 = 0.7

R^2 = 0.9 (for more stringent criteria)

HF = 0.15



plot of the two primary reliability statis- [, CV(NAC) vs. R^2, for 64-house sample ated houses (for the first year of data).

ld reduce the number of reliable cases ɔnly one (labeled "medium fit" in Fig. a scatterplot like the one in Fig. 2 as a the reliable vs. the unreliable cases ent sets of reliability cut-offs.

ɪns, these simple criteria, based only on ıigh R^2, may be sufficient. Numerous PRISM have the high model reliability ;. The median R^2 for PRISM applied to ɔr the sample of 64 is 0.97, the median the median se(τ) is 3.0°F; very similar he second year as well. Overall, 43 out reliable NACs in both years, and only ͺC in neither year. (The remaining 17 ͺC in only one year.) These average vith other studies of gas-heated and ll as electrically heated houses in the Therefore, criteria to separate reli- ͺCs apply to those few cases that fall ɜn, or average.

ʰich it is important to squeeze as of a sample as possible (when the ..ıɑıı, or when NAC information for a ꞇustomer is highly valued), a look at some of the other PRISM statistics (se(τ) and heating fraction)—as described in the reliability sieve box—can add a few additional cases to the "reliable" bin. For the case study, this more elaborate sieve adds two to the original 51 (out of 64) that made it through the reliability sieve.

What if the Results Aren't Reliable?

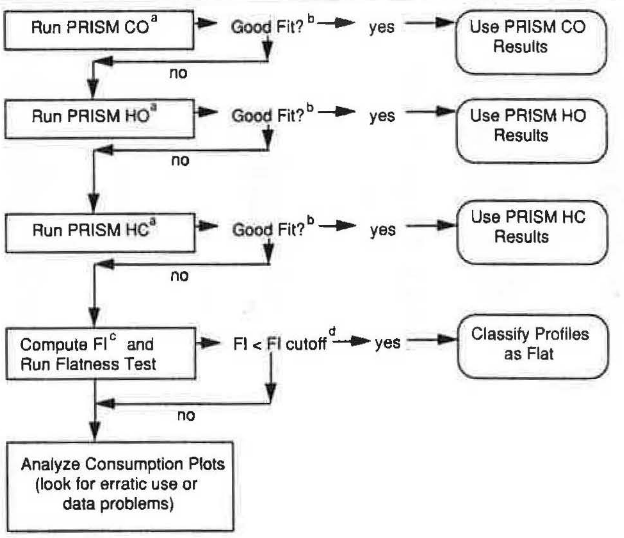
Whether or not special attention should be given to unreliable results depends on the particular application. In an evaluation of a conservation program involving hundreds or thousands of participants, these criteria may be used to create a "good-house" subsample, and comparisons may be made between the characteristics of the "good" houses and the rest of the sample. Time may not allow additional attention to the "bad-house" data. On the other hand, if the sample is small, or if each homeowner in the sample is to receive feedback based on PRISM results (as in an audit, a rating system, or a shared savings program), some extra effort to extract information from the consumption data for those not modeling well may turn out to be worthwhile.

Model Selection

The first step is to test for the appropriate model, as shown in Fig. 3.[3] This mainly applies to electricity, with the possibility of cooling as well as heating. For example, if the standard PRISM model (for heating) does not yield a good fit, the analogous PRISM model for cooling is then applied. If the resulting fit is good (using the same reliability criteria), it is clear that the first model shouldn't have worked. Perhaps the house's HVAC system was misclassified or there was a switch to a heating system using a different fuel; in any event one then has a good model for the data, and a quantitative understanding of the house's space-conditioning use.

Looking at the Consumption Plots

For those cases classified as "unreliable," the next step is to take a closer look at the raw consumption data. Using PRISM output, plots of consumption per day (billing consumption divided by the number of days in the billing period) vs.



[a] CO = Cooling-only; HO = Heating only; HC = Heating-and-Cooling; see note 3.

