Short-Term Metering for Measuring Residential Energy Savings: Concerns and Recommendations from Comparison with PRISM

William W. Hill, Ball State University
Michael Blasnik, GRASP
Kathleen M. Greely and John Randolph, Virginia Polytechnic Institute and State University

The PRInceton Scorekeeping Method (PRISM), the most widely used technique for evaluating residential consumption, has the disadvantage of imposing a long lag time between retrofit and results, due to PRISM's requirement of 10 to 12 months of post-retrofit utility bills. In addition, serious questions have been raised concerning possible attrition bias as a result of houses dropped from the sample for lack of utility data or poor PRISM "fits". Increasingly, researchers and practitioners alike have turned to short-term evaluation methods to obtain more timely results. However, there has been little attention given to assessing the accuracy and reliability of these short-term methods.

One increasingly popular short-term approach utilizes furnace run-time meters to obtain estimates of pre- and post-retrofit energy consumption. In this paper we compare the savings estimated from these run-time meters with savings obtained from PRISM analyses for 13 houses included in Virginia's low income weatherization evaluation. Short-term metering was carried out over the course of a single heating season, with 3 to 10 weeks of run-time data in the pre- and post-retrofit periods. Savings from this run-time data are compared with PRISM's weather-normalized estimates from utility data. Mean savings determined by the two methods, while not differing by a statistically significant amount, do appear to differ in a non-random way. Run-time data appear to result in somewhat higher savings both in absolute and percentage terms.

Various hypotheses/explanations for this difference are examined using both the Virginia data and data from a "well behaved" and sub-metered house in Indiana. The inclusion of data from weeks with few heating degrees appears to be especially problematic. We suggest a number of ways for dealing with this problem including screening criteria for weekly data based on HDD and a calculation method which give greater weight to data from colder weeks. In general, our analyses suggest that careful attention to research design and data quality may be even more important when using short-term metering than when using PRISM.

Introduction

The standard evaluation tool for measuring energy savings in the residential sector is PRISM, the PRInceton Scorekeeping Method (Fels 1986). The principal refinement of PRISM over other regression approaches is that it uses Newton's method to iteratively search for the best reference temperature for a house and then uses this reference temperature to compute the heating degree days (HDD). Perhaps PRISM's biggest attribute is that it provides a standardized methodology for researchers and practitioners alike, ensuring that all are "keeping score" the same way, thus allowing for comparison of results. The principal drawback of PRISM or any method which uses a full year of data is the long lag time between retrofit and results, a consequence of the requirement for 10 to 12 months of post-retrofit utility data. This lag time, while of concern in any evaluation, is especially significant in low-income weatherization where evaluation is increasingly being seen as an integral part of the process, providing feedback to crews and contractors on what worked and what didn't, and thus increasing the rate at which programs move up the learning curve. A more serious concern is that this lag time, when coupled with other normal utility data problems, often results in large sample attrition, especially in low-income populations. This loss of sample is a serious concern if, as found by Blasnik (1989), it is a non-random loss, thus resulting in attrition bias.

Numerous short-term methods, i.e., those that need much less than a year of data, have been proposed over the years (Nadel 1987, Blasnik and Lent, 1988, Pigg 1992) as a way of avoiding the time constraints imposed by PRISM. In recent years, short-term methods have grown increasingly popular in the weatherization community, driven in large part by the desire to get timely feedback to crews and program managers, and also by the need to measure savings in homes with oil-fired or other non-metered fuel furnaces. The approach has also received a boost from new technological developments which make furnace run-time meters relatively inexpensive.[1] The usual approach in short-term metering is to wire an elapsed timer or "run-time" meter to a furnace so that the timer comes on whenever the furnace is firing. By multiplying this run-time for each week by the firing rate and dividing by the number of degree days in that week one obtains a measure of the space heat consumption of the house in Btu/HDD. Dividing this number by the heated area of the house in square feet yields space heat energy intensity in Btu/ft^2-HDD. By doing these calculations immediately, it is possible to identify outliers in real time and immediately check to see if there is justification for throwing out the data point (if, for example, upon call back to the client it is discovered that the energy use is low because of a vacation or heating system malfunction.) This is an improvement over methods such as PRISM in which decisions regarding outliers are made long after the fact and can not be so easily checked.

One of the main arguments in favor of short-term approaches is that they are relatively straightforward and more easily utilized by field-level personnel than is PRISM. Another argument is that they are much less time consuming. While our experience raises questions regarding both of these assertions, anecdotal evidence suggests that short-term methods are beginning to receive widespread use. For this reason, it is important that possible limitations of these methods, as well as data quality and data analysis concerns be addressed.

In this paper we compare results from short-term methods with results from PRISM using 13 houses from a low-income weatherization study conducted by three of the authors in Virginia (Randolph et al. 1991, Greely et al. 1991) and one house in Indiana for which we have two years of both weekly run-time and furnace sub-metering data. Using these data, we examine data quality issues and attempt to develop screening criteria and calculation procedures to increase the reliability of short-term metering methods. We also briefly address concerns regarding the appropriateness of using short-term data to estimate annual energy consumption.

