PERFORMANCE OF THE PRINCETON SCOREKEEPING METHOD FOR ESTIMATING ANNUAL END-USE ENERGY SHARES AS MEASURED WITH METERED END-USE LOAD DATA

Erik D. Westman
Bonneville Power Administration

William H. Monroe
Electronic Data Systems

ABSTRACT

The Princeton Scorekeeping Method (PRISM) is a widely used technique for evaluating the performance of residential energy conservation programs. Estimated energy savings are computed from PRISM estimates of space heating and baseload energy consumption for pre and post program periods. The accuracy of estimated conservation savings is therefore dependent on the accuracy with which PRISM estimates space heating and baseload energy use. This paper reports the results of an evaluation of PRISM's performance at estimating space heating and baseload energy consumption in 40 Pacific Northwest residential households located in Albany and Bend, Oregon. Metered end-use data for an entire year, beginning in April 1983, was used to represent utility billing data in regression models of the specification found in PRISM. Regressions were estimated for each of the 40 households using metered total energy data and local NOAA weather station data. The results of these linear regressions were then compared with the metered end-use data for space heating and baseload end uses. Comparisons were made using a variety of statistical measures. Correlation coefficients for percent errors in PRISM end-use estimates and model inputs and parameters were investigated. Regression residuals were analyzed for behavior of error components, and survey-based characteristics data were used to explore for differences in PRISM's performance distinguishable through dwelling, occupant, and geographic characteristics. Finally, a discussion of the implications of these findings was made with respect to the current uses of PRISM for energy conservation assessment and evaluation.

PERFORMANCE OF THE PRINCETON SCOREKEEPING METHOD FOR ESTIMATING ANNUAL END-USE ENERGY SHARES AS MEASURED WITH METERED END-USE LOAD DATA

Erik D. Westman, Bonneville Power Administration
William H. Monroe, Electronic Data Systems, Inc.

INTRODUCTION

Since the early work by Schrader (1978) and Goldberg (1982), numerous studies of savings resulting from energy conservation programs have applied a simple linear regression model for "normalizing" utility billing records of residential energy usage with respect to weather (Dunsworth et al., 1984; Fels and Goldberg, 1984; Hirst et al., 1984; Hirst, 1986). One such model, developed by Princeton University's Center for Energy and Environmental Studies, is referred to as the Princeton Scorekeeping Method, or PRISM. This model not only derives normalized annual energy consumption, but also predicts the temperature-independent and temperature-dependent end-use shares of annual household energy consumption. These two components are interpreted as reflecting: (1) the baseload energy consumption, which is associated with appliance and lighting uses, and (2) space heating energy consumption, which is the result of dwelling thermal performance, heating system efficiency, and outdoor temperature (Fels, 1986).

Mean differences in normalized annual space heating energy consumption as estimated using the PRISM model are often used to measure the savings due to energy conservation programs. These estimated savings then are used as a measure of program performance. The reliability of this measure is dependent, however, on how accurately PRISM estimates space heating consumption. This is particularly true when the conservation measures are aimed at improving the thermal performance of residential dwellings. A problem arises when the model introduces a systematic error that affects the estimates of end-use energy consumption of program participants differentially. While PRISM accurately yields weather-normalized annual total energy consumption, its estimate of end-use shares is reportedly less robust, and in fact has been found to depart systematically from metered end-use data (Fels, Rachlin, and Socolow, 1985; Hirst and Goeltz, 1986). The present study reports differences in PRISM end-use estimations compared with metered end-use data for two distinct climate zones in the Pacific Northwest and further, describes indications of systematic behavior in PRISM end-use estimation errors characterized as percent errors of annual metered end-use energy consumption.

THE PRISM REGRESSION MODEL

The strong relationship between weather and residential energy consumption is modeled by the Princeton Scorekeeping Method as a simple "change-point" regression of energy use as a function of outdoor temperature, with two regimes, and is described in detail by Goldberg (1982). In a simplistic sense, the two regimes represent a piece-wise function in which one segment has zero slope for the range of outdoor temperatures over which total energy use is dominated by factors other than outdoor temperature. The second piece of the function lies along the range of outdoor temperatures requiring energy above solar and internal gains to maintain an indoor temperature. The point along the outdoor temperature scale at which the two segments intersect is often referred to as the structure's "balance point".