[b] Primary criteria used in sample study to define "Good Fit" were CV(NAC) ≤ 0.06 and R ≥ 0.7; see figure in "The PRISM Reliability Sieve" box.

$$^{c}\text{FI (Flatness Index)} = \frac{\text{standard deviation (mean consumption/day)}}{\text{mean consumption/day}}$$

[d] FI cutoff = 0.20 for sample study

Figure 3. Overall procedure for model testing and examination of problem data.

time are analyzed, and considered in the following three well-defined categories: a) flat consumption profiles, b) presence of possible data errors, and c) unexplained but patterned profiles.

Flat Consumption Profiles

Flat consumption profiles indicate little or no weather dependence of the energy consumption, meaning that the fuel being analyzed is not being used for space conditioning. In PRISM terms, this means that consumption is "base-level" throughout the year. In order to determine the degree of "flatness," we compute a "flatness index" for a house's consumption data (as defined in Fig. 3), which can serve as a useful indicator for why PRISM didn't model well.

Fig. 4 shows an example of a house that modeled well in year one (with flatness index = 0.65) and dramatically changed to flat consumption in year two (second year results for this house are shown in Table 1 as being "bad"); the corresponding value of flatness index is 0.19. An interview confirmed a heating fuel switch—the residents had replaced their electric furnace with a gas-fueled one between the two years.

In the two years of data for the 64-house sample (i.e., for 128 cases), the median flatness index for the entire sample was 0.6. Only seven have flatness indices lower than 0.2. All seven failed the PRISM reliability criteria, clearly because of a lack of strong weather dependence. It is interesting that six of these (including year two for the house in Fig. 4) had extremely low R^2 (below 0.5) but a CV(NAC) within the reliability cut-off of 0.06. These correspond exactly with the cases that made it through the reliability sieve on the basis of a low heating fraction (the reliability sieve allows low R^2 if heating fraction is sufficiently low) and thus have reliable NACs. Not surprisingly, a low flatness index seems to reinforce the test for low heating. Using a cut-off of 0.2, the seventh case with flatness index lower than 0.2 is the only low flatness index case with unreliable NAC (with a CV(NAC) of 0.20). Therefore, only one of the unmodelable cases in this sample would receive a "flat consumption" classification.

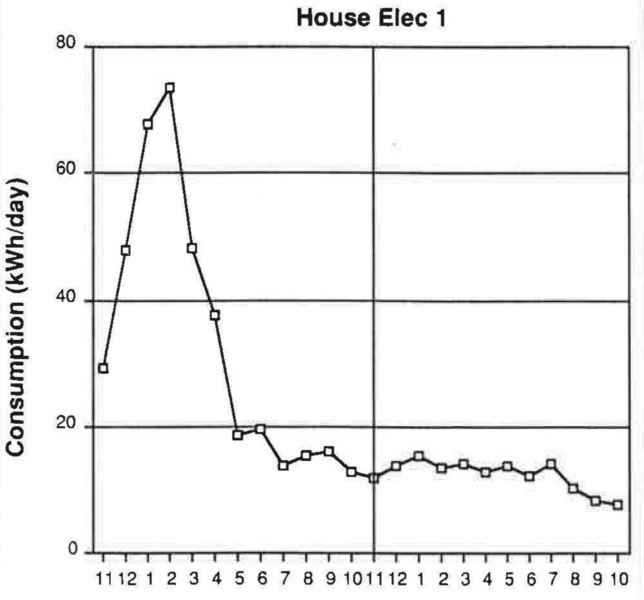


Figure 4. Example house that modeled well in Year 1 and changed to flat consumption in Year 2. (House HER39.)

Table 2. Table of PRISM results before and after combination of pairs of incorrect data points.

