7911141

ANNAN, EMMANUEL

A TEMPERATURE-DEPENDENT STOCHASTIC MODEL FOR POWER SYSTEMS DEMAND FORECASTING.

THE UNIVERSITY OF OKLAHOMA, PH.D., 1978

University Microfilms International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs _______
2. Colored illustrations _______
3. Photographs with dark background _______
4. Illustrations are poor copy _______
5. Print shows through as there is text on both sides of page _______
6. Indistinct, broken or small print on several pages ✓ throughout _______
7. Tightly bound copy with print lost in spine _______
8. Computer printout pages with indistinct print _______
9. Page(s) _______ lacking when material received, and not available from school or author _______
10. Page(s) _______ seem to be missing in numbering only as text follows _______
11. Poor carbon copy _______
12. Not original copy, several pages with blurred type _______
13. Appendix pages are poor copy _______
14. Original copy with light type _______
15. Curling and wrinkled pages _______
16. Other _______

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A TEMPERATURE DEPENDENT STOCHASTIC MODEL FOR POWER SYSTEMS DEMAND FORECASTING

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

By

EMMANUEL ANNAN

Norman, Oklahoma

1978

A TEMPERATURE DEPENDENT STOCHASTIC MODEL

FOR POWER SYSTEMS DEMAND FORECASTING

APPROVED BY

DISSERTATION COMMITTEE

ABSTRACT

The temperature dependency of electrical power demand, i.e., load, was statistically analyzed and a set of temperature versus demand characteristics were used to develop a demand forecasting model.

In this research a statistical distribution analysis of a temperature dependent demand characteristic was made. The conditional distributions of load and temperature were determined. These were applied in the estimation of the probability of occurrence of a load. The results show that the basic demand characteristic of one utility may be representative of the characteristics of several other electric utility systems in the region studied. This spanned Oklahoma, Texas, and Louisiana, which experience summer peaks.

Also, a system demand forecasting model was developed. Annual system peak loads were predicted using the model, as the ultimate goal. Application may be made from one system to another for regions with the basic demand characteristic described. A temperature dependent characteristic, based on temperatures occurring in the spring, summer, and fall seasons, was used in the determination of the forecasting model. The developed model is a compromise between the current methods that use conventional temperature dependent forecasting, and time series analysis approach. An explicit temperature dependent forecasting model was developed by utilizing time series analysis based on Box-Jenkins' models. Real system demand data, comprising peak temperatures, were used to illustrate the methodology. The data was divided into two main parts. The first part was used to estimate the parameters of the forecasting model. The forecasts were then compared to the second part of the data. The relative errors, calculated as a percentage of the actual demand, were small; less than 5% for the system in a six-year forecast horizon.

The results show that the developed model is a promising tool that can be used by researchers and others in long range capital and operational planning, and in the advancement of forecasting methodology.

ACKNOWLEDGMENTS

The author wishes to express his sincere thanks to his wife, Winifred whose understanding, patience, and cooperation made this study possible.

Special thanks are due to Professor M.E. Council who as chairman of the committee helped by advice and encouragement during the entire period of this study. The author also wishes to thank Drs. W. Cronenwett, J. Fagan, B. Foote, T. Gonen, C. Haden, and B. Crain for their guidance and constructive critiques throughout this study.

Further, the author thanks the managements of Oklahoma Gas and Electric Company, Oklahoma City, Oklahoma; Public Service Company of Oklahoma, Tulsa, Oklahoma; Louisiana Power and Light, New Orleans, Louisiana; Dallas Power and Light and Texas Power and Light, Dallas, Texas; for making their respective system demand data available.

The sponsors, the African-American Institute deserve mentioning for their financial support to continue the author's education.

Finally, he wishes to thank Pat Ybarra, typist, and Mary Ellen Kanack, drafter, for the excellent work of producing this volume.

DEDICATION

To my wife, Winifred, son Edwin, and daughter, Lauren.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

SYMBOLS

$f(X)$	Density function
$F(X)$	Cumulative distribution function
$F(X \cap T)$	Cumulative distribution of X intersecting T
$F(X/T)$	Cumulative distribution of X given T
T	Temperature ^{o}F
S	Sample standard deviation or series seasonal length
$H(X)$	Cumulative distribution of assumed statistic
$G(X)$	Empirical cumulative distribution of raw data
H_o	Null hypotheses
H_1	Type one test
Sup	Supremum or least upper bound
σ	Population standard deviation
μ	Population mean
G_{i+}, G_{i-}	Values of the function G_i just before or after G_i
D_i	Difference at index i
$D_\alpha(n)$	Critical Kolmogorov-Smirnov value at α level of significance and index count n
cdf	Cumulative distribution function
$\phi(Z_i)$	normal cumulative distribution function
Γ	gamma function notation
Var	Variance of a sample or population
γ_j	Autocovariance at lag j

ρ_j	Autocorrelation at lag j
B	Back shift operator
ε_t	Error at lag t
ρ_{jj}	Partial autocorrelation at lag j
ρ	Order of autoregressive process
d	Degree of regular differencing
q	Order of moving average process
y_t	Differenced series
ϕ_p	Autoregressive component of order p
θ_q	Moving average component of order q
ESS	Error sum of squares
$X^2(\nu,\alpha)$	Chi-square statistic

A TEMPERATURE DEPENDENT STOCHASTIC MODEL FOR POWER SYSTEMS DEMAND FORECASTING

CHAPTER 1

ELECTRIC POWER DEMAND MODELING

A. Introduction

The necessity of a power system demand model has been evident to system planners for years. However, it is only recently that quantitative specifications for such a model have been needed as planning techniques have become more sophisticated. The purpose of this research was: 1) to determine the parameters of a system demand model, 2) to determine the statistical distributions of the model, and 3) to formulate a normalized model for individual, utility and regional application.

The major application of such a model is in the area of demand forecasting. It will be shown how such a model may be so applied. There are several forecasting techniques [18] and completely specified demand models will greatly aid in consolidating and improving accuracy and confidence in their use.

The need for more accurate and reliable forecasts by electric power system planners assumed higher proportions in the period after the oil boycott of 1973 and subsequent formation of the international oil cartel, which has raised the price of petroleum fuel to levels considered intolerable a few years ago. It is noteworthy that several institutions and agencies are coming up with forecasts of total energy requirements that indicate shortfalls in supply of fuel in the short term. Most of these forecasts have, in many cases, not had the impact they deserve, due partly to the incomplete specifications of the models upon which they are based. Their statistical distributions may not be specified, and the data base may be questionable.

However, forecasting methodology is becoming more important due to the growing complexity of power systems and the demands of utility regulating agencies. Variables that previously were not important have increasingly had more influence lately and will have much more in the future. Some other areas that depend on a demand model are reviewed below. [18]

1) Economic dispatching and unit commitment depend on a demand model used in a short term forecasting model. Knowledge of the demand characteristic is needed to schedule generation or inter-area power exchange. On-line control programs for computer control of scheduling are based on a demand model.

2) Spinning reserve and other system reserves are planned based on forecasts reached using a demand model. An accurate estimation of reserve is important in planning maintenance, future capital expenditures in the system, as well as keeping within reliability limits set by the appropriate reliability councils.
3) Several studies such as load flow, reliability and stability studies are dependent on a load model to some extent.
4) Generation and transmission are major areas of the power system where a good demand model is indispensable. The presence of a load to be served or the anticipation or development of load is the main motivation for expansion.

There are requirements for capital, equipment orders, right of way, system optimum operation, etc., that depend on a demand model. In recent years, power plants have become increasingly larger with resultant complexity in design and high cost. The cost of an installed kilowatt exceeds about $500.00 for coal and nuclear power plants. When the costs of transmission and distribution are accounted for, the decision to build a new plant assumes profound proportions, not to mention the requirements of state and national regulatory agencies as well as highly vocal environmental groups. An error in forecasting demand and capital acquisition can be financially very expensive, considering the long lead times of the order of eight to twelve years.

Forecasting electric power demand requires a great deal of subjective judgment. Planners know very well the difficulty and uncertainty involving a forecast. In the period before 1950, most utilities forecasted demand based solely on judgment by "eyeballing" the available data to obtain a "ball park" forecast. This method was sufficient so far as the demand was dependent on predictable parameters and insensitive to weather and fluctuations in the economy. There were no indications of any price elasticity in demand.

In the period of 1950 through 1972, there was a sharp and diversified increase in the use of electricity. Household appliances, including air conditioning equipment, gained widespread use. The combination of units and their duty cycles at any time varied widely from one customer to another. This created a diversity within the demand for power that increased the uncertainty of a forecast. Statistical methods [12] of data analysis found applications in demand modeling during this period. The increased capability of the electronic computer greatly facilitated statistical applications. There were numerous technical papers published during this time in the area of forecasting.

The following is a more formal but very brief review of the past literature.

B. Review of the Literature and Problem Development

The importance of demand forecasting in the power industry was first recorded by Reyneau [19] in 1918. Load densities were then quite low and most industries were coal or oil fired. From 1918 to 1944 there were only few publications on electric demand forecasting, when Dryar [5] showed the effects of weather on system load. He demonstrated the forecast technique of separating demand into a base load and weather sensitive components. There have been numerous publications investigating various aspects of this technique. The advantage of the method was its large data base that gave more meaning statistically to the results. The underlying mathematical forecast tool in most of these methods was linear least squares regression.

Latham and Nordman [12] extended Dryar's technique further to include a method of probability estimation of forecasts. Probabilities of load occurrence were estimated by calculating the empirical discrete density function of load, taken over the years. The weather function was further used by Stanton [22] in a method for forecasting the weather sensitive component, and combines with the base load forecast to obtain total system demand forecast. The method consisted of curve fitting the weather sensitive component with a straight line; the slope of the line measured in megawatts per degree of temperature calculated for each year in the data set. The slope of the weather sensitive component is predicted by regression. There are numerous variations on these techniques in the literature. In each case an attempt was being made to overcome the drawback of having to use only one data point, the annual peak, per year. The methods use large amounts of data. However, this drawback remained though the forecasts continued to improve. Analysis of load against temperature is made, but the forecasts are time dependent.

Several researchers, such as Christiaanse [2], Gupta [7], Keyhani [11], Vemuri [25] looked into the use of time as an independent variable for predicting future loads. The application of time series models [25,29], using the methods known as the Box-Jenkins models [26], by some of these authors to utility forecasting further widened the scope for the methods currently available.

Concurrent with the development of forecasting methodology as outlined, there were other methods developed that were economic in nature. One such method is the method of land use described by Lazzari [13]. The service area is divided into one mile squares or smaller. Population growth is forecast, and the areas likely to receive the increase in population determined. It is then possible to determine for each square area the likely customer demand, and thus cumulatively arrive at a system forecast. The weather relation was brought into the foreground by Dryar, Heineman [9], Nordman [12], Stanton, Davey [4], and others. The other major approach has time or a combination of time and weather, in a regression model, as the important factor of consideration, described by Christiaanse, Gupta, Vemuri, Mabert, Lijensen [14], Toyoda [24], and others.

The objectives of this research were stated at the beginning of this chapter. In the light of discussion that included the merits of weather and time related forecasting respectively, further explanation of those objectives follow. The peak load, for most electric utility companies shows marked weather dependence, particularly temperature dependence. A temperature dependent demand characteristic will be described.

In this research, a model will be developed for time series analysis that is implicitly temperature dependent. It will thus be a compromise between the two major techniques presently in use. Considerable attention will be given to the development of statistical distributions of the data set used as a means of carrying information about the data set into the future.

C. The Data Base

In order to generalize the results of this research, a rather large amount of data were collected from several utilities servicing a relatively large geographical area. Hence, the data analysis and models derived were intended to have a regional application. The geographical area under study has the characteristic that peak demand for electric power occurs during the summer. This area (see Appendix B) includes most of Oklahoma, Texas and Louisiana, selected solely for availability of their system data for study. There are several other states with a predominant summer peak, however, data from these areas were inaccessible. Specific information about the sources of peak demand data will be held in confidence. Real system data were used in all analysis, unless noted otherwise. The data obtained consist of 1) daily system peak demand in megawatts, and 2) daily peak temperature in $^{\circ}F$ of a major load center within the service area for the period 1967 through 1976, where available.

D. <u>Temperature Data</u>

The only independent variable used is temperature, [3,10,23] which is adequate to demonstrate the model, and possible extensions to several variables. Every system planning engineer or researcher is aware of the problem of using one temperature value taken at one location, such as the geographical load center which is a hypothetical location that may be far from any real concentration of load. Since system load is a widely distributed function, the associated temperature function will be distributed also, because the temperature is usually different from one location to another within the service area. However, it is immediately clear that this is an infeasible proposition, since there is an infinite number of sample points. In this study, temperatures were taken from the most dominant load center within the service area. This showed strong correlation with system peak demand. In the area under study, daily load curves show a rise in demand with increasing temperature that is predictable. However, winter temperatures have not been found to have the same strong effect on demand as do summer temperatures. The greater interest in the annual peak occurring during the summer is sufficient reason to use peak daily temperatures as an input variable of this model.

The data (peak daily loads and temperatures arranged by date), were key-punched where not supplied on magnetic tape. To facilitate accessing, all the data were loaded to computer disk files. The disk file system is on line when the computer is up. There is a file for each data set from each utility, and a composite file containing data for the region. Each record in a file has a date,maximum temperature and peak load. Full documentation to access each file is given in Appendix A.

CHAPTER 2

CHARACTERISTICS OF THE MODEL

A. Introduction

One of the objectives of this study is to determine the statistical distributions of a demand model. In this chapter, the basic form of a demand characteristic is described and the conditional distribution of demand, at a given temperature, is shown. Ultimately, the statistical distributions and the demand model will be used to estimate the probability of load occurrences when the model is used in forecasting future demand. The temperature and demand data used are in computer data storage.

B. The Basic Demand Characteristic

A plot of demand in megawatts (MW) versus temperature in degrees Fahrenheit (^{o}F) yielded the characteristic shown in Figure 2.1. This characteristic indicates three broad regions discussed below.

Region I of the plot shown in Figure 2.1 shows a fairly constant region, where the demand is independent of temperature. This region contains data mainly from the spring and fall seasons, during which time temperatures are

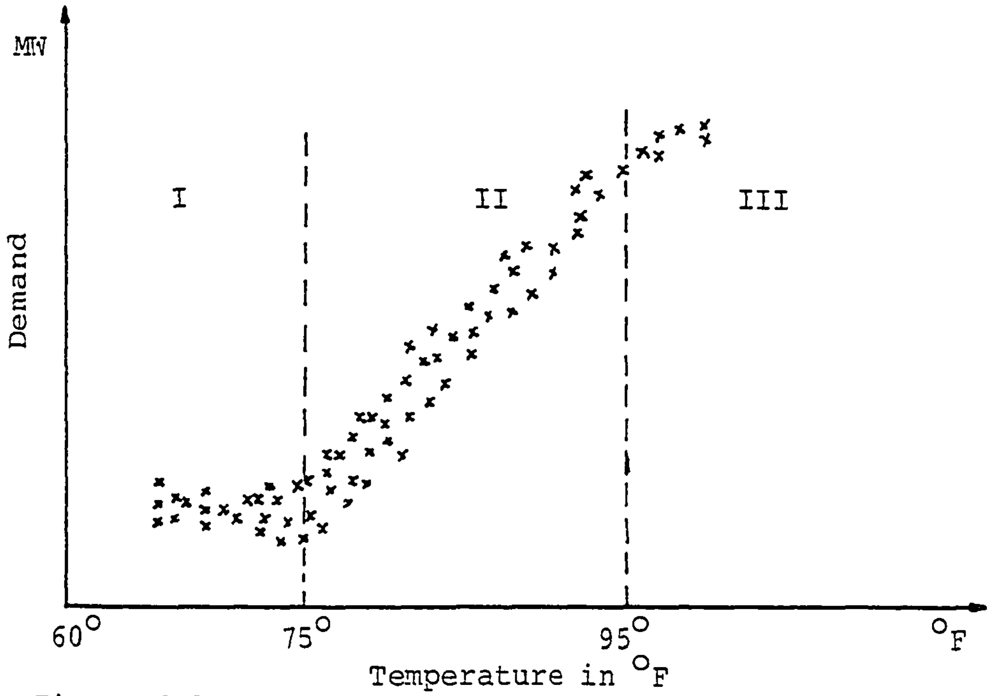


Figure 2.1. Demand Versus Temperature Characteristic

mild and comfortable. It is a "comfort region" or baseload region of daily maximum demands. It is bounded by about 60°F on the low side and about 75°F on the high side. These bounds are approximate, and there may be variations of ±5°F in some cases. In this region, most heat sensitive equipments stay off, or on a minimum of the time.

Heat sensitive equipment, as used here, refers to residential, commercial or industrial weather modification devices such as air conditioning plants, both heating and cooling. Agricultural water pumping equipments that are thermally controlled are also included in this description of temperature sensitive equipment.

Region II of Figure 2.1, shows a marked temperature dependence. It is fairly defined for each year's data plotted. Most heat sensitive equipment will cycle on and off in a steady state mode for any given temperature. The higher the temperature, the higher the rate of cycling and the longer the duration of the on cycle. This region extends approximately from 75°F to 95°F. The slope of this region of the characteristic depends on the density of heat sensitive equipment within the geographical service area. For example, a high use of air conditioning equipment in a service area causes a large number of equipments to cycle on when a rise in temperature occurs. Hence, the load on the power system increases with rising temperature.

Beyond 100°F, the effects of saturation become noticeable. This is Region III, where temperatures reach such high values that most heat sensitive equipments tend to stay on most of the time. The diversity in equipment use has been completely eliminated, hence, demand reaches an approximately constant value again, this time in weather conditions generally accepted as hot .

Hence, there is a region of minimum sensitivity followed by a region of dependence, where sensitivity is a steady state value given a state of the independent variable. Finally, there is a region of saturation where further increases in the independent variable do not produce appreciable corresponding increases in demand.

It is appropriate to restate here that most of the analysis will be directed at summer load. This is important because it is the period during which the system peak load occurs. The geographical region selected for study does not generally experience very severe winter weather conditions. Other sources of winter heating, other than electricity, such as gas, wood and oil are readily available. Heavy clothing worn during the winter season is one more factor to consider. Hence, the winter load is much more difficult to characterize in those states where winters are mild.

C. Statistical Analysis [28]

The conditional distribution of the load at any given temperature, may be defined as

$$F(X/T) = \frac{F(X \cap T)}{f(T)} \tag{2.1}$$

where X is the load or dependent variable and T is temperature or independent variable. Hence,

$$F(X \cap T) = F(X/T) \cdot f(T) \tag{2.2}$$

The calculation of this requires a knowledge of the density function f(T) and of F(X/T). Estimation of f(T) will be shown in a later chapter. The distribution function of F(X/T) is the object of this chapter. Figure 2.2 shows a plot of load and temperature data points, similar to Figure 2.1. The distribution function F(X/T) is shown here as a normal distribution, which is assumed, and will be verified true or false. Figure 2.2 is for the purpose of illustration only.

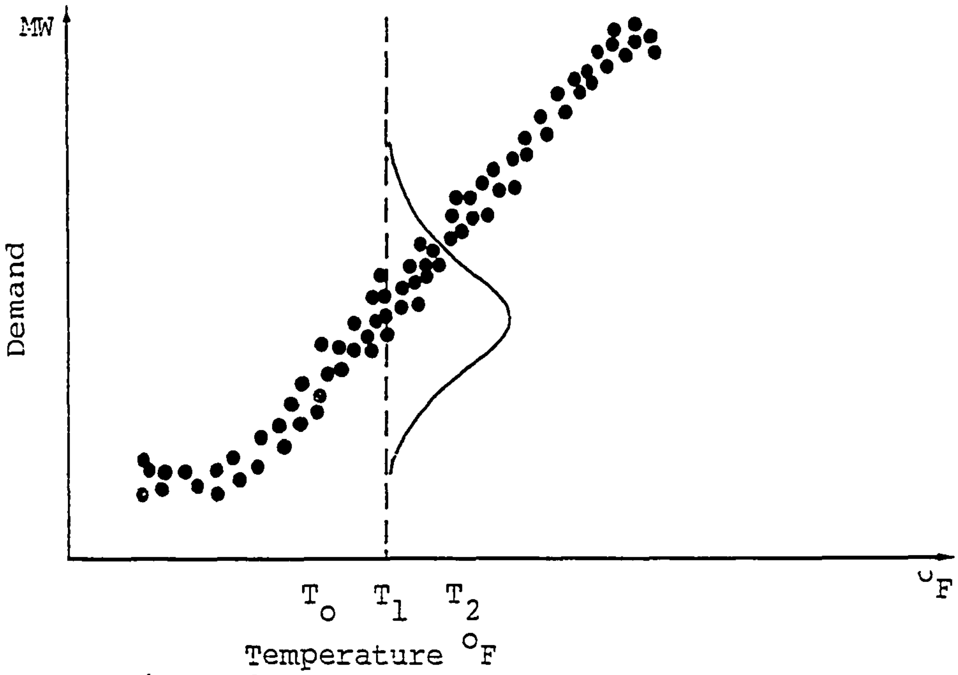


Figure 2.2. Normal Conditional Density Function

Statistical parameters such as mean and standard deviation will be calculated where necessary in the description of the data sets. The distribution of load, at a given temperature, will be calculated for several years' data over the test geographical region in an attempt to describe their distribution functions. Several tests of the data will be made, the basics of which are described presently. Mean and standard deviations are calculated as explained below.