PRISM as applied on an individual household basis, and following the notation of Hirst and Goeltz (1986), has the following model specification:

$$E_{ij} = a_j + b_j HDD_{ij}(Tref) \qquad (1)$$

where, i denotes the ith monthly period, j denotes the jth house, E_{ij} is the average daily energy use by house j during month i, a_j represents the average daily non-temperature dependent energy use, b_j is the rate of temperature-dependent energy usage , and HDD_{ij} denotes the mean daily heating degree days computed for month i and house j. Daily heating degree days (HDD) is computed as:

$$HDD = \max(Tref - Tout, 0) \qquad (2)$$

where, Tref is a selected reference temperature and Tout is the daily average outdoor air temperature for each day. Daily average outdoor air temperature is obtained by locating the midpoint outdoor air temperature between the maximum and minimum temperatures occurring each day. Negative HDD values are always set to zero when using PRISM. The PRISM model is estimated with ordinary least squares at varying levels of Tref in search of each dwelling's balance point as determined by the R-squared best-fit criterion.

Model estimates of annual total load (E_j), baseload (a_j365), and space heating load ($b_jHDD(Tref)$) are derived from the following equation using total annual heating degree days as the predictor variable HDD.

$$E_j = a_j 365 + b_j HDD(Tref) \qquad (3)$$

This equation can be used to compute a normalized annual energy consumption estimate through the use of long-term average temperature data. For the purposes of this study, our principal interest is in PRISM's estimated baseload and space heating end-use energy shares.

A variety of assumptions are inherent in the PRISM model and also in application of the model for conservation program evaluation. The model assumes a linear relationship between space heating consumption and heating degree days. Ideally, non-space heating end uses are assumed independent of heating degree days, or outdoor temperature. Therefore, the regressions on total and space heat metered load data should yield nearly equivalent slope coefficients if the model assumptions are met. As applied to conservation program evaluation, it is assumed that the intercept coefficient is a good approximation of mean daily baseload consumption and that the annual temperature-dependent (correlated) load, computed from the regression slope coefficient, is likewise effective at describing the annual space heating requirements of each household.

METERED END-USE DATA

The data used for this preliminary work on evaluating the performance of the PRISM technique was provided by Pacific Power & Light (PP&L), a large Pacific Northwest electric utility. The data were collected by PP&L for an end-use metering study conducted among their customers located in Albany and Bend, Oregon, during a one year period beginning in April 1983.

The population targeted by PP&L's study was single family, owner occupied households with electric space and water heat. Metered end-use data was collected at 15 minute recording intervals for each of the study's randomly selected households, including total load, space heat load, and water heat load circuits. Because the period of metering was not entirely uniform for all cases, an annual period with the most complete record of end-use data was selected; this was found to be May 1983 through April 1984.

Corresponding hourly temperature data was obtained from the National Oceanographic and Atmospheric Administration (NOAA) weather station sites of Redmond and Salem, Oregon. These sites were used for Bend and Albany, respectively.

The end-use data was aggregated to the monthly level, cases which had missing data or negative kWh end-use load observations were dropped from the analysis. As a result, there were 40 cases retained for the present study, 21 from Bend and 19 from Albany.

PRISM RESULTS COMPARED WITH METERED END-USE RESULTS

Once the 40 households were selected, mean daily total energy consumption was computed for each calendar month. Mean daily heating-degree days (HDD) per month were computed using Equation 2 for temperature references of 40 to 65 degrees Fahrenheit using one degree intervals (reference temperatures in five degree intervals were used outside the 40 to 65 degree Fahrenheit range).

The regression of mean daily total energy (kWh) use per month on mean daily heating degree days per month was used to construct the energy demand model for each household. Regression results and diagnostic statistics were obtained using a standard statistical analysis software package as opposed to the PRISM software. The estimations of total, space heating, and baseload annual consumption derived from the PRISM model were computed from Equation 3, and were subsequently compared with the annual and monthly end-use metered results. Regression residuals were inspected along with model coefficients and diagnostic statistics in order to assess the performance of each household regression.