House*	R^2	CV(NAC)	$\tau(\pm se)$ (°F)	CV(α)	CV(βH_0)
Gas 1					
before	0.728	0.144	64(± 12)	0.83	0.37
after	0.966	0.043	62(± 3)	0.21	0.10
Gas 2					
before	0.790	0.098	50(± 7)	0.22	0.31
after	0.919	0.049	64(± 6)	0.16	0.18
Elec 2					
before	0.777	0.241	33(± 10)	0.43	0.32
after	0.975	0.040	49(± 4)	0.12	0.08

* Houses Gas 1 and Gas 2 are gas-heated houses (S209 and S244) from Dutt et al. (1986); House Elec 2 is an electrically heated house (HER 37) from our 64-house sample (Reynolds and Fels, 1988).

Data Errors

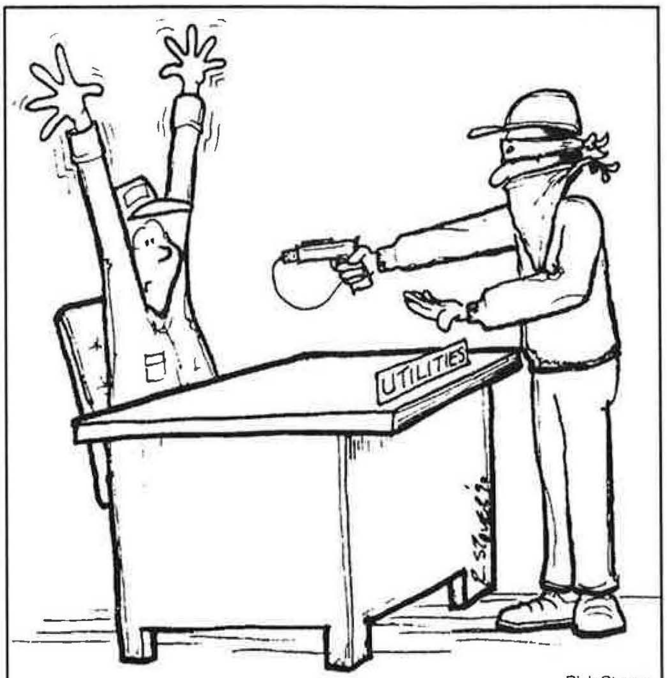
Data errors may result from three sources: error in the original recording of the meter reading (either in an incorrect reading or in a miscoding of an estimated reading as an actual reading), error in the transcription of the data for analysis, or error due to broken equipment (i.e., gas or electric meter). Such errors may be very obvious when the plots are reviewed.

One example of a meter reading error in a gas-heated house is shown in Fig. 5a. The circled "data points" are consecutive and, respectively, considerably lower and higher than the regression line (shown in Fig. 5b) implied by the other consumption data. Combining these two data points, effectively treating the December consumption period as estimated, yields a dramatic improvement in the fit. Table 2 summarizes this and two other examples that meet the reliability criteria as a result of this step in the process.[10]

If this problem occurs exceedingly often (across months and for many houses), the actual meter reading dates could be suspect. The evaluator should be wary of PRISM estimates based on "theoretical" meter reading dates (like the first of the month) that are used because the utility may not have provided the actual meter reading dates for the consumption data. Results from these bad data are likely to fail the PRISM reliability criteria.

Unexplained Patterns

Some profiles may appear to have very well defined patterns of usage that are not weather related. These patterns may be influenced by other variables such as the installation of new equipment, the addition or deletion of a major appliance, or a change in occupant schedules (a new baby arrives; a teenager goes off to college). The consumption plot allows the scorekeeper to explore the



Rick Stover

"OK, I want all your utility bills from 1988 on up—preferrably in small Btus."

reasons for such patterns with the homeowner during an energy audit, by asking a simple question such as, "I noticed that your electricity use changed in March—did you do something different that month?"

Fig. 6 shows two houses from our study sample that had reliable model results in the first year and "unexplained" consumption patterns in the second year. For the house in Fig. 6a, a strange drop in consumption occurred in January, leveling out to the house's base-level consumption. An interview with this homeowner revealed a fuel switch in mid-winter, from electricity to gas. The house in Fig. 6b suggests no or very little air conditioner use in year one, but considerable (though erratic) air conditioner use in year two. Thus the consumption plots can shed some light on what may be occurring in the homes to cause poor PRISM fits.

