MULTIVARIATE STATISTICAL ASSESSMENT OF METEOROLOGICAL INFLUENCES ON RESIDENTIAL SPACE HEATING

Donald L. Hadley and Stan D. Tomich
Pacific Northwest Laboratory

ABSTRACT

The relationship between residential space heating in four residences in the Pacific Northwest and the external meteorological determinants of the space heating variability were investigated using the statistical techniques of principal component analysis (PCA) and multivariate regression analysis. The PCA identified 3 significant, uncorrelated components that were linear combinations of the 6 original meteorological variables. The first component, dominated by the outdoor temperature, accounted for between 65% and 69% of the variability in the original meteorological data and explained between 55% and 85% of the variability in the daily total space heating consumption. The first three principal components combined to explain over 95% of the variability in the data and up to 90% of the variability in the space heating.

MULTIVARIATE STATISTICAL ASSESSMENT OF METEOROLOGICAL INFLUENCES ON RESIDENTIAL SPACE HEATING

Donald L. Hadley and Stan D. Tomich
Pacific Northwest Laboratory

1.0 INTRODUCTION

Applications of statistics to the analysis of residential electrical space heat consumption tend to focus on outdoor air temperature as the single most "important variable" which "best" predicts the space heat consumption variability. Other meteorological factors such as insolation and wind speed have been relegated to positions of secondary importance, however, outdoor temperature variations are not independent of variations in these other meteorological elements. The intercorrelated nature of the meteorological variables results from the fact that many of the routinely measured meteorological variables are actually only different measurements of the same fundamental governing controls. Given this association, it has been difficult to fully account for the influence of the variations in meteorological conditions on the variability in space heat consumption.

Serious uncertainties can result from the instability of regression equations caused by the lack of truly uncorrelated independent variables. One approach to overcoming this difficulty is to reduce the number of independent variables in order to increase the stability of the regression coefficients; as, for example, determining space heat consumption as a function of a single variable such as the indoor/outdoor temperature difference. Unfortunately, such an approach cannot take into account the combined effect of the weather conditions on daily space heating consumption.

On the other hand, principal component analysis (PCA) applied to the correlated meteorological variables produces a set of uncorrelated independent variables that can be used in subsequent regression analysis. Thus we are able to account for the variation in the daily space heat consumption attributed to the variability in the original independent, but intercorrelated, data; not just part of it.

The technique of principal component analysis followed by multivariate regression analysis has been widely used to investigate the meteorological influences on ambient air quality (Lioy et. al 1982; Henry and Hidy 1979; Henry and Hidy 1982; Wolff et. al 1984; Thurston and Spengler 1985) and in other meteorological studies (Clarke and Peterson 1973; Hardy and Walton, 1978). These studies used daily values of the principal components as the independent variables in the regression analysis. By identifying which combinations of variables were significant determinants of the variability in the dependent variable, the results of the analysis provided additional insight into the relationships of point measurements of ambient air quality and the underlying atmospheric/chemical processes.

It is the intent of this paper to report on the initial results of a statistical analysis approach that is unaffected by the intercorrelated nature of the meteorological data and which is more fully considers the influence of the weather on residential electrical space heat consumption. Principal component analysis was applied to the original meteorological variables to derive the underlying weather regime components which were subsequently used in the regression analysis. This initial study was restricted to four residences for the winter season of 1985-86.

2.0 STATISTICAL METHODOLOGY

The statistical analysis technique of multivariate regression is commonly used to investigate the basic relationship between two or more variables. However, considerable problems can arise if the explanatory variables are highly correlated, in fact, for the results of the regression analysis to be correct, the explanatory variables must not be intercorrelated. If they are, the multivariate regression analysis overstates the importance of the joint correlations. (For details on the potential dangers of multiple regression on intercorrelated variables, the reader is referred to Swindel (1974)). One solution is to reduce the number of predetermined variables through the technique of principal components, whereby a group of variables is reduced to a more fundamental set of variables.