The results of the PRISM regressions, summarized in Table I, reveal that individual household regressions generally had very high explanatory power with a mean adjusted R-square of .93 (R-square adjusted for degrees of freedom). Only three of the households had adjusted R-squares below .90. The minimum adjusted R-square was .70. Virtually all of the regression coefficients were significant at $p = 0.001$. The mean PRISM reference temperature for all 40 households was 56.7 degrees Fahrenheit.

The mean of PRISM estimates of annual space heat energy consumption for the 40 selected households was 11,074 kWh (see Table II), whereas the mean for metered space heat energy consumption was 10,220 kWh. The mean of absolute differences between PRISM estimated and metered end-use energy consumption was 854 kWh. Table II also shows the mean of PRISM baseload estimates at 11,376 kWh while the mean of metered baseloads was 12,226 kwh. The mean of absolute differences between PRISM and metered estimates of baseload was -850 kWh.

After an initial comparison of PRISM estimates and metered end-use data we felt that descriptive statistics of kWh estimates provided a less effective portrayal of the size and distribution of PRISM estimation errors across households. Therefore, we computed a statistic that we refer to as percent error which characterizes the estimation error of PRISM as a percent of metered annual energy consumption. Percent error is computed for the PRISM estimates per household using metered end-use data as the basis. Percent error for baseload and space heating were computed as follows:

$$\text{percent error} = ((E_p/E_m)-1)100 \tag{4}$$

where, E_p and E_m refer to the PRISM estimated and metered annual end-use energy consumption for each household, respectively.

In 75% of the households, PRISM overestimated annual space heating consumption. PRISM space heating overestimates ranged from 1% to 226% while underestimates ranged from -1% to -51%. In addition, 60% of the households had percent errors less than 10% in either direction. The distribution of percent errors of PRISM space heating estimates posesses an extreme tail including two scores that overestimate annual space heat kWh consumption by 192% and 226%, respectively. These scores are approximately twice as large as the next highest overestimate at 105%, and about eight times larger than the third quartile score 27%.

Removing the two cases that had the highest percent error scores resulted in a significant change in the means of metered and PRISM estimated space heating energy use. The mean of metered estimates was then 10,596 kWh and the mean of PRISM estimates became 11,157 kWh. If the same two cases were removed as outliers the mean percent error of PRISM space heating estimates changed from 20.9% to 10.9%. Not surprisingly, the two outlier candidates represent a large proportion (48%) of the mean space heating percent error, however, their individual PRISM model fits are still quite good (adjusted R-squares of .96 and .81). In fact, detailed inspection of these cases revealed no apparent cause for their exceptional percent errors of predicted space heating consumption.

The distribution of percent errors of PRISM estimates of baseload energy consumption does not have as large a range or extreme tail as was the case of space heating percent errors. Percent error scores are clustered near the median score of -8.8% with a minimum score of -50.8% and a maximum of 74.9%. There do not appear to be households which can be typified as outliers as was the case with space heating percent error. Trimming extreme scores provided no significant improvement in the mean of baseload percent errors.

Further investigation of PRISM end-use estimation errors revealed an interesting systematic pattern. The percent errors of PRISM estimates of space heating varied systematically with metered annual space heating load and also with the ratio of metered space heating load to metered annual total load (see Figures 1 and 2). Further, the percent errors in estimated baseload appear to have a similar bias, though opposite in sign, in which the percent errors tend to become consistently negative as the baseload energy share increases (see Figure 3). We also noted that the percent errors for PRISM estimates of baseload energy consumption increased, particularly in the positive direction, as metered baseload share decreased. Our results suggest that percent errors in PRISM end-use estimates are proportionately related. This relationship can be expressed as:

$$e_{sh} = -k\ e_{bl} \tag{5}$$

where e_{sh} is either the proportionate (percent) or absolute (kWh) error attributed to space heating and e_{bl} is either the proportionate or absolute baseload error. The constant k is approximately 1.0 for absolute error and is the ratio of metered space heating load to metered baseload consumption for the percent error in PRISM end-use estimates when PRISM closely approximates metered annual total load.