Concerns About Short-term Metering

Extreme Values and Edge Effects

The shorter the time period of analysis, the greater the effect that extreme values can have on results. The relationship between consumption data and temperature is inherently more variable for weekly data than for monthly data. Also, as the time period gets shorter, the data at the beginning and end of each period become more important. Consumption data for an individual day may or may not correlate strongly with average temperature for that day, depending on when the minimum and maximum temperatures occurred; this effect will be more pronounced for weekly data than for monthly data.

The problem of days at the period "edges" or "boundaries" may be exacerbated by mismatches between temperature and meter reading times. For example, meters on the Indiana house discussed below are read once a week at 5:30 p.m., whereas the local weather station records 24 hour min-max temperatures at 7:00 a.m. Correlations between consumption and HDD are improved by calculating daily mean temperatures using the minimum temperature from day "i" and the maximum from day "i+1". With monthly data this "edge effect" is not worth correcting for, but with weekly data it may be important. Unfortunately, in actual field applications of short-term methods, the mis-match between weekly meter data and temperature data may be even more pronounced, owing to uncertainties surrounding the actual read times of meters and non-trivial data management protocols necessary to match meter and temperature data files.

Weeks with Few HDDs

Weeks with few heating degree days can result in poor estimates of annual consumption for a number of reasons. An average temperature for a sunny spring or fall day can be a very misleading indicator of the actual hourly temperature differential across the thermal envelope. Additionally, as the average daily temperature gets closer to the reference temperature (or balance temperature) for the house, the importance of using the correct reference temperature in the calculation of HDD increases. For example, the relative error in using heating degree days computed using a base of 65°F (HDD_{65}) for a house in which the correct reference temperature is 60°F is much larger when the average temperature is close to 65°F than when it is, say, near 30°F. Also, other factors such as mass effects, solar gains, internal gains, ground temperatures, occupant behavior, thermostat set point, etc. become relatively more important in weeks with few heating degree days.

The above problems could be avoided by simply avoiding weeks with too few HDD. Unfortunately, the typical research design utilizing an autumn/early winter pre-retrofit period, followed by mid-winter retrofit/weatherization, with the post-retrofit period in late winter/early spring, often runs into problems. Typically, data for the pre- period are fine, evaluators/researchers simply postpone weatherization until sufficient data from cold weeks have been accumulated. As a result, however, the post- period frequently ends up running into March, which in many climates results in weeks with marginal HDDs.

Seasonality Concerns

There are a lot of other factors besides average daily outside air temperature that determine a house's consumption of energy for heating. Unfortunately, many of these other factors change with the seasons. Ground temperatures tend to be warmer in the fall and early winter than in the late winter and spring. Thus, a house might be expected to use somewhat less energy per HDD in the fall than it would in the spring, depending on the extent of "ground coupling." Solar gains may be expected to vary with the seasons as a result of difference in solar intensity and sun angle, as well as differences in shading from deciduous trees. Wind also may vary significantly with the season.

The seasonality of non-heating energy consumption is well established (Fels et al. 1986). Energy consumption for domestic hot water (DHW), for example, will typically increase in the winter as a result of lower incoming water temperatures, greater standby losses from pipes and hot water tank, and perhaps higher "hot water" temperatures being called for by the occupants to compensate for colder shower enclosures.

PRISM, which treats base load as a constant, ignores this seasonal variation of hot water use. As a result, PRISM will tend to overestimate the space heat fraction and underestimate base load, something that has been shown by other researchers (Fels et al. 1986, Hirst and Goeltz 1986) and documented again in some of the data reported below. (For this reason, there may be some merit to the argument that short-term metering may in fact produce better measures of the change in a building's space heating efficiency than PRISM.)

Others have argued that occupants' thermostat setting behavior may change with the season. The point of all this discussion is that heating energy consumption per degree day can be expected to vary with the season and therefore some caution is called for in research designs in which the pre- and post-retrofit periods do not encompass a whole year. PRISM, or any method which utilizes annual data, should better capture these other factors which affect energy consumption, and thus provide a better estimate of annual energy consumption. How important are these other factors? How reliable are annual estimates based on short-term methods? These are questions which need addressing before evaluators can confidently use short-term methods to estimate annual savings.

Effect on Occupant Behavior--The Hawthorne Effect?

Finally, another concern raised about short-term methods which utilize run-time meters, especially those that require the client/occupant to read and report the meter readings on a weekly basis, is that they are not exactly unobtrusive measures. There may be a tendency on the part of the occupants reading these meters to please the researchers by exhibiting more energy conservation behavior than they would otherwise, thus affecting measured savings. While not the focus of this paper, we do think that the potential for this "Hawthorne effect" (named for the famous study of workers at Western Electric's Hawthorne plant in 1924) to affect the persistence of savings merits serious attention. One way of dealing with this concern is to utilize control groups in which meters are hung and read in a similar group of houses which receive no weatherization. In light of the seasonality concerns discussed above, the use of control groups may be even more important in short-term metering studies than in studies in which annual data are used.