Principal component analysis creates a set of new, uncorrelated variables (orthogonal functions or eigenvectors) that are linear combinations of the original variables. These new principal components explain all of the variability in the original data set; and in most applications, only the first few components are required to explain the majority of that variability. The first principal component accounts for the greatest variability in the data; the second principal component accounts for the maximum variability in the data not explained by the first component, followed by the third component, etc.

In PCA, it is first necessary to transform the independent variables into a set of normalized, dimensionless variables:

$$Z_{ik} = (X_{ik} - \bar{X}_i)/S_i$$

where

X_{ik} = kth value of the ith variable

$\bar{X}_i$ = mean of variable i

S_i = standard deviation of variable i.

This transformation results in a new set of variables with a mean of zero and a standard deviation of one. Without this transformation, the variables with the largest absolute magnitude would dominate in subsequent PCA, biasing the results and interpretation.

Each variable is composed of the sum of the contributions from each of the underlying components. The mathematical form of the PCA model is

$$Z_{ik} = \Sigma A_{ij} P_{jk}$$

for

$i = 1,\ldots,m$ variables

$j = 1,\ldots,n$ underlying components

$k = 1,\ldots,p$ data points

where P_{jk} is the kth value of the jth principal component and A_{ij} is the component loading (scoring coefficients). Interpretation of the component loadings is given in Table I.

Table I. Interpretation of the component loadings.

Value of A_{ij}	Meaning
Near Zero (-.200 to +.200)	Almost no dependence of variable on component.
Large Positive (+.900 to 1.000)	Strong positive dependence. If the component is positive (negative), the variable is above (below) its average; if the component is zero, the variable is near its average value.
Large Negative (-.900 to -1.000)	Strong negative dependence. If the component is positive (negative), the variable is below (above) its average; if the component is near zero, the variable is near its average.
Moderate (+.200 to +.900) (-.200 to -.900)	Same as above, but with a proportionately less strong dependence.

3.0 ANALYSIS OF SPACE HEATING ENERGY CONSUMPTION

3.1 Data Sets

A large data base consisting of measured electrical energy consumption in both residential and commercial buildings, scattered throughout the Pacific Northwest, has been established as part of a Bonneville Power Administration program called the End-Use Load and Conservation Assessment Program (ELCAP). This large-scale data collection effort, which is being conducted by the Pacific Northwest Laboratory, has resulted in the instrumentation of over 400 residences to monitor hourly electrical energy consumption by major end-use category and indoor temperature in the main living area of the residence. Coincident with these measurements, external environmental data are being obtained from 50 meteorological stations installed on strategically located residences. These stations provide site-specific measurement of outdoor temperature, wind speed and direction, and direct solar insolation.

It was the intent of this initial study to focus on a single winter season. The normal climatological winter is defined as the three month period December 1 through February 28, however, for this past winter season, the coldest temperatures for much of the region occurred during the last two weeks of November, when an outbreak of early-season record cold arctic air affected the entire Pacific Northwest. Temperatures for the remainder of the winter were near to above normal over most of the Pacific Northwest. Therefore, time period chosen to represent this particular winter season was November 15, 1985 through February 28, 1986.

Hourly residential space heating consumption and coincident meteorological data for four residences for the winter of 1985-86 were obtained from the ELCAP data base. One residence was located in eastern Washington, one in eastern Oregon, and two of the residences were located in western Montana. Characteristics of the residences selected for this study are shown in Table II. Summary statistics of the mean daily values of the meteorological variables for the four residences are presented in Table III. The hourly data were validated by visual inspection of time series plots of the basic parameters and by intersite comparisons.

TABLE II. Characteristics of residences used in PCA analysis.

Characteristic/Residence	RES 1	RES 2	RES 3	RES 4
Location	E. Wash.	E. Ore.	W. Mont.	W. Mont.
Age	P-75*	P-75*	P-75*	P-75*
Size (sq. ft.)	1600	1850	1200	1150
Heating Consumption				
Daily mean (kW-hrs)	46.2	104.1	81.6	28.3
Std. dev.	11.6	62.6	30.9	11.6

---

*P-75: Built after 1975.

Table III. Average daily values of meteorological variables for the four residences for the Winter 1985-86.