We also note in Figure 2, that the percent errors in PRISM estimates of space heating energy consumption increased significantly as metered baseload consumption took on a more dominant role in the energy consumption pattern of particular households. We believe this to be a function of PRISM's inability to distinguish between seasonally, or temperature, correlated baseload and actual space heating energy use. One cannot, however, infer from total energy use, the likelihood or magnitude of absolute errors (kWh) or percent errors in PRISM estimates of baseload or space heating energy consumption (see Figure 4).

In order to investigate the role of seasonally correlated baseload energy use in PRISM end-use estimation errors, we constructed PRISM regression models in which the metered end-use channel data for space heating, baseload, water heating load, and residual load were used as dependent variables in separate regression models. We distinguish baseload from residual load energy used for purposes other than space heating and water heating. Residual load was computed as the difference of total load and space heating plus water heating loads. Baseload, on the other hand, was computed as the sum of all metered energy use except space heating. These end-use channel models were of the form:

$$E_{ijk} = a_{jk} + b_{jk} HDD_{ij}(Tref_j) \tag{6}$$

where, k refers to the kth end-use channel and the other elements of the equation generally follow Equation 1 described earlier with the exception that $Tref_j$ is obtained from the regression on total load. Each end-use channel was used in this manner to provide insight into the assumptions of the PRISM model regarding linearity and temperature dependence. We reasoned that the PRISM model coefficients can be thought of as linear combinations of temperature-dependent and temperature-independent components of household energy end uses when holding Tref constant for each dwelling. Therefore, the heating slope coefficient is actually the sum of temperature-dependent and seasonally collinear effects within each end use. This implies that:

$$b_j = \sum_{k=1}^{n} b_{jk} \tag{7}$$

where, k represents n end uses and b_{jk} is the temperature-dependent or seasonally collinear slope for each end use within a particular household. The intercept coefficient can be interpreted similarly. This hypothesis is similar to Hirst and Goeltz's (1986) concept of total annual energy use as a linear combination of end uses.

When the PRISM model was applied to the space heating channel data, the results were comparable, in most cases, to those derived using total load channel data. The mean adjusted R-squared was nearly the same (.94) and 38 of the 40 households had estimated models with goodness-of-fit F tests significant at the $p = 0.005$ level. When the model was applied to the water heating, residual load, and baseload channel data, the explanatory power of heating degree days diminished, though not as much as one might have expected. In fact, 25% of the households obtained adjusted R-square results above .70, and temperature-dependent slope coefficients were significant at $p = 0.05$ when the PRISM model was applied to residual load end-use channel data. These same households were associated with heating slope coefficients of baseload energy use greater than 0.5, which suggests overestimation of space heating by PRISM as a result of seasonally collinear components of non-space heating end uses. Similar results were obtained for the water heating and baseload end uses.

We found very high correlation (-.78) between PRISM space heating percent errors and the ratio of heating slopes obtained from regressions on space heating and total load channel data. In particular, households which have very large space heating end-use channel contributions to PRISM's estimated heating slope are strongly associated with low space heating percent errors. We are encouraged by this result since PRISM is theoretically based on a strong association between outdoor temperature and space heating energy demand (Schrader, 1978; Goldberg, 1982).

In a similar fashion,the ratio of heating slopes obtained from the baseload and total channel regressions was used to measure the contribution of baseload to PRISM's predicted heating slope. This measure was strongly correlated (.65) with space heating percent error. This association suggests that PRISM systematically fails to differentiate between seasonally correlated baseload and space heating end-uses.

Having identified a systematic pattern in PRISM end-use estimation errors, we sought to identify means of detecting cases which might be expected to yield biased estimates of end-use shares when PRISM was applied to monthly billing record data. Our efforts concentrated on investigating correlations between PRISM model parameters and PRISM end-use percent error estimates; and also on analyzing, with exploratory analysis techniques, dwelling and occupant characteristics.