D. Mean

This is estimated as the sample mean of data at a given temperature, or a band of temperatures. The particular case will be specified. Referring to Figure 2.2, the sample mean $\overline{X}$ is calculated as

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{2.3}$$

where X_i is the daily peak demand at T_i. ^{o}F, n the number of occurrences of X_i's.

E. Standard Deviation

The sample standard deviation of the data at a given temperature or band of temperatures is calculated as

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \tag{2.4}$$

where S is the sample standard deviation, $\bar{X}$ is the sample mean as calculated above, X_i is demand occurrence at T_i ^{o}F and n is the sample size; number of X_i's.

Several statistical distribution functions were selected for testing. The test procedure used is the Kolmogorov-Smirnov test of goodness of fit.

F. Kolmogorov-Smirnov Test of Goodness of Fit [28]

This is a non-parametric statistical test. It does not depend on any particular distribution. It compares the empirical cumulative distribution of the data set being tested to that of an assumed distribution by comparing the deviation at all points of the two distributions against an acceptance limit. This test was selected over other tests because of its power to detect differences in cumulative distributions. Most statistical tests require large sample sizes for any conclusive descriptions of their nature. However, the Kolmogorov-Smirnov test remains valid for small sample sizes. In a test of hypothesis, the Kolmogorov-Smirnov test provides grounds for rejecting or not rejecting the assumed distribution. A non-rejection region, however implies acceptance of the assumed distribution only to the extent that other distributions might equally describe the same data set. The test helps to narrow down the number of trial distributions, which may then be more explicitly tested. The following and figure 2.3 illustrate the test procedure.

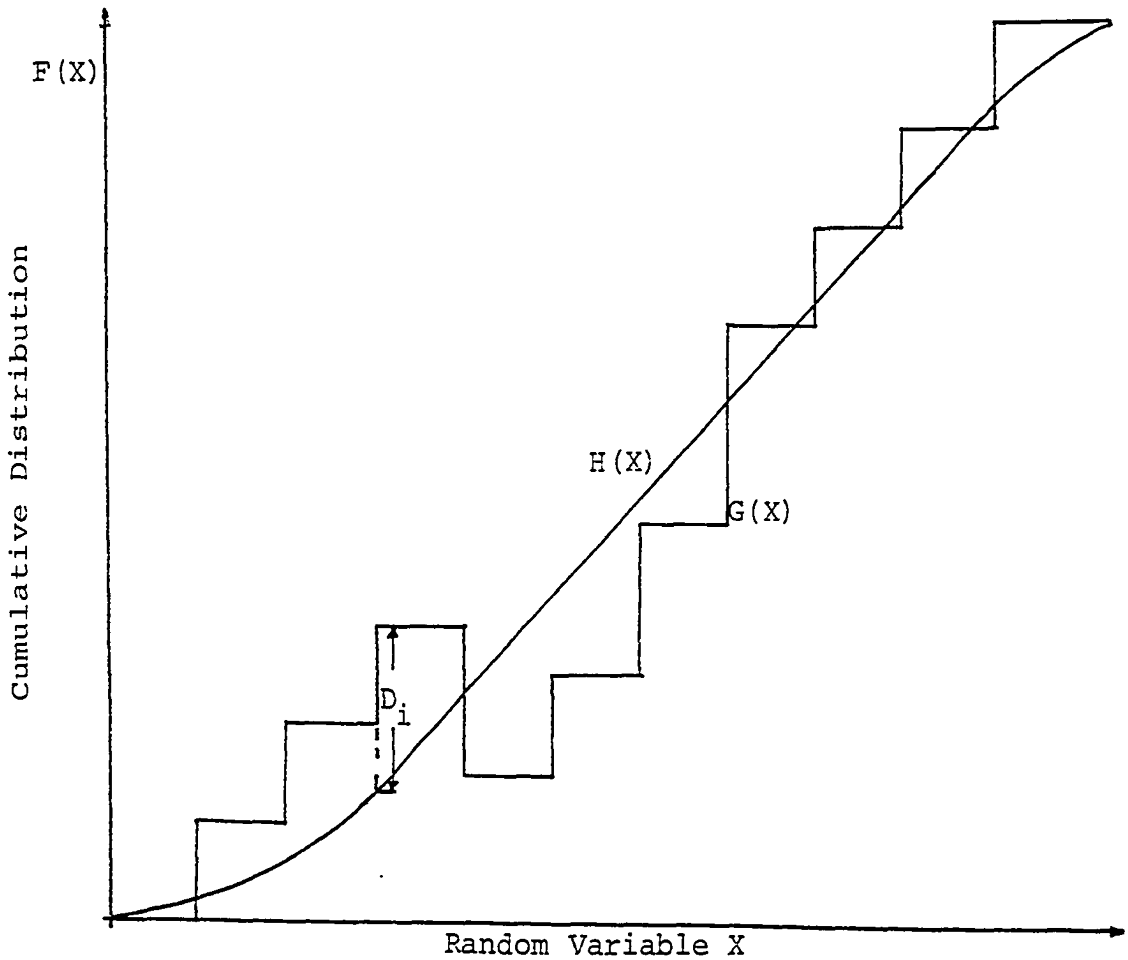


Figure 2.3. Cumulative Distribution for Comparing Assumed Distribution to Empirical Distribution

Figure 2.3 shows a plot of the empirical cumulative distribution function, (cdf), G(x) and the assumed cdf H(x), which is assumed to describe the sample. The deviations at all the breakpoints of the empirical cdf are computed. In the following test of hypothesis; sup represents least upper bound.

$$H_0: \text{ reject normality if sup } |D_i| > D_{\alpha o}$$
$$H_1: \text{ accept normality if sup } |D_i| < D_{\alpha o} \tag{2.5}$$

where sup$|D_i| > D_{\alpha o}$ is the minimum value in the set $\{D_i\}$ that exceeds $D_{\alpha o}$.

Table 2.1. Kolmogorov-Smirnov Test, Assuming Normal Distribution for Data Set

1	2	3	4	5	6	7
i	X_i	$\sum_{i=1}^{n} f_i$	$G(X)=\sum_{i=1}^{n}$	$Z=\frac{X_i-\overline{X}}{\sigma}$	$F(X)=\Phi(z)$	D_i
1	744	1	0.077	-1.56	0.059	0.059
2	787	2	0.154	-1.24	0.108	0.046
3	803	3	0.231	-1.12	0.131	0.100
4	858	4	0.308	-0.70	0.242	0.066
5	860	5	0.385	-0.69	0.245	0.140
6	922	6	0.460	-0.23	0.409	0.051
7	958	7	0.538	0.04	0.484	0.054
8	1002	8	0.615	0.37	0.644	0.106
9	1051	9	0.692	0.74	0.770	0.155
10	1054	10	0.769	0.76	0.776	0.084
11	1077	11	0.846	0.93	0.824	0.055
12	1123	12	0.923	1.28	0.900	0.054
13	1142	13	1.000	1.42	0.922	0.078

Column 1 is a row index of a monotonically increasing data set X_i in Column 2. Column 3 is the cumulative frequency of occurrence. There may be more than one entry in any row i. Column 4 is the empirical cumulative distribution function of the data set. The cumulative distribution function of the data set, assuming a normal distribution as shown in Column 6, is derived by using Column 5 and standard normal cdf tables. Here, D_i is the calculated deviation at point i and $D\alpha_o$ is the Kolmogorov-Smirnov acceptance limit at a significance level of α_o for the data sample size.

Referring to Figure 2.3, the deviation D_i is calculated as

$$D_i = G_i(X) - F_i(X) \tag{2.6}$$

However, the deviation at i has two values due to the discontinuity at i. Hence, the deviation is stated as

$$D_i = \max \ |G_{i+}(X) - F_i(X)| \ , \ |G_{i-}(X) - F_i(X)| \tag{2.7}$$

Some of the distributions assumed for testing are reviewed below, and sample tests follow.

G. <u>The Normal Distribution</u>

This is a continuous distribution function. It is described by the general density function given by

$$f(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \tfrac{1}{2} \left[\frac{X-\mu}{\sigma}\right]^2 \tag{2.8}$$

A plot of the normal distribution is the familiar bell shaped figure shown below.

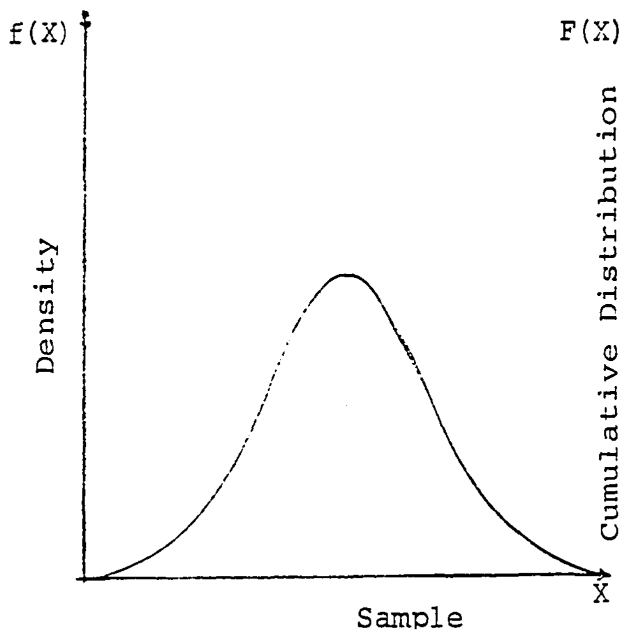


Figure 2.4. Normal Density Function

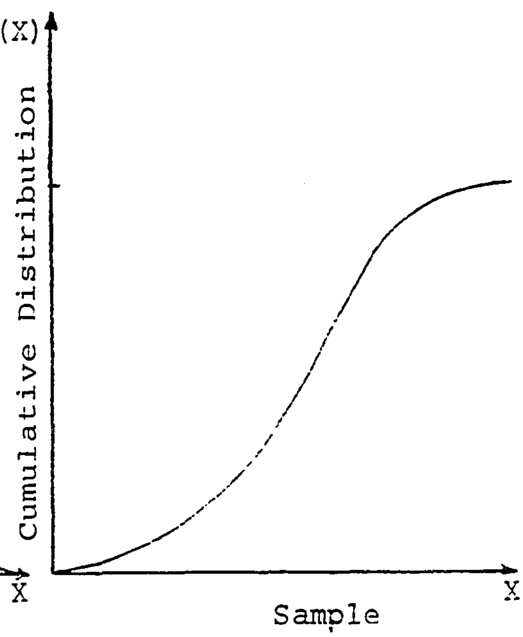


Figure 2.5. Normal Cumulative Distribution Function

The cumulative distribution of the normal density function is shown in Figure 2.5.

The difference between the empirical cdf and the assumed normal cdf is calculated and entered in Column 7, of table 2.1. The method of calculation is by Equation 2.7.

The test requirement is that the assumed distribution is acceptable if

$$D_\alpha(n) \leq \sup_{\text{all } X} |D_i|$$

where n is the sample size.

In Table 2.1, n=13, which yields an acceptance limit for the Kolmogorov-Smirnov test of goodness of fit at 5 percent significance level as $D_{.05}(13) = 0.361$. Hence, no entry in Column 7 may exceed this value if the normal distribution is to be included in the class of acceptable distributions. In Table 2.1, Column 7, all entries are less than $D_{.05}(13)$. Hence, the normal distribution has been accepted.

The next distribution function assumed is the uniform distribution. The results are shown in Table 2.2.

Figure 2.6 shows the application of the uniform density function to the given data set.

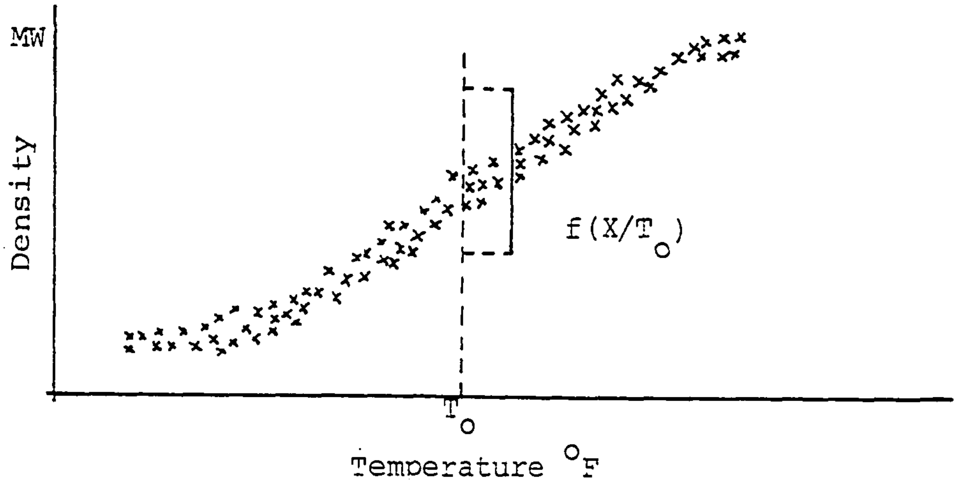


Figure 2.6. Uniform Conditional Density Function

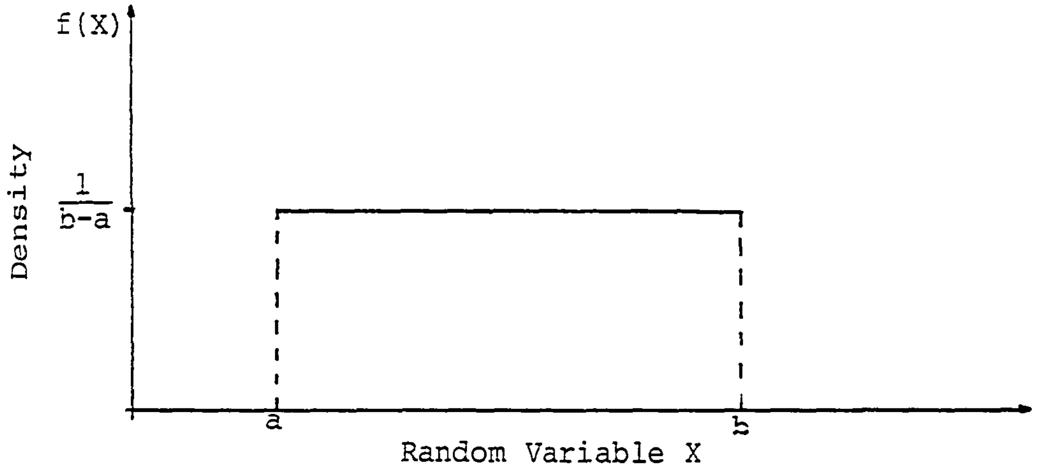


Figure 2.7a. Uniform Density Function

The plot, which is similar to Figure 2.1, is now assumed to be a uniform distribution. The density function is given by

$$f(X/T) = \frac{X}{b-a} \qquad a \le X \le b$$

Hence, the cumulative distribution function is given by

$$F(X/T) = \frac{x-a}{b-a} \qquad a \le X \le b$$

where a and b are the extremities of the function as shown in Figures 2.7a and 2.7b. All other definitions remain as before.

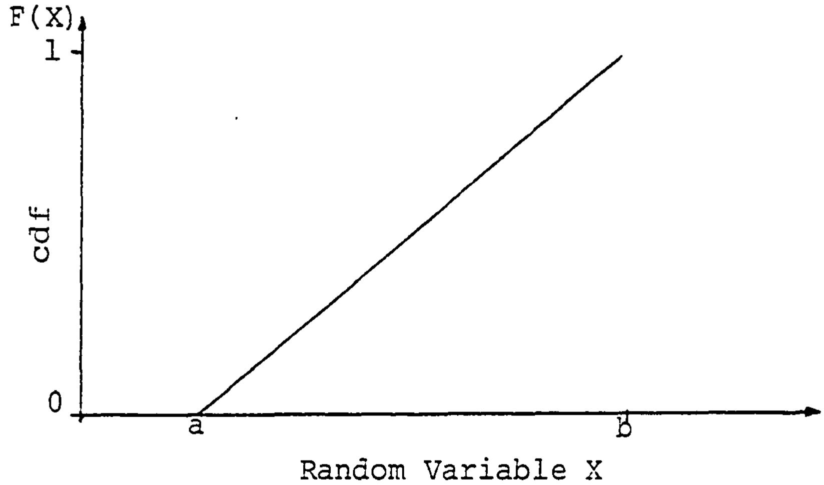


Figure 2.7b. Uniform Cumulative Distribution Function

The criterion of acceptance remains at $D_{.05}(13) = 0.361$. Hence, the uniform distribution is also acceptable to describe the data set. No entry in Column 6 of Table 2.2 is greater than .361. Further testing of data sets from the population are required. Tables 2.3a and 2.3b show a data set tested on normal and uniform distributions. Table 2.3a tests against normality. The limit of acceptance $D_{.05}(16) = .328$ is barely satisfied at $i = 10$, for acceptance of normality.

Table 2.2. Kolmogorov-Smirnov test
Assuming Uniform Density Function
For Data Set

(1) i	(2) X_i	(3) $\sum_{i=1}^{n} f_i$	(4) $G(X) = \frac{\sum_{i=1}^{n} f_i}{n}$	(6) $F(X_i)$	(7) D_i
1	744	1	0.077	0.000	0.077
2	787	2	0.154	0.108	0.046
3	803	3	0.231	0.148	0.083
4	858	4	0.308	0.362	0.131
5	860	5	0.385	0.291	0.094
6	922	6	0.460	0.447	0.062
7	958	7	0.538	0.538	0.078
8	1002	8	0.615	0.648	0.110
9	1051	9	0.692	0.771	0.156
10	1054	10	0.769	0.779	0.087
11	1077	11	0.846	0.837	0.068
12	1123	12	0.923	0.952	0.106
13	1142	13	1.000	1.000	0.077

Table 2.3a. Kolmogorov-Smirnov Test For Marginal Acceptance of Normality

i	X_i	$\sum_{i=1}^{n} f_i$	$G(X)=\frac{\sum_{i=1}^{n} f_i}{n}$	$\frac{X_i-\hat{\mu}}{\sigma}$	$F(X)=\Phi(X)$	D_i
1	577	1	0.0625	-0.93	0.1762	0.1762
2	578	2	0.1250	-0.92	0.1788	0.1163
3	599	3	0.1875	-0.70	0.2420	0.1170
4	603	4	0.2500	-0.66	0.2546	0.0671
5	606	5	0.3125	-0.63	0.2643	0.0482
6	611	6	0.3750	-0.58	0.2810	0.0940
7	616	7	0.4375	-0.52	0.3015	0.1360
8	622	8	0.5000	-0.46	0.3228	0.1772
9	625	9	0.5625	-0.43	0.3336	0.2289
10	633	10	0.6250	-0.35	0.3632	0.2618
11	666	11	0.6875	0.00	0.5000	0.8750
12	706	12	0.7500	0.42	0.6628	0.0872
13	735	13	0.8125	0.72	0.7642	0.0483
14	756	14	0.8750	0.94	0.8264	0.0486
15	800	15	0.9375	1.40	0.9192	0.0442
16	924	16	1.0000	2.70	0.9965	0.0590

Table 2.3b. Kolmogorov-Smirnov Test for Rejection of Uniform Distribution.

i	X_i	$\sum_{i=1}^{n} f_i$	G(X)	F(X)	D_i
1	577	1	0.0625	0.0000	0.0625
2	528	2	0.1250	0.0029	0.1221
3	599	3	0.1875	0.0634	0.1241
4	603	4	0.2500	0.0749	0.1751
5	606	5	0.3125	0.0836	0.2289
6	611	6	0.3750	0.0980	0.2770
7	616	7	0.4375	0.1124	0.3251
8	622	8	0.5000	0.1297	0.3703
9	625	9	0.5625	0.1383	0.4242
10	633	10	0.6250	0.1614	0.4636
11	666	11	0.6875	0.2565	0.4310
12	706	12	0.7500	0.3718	0.3782
13	735	13	0.8125	0.4553	0.3572
14	756	14	0.8750	0.5158	0.3592
15	800	15	0.9375	0.6390	0.2985
16	924	16	1.0000	1.0000	0.0625

However, in the test against the uniform distribution shown in Table 2.3b, the acceptance limit $D_{.05}(16)$ is exceeded at the points indicated by an asterisk. Hence, for this data set, the uniform distribution has been rejected. Normality continues to be accepted.