PARAMETER	UNIT	RES 1		RES 2		RES 3		RES 4	
		MEAN	STD. DEV.	MEAN	STD. DEV.	MEAN	STD. DEV.	MEAN	STD. DEV.
OAT_{avg}	^{o}C	-1.6	6.1	-1.8	6.5	-6.6	8.0	-5.1	7.3
OAT_{max}	^{o}C	1.0	6.7	1.5	6.8	-1.2	7.9	-0.1	7.1
OAT_{min}	^{o}C	-3.9	5.8	-4.7	6.6	-12.0	8.7	-10.0	8.1
IAT_{avg}	^{o}C	20.5	0.5	20.0	1.6	23.3	3.1	18.8	4.1
ΔT	^{o}C	22.1	5.9	21.9	6.5	29.9	7.1	23.9	8.4
TOTSOL	$Watts/m^2$	1208	782	1401	681	1698	809	1780	708
SOL13	$Watts/m^2$	226	147	249	124	307	134	326	125
WS_{avg}	mps	1.6	1.5	1.8	1.4	1.0	0.9	1.8	1.3
IRH_{avg}	%	31	5.0	44	2.7	24	4.8	40	2.9

KEY

OAT_{avg} = average daily outdoor temperature
OAT_{max} = average daily maximum temperature
OAT_{min} = average daily minimum temperature
IAT_{avg} = average daily indoor temperature
ΔT = average daily indoor/outdoor temperature difference
TOTSOL = average daily total solar insolation
SOL13 = average daily 1 pm solar insolation
IRH_{avg} = average daily indoor relative humidity

Smoothed time series plots of the daily average values of the outdoor temperature, insolation and space heating loads for the eastern Washington residence are shown in Figure 1. A nonlinear smoothing function using running means was applied to the data before plotting, to highlight persistent changes to the weather regimes and to de-emphasize the minor day-to-day fluctuations in the daily means.

3.2 Results

In order to determine the dependence of the space heating on the principal components, PCA was applied only to the meteorological variables given in Table IV; this was followed by multivariate regression analysis, with the individual daily principal component (PC) values as the independent variables and the daily total space heating consumption as the dependent variable. The individual daily PC values (Pjk) were calculated from the inversion of equation (2);

$$P_{jk} = \Sigma(A_{ij}/E_j)Z_{ik},$$

where E_j is the eigenvalue associated with a particular component (P_j).

The set of meteorological parameters used in the analysis are shown in Table IV along with an explanation for including the parameter. Results of PCA and subsequent regression of the principal components on space heat consumption are summarized in Table V and detailed in Tables VIa-c.

At all sites, the daily values of the first principal component were equally dependent on the four temperature variables and marginally dependent on insolation and wind speed (Table VIa). This principal component was found to explain between 65% and 69% of the variability in the original meteorological data set. Based on examination of the detailed time series plots of the temperature and first principal component values, it is evident that the first principal component is a reflection of major changes in air mass characteristics affecting the site, dominated by significant changes in mean daily temperatures (e.g., polar/arctic air mass with below normal temperatures vs. maritime air mass with near or above normal temperatures). Consequently, the first principal component is referred to as the air mass component.

The air mass component accounted for 55% (RES4) to 85% (RES2) of the variability in the daily space heating. The negative regression coefficient, or the slope of the regression line, indicates that the greater the magnitude of the air mass component value (higher temperatures) on a particular day, the lower the daily space heating consumption.

The second principal component is called the solar component, owing to the moderately strong dependence of the component on average daily total solar insolation (Table VIb). The component is also marginally dependent on lower than normal minimum temperatures at two of the four residences (RES1 and RES3). This is not unexpected as the clear sky conditions of this component are conducive to rapid nighttime radiational cooling after sunset, resulting in minimum temperatures that are lower than would normally occur on overcast days