PRISM Computation of Standard Error of Savings

$$\text{Savings} = NAC_{pre} - NAC_{post}$$

$$se(\text{Savings}) = \sqrt{se^2(NAC_{pre}) + se^2(NAC_{post})}$$

$$se(\%\ \text{Savings}) = \sqrt{\frac{(NAC_{post})^2\, se^2(NAC_{pre})}{(NAC_{pre})^4} + \frac{se^2(NAC_{post})}{(NAC_{pre})^2}} * 100$$

where $\%\ \text{Savings} = \dfrac{\text{Savings}}{NAC_{pre}}$

Reliability criteria for savings are that NAC_{pre} and that NAC_{post} individually meet their criteria and savings satisfy the following:

$$|\text{Savings}/se(\text{Savings})| > 1.8$$

(See Appendices to Fels, 1986, and Reynolds and Fels, 1988.)

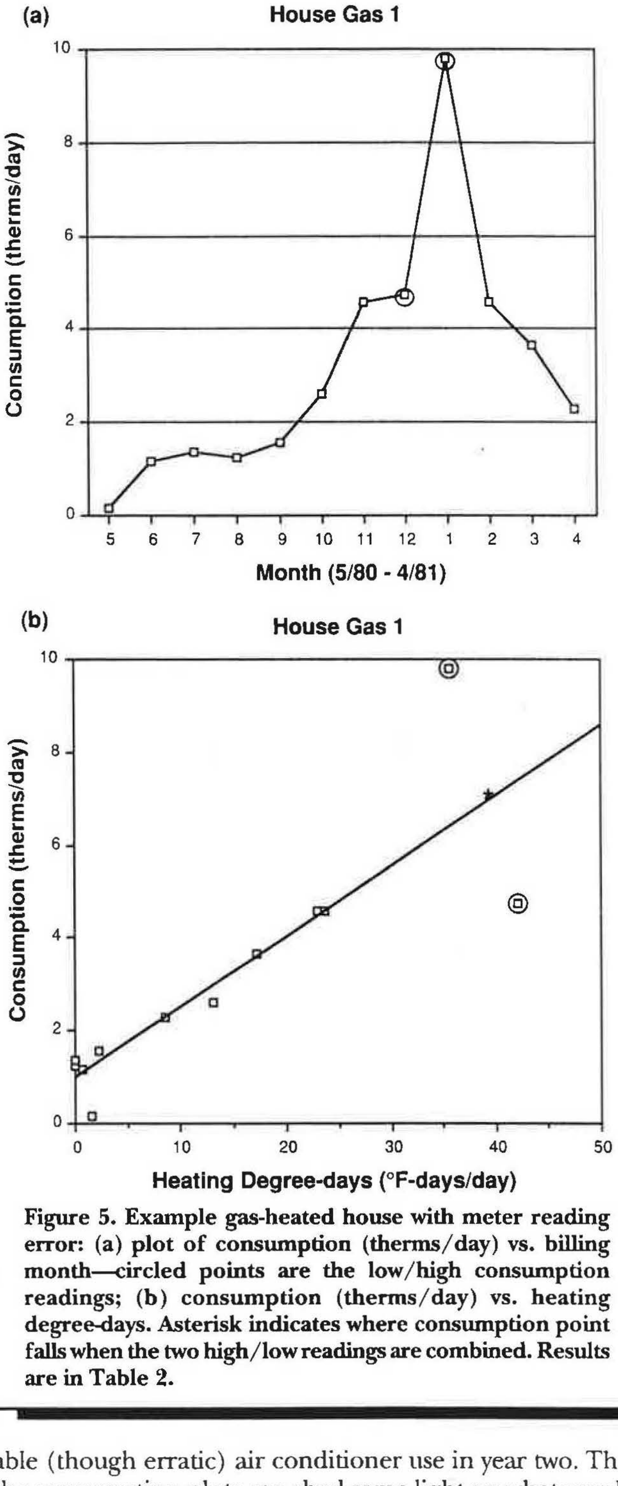


Figure 5. Example gas-heated house with meter reading error: (a) plot of consumption (therms/day) vs. billing month—circled points are the low/high consumption readings; (b) consumption (therms/day) vs. heating degree-days. Asterisk indicates where consumption point falls when the two high/low readings are combined. Results are in Table 2.