Figure 2.8 shows a comparison of the empirical cdf and the assumption of normality on the data set in column 2 of Table 2.1. The two curves show a close fit for the amount of data at hand. In conclusion, the conditional density function of the load, at a given temperature, may be described by a normal density function.

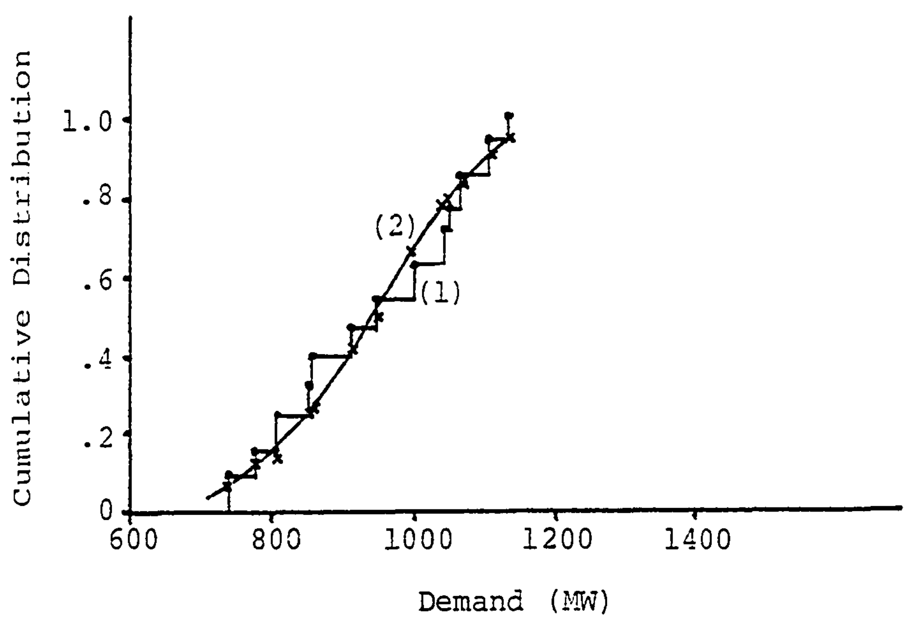


Figure 2.8. (1) Empirical Cumulative Distribution,
(2) Normal Cumulative Distribution

CHAPTER 3

TEMPERATURE DISTRIBUTION

A. Introduction

In Chapter 2, a distribution of temperature f(T) was described in Equation 2.2. In deriving this distribution, an assumption had to be made about the nature of f(T) from year to year. In this study, it will be assumed that the earth's atmosphere is neither cooling nor warming appreciably. Hence a time stationary distribution, is assumed.

Several definitions of f(T) may be made, each requiring a different data set. Each definition depends on the time frame being considered. If the temperature data set is the set of daily maximum temperatures, then f(T) describes the probability of T being the daily maximum temperature. If the data set is the set of all annual peak temperatures, f(T) describes the probability of T being the yearly maximum temperature. Daily maximum temperature data from 1962 through 1976 were used.

Table 3.1 shows for area 1, (see appendix B) all positive temperatures between 0 and 110°F. The table is self-explanatory. The column are as defined previously in Chapter 2. The table shown here may be tested for normality.

Table 3.1. Test of Normality of Empirical Temperature Distribution

i	X_i	n_i	$\sum_{i=1}^{i=22} n_i$	$\frac{\sum n_i}{n}$	$Z_i=\frac{X_i-\overline{X}}{\hat{S}}$	$\phi(Z_i)$	D_i
1	0- 5	0	0	0.0000	-3.18	0.0007	.0007
2	6-10	0	0	0.0000	-3.18	0.0007	.0007
3	11-15	9	9	0.0016	-2.93	0.0017	.0017
4	16-20	17	26	0.0047	-2.67	0.0038	.0022
5	21-25	32	58	0.0106	-2.42	0.0078	.0031
6	26-30	54	112	0.0204	-2.17	0.0150	.0054
7	31-35	101	213	0.0388	-1.92	0.0274	.0114
8	36-40	173	386	0.0704	-1.67	0.0475	.0229*
9	41-45	209	595	0.1085	-1.42	0.0778	.0307*
10	46-50	283	878	0.1602	-1.17	0.1210	.0591*
11	51-55	340	1218	0.2221	-0.92	0.1788	.0433*
12	56-60	378	1596	0.2911	-0.66	0.2546	.0365*
13	61-65	377	1973	0.3598	-0.41	0.3409	.0498*
14	66-70	481	2454	0.4467	-0.16	0.4364	.0766*
15	71-75	488	2942	0.5366	0.09	0.5359	.0892*
16	76-80	460	3402	0.6205	0.34	0.6331	.0965*
17	81-85	586	3988	0.7273	0.59	0.7224	.1019*
18	86-90	584	4572	0.8339	0.84	0.7996	.0723*
19	91-95	520	5092	0.9287	1.09	0.8621	.0666*
20	96-100	300	5367	0.9788	1.35	0.9082	.0706*
21	101-105	106	5473	0.9982	1.60	0.9452	.053*
22	106-110	10	5483	1.0000	1.85	0.9678	.0322*

The acceptance limit at a significance level of 0.05 is $D_{.05}$ = .0184, which is exceeded at i = 8 and all asterisked points. Hence, normality is rejected. Such a statistical test determined that the data set may not be described under a normal distribution. Further analysis, as shown in Figure 3.1, revealed that a normal distribution is a very poor fit to the empirical distribution, leaving room for further improvement. However, the shape of the boundary of the density function f(T) indicated possible fit using well-defined standard statistical functions. The densities in Figure 3.1 show a skewness to the left.

Upon comparison with several standard statistical functions, the beta function was selected for trial. Selection was based on the fact that the same left-handed skewnew of Figure 3.1 can be approximated by the beta function. The following is a brief review of the beta density function.

B. Beta Density Function

The beta density function is given by

$$f(X) = \frac{1}{B(\alpha,\beta)} X^{\alpha-1}(1 - X)^{\beta-1} \qquad 0 \leq X \leq 1$$

where $B(\alpha,\beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta)}$ a constant for a given α, and β;

and where Γ is a gamma notation, and α and β are parameters of the beta density function.

The shape of the density function depends on the choice of α and β. For example, when $\alpha = \beta = 1$, then $f(X) = 1$, a constant, which is the uniform density function.

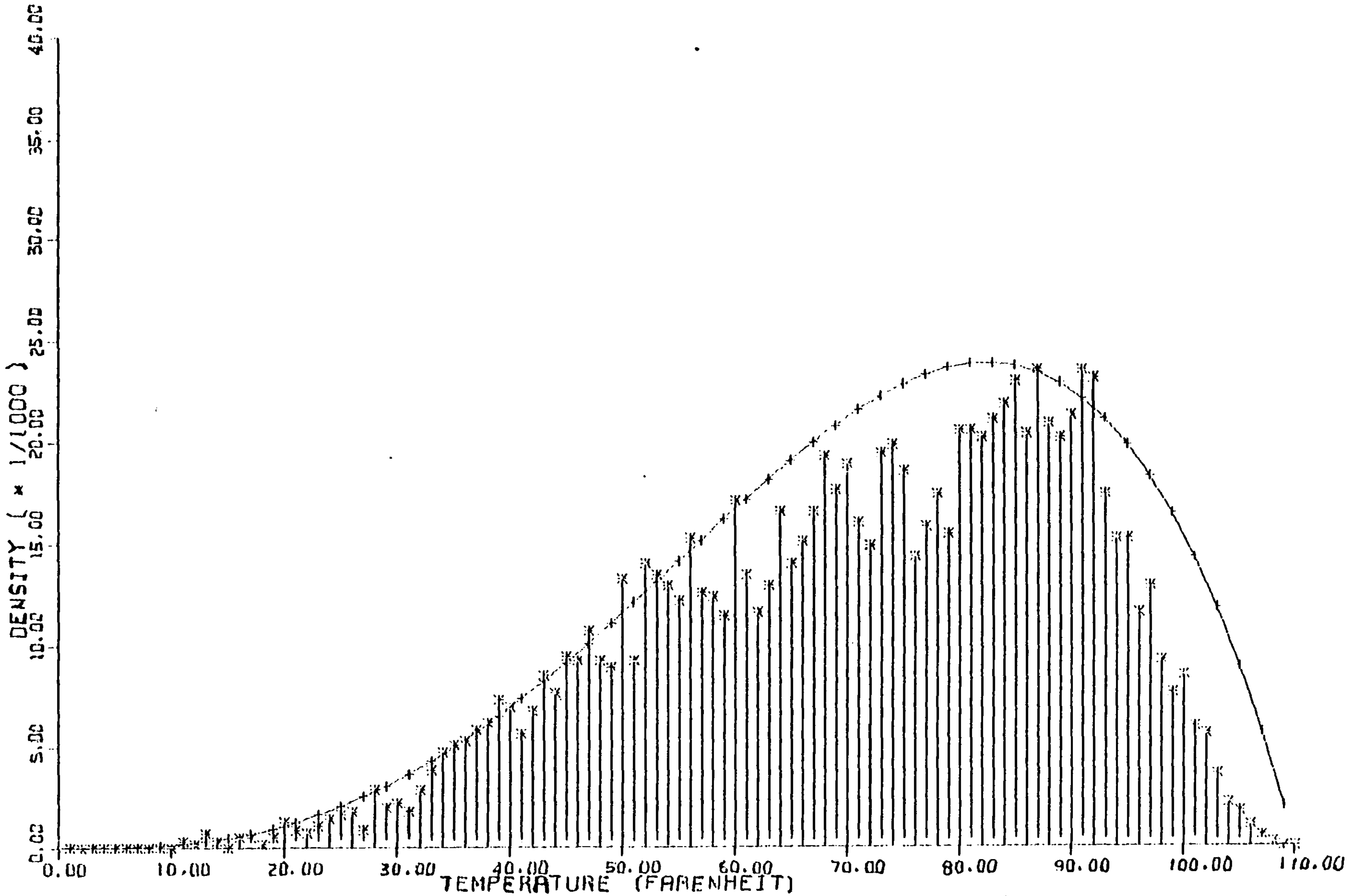


FIGURE 3-1 Temperature Density Function for Area 1

The mean, μ, and varance, Var(X), of the beta density function were found using the moment generating function, thus

$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

A choice of $\alpha > \beta > 1$ yields a left skewed beta density function of the form shown in Figure 3.2.

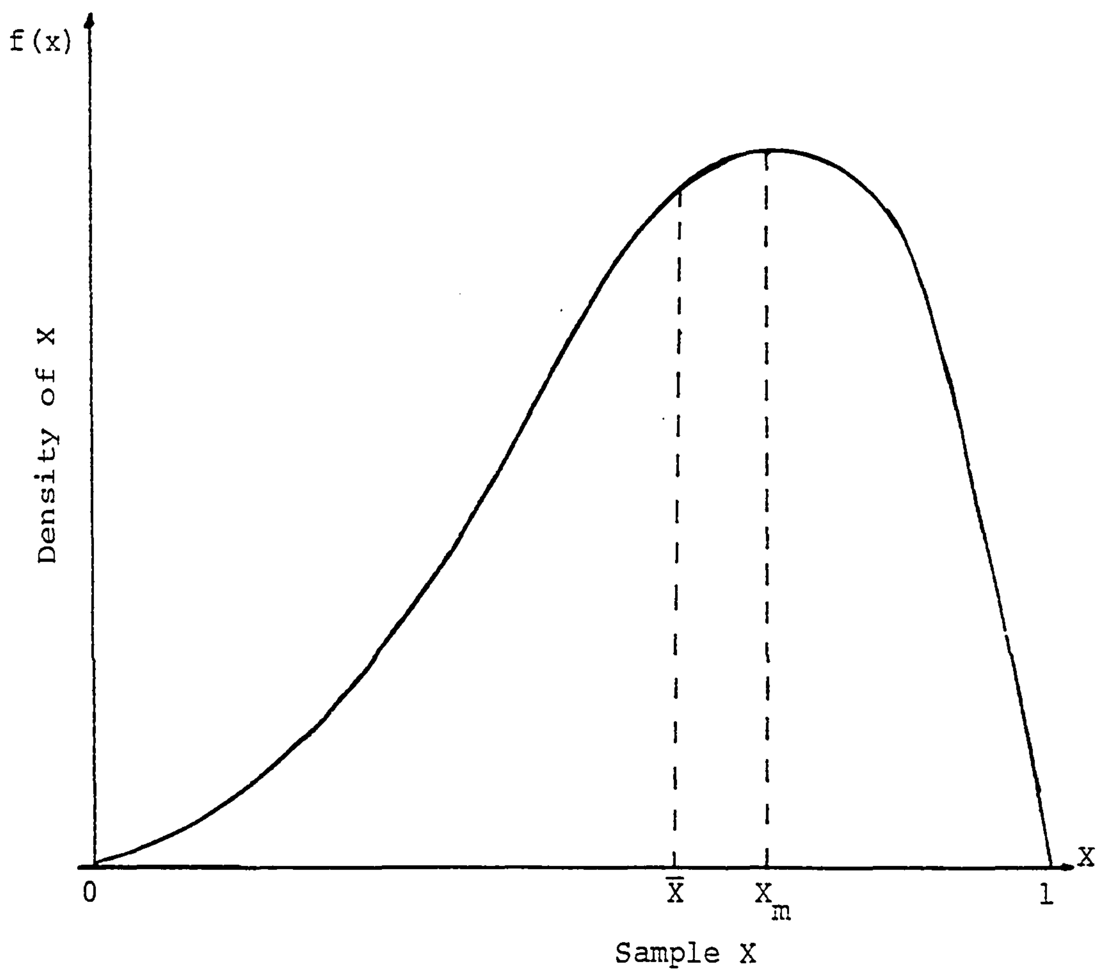


Figure 3.2. A Typical Beta Function for $\alpha > \beta > 1$.

The point of maximum density X_m is given by

$$X_m = \frac{\alpha - 1}{\alpha + \beta - 2}$$

and is to the right of the mean value $\overline{X}$.

The density function shown in Figure 3.1, was approximated using a trial and error method in the selection of values for α and β. The smooth curve shows this approximation, using $\alpha = 4$ and $\beta = 2$. These values for α and β yield a density function given by

$$f(X) = 20\, X^3\, (1 - X) \qquad 0 \leq X < 1$$

Since the beta function lies between 0 and 1, a transformation of variables has to be made in order to satisfy the boundary conditions. In this example, the transformation used is given by

$$X = \frac{T}{110} \qquad 0 \leq T < 110$$

where T is temperature in $^{\circ}$F.

Having defined the density function by an approximate curve fit, the cumulative distribution function F(T) is easily calculated as

$$F(T) = \int_0^{\frac{T}{110}} 20Z^3(1 - Z)\,dZ$$

$$= \left(\frac{T}{110}\right)^4 \cdot \left(5 - \frac{4T}{110}\right)$$

or $F(T) = X^4(5 - 4X)$

where X is as defined previously.

The above example was done using data from one of the areas under study. Figures 3.3 through Figures 3.6 show the temperature density functions for the other areas. Figure 3.7 shows an overlay, or composite density distribution function, which is an average of all the areas together. It is a hypothetical distribution function. The estimation of f(T), using an appropriate curve fit as demonstrated above, completes the calculation of the right-hand side of the equation,

$$F(P \cap T) = F(P/T) \cdot f(T)$$

where P is demand in megawatts and T is temperature, described in Chapter 2.

The objective to this point has been to define the natures of F(P/T), the conditional distribution of a load level P, at a given temperature T, and of f(T), the distribution function of temperature T. Hence, the distribution function of load P and temperature T is known.

$F(P \cap T)$ may be calculated at temperatures T_o to T_n, as shown in Figure 3.8.

The probability of reaching and exceeding load level P megawatts is thus given by