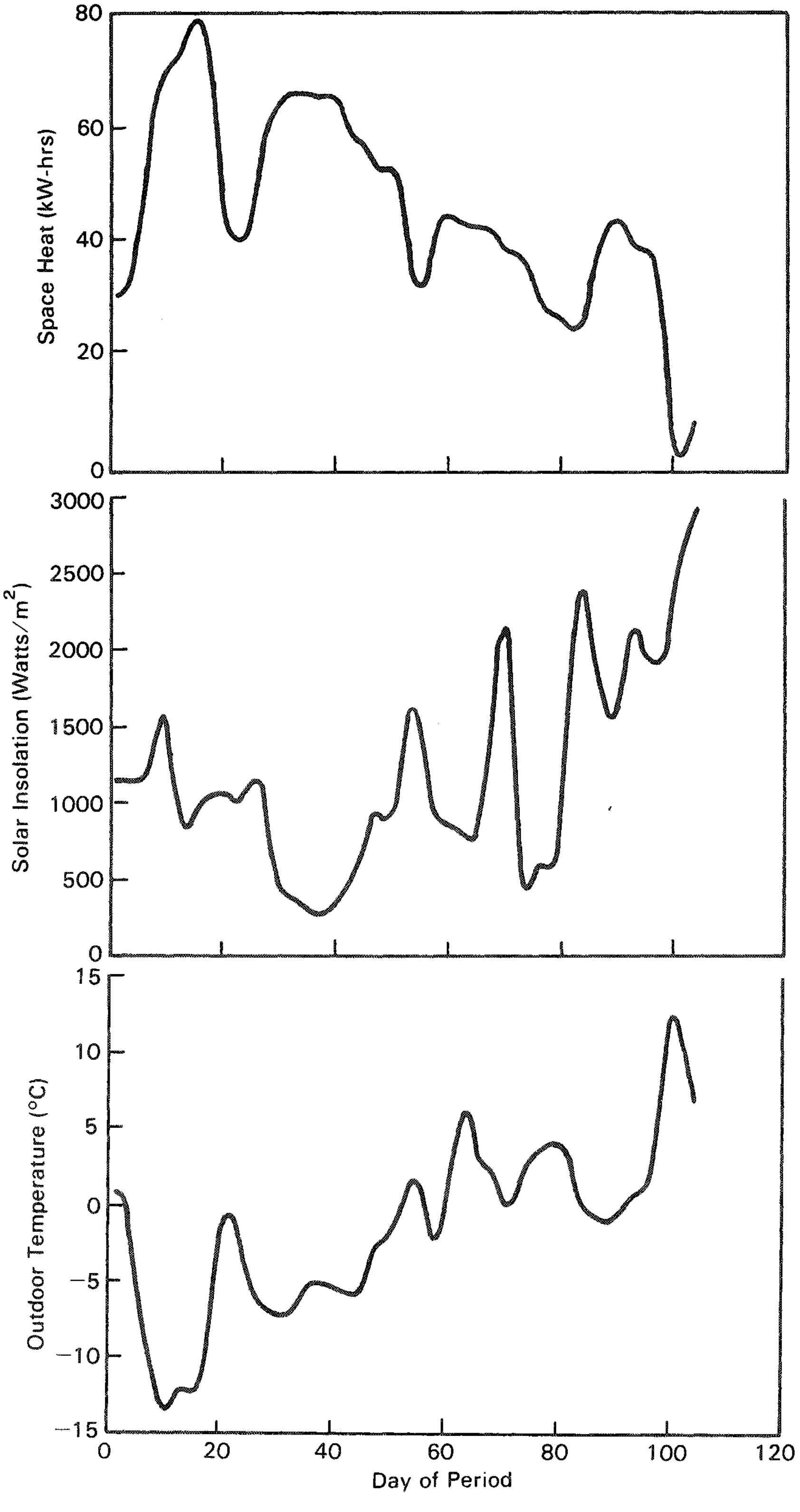

Figure 1. Time-series plot of Daily Average Space Heat Consumption (top), Insolation (middle), and Outdoor Temperature (bottom) of RES1 for the period November 15, 1985 through February 28, 1986.

Table IV. Summary of meteorological variables used in the principal component analysis and space heating regression analysis.