Now That I've Got Reliable Results...

The next step is to calculate savings and determine whether those savings are reliable. Typically, PRISM is run twice on a single house (Fig. 1), for a pre-retrofit period and (skipping the month or months in which the retrofit was performed) for a post-retrofit period, giving respectively NAC_{pre} and NAC_{post}, with associated statistics for determining reliability. The savings and their standard error are readily computed from the standard errors of NAC, as shown in the box on the previous page.

If the objective is to evaluate savings in a group of houses, the individual-house savings results may be run through a reliability sieve with predetermined cut-off values appropriate to the application. The resulting "reliable" cases would then be run through a standard statistics package to find the average (median or mean) savings in the treatment vs. control group and the corresponding distributions (see p.25, "Elements of a Weatherization Evaluation").[11] Additional statistics can then determine

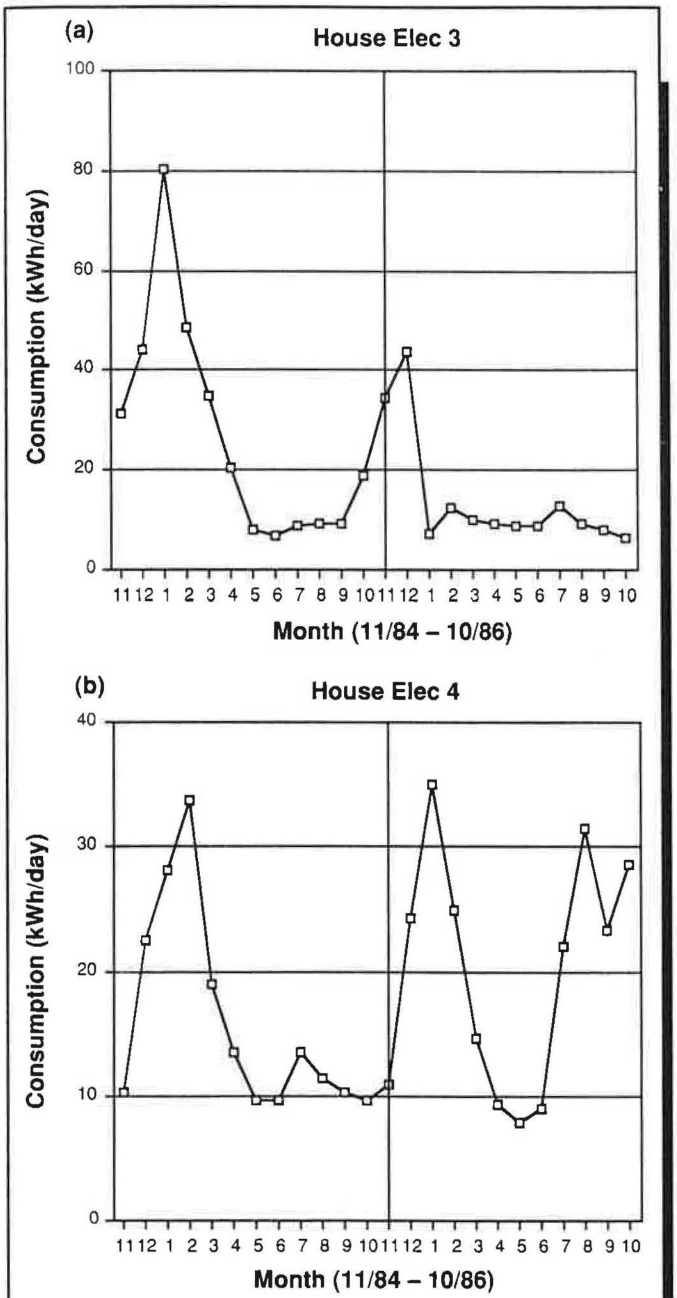
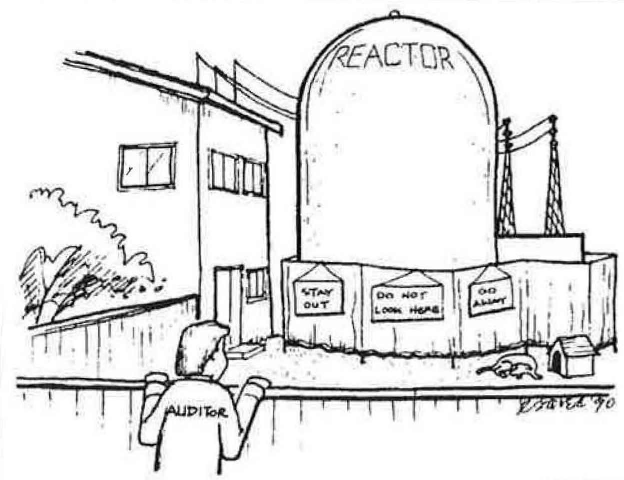