$$F(P) = \sum_{T=T_o}^{T=T_n} F(P \cap T)$$

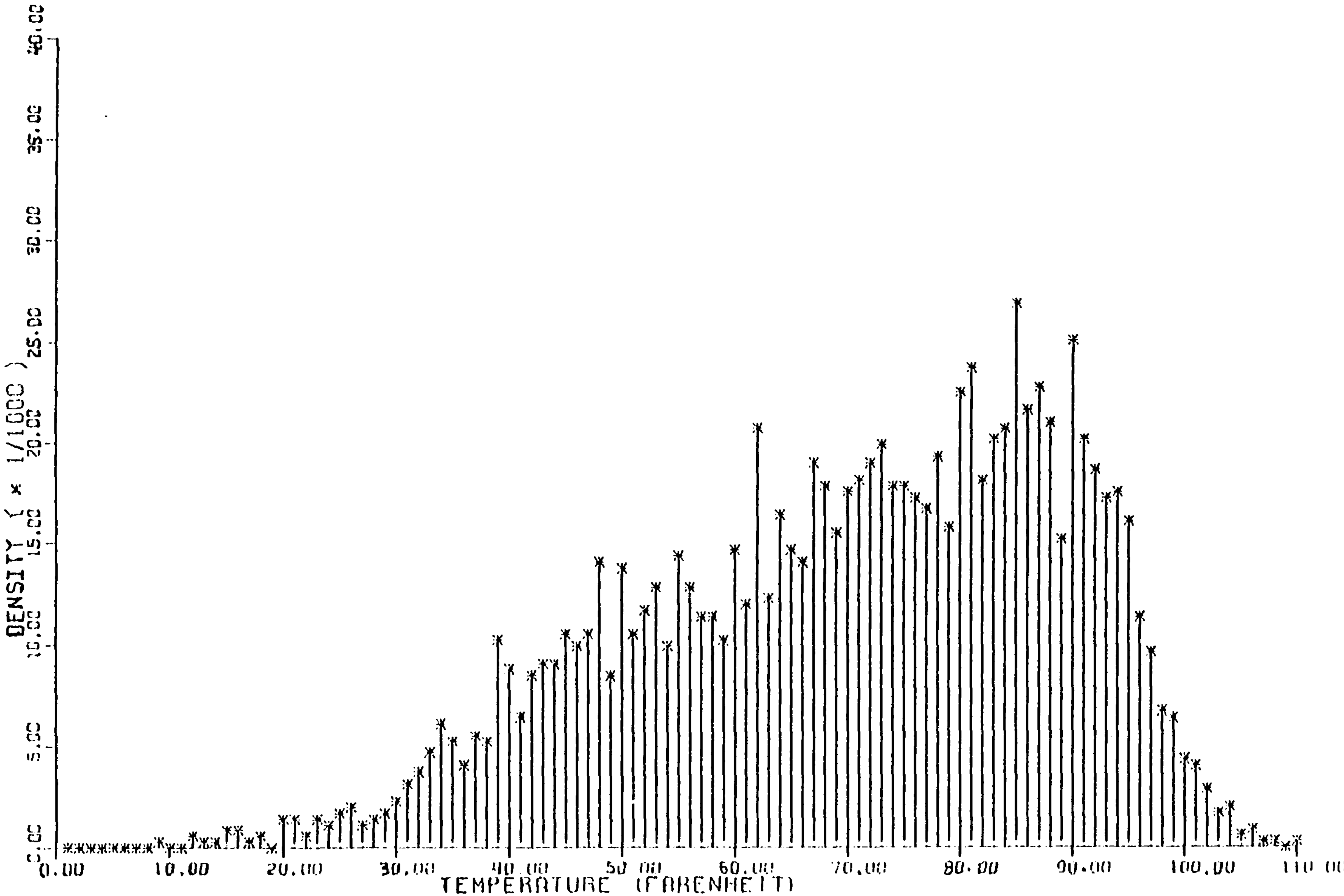


FIGURE 3-3 Temperature Density Function for Area 2

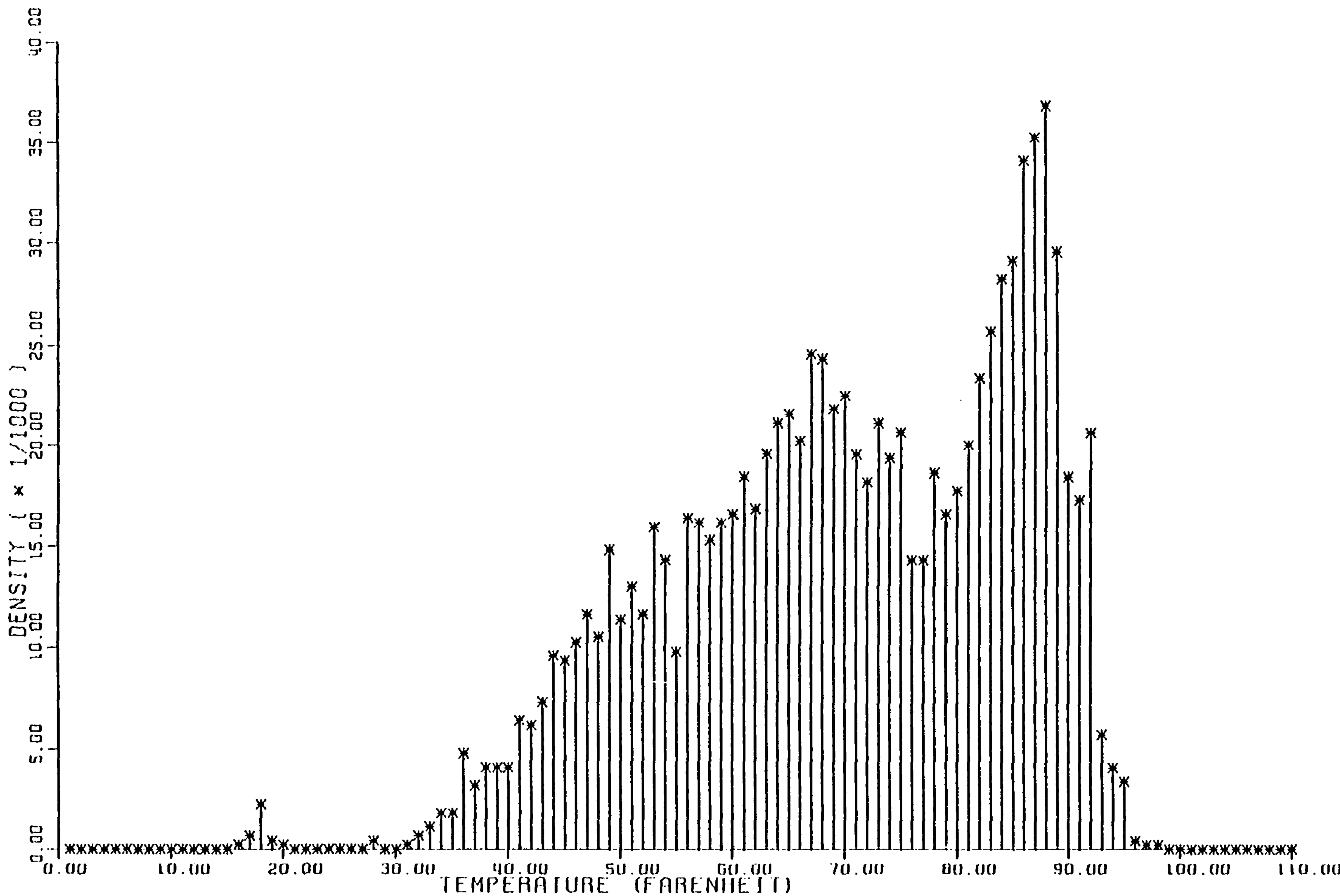


FIGURE 3-4 Temperature Density Function for Area 3

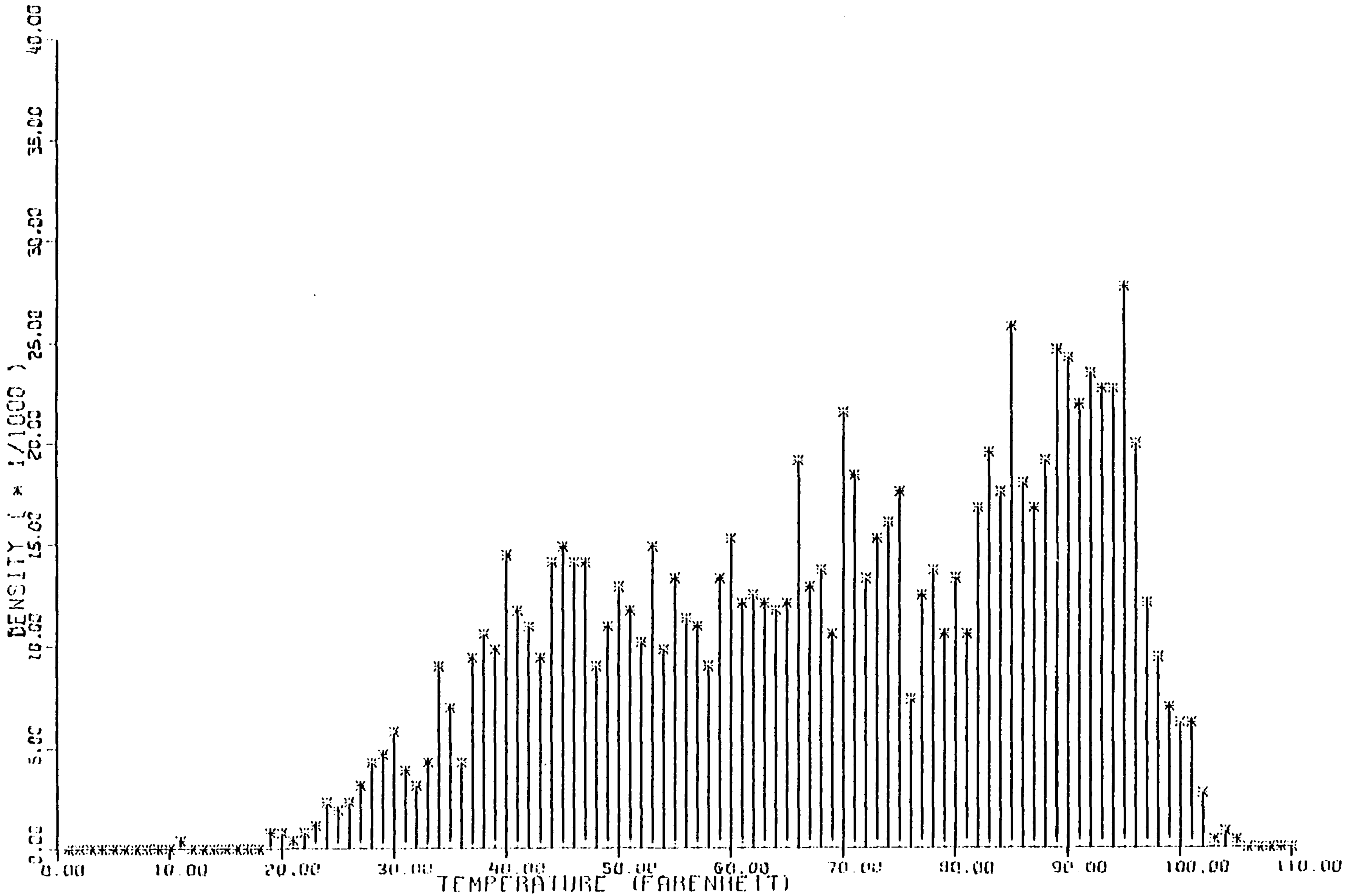


FIGURE 3-5 Temperature Density Function for Area 4

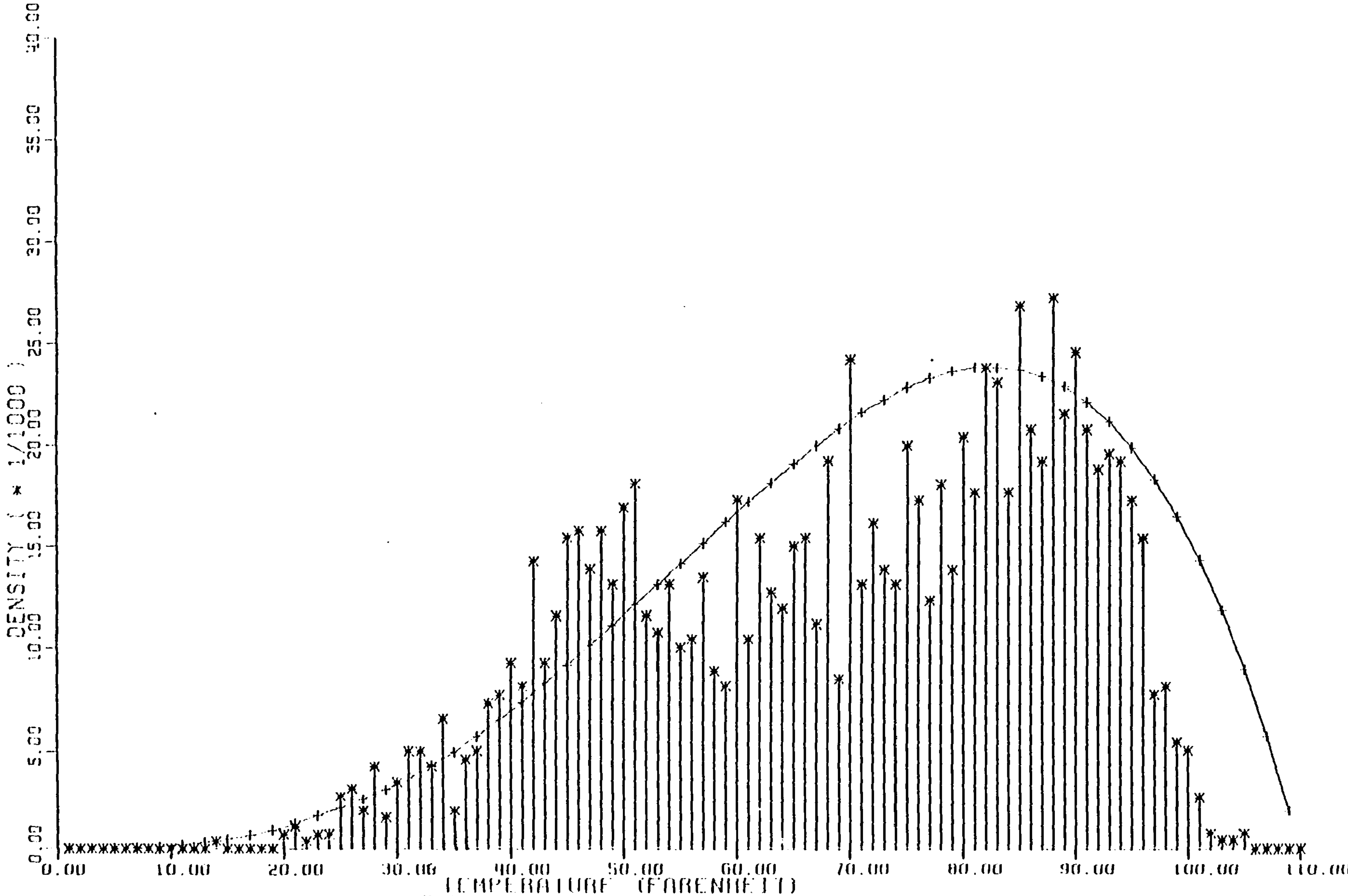


FIGURE 3-6 Temperature Density Function for Area 5

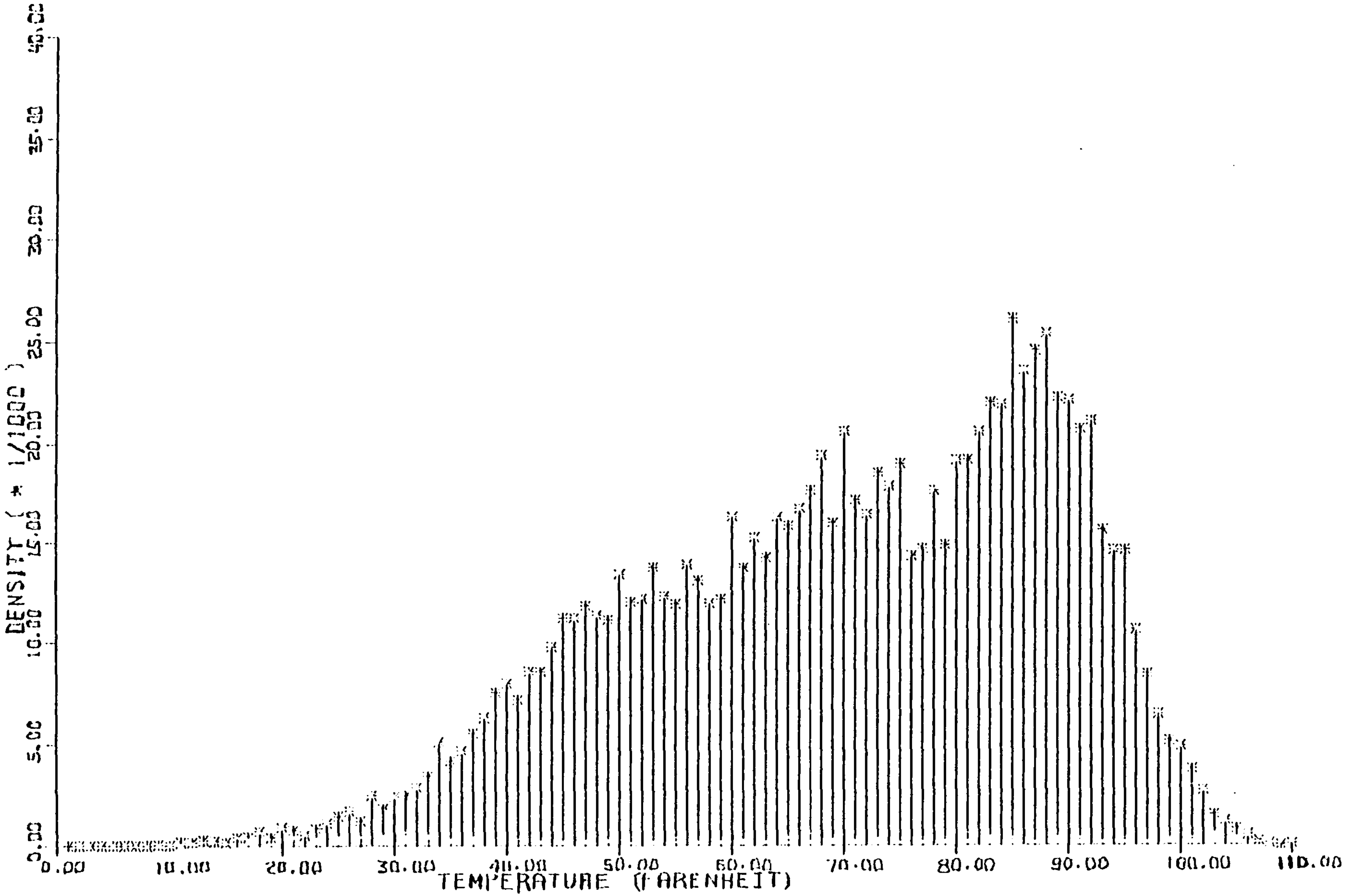


FIGURE 3-7 Temperature Density Function for Whole Region

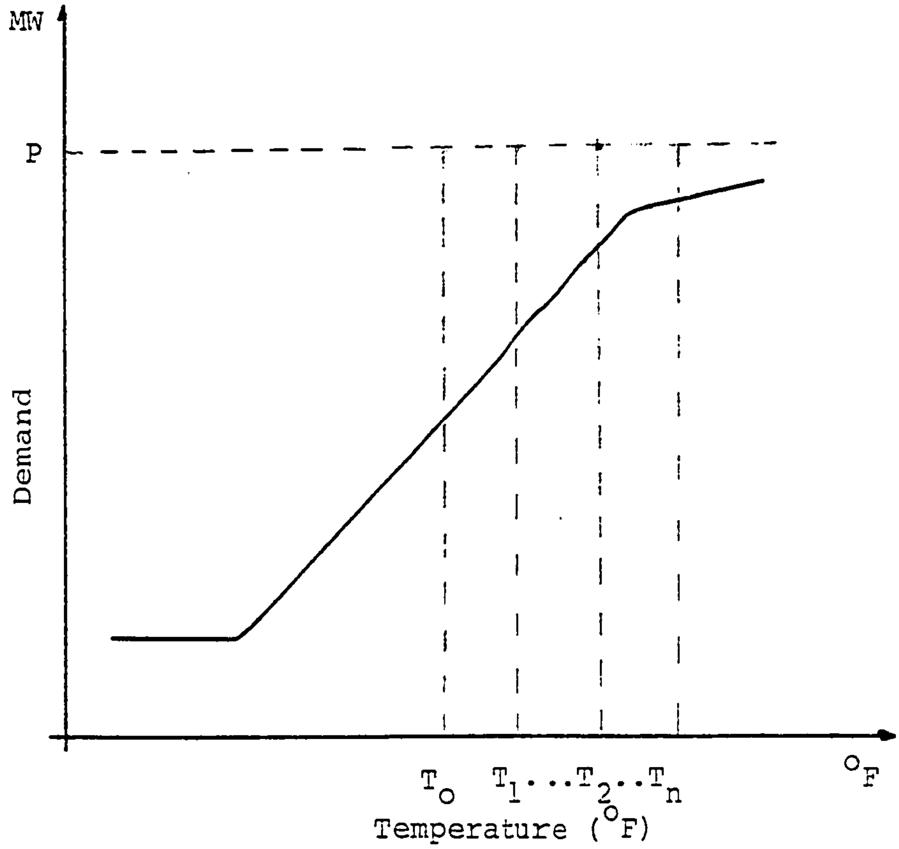


Figure 3.8. Sample Points for the Calculation of $F(P \cap T)$.

or $$F(P) = \sum_{T=T_o}^{T=T_n} F(P/T) \cdot f(T)$$

The probability densities of temperature f(T) and the average loads or forecasts for the year being studied, were determined and stored on disk. These were used in the computer program listed in Appendix A to calculate the probabilities of load occurrence, assuming the standard normal distribution previously accepted. For example, let the peak demand forecast for 1970 be 2163 MW. The program calculation showed that this level of load has a probability of 0.0157 of being reached or exceeded in that year. This level of load has a low chance of occurrence, and the supply situation is the criterion for acceptance of this load level for planning purposes.

Factors such as the availability of adequate external supplies when needed, and the cost of capital will aid in the determination of the most feasible supply option. Thus, the probability levels are measures of risk, as well as an index to capital expenditures.

CHAPTER 4

MULTI-AREA DEMAND CHARACTERISTICS

A. Introduction

In this chapter, data from each area under study are discussed in greater detail. An attempt is then made to state, in broad terms, what the characteristics of most utilities in similar areas can be expected to be. Finally, a composite characteristic of all regions is developed, and its statistical parameters calculated and checked for conformity with the results of Chapter 2.

Figures 4.1 through 4.5 are representative plots of peak daily load versus peak daily temperature for geographical areas 1 through 5 which are climatically very similar geographical areas. Area 1 and 2 interleave, as do areas 3 and 4. Area 5 has a warmer weather system during most of the year. Descriptions of the characteristics of these plots are as stated at the beginning of Chapter 2. Sections of base load, temperature dependence and saturation are evident on most of these plots.

Composites of these plots will be built in stages. Two approaches may be used in the combination of the daily data points to yield a total multi-area demand.

B. Summation by Temperature

Plots, such as those shown in Figure 4.1 and Figure 4.2 for the same year, are overlaid. Such an overlay is shown in Figure 4.6. The points of match in the overlay are temperatures. The demand obtained in this case is not the sum of the demands in the two areas, but an average demand for both areas. It is obvious that this approach is useful only when the demands and weather sensitivity of the demand in the two areas are quite similar. If this condition is not met plots of the data points will be so widely spread out that no meaningful trends will be noticeable. An advantage of this method is that the number of data points at each temperature is significantly increased, giving a more solid base for statistical analysis.