Parameter	Explanation
Independent Variables	
Average Outdoor Temperature	Primary Determinate of Heating Demand
Maximum Temperature	Related to Conditions at Time of Minimum Heating Demand
Minimum Temperature	Related to Conditions at Time of Maximum Heating Demand
Average Indoor-Outdoor Temperature	Primary Determinate of Heating Demand
Total Solar Insolation	Secondary Determinate of Heating Demand
Wind Speed Squared	Infiltration Factor
Dependent Variable	
Space Heating	Dependent Variable

Table V. Summary of principal component analysis and space heating regression for four houses.

	First 3 Components		All Components
Residence ID	% Total Variability	% of Space Heating Variability Explained	% of Space Heating Variability Explained
RES 1	95	90	91
RES 2	97	87	88
RES 3	96	78	82
RES 4	95	61	71

Table VI

a. Results of the air mass PCA and space heat regression.

Variable	Residence ID			
	RES 1	RES 2	RES 3	RES 4
Daily Average Temperature	.469	.484	.498	.500
Daily Maximum Temperature	.468	.472	.486	.481
Daily Minimum Temperature	.444	.468	.481	.490
Daily Ave. Indoor/Outdoor Temperature	-.468	-.476	-.487	-.461
Daily Total Solar Insolation	.258	.221	.141	.038
Daily Ave. Wind Speed	.279	.216	.165	.253
% Total Variance	69	69	65	65
R^2 for Space Heat	.80	.85	.67	.55
Space Heat Regression Coefficient	-8.0	-23.3	-12.8	-4.4

b. Results of the solar insolation PCA and space heat regression.

Variable	RES 1	RES 2	RES 3	RES 4
Daily Average Temperature	-.174	.016	-.097	-.057
Daily Maximum Temperature	-.045	-.049	.028	.008
Daily Minimum Temperature	-.337	-.098	-.217	-.081
Daily Ave. Indoor/Outdoor Temperature	.178	-.033	.005	.144
Daily Total Solar Insolation	.715	-.794	.969	.908
Daily Ave. Wind Speed	.555	.597	.037	.380
% Total Variance	16	14	16	18
R^2 for Space Heat	$<.01$	.02	.05	$<.01$
Space Heat Regression Coefficient	0.7	9.0	-7.1	-0.5

c. Results of the ventilation PCA and space heat regression.

Variable	RES 1	RES 2	RES 3	RES 4
Daily Average Temperature	.014	.154	.065	.051
Daily Maximum Temperature	.104	.107	.164	.232
Daily Minimum Temperature	-.018	.204	.070	.026
Daily Ave. Indoor/Outdoor Temperature	-.026	-.145	-.016	-.125
Daily Total Solar Insolation	.614	.553	.055	.391
Daily Ave. Wind Speed	-.781	-.722	-.980	-.880
% Total Variance	10	14	16	13
R^2 for Space Heat	.10	$<.01$	.07	.06
Space Heat Regression Coefficient	-7.5	-4.8	-8.1	-3.3

with lower than average daily total solar insolation. This component is also moderately dependent on higher than average wind speed at three of the four residences. The solar component accounted for between 14% and 18% of the total variability in the original meteorological data set, but accounted for 5% or less of the daily space heating variability at the four residences.

The third principal component accounted for between 10% and 16% of the total variability of the original data set and as much as 10% of the variability in the daily space heating. The component has a moderate to strong dependence on the wind speed (Table VIc) and is, therefore, referred to as the ventilation component. The component is also moderately dependent on daily total insolation. It is interesting to note that the percentage of space heat variability explained by the ventilation component is larger than was expected, but because of the limited data involved in this study, no real significance is associated with this observation.

This approach has several advantages, besides numerical stability, over traditional multivariate regression techniques or the use of only a single variable (i.e., outdoor temperature) as the explanatory variable. (Henry and Hidy, 1979). First, the principal components are mathematically independent; thus, any number of terms may be dropped from the regression while retaining its validity. Second, only a few principal components are required to explain most of the variability in the dependent variable. Third, the eigenvectors with very small eigenvalues can be used to examine the exact nature of the intercorrelation in the data set.