Figure 6. Example houses with reliable fit in first year and unreliable fit in second year. (a) a house with a fuel switch in mid-winter, and (b) a house that had increased air conditioning use in the second summer. (House Elec 3= HER 12; House Elec 4 = HER 42



Rick Stover

Auditor Joe was stunned when, upon closer inspection, he found the Hendersens heating their home with waste heat from a private nuclear reactor.

whether the distributions of savings in the different groups being compared are significantly different from each other.

If, on the other hand, the program under analysis involves individual feedback to the homeowner, the auditor wants to be confident that the numbers presented to each homeowner are accurate, and thus the statistical significance of each savings estimate becomes relevant. For an auditor, running PRISM on a house's billing data before the actual audit can sharpen questions for targeting sources of energy efficiency measures (e.g., enrollment in a time-of-use rate if base-level use is high, or furnace efficiency improvements if heating use is high or, for houses that don't model well, sources of heating other than the fuel designated as the heating fuel). For savings measured after a conservation action, our criteria for reliability and statistical significance of the savings (or any change in consumption) are first that NAC_{pre} and NAC_{post} individually meet their reliability criteria, and then that savings meet the criterion shown in the box.

Recommendations

For the scorekeeper, the procedure outlined here involves the following steps:

1) Run PRISM on pre- and post-retrofit billing data.
2) Determine reliability criteria for the particular application (suggestion: use cut-offs suggested in the reliability sieve as a starting point, and plot CV(NAC) vs. R^2 to see how the results cluster).
3) Categorize PRISM results for each period for each house as "reliable" and "unreliable" (using reliability sieve with cut-offs established in previous step).
4) From NAC in the pre- and post-retrofit periods, calculate savings for each house, the associated standard error of savings, and whether the savings are statistically significant.
5) For those homes and time periods with unreliable results, examine the raw consumption data:
   a) Calculate the flatness index. If it is lower than an established cut-off (e.g., lower than most of the flatness indices for cases that model well), classify profile as "flat."

b) For unreliable cases with a flatness index greater than the cut-off, examine consumption plots for data errors, indicated by consecutive high / low consumption data. Re-run PRISM with combined data points and repeat reliability tests.

c) For remaining unreliable cases, examine consumption plots for possible non-weather-related patterns.

Although PRISM is a simple method, interpretation of the results requires a solid understanding of the assumptions and statistical underpinnings of PRISM. This is true for someone reading or writing a report on a PRISM-based evaluation as well as for the PRISM analyst or scorekeeper. It is important to keep in mind that application of PRISM to determine savings is just the first step. Standard errors should be presented alongside the savings estimates, so that the reader can know how much confidence to place in the results. Only with reliability statistics can savings estimates from different groups or different types of weatherization work be meaningfully compared. With the availability of individual-house results, different groups—of "good" houses vs. the complete sample, or of a selected income group or house size vs. all participants, etc.—can be compared to provide enhanced insight into the level of success of the program being evaluated.