Adjusted R-squared was not strongly correlated with the accuracy of PRISM model estimates of end-use energy consumption (Table III). Since most adjusted R-square scores are quite high (over 90% were above .80), we are not able to infer potential for error among PRISM predictions of end-use shares from the R-squared criterion for goodness-of-fit. It appears that the adjusted R-squared statistics do not provide much additional information regarding the performance of the model in predicting annual space heat consumption. Reference temperature, on the other hand, exhibited a higher degree of correlation with the percent error of baseload estimations, though not for space heating percent error. Underestimations of baseload occurred only for individual household PRISM models with reference temperatures above 50 degrees Fahrenheit, whereas all cases were overestimated when the temperature reference was below 50 degrees.

Coded scatterplots and multiple box and whisker plots were used to investigate the interaction of demographic and household structural characteristics with space heating and baseload percent errors. Geographic location, or climate zone, did show significant differences in the variation of PRISM space heating and baseload percent errors. Households located in Bend (6,891 annual HDD) showed larger variation in the percent error distribution than was the case for households from Albany (4,831 annual HDD). This appears consistent with, and is perhaps related to, the differences in daily mean temperature distributions for these two locations.

Inconclusive results were obtained from most of the exploratory analyses. Characteristics such as size of dwelling, heating system type, presence of supplemental wood heat or air conditioning, numbers of major appliances , and number of occupants failed to reveal systematic differences in space heating or baseload percent error. We believe that such results are likely to be the outcome of our small sample size and that factors contributing strongly to the proportion of temperature-dependant to temperature-independant energy usage will be revealed in end-use metering studies with larger sample sizes.

SUMMARY AND CONCLUSIONS

The Princeton Scorekeeping Method is used extensively to evaluate the performance of residential energy conservation programs. It relies on utility billing data and NOAA weather data to: (1) weather normalize energy consumption, (2) disaggregate baseload and space heating components of energy use, and (3) provide a means of normalizing billing periods. The efficacy of this technique relies heavily on the independence of space heating and baseload energy consumption.

This study focuses on a systematic error in PRISM estimates of end-use energy consumption. We found that percent errors in space heating and baseload PRISM estimates were proportionately related and opposite in sign; and that the coefficient of proportion was equal to the ratio of metered space heating load to metered baseload energy use. In addition, our findings suggested that the space heating end-use share was overestimated by PRISM and that the degree of overestimation was directly associated with the proportion of total energy consumption attributable to baseload, or non-temperature dependent, energy use. This association can be explained by the fact that baseload consumption fluctuates seasonally as has been previously found (Fels, Rachlin, and Socolow; 1985), and consequently, is correlated with the heating degree day regressor which contributes to the estimation of space heating energy consumption.

These results suggest that PRISM has conditional applicability for use in estimating end-use shares and thus for energy conservation program evaluation. PRISM appears to be most accurate when applied to households with at least 50% of their energy consumption attributed to space heating. If the actual end-use shares of existing and new housing stock shift toward non-space heating energy uses, PRISM may perform less accurately in measuring energy conservation program savings. PRISM remains a very useful technique. We merely recommend caution in applying PRISM to households that are characterized as having annual energy consumption dominated by baseload end uses.

Clearly, additional investigation of alternative model specifications and estimation techniques are warranted. Model specifications allowing estimation of non-heating season slope and intercept coefficients in addition to heating season coefficients may provide improved estimates. Further efforts are also needed to identify those end uses that contain large temperature correlated, or seasonally correlated, components of energy use that result in significant PRISM end-use estimation errors.

REFERENCES

Dunsworth, T.S., T.A. Miller, and M.J. Hewett. 1984. Empirical Guidelines for Using the Princeton Scorekeeping Model. Minneapolis Energy Coordination Office, Technical Report No. 84-7 (draft), Minneapolis, MN.

Fels, M.F. 1985. The Princeton Scorekeeping Method: An Introduction. Princeton University, Center for Energy and Environmental Studies, Report No. 163, Princeton, NJ.

Fels, M.F. and M.L. Goldberg. 1984. Using billing and weather data to separate thermostat from insulation effects. Energy 9(5):439-445.

Fels, M.F. and M.L. Goldberg. 1986. Using the scorekeeping approach to monitor aggregate energy conservation. Energy and Buildings 9(2).