The scatter plots shown in Figure 2 illustrate daily total space heat consumption at RES1 as a function of the indoor/outdoor temperature difference (top) and as a function of the first principal component (bottom). The solid line is the linear least-squares fit to the data points. The r-squared values for the data plotted in Figure 2 and for the other three residences are given in Table VII. If only the first principal component is considered, no advantage is evident in applying PCA to the independent variables. However, when all components are included in the regression analysis, the amount of variability explained in the daily space heat consumption can increase considerably over the single correlate approach.

4.0 CONCLUSION

Before the conclusions are presented, a few important precautionary comments are in order. The results must be considered as preliminary in that they are case studies of only four residences for one winter heating season. The two climate zones included in the case studies did not include the major population areas west of the Cascade Mountains. The results are site-specific, and it must be emphasized that there should be no attempts to extrapolate from these results or conclusions to the entire housing stock of the Pacific Northwest.

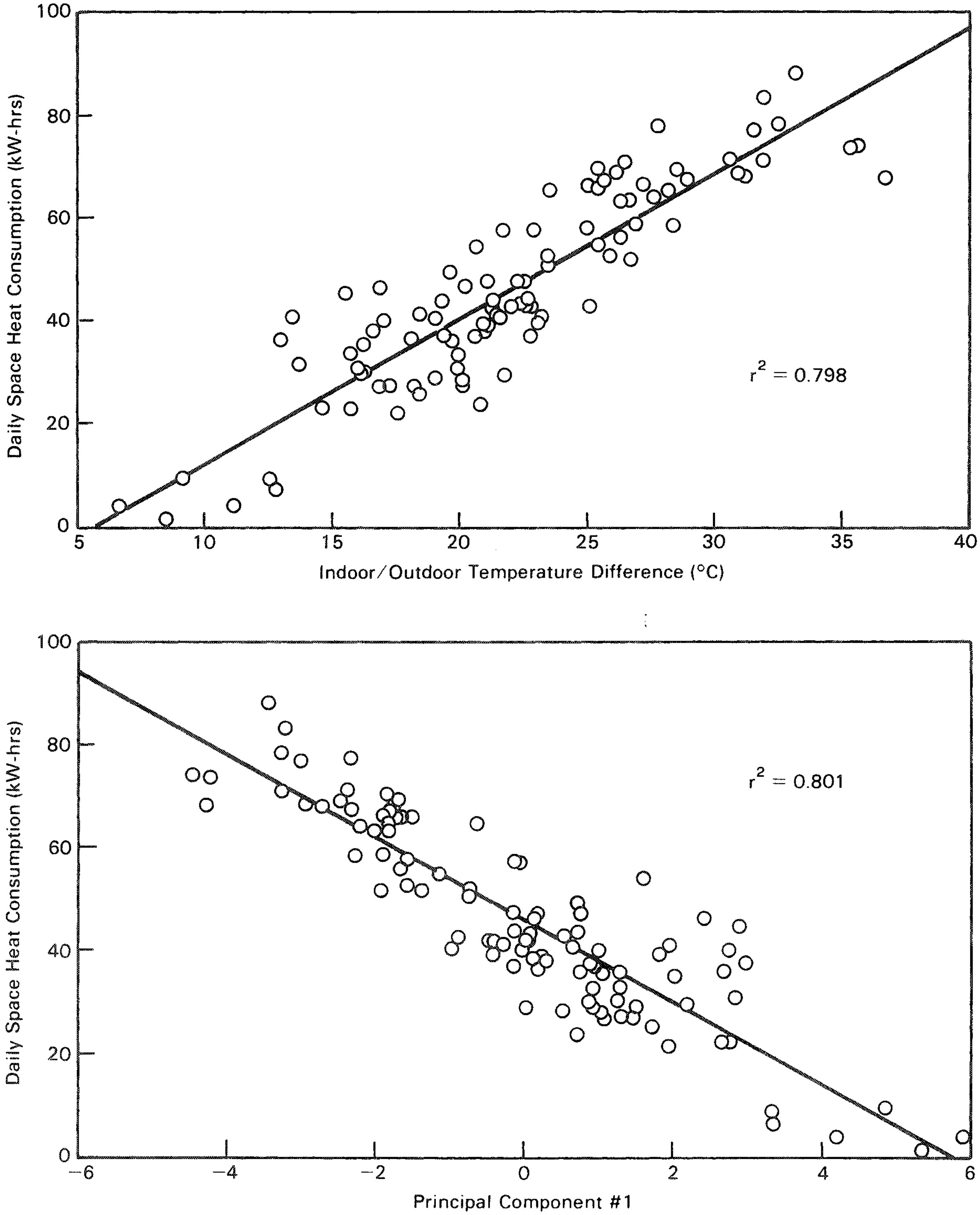

Figure 2. Scatter Plot of Daily Space Heat Consumption at RES1 as a Function of Indoor/Outdoor Temperature Difference (top), and as a function of the First Principal Component (bottom).

Table VII. R-square statistics for space heat as a function of the indoor/outdoor temperature difference and as a function of the principal components.

	IAT/OAT Difference	PC 1	All PCs
RES 1	.798	.801	.906
RES 2	.831	.848	.884
RES 3	.702	.668	.817
RES 4	.655	.545	.717

Principal component analysis has been shown to be a useful technique to further our understanding of the influences of the underlying meteorological factors on the variability of the daily space heating electrical energy consumption in residences. It has been demonstrated in other studies that outdoor temperature is the dominant factor affecting the space heating requirements for residences, with solar insolation and the other meteorological variables having a minor secondary effect. This investigation provides a different approach to determining the combined effect of various meteorological conditions on space heating consumption. The results do not differ from previous conclusions, but do add significantly more information to the total effect of weather on the space heating variability throughout the winter season.