We realize that we've subjected the reader to a lot of statistical jargon. The important points to remember, however, revolve around the statistics PRISM uses. Scorekeepers choose to use PRISM for analysis of billing data not only because of the program's simplicity but because they want to know that the savings calculations can be relied upon. When utility executives read a report on savings from a particular in-house weatherization program that used PRISM, they can know that the group-level savings estimates were based on a distribution of savings calculated at the individual-house level and subjected to tests of desired levels of reliability. PRISM provides the information to make this possible. ■

Endnotes

1. Fels, M.F., ed., 1986, "Special Issue devoted to Measuring Energy Savings: The Scorekeeping Approach," *Energy and Buildings*, Vol. 9, #1–2.
2. Fels, M.F., 1986, "PRISM: an Introduction," *Energy and Buildings*, Vol. 9, #1–2, pp. 5–18.
3. The PRISM model discussed throughout this article is the Heating-Only (HO) model. The PRISM Cooling-Only (CO) model (incorporated as an option in the software) is a direct cooling analogue of the HO model, for application to electricity used for cooling and not heating. Its parameters include a base level, cooling slope, and cooling reference temperature. Ways to interpret the CO model results may be inferred from the discussion in this article that emphasizes the HO model. A third and more complicated model, PRISM HC (Heating-and-Cooling), which has been developed for research purposes only at CEES, is useful for analyzing fuel used for both heating and cooling. For a description of the HC model, see Stram, D.O. and M.F. Fels, 1986, "The Applicability of PRISM to Electric Heating and Cooling," *Energy and Buildings*, Vol. 9, #1–2, pp. 101–110.
4. Goldberg, M.L., 1984, "A Geometrical Approach to Nondifferentiable Regression Models as Related to Methods for Assessing Residential Energy Conservation," Ph.D. thesis, Center for Energy & Environmental Studies Report #142, Princeton University, Princeton, NJ.
5. Layne, L.L. et al., 1988, "Design Criteria for a Consumer Energy Report: A Pilot Field Study," Center for Energy & Environmental Studies Report # 220, Princeton University, Princeton, NJ; and Reynolds, C.L. and M.F. Fels, 1988, "Reliability Criteria for Weather Adjustment of Energy Billing Data," *Proceedings of the 1988 ACEEE Summer Study on Energy Efficiency in Buildings*, the American Council for an Energy-Efficient Economy (ACEEE), Washington, DC.
6. The earlier report (Reynolds and Fels, 1988) used a variation of PRISM ("Robust" PRISM, under development), which downweights anomalous data points. For this article we report results obtained from the "regular," readily available version of PRISM.
7. Dunsworth, T.S., T.A. Miller, and M.J. Hewett, 1984, "Empirical Guidelines for Using the Princeton Scorekeeping Model," Minneapolis Energy Office Technical Report #84-7.
8. Fels, M.F., ed., 1986.
9. See Note 3. For an application of this procedure, see Reynolds, C.L., P. Komor, and M.F. Fels, 1990, "Using Monthly Billing Data to Find Energy Efficiency Opportunities in Small Commercial Buildings," *Proceedings of the 1990 ACEEE Summer Study on Energy Efficiency in Buildings*, Washington, DC.
10. Fels, M.F. et al., 1985, "PRISM: A Conservation Scorekeeping Method Applied to Electrically Heated Houses," EPRI Report EM-4358, Palo Alto, CA.
11. Dutt, G.S. et al., 1986, "The Modular Retrofit Experiment: Design, Scorekeeping and Evaluation," *Energy and Buildings*, Vol. 9, #1,2, pp. 21–33.

Further Reference

See also "Performance Measurement and Analysis," Volume 10 of *Proceedings of the 1990 ACEEE Summer Study on Energy Efficiency in Buildings* (in particular, papers in PRISM session), Washington, DC.