Fels, M.F., J. Rachlin, and R.H. Socolow. 1985. Seasonality of non-heating consumption and its effect on PRISM results. Energy and Buildings 9(2).

Goeltz, R., E. Hirst, and J. Trumble. 1986. Electricity Savings One, Two, and Three Years After Participation in the BPA Residential Weatherization Pilot Program. Oak Ridge National Laboratory, Report No. ORNL/CON-194, Oak Ridge, TN.

Goldberg, M.L. 1982. A Geometrical Approach to Nondifferentiable Regression Models as Related to Methods for Assessing Residential Energy Conservation. Princeton University, Center for Energy and Environmental Studies, Report No. 142, Princeton, NJ.

Goldberg, M.L. and M.F. Fels. 1986. Refraction of PRISM results into components of saved energy. Energy and Buildings 9(2).

Hirst, E. 1986. Actual energy savings after retrofit: electrically heated homes in the Pacific Northwest. Energy 11(3):299-308.

Hirst, E. and R. Goeltz. 1986. Electricity Use for Residential Space Heating: Comparison of the Princeton Scorekeeping Method With End-Use Load Data. Oak Ridge National Laboratory, Report No. ORNL/CON-203, Oak Ridge, TN.

Hirst, E., R. Goeltz, and D. White. 1984. Use of Electricity Billing Data to Determine Household Energy Use "Fingerprints." Oak Ridge National Laboratory, Report No. ORNL/CON-164, Oak Ridge, TN.

Rachlin, J. 1984. "Finding Signitures of Physical Effects on Residential Energy Consumption Data." MS Thesis for The Department of Civil Engineering, Princeton University, Princeton, NJ.

Rachlin, J., M.F. Fels, and R.H. Socolow. 1985. The stability of PRISM estimates. Energy and Buildings 9(2).

Schrader, T.F. 1978. A Two-Parameter Model for Assessing the Determinants of Residential Space Heating. Princeton University, Center for Energy and Environmental Studies, Report No. 69, Princeton, NJ.

Table I. PRISM Model Regression Coefficients and Statistics

	R-Square	Adjusted R-Square	Baseload 1/ Intercept (kWh/day)	Heating 2/ Slope (kWh/HDD)	Temperature Reference (Deg. F)
Mean	.939	.932	31.52	3.74	56.73
Median	.959	.955	28.58	3.44	57.50
Std Dev	.064	.071	16.38	2.01	4.36
Max	.984	.983	88.67	11.70	65.00
Min	.704	.674	7.12	1.52	42.00

1/ 30 cases significant at p = 0.001
2/ 40 cases significant at p = 0.0001

TABLE II. Comparison of PRISM end-use estimates with metered end-use energy consumption from 40 households located in Albany and Bend, Oregon.

Annual Space Heating Energy Consumption

	Metered (kWh)	PRISM (kWh)	Absolute Difference (kWh)	Percent 1/ Error
Mean	10,220	11,074	854	20.9
Median	9,303	9,630	906	8.0
Std Dev	4,343	4,289	2,967	54.5
Max	20,127	22,748	6,672	226.3
Min	2,949	4,809	-6,373	-51.0

Annual Baseload Energy Consumption

	Metered (kWh)	PRISM (kWh)	Absolute Difference (kWh)	Percent 1/ Error
Mean	12,226	11,376	-850	-5.0
Median	11,635	10,416	-893	-8.8
Std Dev	5,939	5,940	2,962	26.4
Max	32,843	31,547	6,382	74.9
Min	4,496	2,599	-6,685	-50.8

1/ Percent error is computed for each household using Equation 4.

Table III. Pearson's correlation coefficients for selected variables associated with percent errors in PRISM estimates of end-use energy consumption.