C. Summation by Date of Occurrence

In this case, the number of data points stays the same for any given year. However, the ordinate of the graph is the sum of demands in the two areas on any given date. The problem here is that two temperature values, one for each area, are now to be considered. A direct overlay of plots is not possible. A new temperature for the composite load has to be calculated and weighted to account for differences in the size of the loads. It has been noted before that the total area system load has a base load and heat sensitive component expressible as

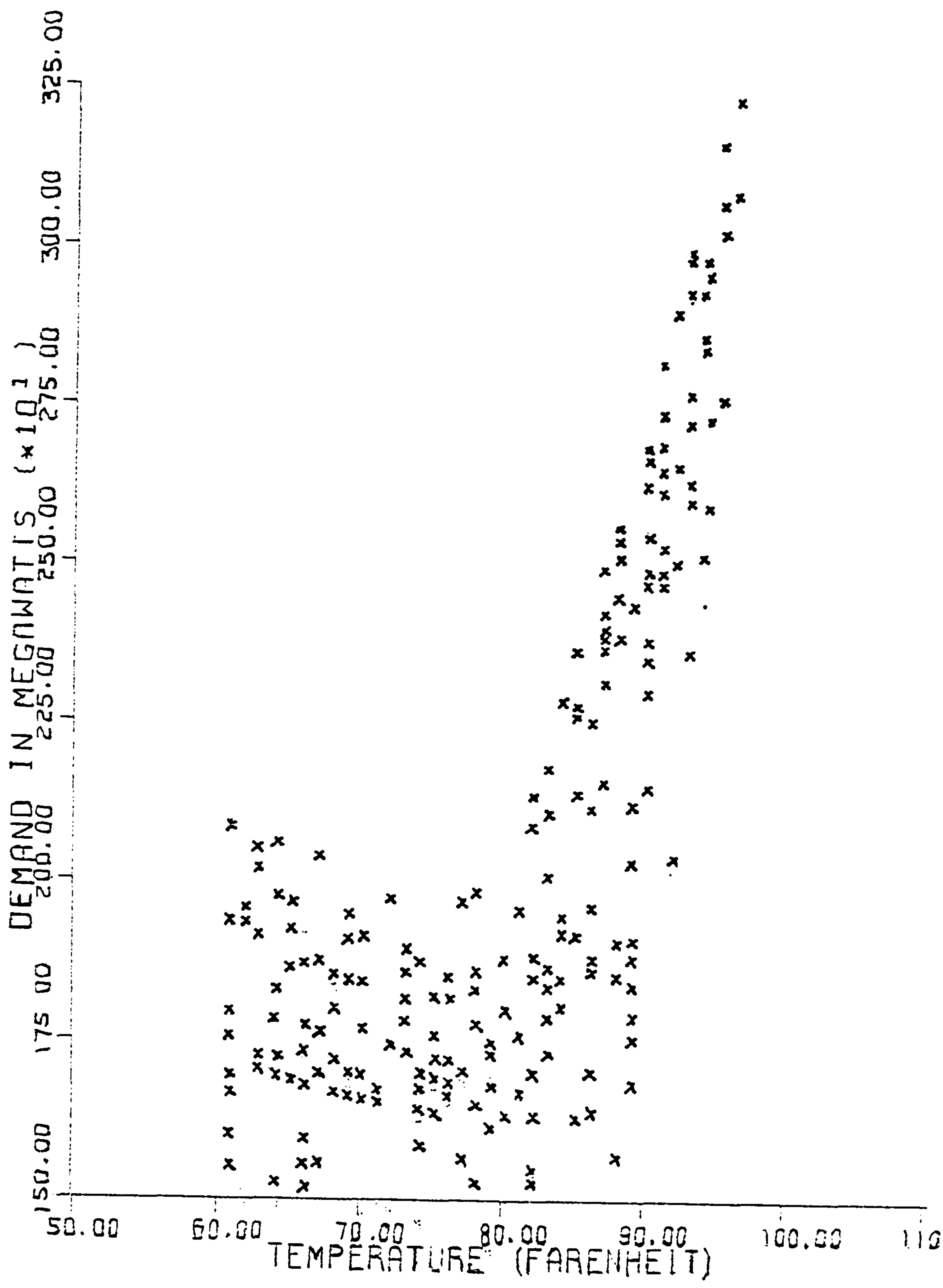


Figure 4.1. Demand Characteristic Area 1.

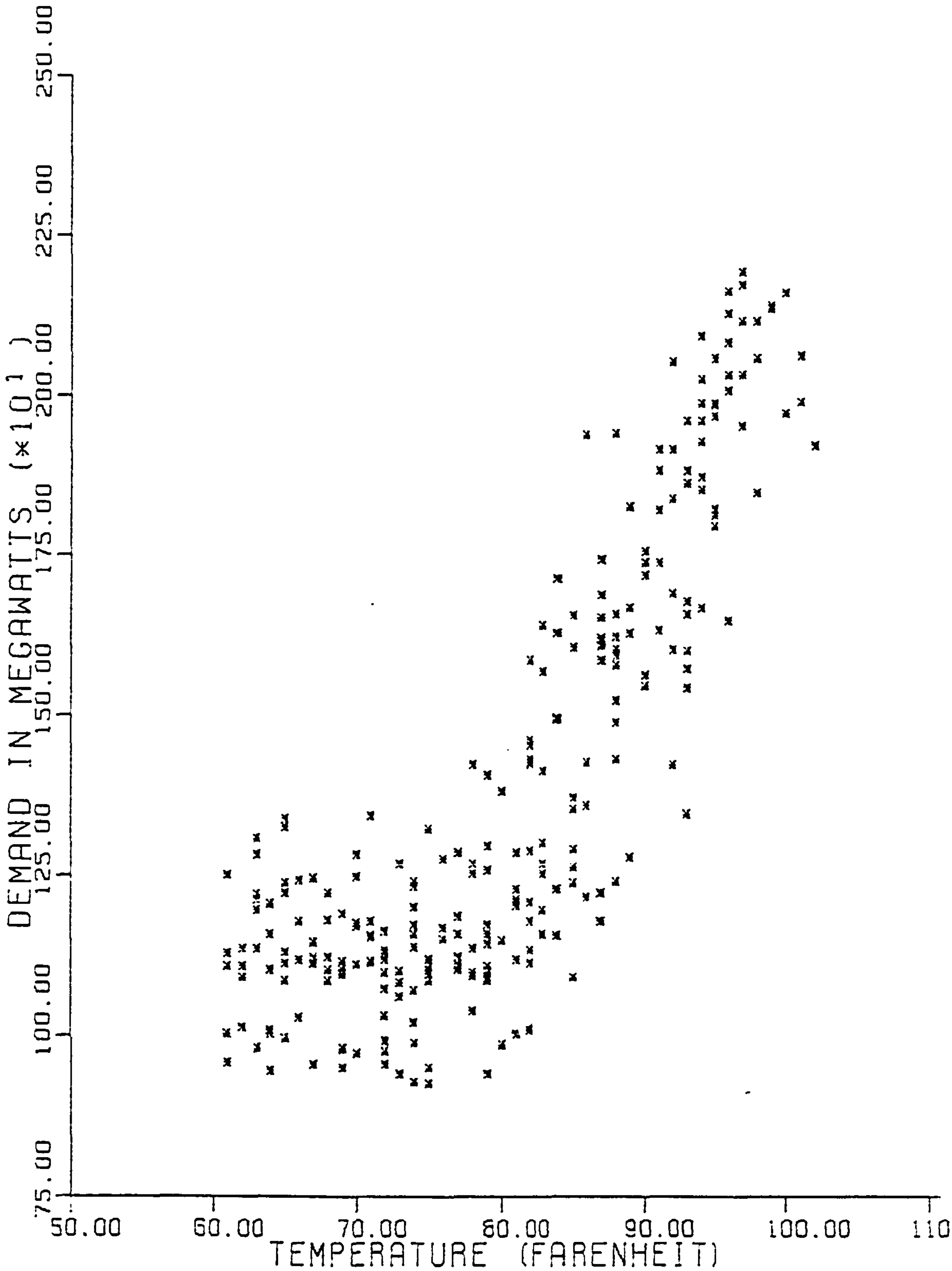


Figure 4.2. Demand Characteristic Area 2.

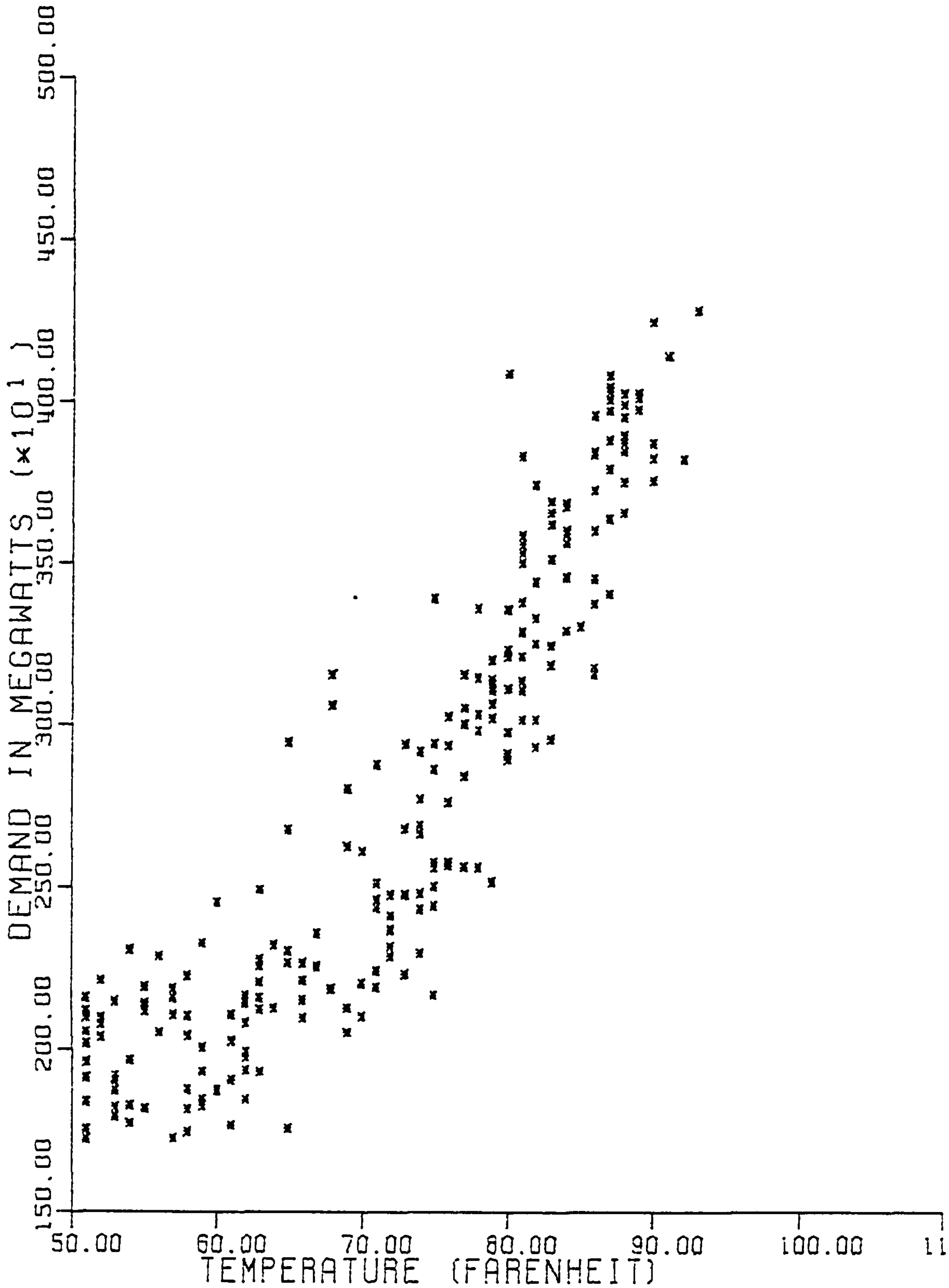


Figure 4.3. Demand Characteristic Area 3.

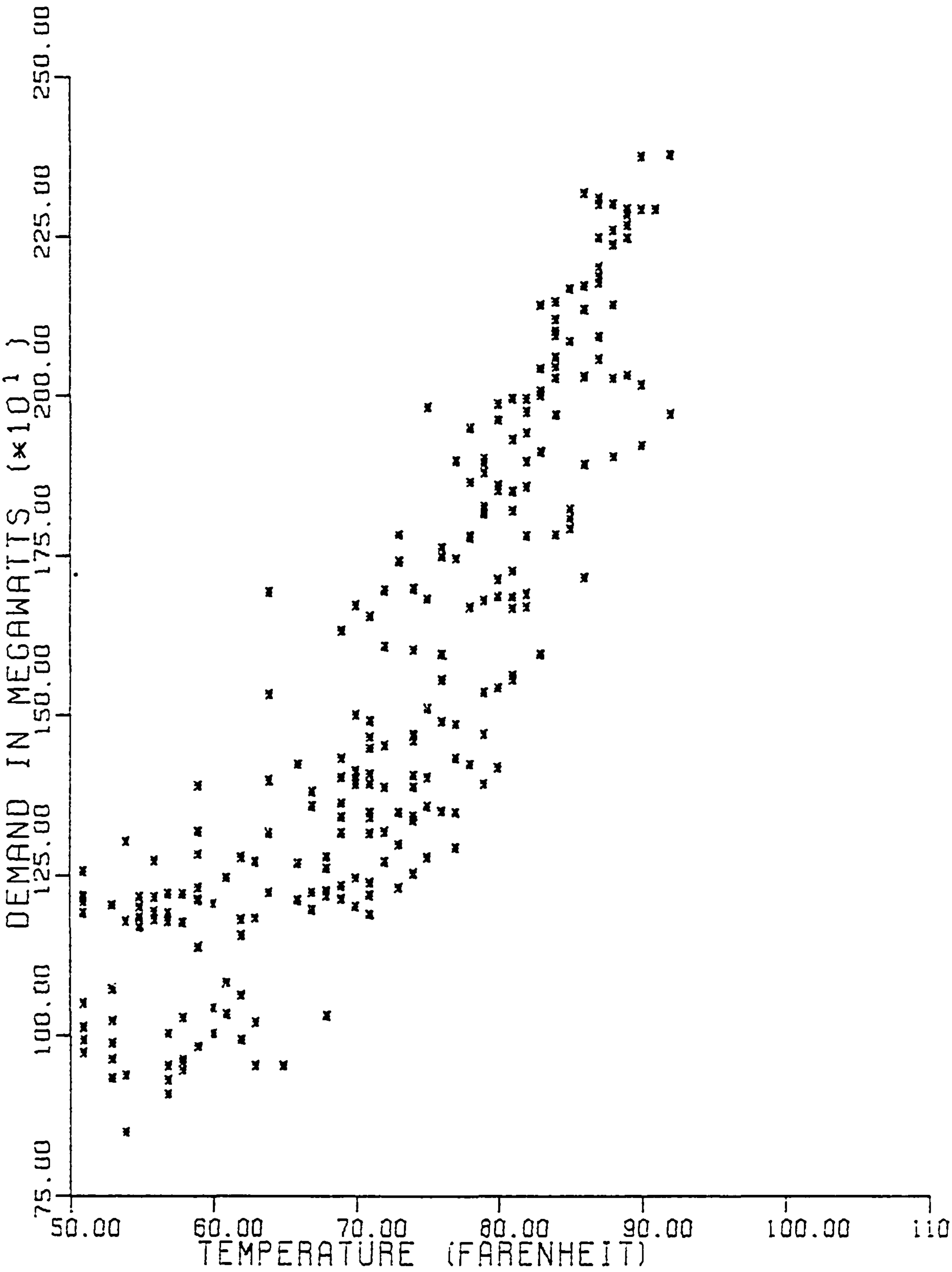


Figure 4.4. Demand Characteristic Area 4.

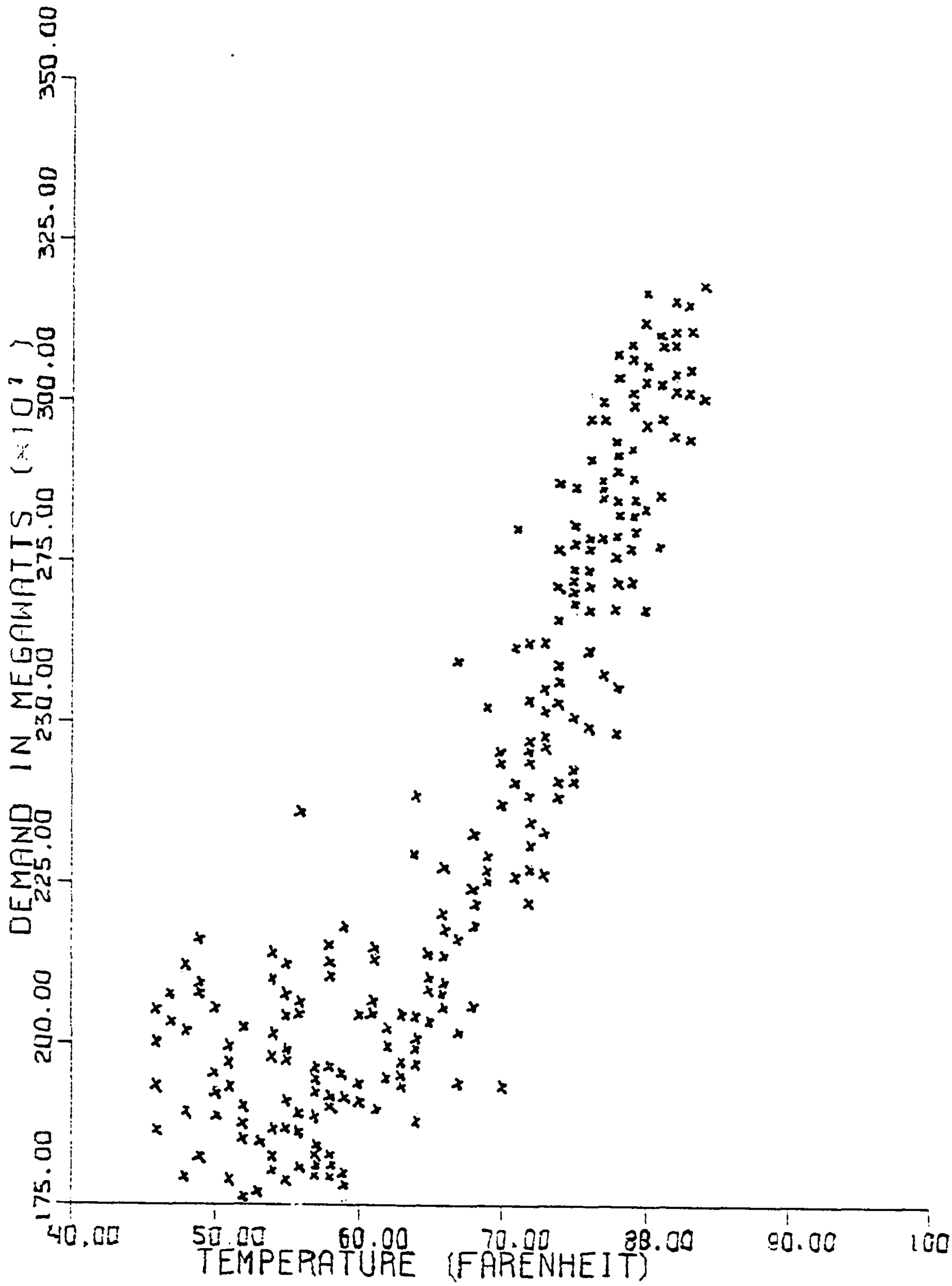


Figure 4.5. Demand Characteristic Area 5.