Measured Savings--Comparison of Results of Different Methods

The Virginia Low-Income Weatherization Evaluation Project

As part of an evaluation of the Virginia Low-Income Weatherization Program (Randolph et al. 1991, Greely et al. 1991), 59 single-family and mobile homes were fitted with furnace run-time meters and weatherized with innovative conservation measures during the winter of 1989/90.[2] Run-times and daily average temperatures were collected at approximately weekly intervals, with weatherization taking place between December and early March. The earliest run-time data were collected in

September, with data collection running through the end of April in some homes. Run-time data from an average of 11 weeks preceding weatherization and 8 weeks in the post- period were collected; 90% of the homes had at least 5 weeks of data in each of the pre- and post- periods.

Energy consumption for each measurement period was determined by multiplying the run-time by the furnace's firing rate. Furnace firing rates were determined by weatherization crews at the time the run-time meter was installed; installers timed one or two revolutions of the one or two cubic foot dial on the gas meter while nothing other than the furnace was on. Each week, telephone calls were made to each home to obtain weekly run-time readings. The space-heating energy intensity was then calculated in Btu/ft^2-HDD.

The mean Btu/ft^2-HDD for the pre- and post weatherization period for each house was then computed by summing these weekly Btu/ft^2-HDD figures and dividing by the number of weeks in each period, the method referred to below as the "average ratio method." Weeks with anomalous data, as revealed in client interviews (e.g., house unoccupied for a week), were excluded from the average, as were weeks with Btu/ft^2-HDD differing from the mean by more than 50%. Outliers identified by this rule usually corresponded with weeks that had extremely mild weather; the rule typically excluded one or two data points for each house. Annual energy consumption was calculated from the mean space-heating energy intensity by multiplying by the heated area of the house and the long-term average annual HDD_{65}.

Measured Savings--PRISM Versus Run-Time Results

Of the 59 pilot homes, sufficient utility billing data to perform a PRISM analysis were available for only 15 houses. (About half of the homes were oil-heated, with no consumption data available; gas utilities were unable to provide sufficient historical data for many of the homes, primarily due to a merger of gas companies which eliminated access to customer records.) All of the homes with utility data were single-family structures which used gas for space heating only, or for space heating and hot water. One of these 15 homes was dropped from further analysis because its consumption was poorly fitted by PRISM (R^2=0.3); for the other 14 homes, R^2 was $\geq$ 0.9, average coefficient of variation of NAC (CV(NAC)) was < 4%, and average coefficient of variation of normalized annual heating consumption (CV(NAHC)) was < 7%. Another home was dropped from the summary statistics reported below because it had only 3 run-time meter readings in the post- period.

Using PRISM we found mean savings of 16% for these 13 houses, with average reference temperatures of 64°F in the pre-retrofit period and 63°F in the post period.[3] Mean savings of 22% were found for these same 13 houses using run-time data and the "average ratio method" discussed above. While this difference is not statistically significant (as might be expected given the small sample) absolute savings for specific houses varied widely, as shown in Figure 1. The median absolute deviation of savings (MAD) is an indicator which measures the typical absolute value of the discrepancy between savings estimated using PRISM and run-time metering. The MAD for these data using the average ratio method and HDD_{65} was 9.6 MBtu/yr, or 40% of the average savings.

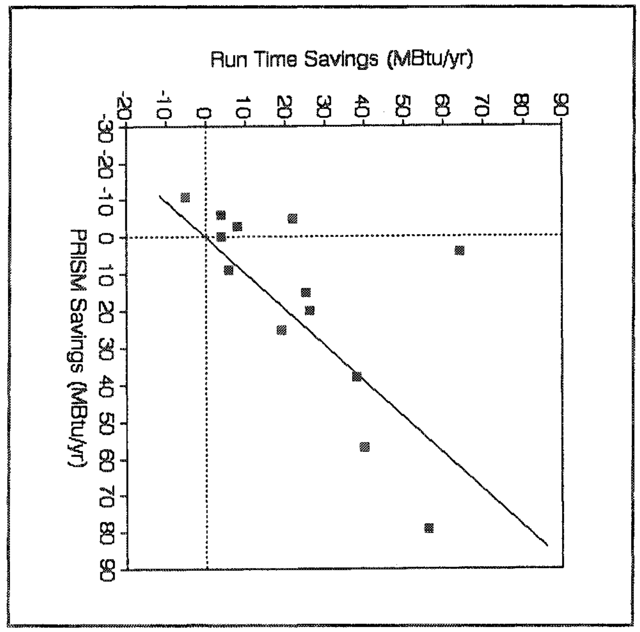

Figure 1. Annual Savings (in MBtu/yr) as Estimated by Run-Time Meter Using Average Ratio Method, Versus Annual Savings Estimated by PRISM. Data from 13 Houses in Virginia

Possible Explanations for the Difference in Measured Savings

We explored a number of hypotheses that might explain the difference in individual consumption estimates between the run-time metering and PRISM approaches. PRISM, as discussed above, produces a systematically biased estimate of space heating consumption because it loads seasonal non-heating usage onto the space heating parameter (Fels et al. 1986, Hirst and Goeltz 1986). This may partially account for the difference between PRISM and run-time meter estimates. However, cases with heating as the only gas use showed no better correspondence between PRISM and run-time estimates than cases with gas baseload uses.