Given these caveats, we present the following conclusions for this initial study:

1. PCA identifies a few linear combinations of the original, larger meteorological data set, which can be used to summarize the original data with minimal loss of information.
2. The principal components selected have been shown to represent the fundamental, underlying weather controls.
3. The periodic, region-wide changes in air mass characteristics over the region east of the Cascade Mountains account for up to 87% of the variability of the daily space heating consumption during the winter season.
4. The effect of the air mass component was consistent among the four residences investigated.
5. The daily variation in the solar component contributed insignificantly to the variability of the daily space heating consumption.
6. The daily variation in the ventilation component accounted for up to 10% of the variability in the daily space heating consumption.

5.0 REFERENCES

Christensen, W.I., and R.A. Bryson. 1966. "An Investigation of the Potential of Component Analysis for Weather Classification." Mon. Wea. Rev., 94:697-709.

Clarke, J.F., and J.T. Peterson. 1973. "An Emperical Model Using Eigenvectors to Calculate the Temporal and Spatial Variations of the St. Louis Heat Island." J. Appl. Meteor., 12:195-210.

Hardy, D.M., and J.J. Walton. 1978. "Principal Components Analysis of Vector Wind Measurements." J. Appl. Meteor., 17:1153-1162.

Henry, R.C., and G.M. Hidy. 1979. "Multivariate Analysis of Particulate Sulfate and Other Air Quality Variables by Principal Components - I. Annual Data from Los Angeles and New York. Atmos. Environ., 13:1581-1596.

Henry, R.C., and G.M. Hidy. 1982. "Multivariate Analysis of Particulate Sulfate and Other Air Quality Variables by Principal Components - II. Salt Lake City Utah and St. Louis Missouri." Atmos. Environ., 16:929-942.

Lioy, P.J., R.P. Mallon, M. Lippmann, T.J. Kneip, and P.J. Sampson. 1982. "Factors Affecting the Variability of Summertime Sulfate in a Rural Area Using Principal Component Analysis." J. Air Polut. Control Assoc., 6:791-802.

Swindel, B.F. 1974. "Instability of Regression Coefficients Illustrated." Am. Statistic., 28:63-65.

Thurston, G.D., and J.D. Spengler. 1985. "A Multivariate Assessment of Meteorological Influences on Inhalable Particle Source Impacts." J. Climate Appl. Meteor., 24:1245-1256.

Wolff, G.T., M.L. Morrissey, and N.A. Kelly. 1984. "An Investigation of the Sources of Summertime Haze in the Blue Ridge Mountains Using Multivariate Statistical Methods." J. Climate Appl. Meteor., 23:1333-1341.