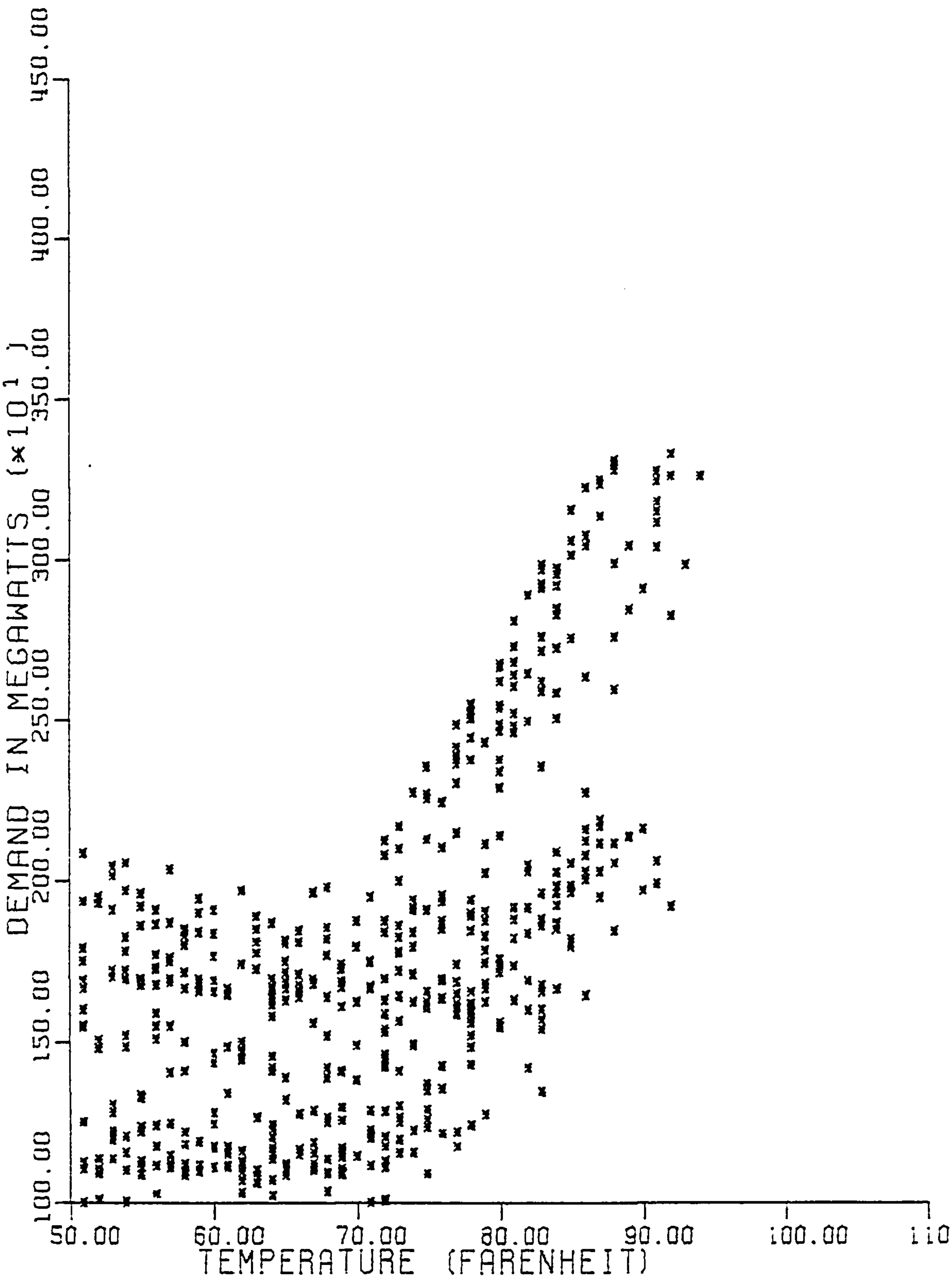


Figure 4.6. Demand Characteristic Area 1 and 2 by temp.

$$P = B + H(T) \tag{4.1}$$

where P is the total area load, B is the base load, H(T) is the temperature sensitive component and T is the temperature in ^{o}F.

In deriving the temperature to use, only the heat sensitive component will be used. The parameters that determine the level of base load are largely independent of temperature or weather. In the calculation of the resultant temperature for the two areas, it is necessary to give more weight to the area with the higher temperature sensitive demand. Given a date and two areas, 1 and 2, the two-area temperature is calculated as

$$T = \frac{H_1(T_1) \cdot T_1 + H_2(T_2) \cdot T_2}{H_1(T_1) + H_2(T_2)} \tag{4.2}$$

The resultant temperature T, is weighted in the direction of the area with the higher heat sensitive component, in the transformation which may also be used to obtain a temperature, such as T_1 or T_2, for an area using demands and temperatures from sub-areas on a distributed basis.

The total demand is retrieved as

$$P = B_1 + B_2 + H_1(T_1) + H_2(T_2) \tag{4.3}$$

$$P = P_1 + P_2 \tag{4.4}$$

Using the data sets of Figure 4.1 and Figure 4.2, the resultant two-area demand is shown plotted in Figure 4.7.

A comparison of Figure 4.7 to Figures 4.1 and 4.2 shows a strong similarity in characteristics. These figures show sections of base load followed by linear dependence. This characteristic describes the combined demand within the two service areas taken as one. In order to verify the hypothesis that a normal distribution function can be used to describe the data set of a given temperature, a Kolmogorov-Smirnov test was made and the results are presented in Table 4.1 for a given example. Further analysis on probability graph paper yielded an approximate straight line, as shown in Figure 4.8, which reinforced the test results of Table 4.1. The composite model is thus derived from its components.

The above analysis was repeated, using data from areas 3 and 4, which are adjacent to each other, and serve a similar customer mix. Figure 4.9 is a plot of the composite two-area model for areas 3 and 4. Again, there is an evident similarity in characteristics (see Appendix C). Finally, all five areas were combined to produce a five-area composite demand characteristic, shown in Figure 4.10. Total demand continued to conform to the separate demands for each utility. Figure 4.10 shows a five-area composite characteristic. Methods developed for the analysis of the separate demands can, therefore, be applied to the composite model without loss of generality.

This yields an effective tool for system planners on

TABLE 4.1
Kolmogorov-Smirnov Test of Composite Data Set

(Critical acceptance limit at 5% significance level is 0.391.)

i	X_i	n_i	$\sum_{i=1}^{n=10} n_i$	$\frac{\sum_{i=1}^{n} n_i}{n}$	$z_i = \frac{X_i - \overline{X}}{\hat{\sigma}}$	$\phi(z)$	$\lvert D \rvert$
1	1790	1	1	0.091	-1.193	0.117	0.117
2	1848	1	2	0.182	-0.864	0.195	0.104
3	1878	1	3	0.273	-0.693	0.245	0.063
4	1885	1	4	0.364	-0.653	0.258	0.106
5	1941	1	5	0.455	-0.335	0.367	0.088
6	1969	2	7	0.636	-0.176	0.433	0.203
7	2014	1	8	0.727	0.079	0.528	0.199
8	2026	1	9	0.818	0.148	0.560	0.258
9	2310	1	10	0.909	1.760	0.961	0.143
10	2335	1	11	1.000	1.900	0.971	0.062

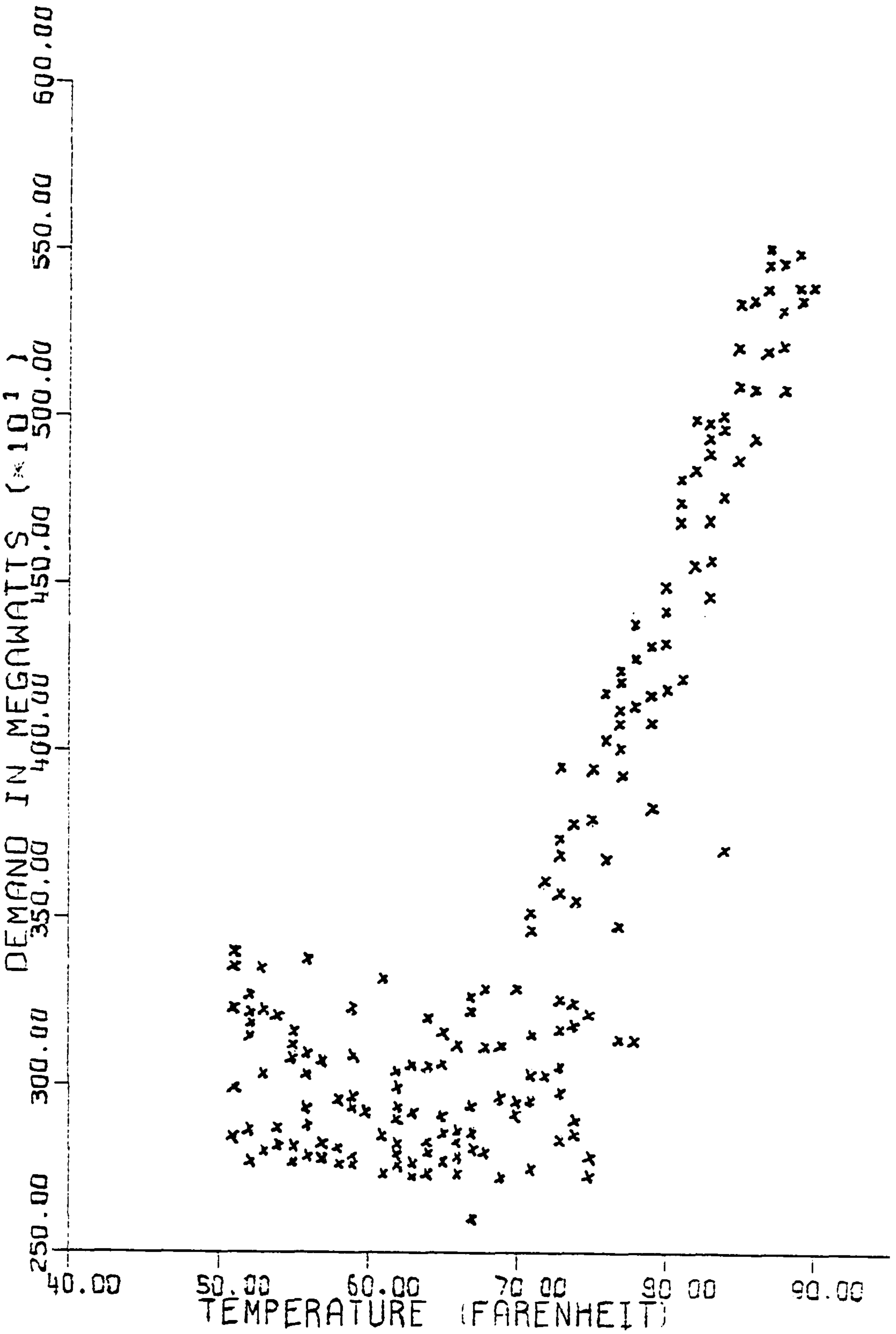


Figure 4.7. Demand Characteristic Area 1 and 2 by date.

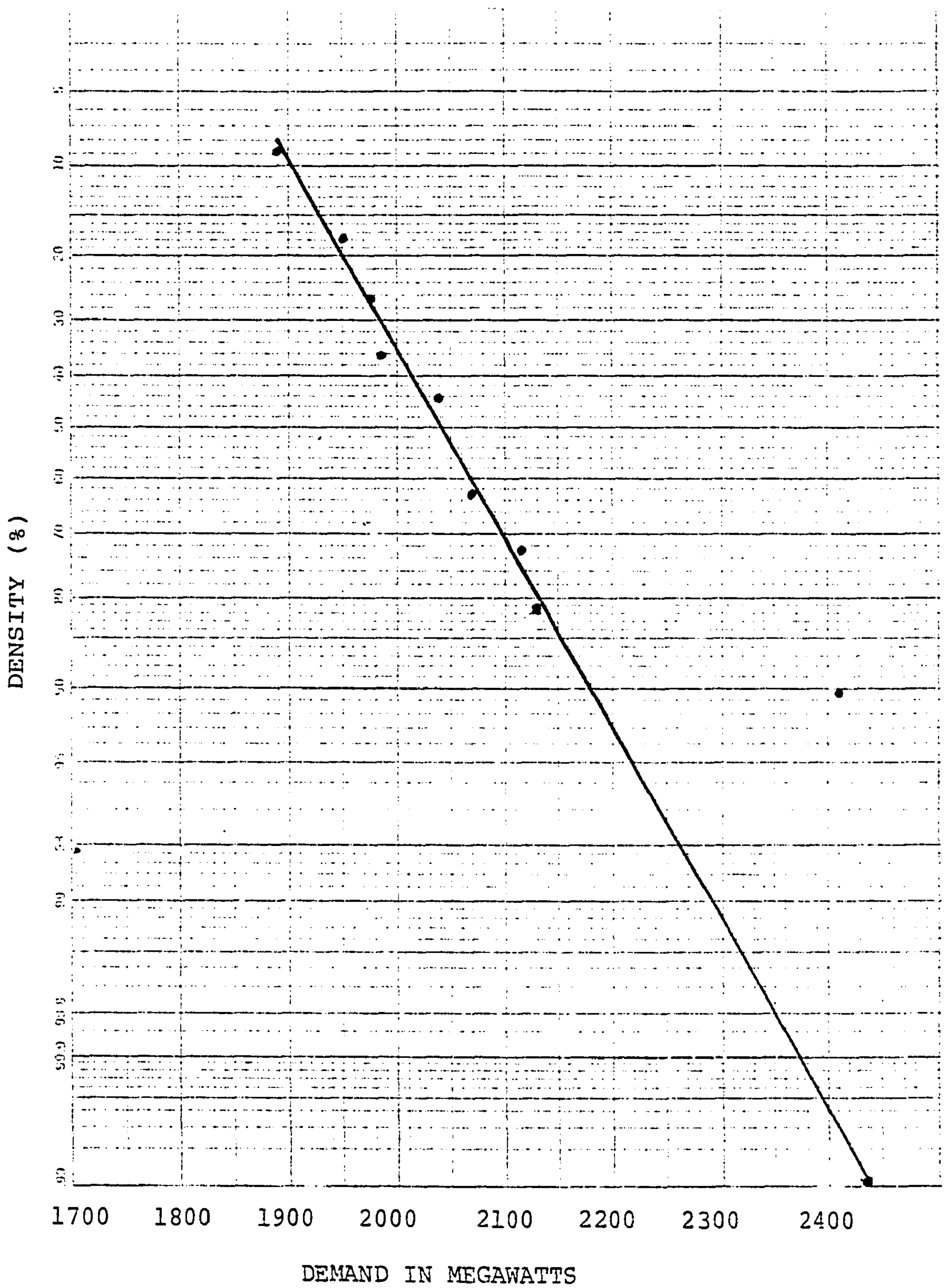


Figure 4.8. Probability Graph of Empirical Data in Table 4.1

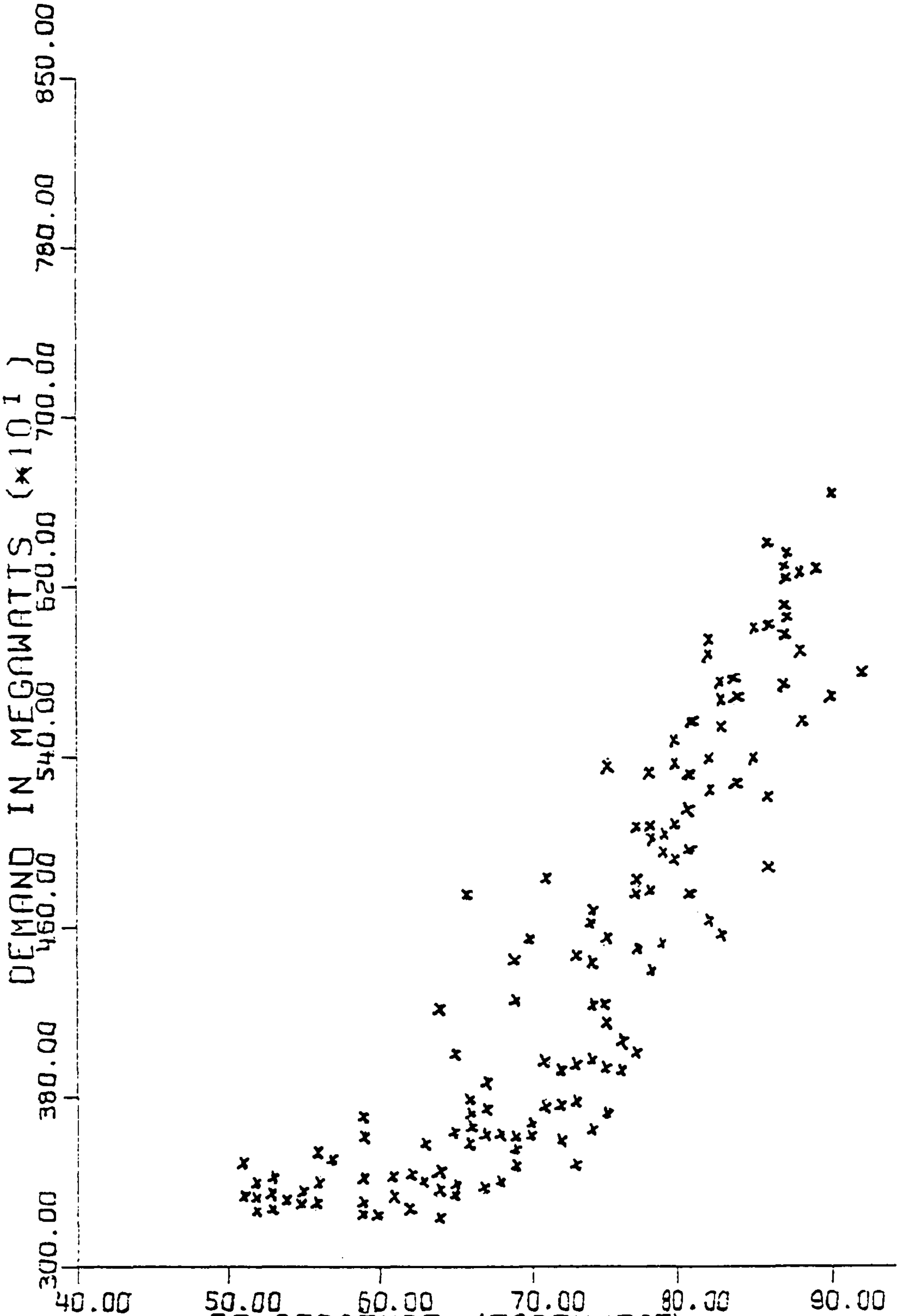


Figure 4.9. Demand Characteristic Area 3 and 4 by date.

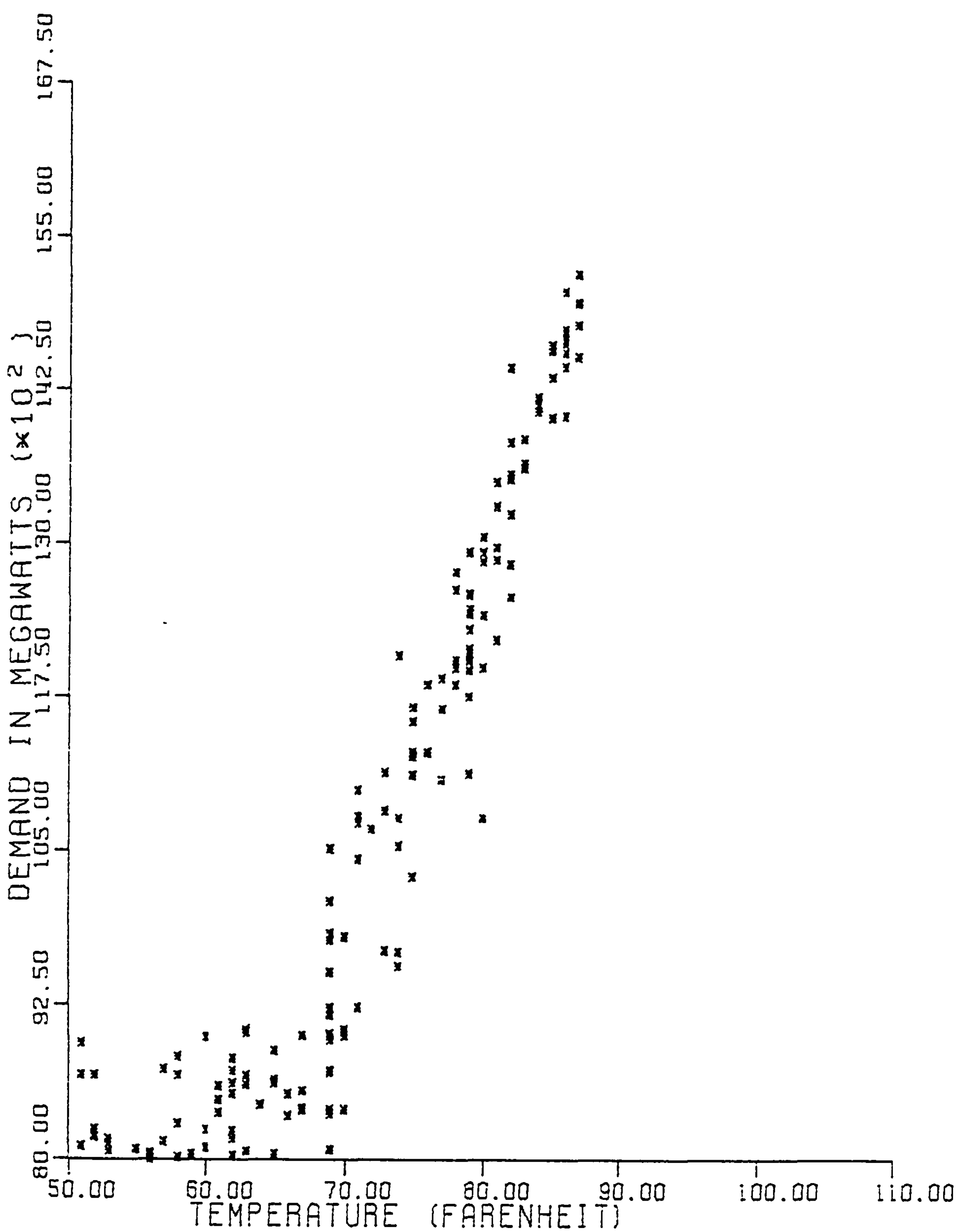


Figure 4.10. Demand Characteristic all areas by date.

a regional basis. Although utilities usually do not plan on a regional basis, they are all members of power reliability pools, where some degree of regional planning is done. The composite demand of any given power pool with the characteristics described for the five-area composite model, described above, can be similarly analyzed. The ever increasing size of generating plants and transmission voltages, and the mounting pressures of environmental groups will make regional planning necessary in the future.