A closer examination of the data during the preparation of this paper revealed some possible sources of error which may have influenced these results. In comparing PRISM daily temperature files (from NOAA weather tapes) with daily temperature data collected at the time of the weatherization evaluation (collected from local newspapers by local agency personnel) a number of errors were discovered in the local agency's temperature data for one city. Preliminary analyses of data from an ongoing evaluation study in Indiana which also utilized this short-term approach has uncovered similar temperature data problems. It may be that the apparent advantage of each agency using temperature data from a nearby weather station is outweighed by the poorer quality of data from these smaller stations and the increased opportunities for introducing errors. (Data from the principal weather stations is more carefully recorded and is verified against other weather stations before release.)

The measurement of the furnace firing rate introduces another potential error in short-term methods. In some houses in the Virginia study, the firing rate was measured twice, with discrepancies between the two readings of about 20%. Inaccuracies in the measurement of the firing rate, while not affecting the percent savings, would affect absolute savings.

Another possible factor affecting these data is the use of supplemental heating fuels, a fairly widespread practice among this Virginia low-income population. Interviews with the occupants of these homes, conducted as part of the weatherization evaluation, revealed that electric or kerosene heaters had been used in six of the homes. While these occupants promised not to use these supplemental heaters during the evaluation period (i.e., during the short-term metering study), the fact that they had used them prior to the study would affect the savings calculated by PRISM. If supplemental heating was more widespread in the pre-retrofit year than in the post, a reasonable assumption given the effectiveness of the weatherization performed, this would have the effect of lowering estimates of savings by PRISM.

Warm Weather and Reference Temperature Concerns

We suggested above that using the correct reference temperature in the calculation of HDD would probably be especially important in warmer weeks with few HDD. To examine the effect of the reference temperature, we calculated savings using a number of different degree day bases. Reanalyzing our data using a 60°F reference temperature (HDD_{60}) resulted in a decrease in average run-time meter savings from 22% to 18%, primarily as a result of higher estimates of consumption in the post-retrofit period relative to post-retrofit consumption calculated using HDD_{65}. Part of the post-retrofit period for some houses was marked by unseasonably mild weather, with daily average temperatures hovering around 50°F. If the post- period were in the middle of a typical Virginia winter, a change in the heating degree day base might not be expected to have such a strong influence on usage, but in milder weather the reference temperature appears to be quite important.

Given this evidence for the strong influence of the reference temperature on the resulting savings, we next tried the average ratio method with a "floating reference temperature." Average Btu/ft^2-HDD were calculated for each pre- and post-retrofit period using heating degrees base 55, 60, 65 and 70. For each home, the HDD base which yielded the smallest relative standard deviation of weekly Btu/ft^2-HDD was chosen as the best reference temperature. This method yielded both higher savings (25%) and a lower MAD (24% of absolute savings); that is, the average savings were higher than the PRISM estimates, but the typical discrepancies between PRISM and run-time meter savings for individual houses were smaller. While the above analyses made use of only four discrete reference temperatures, it would be optimal to consider all possible reference temperatures, finding the one which minimized the standard deviation through an iterative process similar to that utilized by PRISM, and calculating savings based on it.

Another Approach to Calculating Savings

Another approach to estimating run-time meter savings was tested in an effort to find a method which would give less weight to warm-weather periods. The average ratio method used above weights each weekly period's Btu/ft^2-HDD ratio equally, regardless of the severity of the weather during the period. However, since most heating energy consumption takes place during cold periods and Btus per degree day tend to be more stable in cold weather, it makes sense to use an averaging method which weights degree days equally, rather than weeks. A method called "Ratio of Sums" or "R-Sums" was devised, in which all the Btus consumed over the entire pre- or post-retrofit period are summed, and then divided by the sum of all the HDDs in the period. The average ratio and R-sums methods can both be classified as weighted least squares estimators of consumption without an intercept (i.e., baseload). The average ratio method is the optimal weighted least squares estimator if one assumes that the standard deviation of consumption is proportional to degree days, while the R-sums estimator is optimal if one assumes the standard deviation of consumption is proportional to the square root of degree days (which gives greater weight to colder periods). Using this R-Sums method with HDD_{65}, we found average savings to be the same as with the average ratio method with HDD_{65}, but the MAD is lower (30% of absolute savings). When using R-Sums with a floating reference temperature (HDD_{float}), the results (24% savings, MAD = 26% of absolute savings) were very close to the average ratio method using HDD_{float}.