Selected Variable	Baseload Percent Error	Space Heating Percent Error
Metered total load (kWh/yr)	-0.001	0.063
Metered baseload (kWh/yr)	-0.158	0.335
Metered space-heating load (kWh/yr)	0.290	-0.553
Maximum PRISM R-square	-0.476	0.105
Tref at maximum R-square regression (Deg. F)	-0.745	0.340
Ratio of heating slope coefficients obtained from space heating and total energy regressions	0.295	-0.782
Ratio of intercept coefficients obtained from baseload and total energy regressions	-0.290	0.645

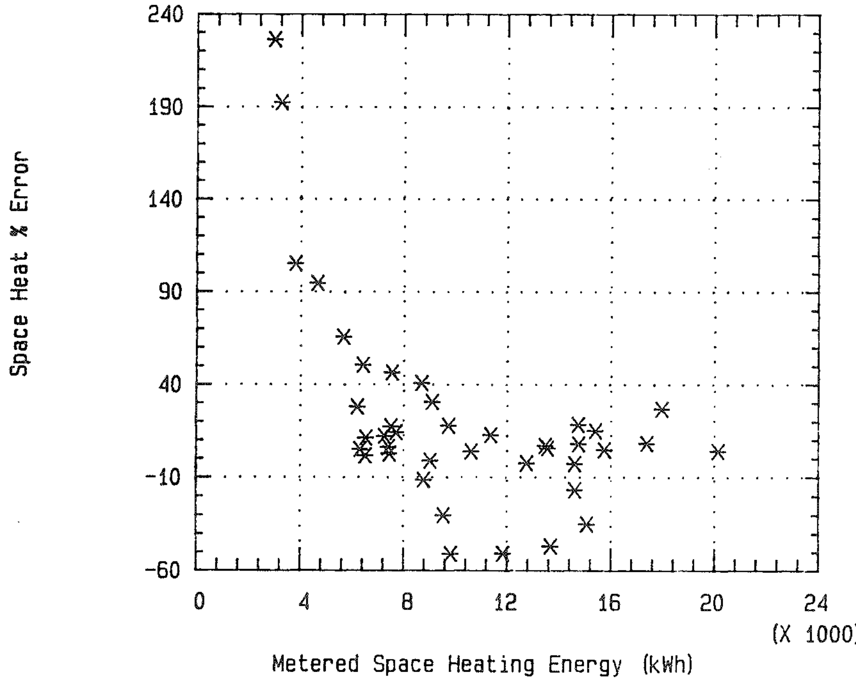

Figure 1. Percent error of PRISM estimates of space heating as a function of annual metered space heat energy use.

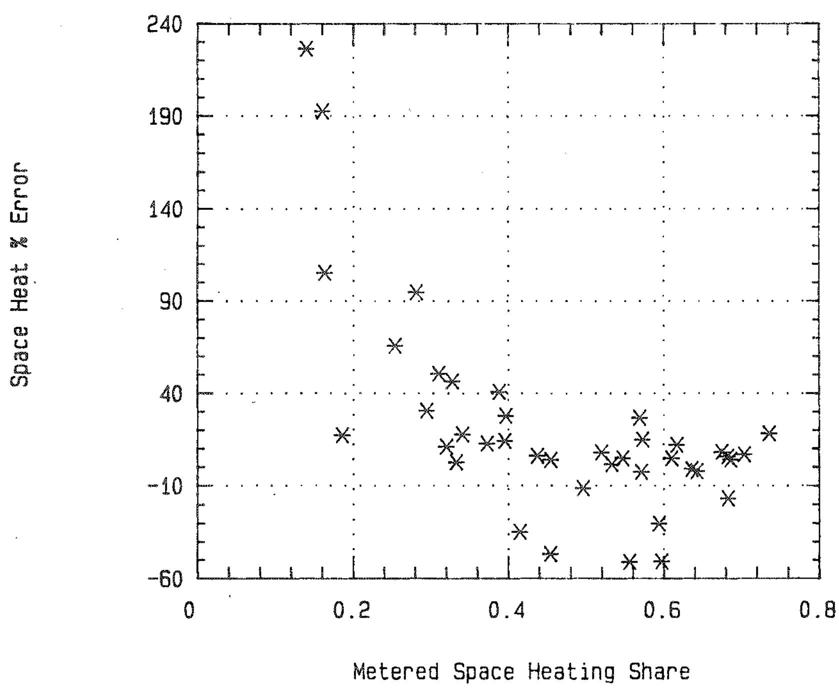

Figure 2. Percent error of PRISM estimates of space heating as a function of metered space heating share.