CHAPTER 5

THE FORECASTING MODEL

A. Introduction

The modeling technique used is described in this chapter. A data series is developed using the demand characteristics previously discussed and the theory of time series analysis, particularly Box-Jenkins models, applied to forecasting. Most forecasting methods extract only one coordinate point for each year's data set, thus losing information available in the entire characteristic.

B. Model Development

In previous chapters the form of the demand characteristic was discussed. If the plots of demand versus temperature are placed side by side, in succession, the series shown in Figure 5.1 results. Axis number 1 has temperature as the independent variable extending from $60^{\circ}F$ to $110^{\circ}F$ for each year's data set. Time series analysis is applicable only to a monotonically increasing axis. Axis number 2 transforms axis number 1 to a monotonically increasing axis, with each year spanning a period index of 50 units.

A mathematical model will be formulated for this

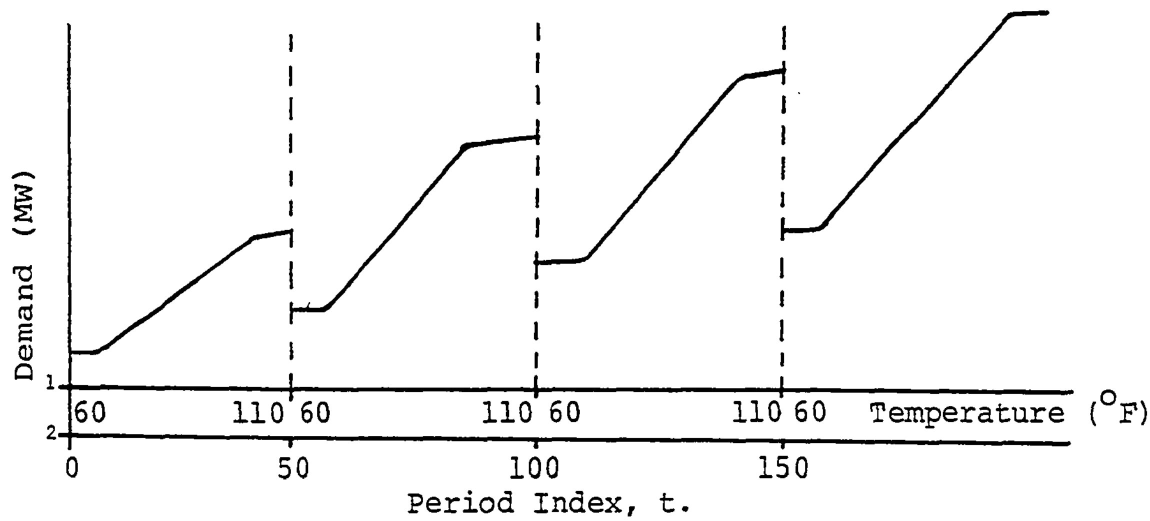


Figure 5.1. Sample Series

series that permits the forecast of the next annual demand characteristic.

A time series may be represented by the linear filter form given as

$$Z_t = \delta + \psi_o u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \ldots + \psi_j u_{t-j} \quad (5.1a)$$

or

$$Z_t = \delta + \sum_{j=o}^{j} \psi_j u_{t-j} \quad (5.1b)$$

which is a discrete linear stochastic process, where δ is the mean of the series. ψ_j are weights associated with u_{t-j}; independent random noise components of the series.

Hence the value of the function Z_t at time t is equal to the mean of the function and a sum of fractions ψ_j, of deviations, u_{t-j}, from a mean δ, at j preceding time periods.

In the analysis Z_t is a time series whose parameters will be identified to obtain a demand model. The method used was initially developed by Box-Jenkins [26]. The method consists of three main parts, namely

1. identification of the nature of the time series, its components, and the order of the components.
2. estimation of the model parameters to completely describe Z_t.
3. diagnostic checking of the model for goodness of fit, against the actual data used, for acceptance or rejection.

The model obtained on acceptance may then be used to forecast.

C. Autocovariance and Autocorrelation [26,30,31]

These are statistical parameters necessary in the identification phase of the model. The prefix "auto" is used to signify the covariance or correlation between observations at different points in the same series, for example, the covariance or correlation between Z_t, Z_{t-1}, or Z_t, Z_{t-j}.

Autocovariance is defined by

$$\gamma_j = E[Z_t - E(Z_t)][Z_{t+j} - E(Z_{t+j})] \quad (5.2a)$$

$$= E[(Z_t - \mu)(Z_{t+j} - \mu)] \quad (5.2b)$$

where

$$\mu = E[(Z_t],$$

the population mean of the function.

The autocovariance depends only on the span, j, between the two observations. γ_j is called the autocovariance of the series at lag j. Variance is a measure of dispersion, and for the analysis of series, is more useful when referred to a reference. γ_o is made the reference, where γ_o is given by

$$\gamma_o = E[(Z_t - \mu)(Z_t - \mu)] \quad \text{i.e., } j = 0 \qquad (5.3)$$

Autocorrelation is defined as

$$\rho_j = \frac{\gamma_j}{\gamma_o}$$

$$= \frac{E[(Z_t - \mu)(Z_{t+j} - \mu)]}{E[(Z_t - \mu)^2]} \qquad (5.4)$$

A plot of the autocorrelation function is called a correlogram.

Since $\mu = E(Z_t)$ may not be easily obtained due to limitations of sample size, the sample average is used.

$$\overline{Z} = \frac{1}{T} \sum_{t=1}^{T} Z_t$$

and the sample autocorrelation is given by

$$\rho_j = \frac{\frac{1}{N-1} \sum_{t=1}^{N-j} (Z_t - \overline{Z})(Z_{t+j} - \overline{Z})}{\frac{1}{N} \sum_{t=1}^{N} (Z_t - \overline{Z})^2}$$

for $j = 0,1,\ldots K$

where $K \leq N/4$

Further information may be obtained by decoupling the autocorrelation function into partial autocorrelation function, between Z_t and Z_{t+k} with all observations bounded by Z_t and

Z_{t+k} neglected. The partial autocorrelation function is given by

$$\phi_{11} = \rho_1$$

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j} \cdot \rho_{k-j}}{\sum_{j=1}^{k-1} \phi_{k-1,j} \cdot \rho_j}, k = 2,3,\ldots K \qquad (5.6)$$

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk} \cdot \phi_{k-1,k-j} \quad \text{for } j = 1,2,\ldots k-1 \qquad (5.7)$$

where ϕ_{kj} is j^{th} coefficient in an autoregressive process of order k. ρ_k is the autocorrelation estimate.

Most time series may be represented by autoregressive processes, moving average processes or combinations of the two. The following is a brief discussion of the two processes.

D. Autoregressive Processes

An autoregressive process is given by an equation of the form

$$Z_t = \delta + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + \varepsilon_t \qquad (5.8)$$

where δ, ϕ_p are parameters of the autoregressive process.

Z_{t-p} are values of the function at time lags p. The value of the function at t depends on previous values of the function at time lags p. ε_t is the error at t.

Equation 5.8 may be written as

$$Z_t = \delta + (\phi_1 B + \phi_2 B^2 + \ldots \phi_p B^p) Z_t + \varepsilon_t \text{ or}$$

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) Z_t = \delta + \varepsilon_t$$

$$\phi_p(B) Z_t = \delta + \varepsilon_t \tag{5.9}$$

where $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ and B is the back-shift operator such that $Z_{t-1} = BZ_t$.

The first order autoregressive process, AR(1) is given by equation 5.9

with p = 1

$$\phi_1(B) Z_t = \delta + \varepsilon_t$$

$$(1 - \phi_1 B) Z_t = \delta + \varepsilon_t$$

$$Z_t = \delta + \phi_1 Z_{t-1} + \varepsilon_t \tag{5.10}$$

By successive application of equation 5.10 it can be shown that

$$Z_t = \delta + \phi_1 \delta + \phi_1^2 Z_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t$$

$$= \delta + \phi_1 \delta + \phi_1^2 \delta + \phi_1^3 Z_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t$$

or in general

$$Z_t = \delta \sum_{j=0}^{N-1} \phi_1^j + \phi_1^N Z_{t-N} + \varepsilon_t \sum_{j=0}^{N-1} \phi_1^j B^j$$

For the series to converge it is required that

$$|\phi_1| < 1$$

Now $\phi_1^N Z_{t-N} \to \infty$ as $N \to \infty$

and

$$Z_t = \delta \cdot \sum_{j=0}^{\infty} \phi_1{}^j + \varepsilon_t \cdot \sum_{j=0}^{\infty} \phi_1{}^j B^j \tag{5.11}$$

The mean of the series may be obtained by taking the expectation of Equation 5.11, given by

$$\mu = E[Z_t] = E[\delta \cdot \sum_{j=0}^{\infty} \phi_1{}^j + \varepsilon_t \cdot \sum_{j=0}^{\infty} \phi_1{}^j B^j]$$

or

$$\mu = \delta \cdot \sum_{j=0}^{\infty} \phi_1{}^j + E(\varepsilon_t) \cdot \sum_{j=0}^{\infty} \phi_1{}^j B^j$$

Now $E(\varepsilon_t) = 0$ since ε_t is random noise about the mean.

$$\therefore \mu = \delta \sum_{j=0}^{\infty} \phi_1{}^j$$

$$\mu = \frac{\delta}{1 - \phi_1} \tag{5.12}$$

The variance of the series may be found by the procedure outlined as follows.

$$Z_t - E(Z_t) = [\delta \sum_{j=0}^{\infty} \phi_1{}^j + \varepsilon_t \sum_{j=0}^{\infty} \phi_1{}^j B^j] - \delta \sum_{j=0}^{\infty} \phi_1{}^j \tag{5.13}$$

$$= \sum_{j=0}^{\infty} \phi_1{}^j B^j \varepsilon_t$$

The variance is given by the expectation of the square of equation 13.

$$E[Z_t - E(Z_t)]^2 = E[\sum_{j=0}^{\infty} \phi_1{}^j B^j \varepsilon_t]^2$$

$$= E[\varepsilon_t^2 + \phi_1^2 \varepsilon^2_{t-1} + \phi_1^4 \varepsilon_{t-2}^2 + \ldots + \text{cross products}]$$

$$= \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} \tag{5.14}$$

where $\sigma_\varepsilon^2 = E(B^j \varepsilon_t^2)$

which is the variance of the series. Thus the autocovariance at lag k is given by

$$\gamma_k = \frac{\phi_1^k}{1 - \phi_1^2} \cdot \sigma_\varepsilon^2 \tag{5.15}$$

The autocorrelation function of the first order autoregressive process is given by the ratio of

$$\rho_k = \frac{\gamma_k}{\gamma_o} = \phi_1^k \tag{5.16}$$

Since $|\phi_1| < 1$, the autocorrelation function of the first order autoregressive process decays exponentially for increasing lags as shown in equation 16. This is important in the identification of an autoregressive process.

E. Moving Average Processes

A moving average process is given by an equation of the form

$$Z_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \tag{5.17}$$

where μ, θ_q are parameters of the moving average process.

ε_t is random noise or error.

The value of the function at time t depends on previous errors.

The first order moving average process MA(1) is given by

$Z_t = \mu + \theta_p(B)\varepsilon_t$ where

$\theta_p(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_p B^p$

with p = 1

$$Z_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} \tag{5.18}$$

The mean of the series is given by

$E(Z_t) = \mu$

Since ε_t is distributed randomly. The autocovariance at lag j is calculated from

$$\gamma_j = E\ \{[Z_t - E(Z_t)][Z_{t-j} - E(Z_t)]\}$$

$$= E\ [(\varepsilon_t - \theta_1\varepsilon_{t-1})(\varepsilon_{t-j} - \theta_1\varepsilon_{t-j-1})]$$

$$\gamma_j = E\ [\varepsilon_t\varepsilon_{t-j}] - \theta_1 E[\varepsilon_t\varepsilon_{t-j-1}] - \theta_1 E[\varepsilon_{t-1}\varepsilon_{t-j}] + \theta_1^2 E[\varepsilon_{t-1}\varepsilon_{t-j-1}] \tag{5.19}$$

At j = 1, $\gamma_1 = -\theta_1\sigma_\varepsilon^2$

where $\sigma_\varepsilon^2 = E[\varepsilon_t\varepsilon_t] = E[\varepsilon_{t-1}\varepsilon_{t-1}]$

and $0 = E[\varepsilon_t\varepsilon_{t-1}] = E[\varepsilon_{t-1}\varepsilon_t]$

The autocorrelation at lag 1 is given by

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-\theta_1}{1 + \theta_1^2} \qquad \text{for } j = 1$$

$$= 0 \qquad \text{for } j > 1 \tag{5-21}$$

The autocorrelation function in a first order moving average process cuts off at lag 1. The function converges for any value of θ_1.

F. Mixed Models

The foregoing analysis describes stationary series. However most real series, such as the data used in this research are non-stationary. It is necessary to transform the non-stationary series to a stationary one. Time series analysis may then be applied to characterize the series. Most real series are a mix of autoregressive and moving average processes of the form

$$\phi_p(B) Z_t = \mu + \theta_q(B)\varepsilon_t \tag{5.21}$$

The first order mixed autoregressive and moving average process ARMA(1,1) is expressed as

$$\phi_1(B) Z_t = \mu + \theta_1(B)\varepsilon_t$$

or

$$Z_t = \mu + \phi_1 Z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \tag{5.22}$$

As an example, the series shown in Figure 5.2a below is non-stationary. The value of the function is not stationary

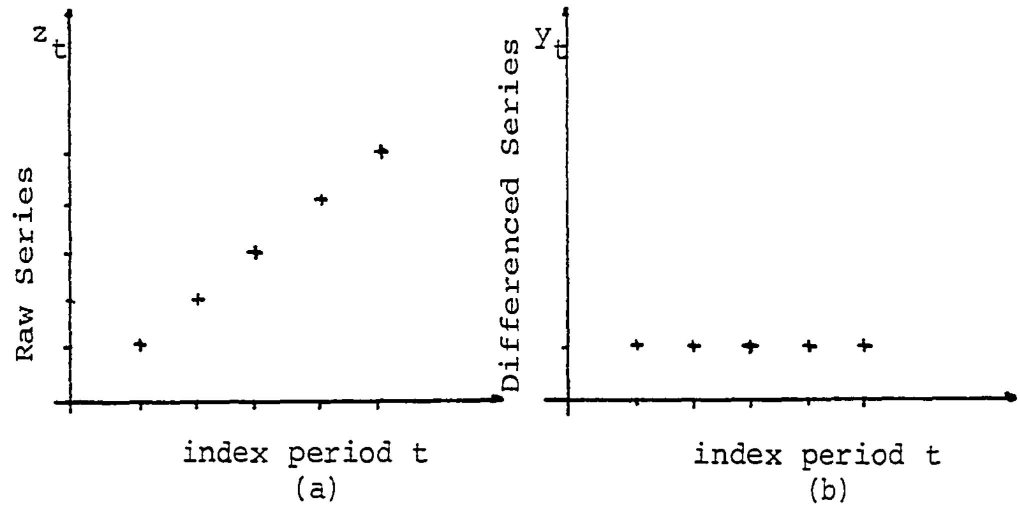


Figure 5.2 (a) Sample Non-Stationary Series
(b) Series After First Regular Differencing

about some constant mean. Most series begin to show stationary characteristics, such as the simple series shown in Figure 5.2b, upon successive differencing, given by

$$y_t = z_t - z_{t-1} \tag{5.23}$$

The order of differencing may be introduced into the model equation as shown below

$$\phi_p(B) \cdot (1 - B)^d z_t = \mu + \theta_q(B) \varepsilon_t$$

where d is the order of differencing.

Most raw series tend to exhibit a seasonal characteristic which may be introduced into the model equation as

$$\phi_p(B) \cdot (1 - B)^d (1 - B^S) z_t = \mu + \theta_q(B) \varepsilon_t \tag{5.24}$$

where S is the seasonal length.

The determination of values for p, d, q and S completes the identification process. This is the most critical phase of the modeling process. The minimum values for p, d, q and S that adequately describe the series are selected. The aim is to develop parsimonious models.

G. Identification of the Data Series

Figure 5.3 shows a plot of the raw data, extending from 1969 through 1976; megawatts demand versus a monotonically increasing index t. The series shows a periodic characteristic. However, there is a growth factor that may be determined. The three regions of base load, temperature dependent load, and saturation, are easily defined for each segment of the series. The following describes an application of time series analysis to the raw series in Figure 5.3. The analysis is aided by use of computer programs PDQ, ESTIMATE, and FORECAST [31], available in the University of Oklahoma IBM 370 computer system library.

In phase one of the analysis, values of p,d,q and S described as in equation 24, will be found. The raw series as shown in figure 5.3, is non-stationary. Hence, differencing is required to achieve a stationary series.

$$y_t = (1 - B)^1 z_t = z_t - z_{t-1}$$

The above equation called a first order difference equation yielded a series that was stationary. Further differencing is not necessary, because once stationarity is reached more differencing continues to yield stationary series. Figure 5.4 shows first order differencing of the raw series. The autocorrelation function for this series, Figure 5.5 shows strong spikes at lag 1 and lag 43. A seasonal component of

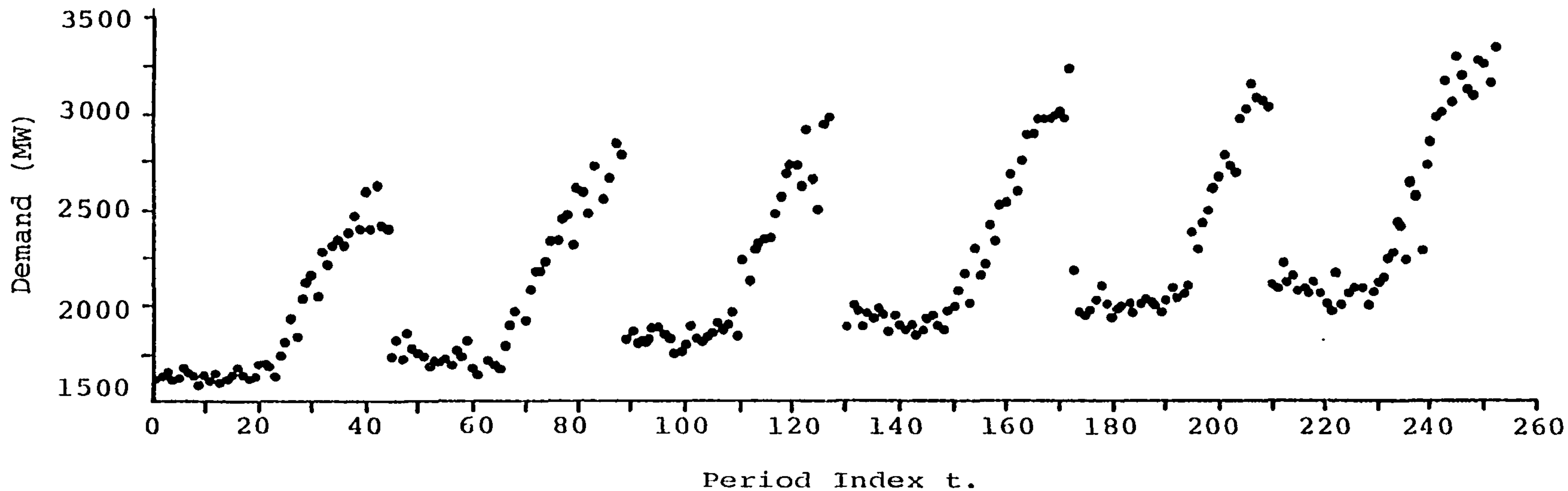


Figure 5.3. Sample Actual Raw Series.