Table 1 summarizes the savings calculated using all the various analysis methods discussed above. Average savings for these 13 houses range from 16% to 25%, with PRISM savings estimates at the bottom of the range. The results indicate that mild weather during the monitoring period can, depending on the analysis technique used, have a greater or lesser effect on the savings estimates. The use of a procedure to find the best reference temperature, and/or the R-Sums averaging method, result in less discrepancy between PRISM and run-time metering results for individual houses. In this particular evaluation study, the average HDD/day was fairly low (15 HDD_{65}/day) in the post-retrofit period. This problem is common to many evaluations using a split heating-season design, and is especially troublesome in mild climate states like Virginia (approximately 4300 HDD_{65} for the location of these homes). Another approach to this problem is to employ a screening procedure to eliminate periods with HDD/day lower than some cut-off criterion. This is further explored in the section below.

A Closer Look at the Variation of Run-Time Data With Weekly Weather

What Can We Learn from One "Well-Behaved" House in Indiana?

The analyses above suggest that data from weeks with few HDDs need to be used with caution. In this section we look at the data from a "well-behaved" house in Indiana to see if we can shed some more light on this issue. The house, located in Muncie, Indiana, is one of three abandoned houses rehabbed by a private developer under HUD's rental rehab program in 1987 and 1988. The senior author provided advice to the developer regarding energy conservation measures to install and has been monitoring the houses on a weekly basis in the years since. In January of 1990, run-time meters and event counters were installed on two of the houses. One of the houses is also sub-metered with a separate gas meter for the furnace; the data below are from that house.

This house is divided into two units (upstairs and down) and is rented to two low-income families under HUD's Section 8, with the developer/landlord paying all utilities.

Table 1. Comparison of PRISM and Short-Term Metering Results for 13 Houses in the Virginia Low-Income Weatherization Pilot Study (Consumption and Savings in MBtu/yr)

	Avg. Pre-Use	Avg. Post-Use	Avg. Abs. Savings	95% Conf. Intvl.	Avg. % Savings	MAD of Abs. Savings
PRISM						
NAHC	107	90	17	±16	16%	
Reference Temperature	64°F	63°F				
Short-Term Metering						
Avg. Ratio Method, HDD_{65}	109	84	24	±13	22%	9.6
Avg. Ratio Method, HDD_{60}	109	89	20	±13	18%	10.8
Avg. Ratio Method, HDD_{float}	111	83	28	±16	25%	6.6
R-Sums Method, HDD_{65}	106	82	23	±16	22%	6.9
R-Sums Method, HDD_{float}	109	83	26	±18	24%	6.9
HDD_{65}/day	25	17				

There is one thermostat which is fixed so that it can not be set above 72°F, and periodic visits and interviews with the occupants confirm that, as expected with the landlord paying the gas bill, the thermostat stays set at 72°F year 'round. This constant thermostat setting, together with the fact that the house is well insulated and has a furnace with no pilot light, leads us to expect it to be very "well-behaved" in terms of the relationship between energy consumption and outside temperature.

PRISM, Run-Time and Sub-Metering Compared

The furnace sub-meter allows us to examine not only how well PRISM and run-time data compare but how accurately each predicts actual space heat consumption. Table 2 shows annual space heat consumption for this house, as estimated by PRISM from monthly main gas meter data (which includes heating and hot water), and as measured by the furnace run-time meter and the gas sub-meter on the furnace. As expected, the PRISM analysis show the house to be very well-behaved, with an R^2 of .997, CV(NAC) of 1.5%, and CV(NAHC) of 2.8%. PRISM finds a reference temperature of 64.7 ± 1.2 °F, so using HDD_{65} to estimate annual consumption from run-time data should introduce little error. Actual metered gas for space heating was normalized by multiplying by the ratio of heating degree days in a "normal year" to those in this year, a minor correction since actual HDD_{65} for the period in question was only 4% less than the normal year as determined by PRISM (from HNORM, the file of normalized degree days).

The PRISM estimate of NAC is nearly "dead on," less than 0.3% from the metered consumption. The PRISM estimate of space heating, NAHC, is about 6% high, slightly higher than its estimated error, but in the direction predicted in the discussion of seasonality above.[4] This one data point, along with previous research on PRISM's heating component estimates (Fels et al. 1986) suggests that attempts to compare PRISM with run-time estimates of space heating will be plagued by PRISM's bias in estimating NAHC. Even in this well-mannered house, with relatively tight confidence limits on NAHC, the estimate is off by 6%.

The (non-normalized) estimate of annual space heat consumption as determined from the run-time meter is 128 MBtu, nearly 8% higher than the consumption as measured by the furnace sub-meter. An important question is whether this 8% difference is a result of a mis-calibration of the run-time meter (determined by measuring the firing rate) or whether it is a result of some other, perhaps non-linear error. Figure 2 shows a plot of weekly furnace consumption as measured by the run-time meter versus the sub-meter. As clearly shown, the relationship is very strong (R^2=0.999) and shows no evidence of change with temperature (inferred here from changes in MBtu/day) as would be expected if the error were a result of inaccuracies of the gas regulator at low temperatures, for example. It would appear that the error has to be in the firing rate, even though this was carefully measured on four separate occasions over as many months, with a standard deviation on the order of 1%.