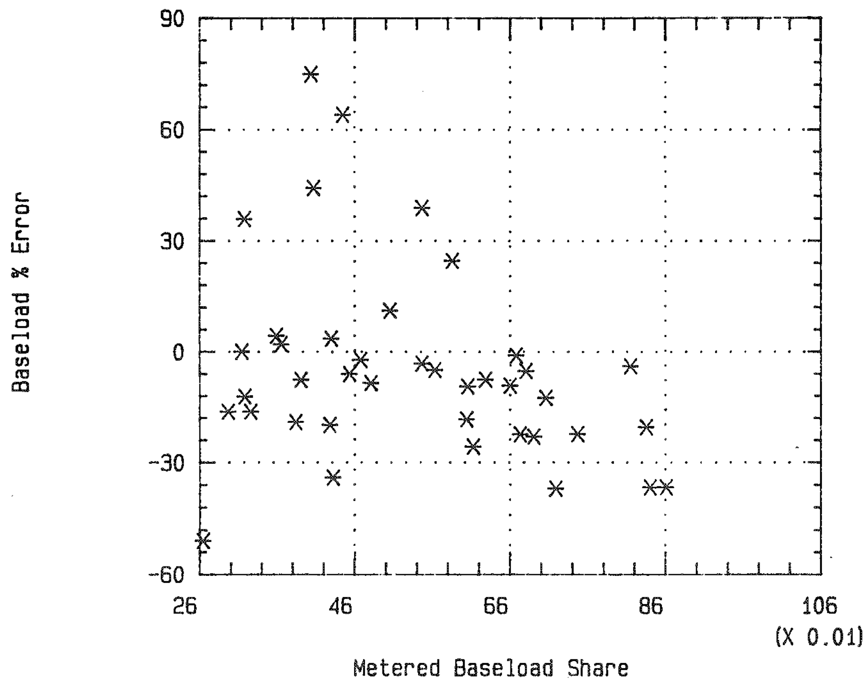

Figure 3. Percent error of PRISM estimates of baseload as a function of metered baseload share.

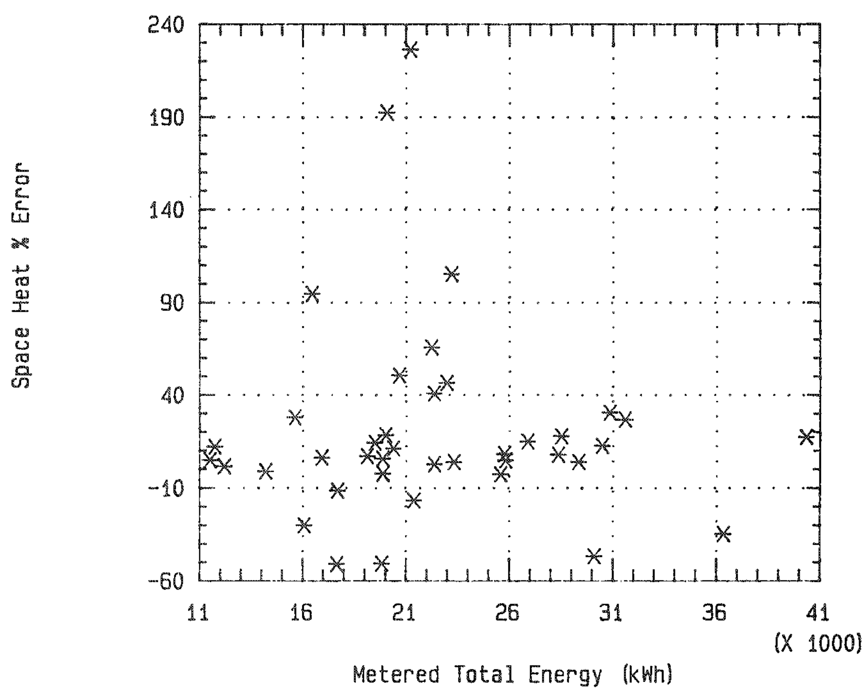

Figure 4. Percent error of PRISM estimates of space heating as a function of annual metered total energy use.