The Effect of Weeks with Few HDDs

Figure 3, which plots weekly space heat energy consumption in MBtu/HDD and HDD/day for this house over time, demonstrates how the relationship between space

Table 2. Actual Metered Gas Consumption (in MBtu/yr) Compared to PRISM and Run-Time Estimates--for House in Indiana

	Total Gas	Heat Only	Hot Water	Heating Fraction
Gas Meters	148	119	29	80%
Normalized Gas Meters[1]	153	124	29	81%
PRISM	152 ±2	131 ±4	21 ±4	86%
Furnace Run-time Meter		128		
Normalized Run-time[1]		133		

[1] Space heat consumption was normalized by the ratio of HDD_{65} in a "normal year" to HDD_{65} in this year (5688/5462).

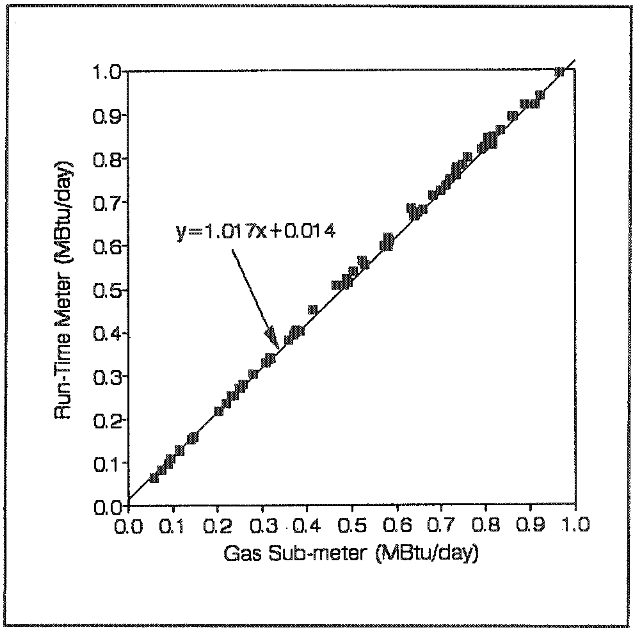

Figure 2. Average Weekly Furnace Consumption (in MBtu/day) as Determined by Run-Time Meter Versus Same Thing Obtained from Furnace Sub-Meter, Showing Excellent Correlation (R^2 = 0.999) Across Whole Range of Consumption (and Hence Temperature). Data for House in Indiana

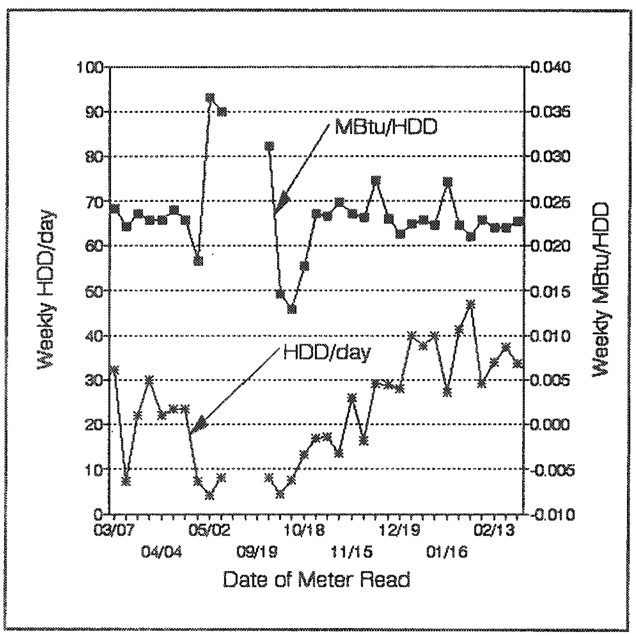

Figure 3. Weekly Heating Consumption (in MBtu/yr) and Weekly Weather Data (HDD/day) Over a One-Year Period, Showing How the Relationship Between Heating Consumption and HDD Breaks Down in Weeks with Few HDD. Data from House in Indiana

heat consumption and degree days deteriorates in the weeks with warmer weather. The values of MBtu/HDD (analogous to the heating slope, beta, in PRISM) stay within a narrow band centered on 0.022 MBtu/HDD for most of the period of record. The weeks in which it deviates most widely from this are those with few heating degree days (generally less than 10 or 12 HDD/day).

Figure 3 suggests that the reliability of the annual estimates determined from several weeks of run-time data might be significantly improved by screening out those weeks with HDD less than some minimum. Table 3 shows the results of applying different screening criteria to all weeks of usable data for this house. As one would expect, the standard deviation when looking at single weeks is quite large. Since typical field protocols for short-term metering call for a minimum of six weeks of data, we analyzed these data using six-week moving averages. The middle column shows the mean and range of annual estimates calculated using the average ratio method, while the column on the right does the same for the R-Sums method discussed above.

The first observation to draw from Table 3 is that, if the average ratio method is employed and data for warm weather is included, the estimates of annual consumption, and hence savings, could vary widely. For example, when using all weeks with HDD ≥ 1, the "right choice" of high and low six-week "pre-" and "post-" periods could result in 34% savings in this house in which no energy conservation measures were installed! The R-Sums method appears to tighten the range of estimates somewhat, as evidenced by the fact that the extremes are somewhat closer to the mean of these moving averages (highest and lowest 6-week means now yield only 20% savings).

If a minimum screen of 5 HDD/day is used, the standard deviation of the average ratio method decreases somewhat, but the R-Sums approach still appears to yield somewhat better results. When weeks with HDD/day of 10 or less are excluded, the precision of the estimates improves significantly. With this screen the R-Sums method appears to yield little advantage.

This analysis reinforces the findings from the analysis of the Virginia data, that there is a need for some sort of minimum HDD/day criterion to use in determining weeks to throw out of the analysis. The criterion suggested by this one-house data set would appear to be somewhere in the neighborhood of 10 to 15 HDD/day. While the data in Table 3 show no improvement in going from 10 to 15 HDD/day, it might be wise to err on the side of

Table 3. Annual Space Heat Consumption Estimates (in MBtu/yr) Using Various Screening Criteria and Different Calculation Methods for House in Indiana for Period 7/18/90 to 7/17/91[1]

	Ave. Ratio Method All Weeks	Ave. Ratio Method 6-Week Moving Avg.	R-Sum Method 6-Week Moving Avg.
Weeks with HDD/day > 1 (N=30)			
Mean	127 ± 39	120 ± 9	119 ± 5
High	279	148	132
Low	65	98	106
Weeks with HDD/day > 5 (N=28)			
Mean	123 ± 36	118 ± 7	118 ± 5
High	139	126	127
Low	65	98	106
Weeks with HDD/day > 10 (N=22)			
Mean	119 ± 11	120 ± 3	120 ± 4
High	139	126	127
Low	96	113	113
Weeks with HDD/day > 15 (N=20)			
Mean	121 ± 9	120 ± 3	120 ± 4
High	129	126	127
Low	99	113	113

[1] Estimates in this table are based on a corrected firing rate; i.e., the firing rate was adjusted to agree with actual annual use as determined by the furnace sub-meter.

caution until further research on a much larger number of houses can better determine an appropriate minimum cutoff value.

Summary and Conclusions

Short-term data from run-time meters on 13 houses in a Virginia Weatherization evaluation study, analyzed using the usual average ratio method, produced estimates of mean annual savings of 22% compared to 16% as determined by PRISM. Other computation methods applied to the same run-time data produced slightly different estimates of savings, but all were slightly higher than the PRISM estimates. The difference in savings estimates produced by the different computation methods suggests a need to bring some standardization to short-term metering approaches. The difference in savings estimates resulting from short-term methods versus PRISM, however, while not statistically significant, may nevertheless be an indicator of a more serious issue relating to the use of short-term data to measure annual savings. Suggestions for improving short-term methods are summarized immediately below. The thornier issue of short-term versus annual data is saved for the final section.

Influence of Warm Weather

The relationship between heating consumption and HDD can deteriorate in weeks with few heating degree days. Our examination of this relationship for one house suggests that 10 to 15 HDD/day might be a reasonable cutoff criterion. Data from these warmer weeks should be used with caution, perhaps only if absolutely necessary. If these weeks are used, consideration should be given to using a calculation method such as the R-Sum method discussed above, rather than the average ratio method typically employed.

Choice of Reference Temperature

The reference temperature used in determining HDDs can be extremely important, especially in looking at individual houses. The difference between base 65 and base 60 heating degree days can make a big difference in calculated savings, particularly if the weather is warm. PRISM, which finds the best reference temperature, has a clear and important advantage here. Calculation methods for run-time data which search for the best reference temperature, such as a modified version of PRISM or the method illustrated in this paper, are probably warranted.

Temperature Data

One of the purported advantages of short-term approaches is the ease of using temperature data from weather stations closer to the house being metered. Our research suggests that the tradeoffs between proximity and data quality need to be carefully weighed. Temperature data from local weather stations reported in local newspapers and then recorded by local agency personnel can be of questionable quality. The increased reliability of temperature data obtained from major weather stations may more than compensate for any loss of accuracy resulting from the greater distances involved, though this would be expected to depend on topography and local weather patterns, and more research is needed. The correct reference temperature is probably much more important than a close weather station; calculation methods which find a best reference temperature would be easier to manage with just a few major temperature files.

Control Groups

Good research design calls for the use of control groups in measured energy savings studies. Seasonality concerns--i.e., the possibility of non-random variations in houses' heating consumption per HDD over the year--suggest that control groups may be even more important when using short-term methods than when using annual data. Concerns regarding the obtrusiveness of the measurement technique--i.e., the possibility that the very act of reading a run-time meter on a weekly basis may result in some conservation behavior on the part of the occupant--provide an additional argument for the use of control groups.

Choice of Analysis Method

The ratio of sums (R-Sums) method utilized herein may be a reasonable alternative to the usual average ratio method for computing mean savings. It appeared to minimize problems caused by poor choice of reference temperature or warm weather. While its use may not be warranted if the data do not include weeks with too few HDD, its use doesn't hurt anything either. It might be useful for other researchers to apply this, and perhaps other computational methods, to larger data sets with the aim of eventually agreeing on a single method.

Needed: A Standardized Run-Time Data Software Package

One of the principal attributes of PRISM is its standardized approach. Not only does this enable evaluators to easily compare results, but it has also resulted in a broad base of experience which is very useful in understanding common pitfalls and unusual situations. If short-term methods are to be widely utilized in energy savings evaluations there is a need to standardize the methodological approach. There is no reason why the statistical rigor embodied in PRISM could not also be brought to bear on short-term methods.

Extrapolating Annual Savings from Short-Term Data

Bringing increased rigor to short-term methods may not solve the problem inherent in trying to measure annual savings using two six-week periods. While short-term methods definitely have their place, some caution is in order when data from short-term methods are extrapolated to annual savings. While the results in this paper are based on far too small a sample (N=13) to make any generalizations, the tendency of our short-term data to yield slightly higher savings than PRISM may be cause for concern. Before short-term methods are adopted on a wider scale, researchers need to more carefully examine the seasonality issue to be sure that the savings are not being biased by a research design in which the pre-retrofit period routinely occurs in the fall or early winter and the post-retrofit period occurs in late winter or spring. A first step is to use control houses in all short-term evaluation studies. This would not only improve the reliability of individual studies, but data collected from all these control groups might also be used to improve our understanding of this seasonality issue.

Acknowledgments

We would like to extend our appreciation to the Virginia Weatherization Program and to its local agencies for supporting this evaluation study. We also owe a big debt to the anonymous reviewers of the draft of this paper, whose comments and questions forced careful examination of a number of issues.

Endnotes

1. For example, certain electronic setback thermostats are capable of storing data on furnace run-time (more precisely, thermostat "on time") for various time periods, one of which is the previous week. Thus, for the price of a thermostat, a reimbursable material expense under weatherization, the house is also fitted with a meter to use in measured energy savings evaluations.

2. Water heater insulation was the only measure installed in the pilot project which would be expected to influence baseload usage. Savings from this measure would not be captured by run-time meters. The effect on whole house savings measured by PRISM is assumed to be minor.

3. In 8 of these homes, gas was reported as being used for space heating plus other end uses (typically water heating and/or cooking). Several of the other 5 homes, although reporting gas being used only for space heating, showed steady gas consumption during the summer months. Therefore, PRISM's estimate of the space heat fraction (NAHC) was used to approximate space heating in all 13 homes.

4. A quick analysis of weekly hot water use for this house demonstrates the pronounced seasonality of energy consumption for water heating. Energy consumption increases and water heater efficiency decreases in the colder winter months due to colder incoming water temperatures and greater losses.

References

Blasnik, Michael. 1989. "Attrition Bias in Fuel Savings Evaluations of Low-Income Energy Conservation Programs." *Energy Program Evaluation: Conservation and Resource Management, Proceedings of the Fourth International Conference in Chicago*, pp.211-217.

Blasnik, Michael and Tom Lent. 1988. "Quick Savings Estimates with GSTEAM: The GRASP Short Term Energy Analysis Method." *Proceedings of the 1988 Summer Study on Energy Efficiency in Buildings*, Volume 10, pp.10.42-10.45. American Council for an Energy-Efficient Economy, Washington, D.C.

Dunsworth, Timothy S. and Martha J. Hewett. 1985. *Data Quality Considerations in Using the PRISM Program.* Report MEO/TR85-4-AM, Minneapolis Energy Office.

Fels, Margaret F. 1986. "PRISM: An Introduction." *Energy and Buildings*, Volume 9, Numbers 1 and 2, pp. 5-18.

Fels, Margaret F., James Rachlin and Robert H. Socolow. 1986. "Seasonality of Non-heating Consumption and Its Effect on PRISM Results." *Energy and Buildings*, Volume 9, Numbers 1 and 2, pp. 139-148.

Greely, Kathleen M., John Randolph and William W. Hill. 1991. "Improving Virginia's Low-Income Weatherization Program: Evaluating the Performance of New Techniques." *Energy Program Evaluation: Uses, Methods and Results--Proceedings of the Fifth International Conference in Chicago*, pp. 425-433.

Hirst, Eric and Richard Goeltz. 1986. *Electricity Use for Residential Space Heating: Comparison of the Princeton Scorekeeping Method with End-Use Load Data.* ORNL/CON #203. Oak Ridge National Laboratory, Oak Ridge, TN.

Nadel, Steven. 1987. "Timely Evaluations: A Discussion of Methods Used by the Massachusetts Audubon Society to Evaluate Fuel Savings in Under One Year." *Energy Conservation Program Evaluation: Practical Methods, Useful Results, Proceedings of the 1987 Chicago Conference*, Volume 2, pp. 73-83.

Pigg, Scott. 1992. "Short-Term Estimates of Energy Savings from Billing Data: What Early Results Can Tell You." (Presentation at Affordable Comfort VI Conference, Pittsburgh, PA). Wisconsin Energy Conservation Corporation, Madison, WI.

Randolph, John, Kathleen M. Greely and William W. Hill. 1991. *Evaluation of the Virginia Weatherization Program.* Virginia Center for Coal and Energy Research, Blacksburg, VA.