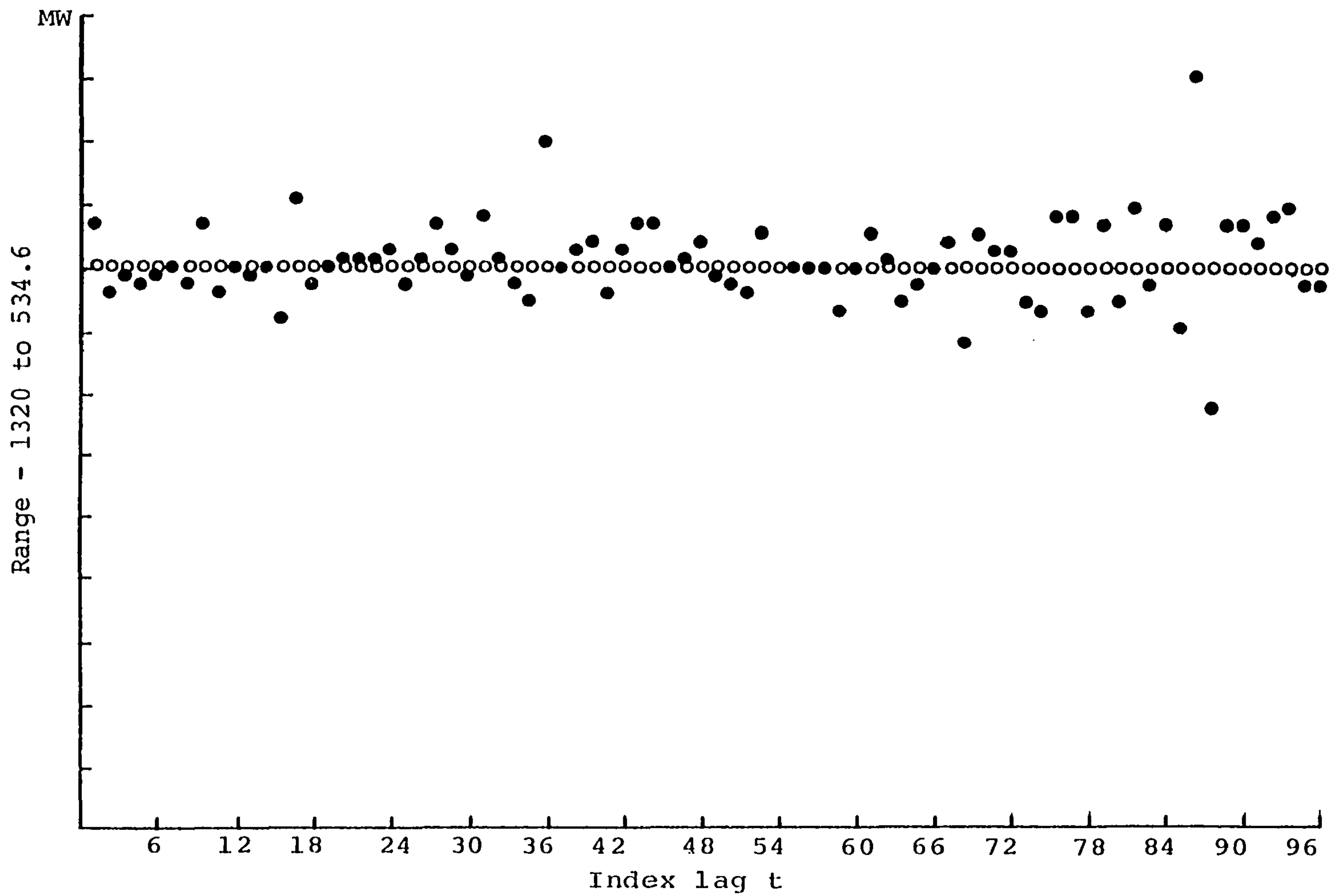


Figure 5.4. First Differencing of Actual Raw Series.

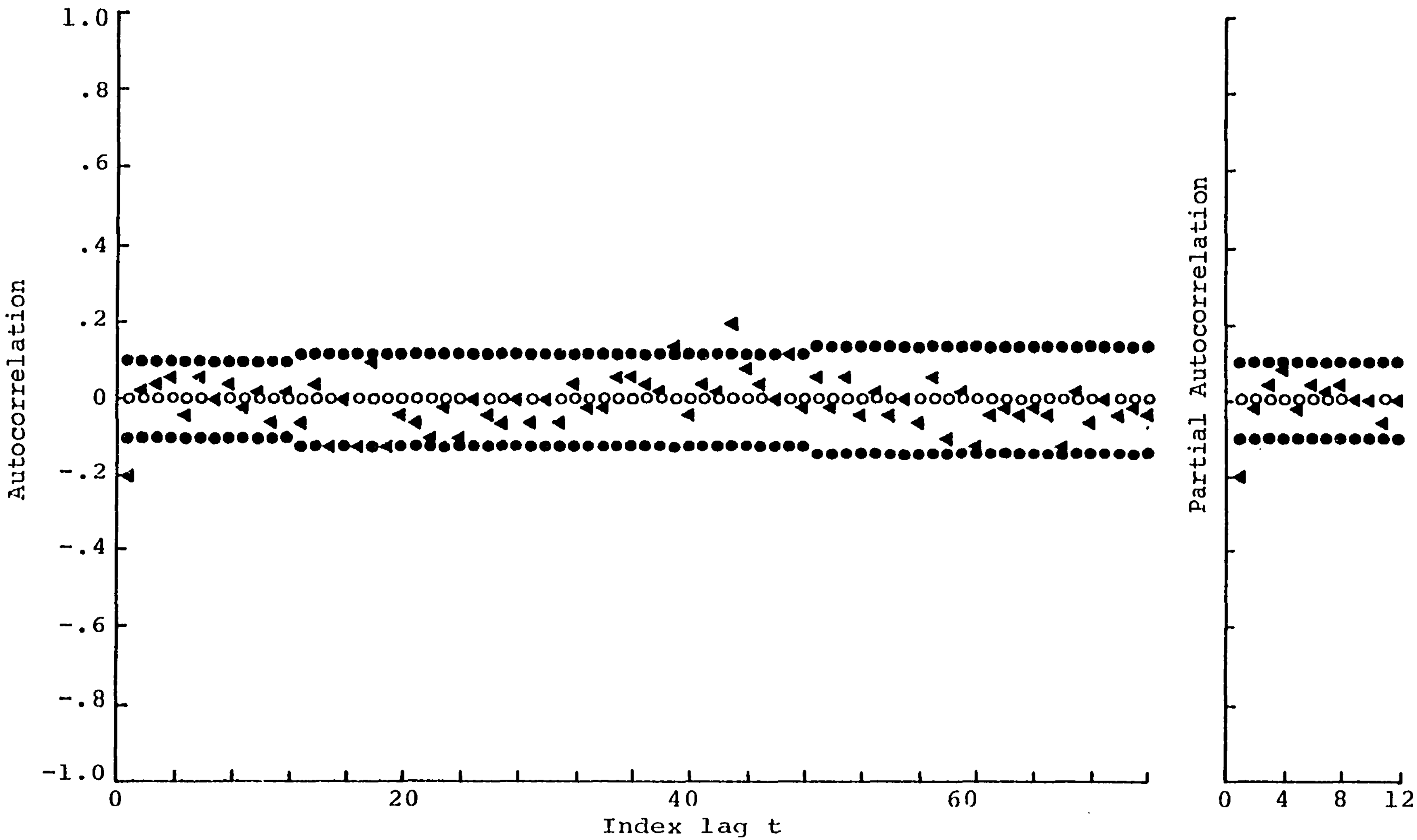


Figure 5.5. Autocorrelation Function and Partial Autocorrelation function for $(1-B)^1(1-B^{44})^0$.

length 43, 44 or 50 was suspected. These values were tried for the best choice. S was set at 44 in the following equations.

The model equation is now

$$\phi_p(B)(1 - B)(1 - B^{44})Z_t = \mu + \theta_q(B)\varepsilon_t$$

The autocorrelation of the series after first order regular and seasonal differencing shown in Figure 5.6 shows a predominantly strong spike at lag 1 repeating at lag 44. Hence the order of the moving average component q is set at unity. Equation 24 thus becomes

$$\phi_1(B)(1 - B)(1 - B^{44})Z_t = \mu + \theta_1(B)\varepsilon_t$$

or

$$(1 - \phi_1 B)(1 - B)(1 - B^{44})Z_t = \mu + (1 - \theta_1 B)\varepsilon_1 \quad (5.25)$$

On expanding equation 5.25 the following is obtained

$$Z_t = [(1 + \phi_1)B - \phi_1 B^2 + B^{44} - (1 + \phi_1)B^{45} + \phi_1 B^{46}]Z_t + \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} \quad (5.26)$$

or

$$Z_t = (1 + \phi_1)Z_{t-1} - \phi_1 Z_{t-2} + Z_{t-44} - (1 + \phi_1)Z_{t-45} + \phi_1 Z_{t-46} + \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} \quad (5.27)$$

H. Model Parameter Estimation

The order of the components of the time series model has been determined. The next step is the calculation of the parameters ϕ_1, μ and θ_1, for the autoregressive and

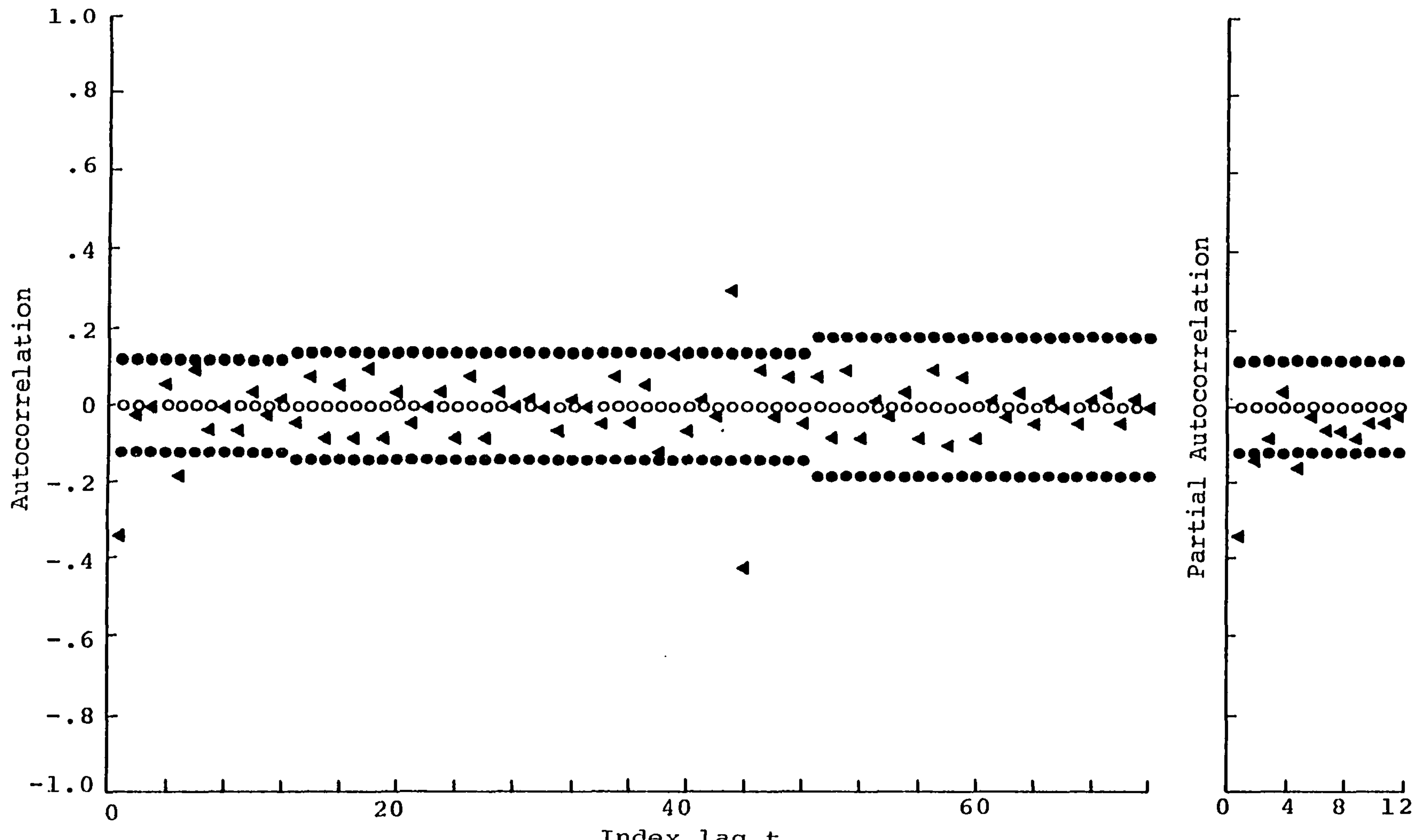


Figure 5.6. Autocorrelation Function and Partial Auto-correlation function for $(1-B)^1(1-B^{44})^1$.

moving average components respectively. The reference equation is given in 5.26. This function is non-linear, and hence the exact methods of linear least squares regression may not be used. Non-linear regression [27] is used to determine the parameters, from the actual data series. Initial estimates of ϕ_1, μ and θ_1 are made, and then used iteratively. The square of the error function is given by

$$e^2 = [Z_t - \hat{Z}_t]^2 \tag{5.28}$$

where Z_t is the actual series value, and $\hat{Z}_t$, the estimate of that value at lag t, for the current values of the parameters.

The total error sum of squares is given by

$$ESS = \sum_{j=1}^{N} e_j^2 ;\ N \text{ is the number of data points.}$$

It is desired to find values of ϕ_1, μ and θ_1 such that the error sum of squares is minimized. The computer program called ESTIMATE [31] is used to determine the optimum parameter values. It uses a method first developed by D.W. Marguardt [15]. Marguardt's method is a compromise between the Gauss-Newton method, and the method of steepest descent. Gauss-Newton requires good initial estimates, and converges quickly to a solution. The method of steepest descent may start with poor initial parameter estimates of the objective function, but converges rather slowly to a solution. Marguardt's method bridges the gap between the two methods. A factor of convergence is introduced such that poor starting values may be used, as in steepest descent while convergence is rapidly approached as in Gauss-Newton.

The parameter set is modified for trial at each iteration, and the error sum of squares calculated. Convergence is reached when the relative change between the last value of the objective function and its immediate previous value is less than some designated constant.

$$\left| \frac{ESS_{i+1} - ESS_i}{ESS_i} \right| \leq C$$

where i is the iteration count and C is the criterion of convergence. In the program ESTIMATE $e = 10^{-6}$. Table 5.1 shows an iterative progression to a solution. The last row in Table 5.1 shows the values of the parameters that minimize the objective error sum of squares function:

$$\phi_1 = .2046; \qquad \theta_1 = .6289; \qquad \delta = -.4767$$

Hence equation 6.26 becomes

$$Z_t = 1.2046Z_{t-1} - 0.2046Z_{t-2} + Z_{t-44} - 1.2046Z_{t-45} +$$

$$0.2046Z_{t-46} - .4767 + \varepsilon_t - .6789\varepsilon_{t-1} \qquad (5.29)$$

I. Diagnostic Checking

After the identification and parameters estimation phases of the modeling process, it is necessary to check the adequacy of the model obtained. This is done by statistical analysis of the error function.

$$e_t = Z_t - \hat{Z}_t$$

Table 5.1. Model Parameter Estimates

Iteration Count	Parameter Estimates ϕ_1	θ_1	μ	ESS $X10^7$	$\frac{\Delta ESS}{ESS}$
0	-0.1053	.0775	3.711	2.127000	
1	-0.1164	.2558	-2.300	2.072930	0.054383
2	0.0203	.4438	-0.9461	1.996629	0.010349
3	0.0690	.4996	-0.8836	1.994213	0.001211
4	0.0999	.5306	-0.7597	1.993397	0.000409
5	0.1213	.5513	-0.7196	1.993016	0.000191
6	0.1370	.5663	-0.6546	1.992814	0.000101
7	0.1495	.5781	-0.9308	1.992710	0.000052
8	0.1600	.5878	-0.5945	1.992622	0.000044
9	0.1689	.5961	-0.7005	1.992570	0.000026
10	0.1764	.6030	-0.5400	1.992526	0.000022
11	0.1831	.6093	-0.6534	1.992515	0.000006
12	0.1892	.6148	-0.5285	1.992498	0.000009
13	0.1947	.6198	-0.5254	1.992488	0.000005
14	0.1998	.6245	-0.5915	1.992480	0.000004
15	0.2046	.6289	-0.4767	1.992478	0.000004

The error function is examined for random behavior. A test of a time series model is that its errors are randomly distributed, and hence the expectation of the errors is zero; $E[e_t] = 0$. One method of testing for random behavior is to calculate the autocorrelation of the error function, $\rho_k(e)$. A Chi-square test of the residue autocorrelations will reveal grounds for rejection or acceptance of the model.

The test statistic is given by

$$Q(K) = (N - b) \cdot \sum_{K=1}^{K} \hat{\rho}_k^2(e) \tag{5.30}$$

$$Q(K) < x^2(\nu,\alpha) \tag{5.31}$$

where Q(K) is the calculated statistic, K is the number of autocorrelations; e is the error at lag K; $x^2(\nu,\alpha)$ is the Chi-square statistic from tables, of ν degrees of freedom and α level of significance. N is the number of data points in the series, b is the highest order in back shift operation in the model. Table 5.2 shows the sample autocorrelations of residuals for 36 lags.

Table 5.2. Sample Autocorrelation of Residuals.

Lag K	$\rho_k(e)$	k	$\rho_k(e)$	k	$\rho_k(e)$	k	$\rho_k(e)$
1	0.00	10	-0.01	19	-0.06	28	-0.00
2	0.06	11	-0.04	20	0.00	29	0.01
3	0.02	12	-0.01	21	-0.05	30	-0.02
4	0.02	13	-0.04	22	-0.02	31	-0.07
5	-0.17	14	0.04	23	0.00	32	0.00
6	0.01	15	-0.05	24	-0.07	33	-0.02
7	-0.09	16	0.04	25	0.04	34	-0.03
8	-0.06	17	-0.04	26	-0.06	35	0.08
9	-0.08	18	0.07	27	0.01	36	-0.01

The only autocorrelation of significance occurs at lag 5. The number of degrees of freedom $\nu = k - p - d - q$; $\nu = 33$. From mathematical tables

$$X^2(33,.05) = 50.22 \qquad \text{at 5\% significance level}$$

$$X^2(33,.1) = 47.11 \qquad \text{at 10\% significance level}$$

From Table 5.3 and equation 5.30,

$$Q(36) = 28.71$$

Hence the inequality 5.31 is satisfied. The test gives no reason for rejection of the model. Alternative models were tried before an acceptable model could be found. This phase of the modeling process ends with information for the model summarized as follows.

Order of autoregressive component, $p = 1$

Order of moving average component, $q = 1$

Degree of regular differencing, $d = 1$

Degree of seasonal differencing, $s = 1$

Length of seasonal differencing, $S = 44$

Autoregressive parameter $\phi_1 = .2046$

Moving average parameter $\theta_1 = .6289$

Constant of model $\mu = 0.4767$

and the model is as stated in equation 5.29. The standard errors and 95% confidence limits on the estimated values are summarized in Table 5.3.

Table 5.3. Summary of Standard Errors and 95% Confidence Limits on the Estimated Parameter Values

	Model Parameter Estimates		
	ϕ_1	θ_1	μ
Estimated Value	0.2046	0.6289	-00.4767
Standard Error	0.1140	0.0904	05.7010
Upper Confidence Limit	0.4327	0.8097	10.9300
Lower Confidence Limit	-0.0234	0.4480	-11.8800

The output of program ESTIMATE produces a comprehensive listing of statistical data about the data series. The model may now be used in forecasting.

J. The Forecast

The main purpose of any forecasting model is to estimate future values of the data series, with the least possible error margin. Subjective judgment cannot be completely eliminated, due to changing conditions which introduce new variables for which historical precedents may not be easily available. However, a high degree of objectivity greatly improves any applied subjective judgment. The following shows results of the method discussed previously.

The forecast model is stated in equation 5.29. A forecast may be made for ℓ steps ahead, with the origin at the present index t, as shown in equation 5.32.

Equation 5.33 is the model formulation for the historical data set from 1969 through 1976. Using the same approach and historical data from 1962 through 1969, the model is described by equation 5.32 below.

$$Z_{t+\ell} = 1.7668Z_{t+\ell-1} - .7668Z_{t+\ell-2} + Z_{t+\ell-44} -$$
$$1.7668Z_{t+\ell-45} + .7668Z_{t+\ell-46} + .0757 + \varepsilon_t - .9936\varepsilon_{t+1} \tag{5.32}$$

$$Z_{t+\ell} = 1.2046Z_{t+\ell-1} - .2046Z_{t+\ell-2} + Z_{t+\ell-44} -$$
$$1.2046Z_{t+\ell-45} + .2046Z_{t+\ell-46} - .4767 +$$
$$\varepsilon_{t+\ell} - .6289\varepsilon_{t+\ell-1} \tag{5.33}$$

Figure 5.7 shows the raw series and a forecast of the series using equation 5.32. The forecast tracks the actual series closely. The peak values of the forecast characteristics are determined and tabulated in Table 5.4. The table shows peak forecasts obtained by forecasting 190 steps ahead. This covers a period of about 4 years, from 1970 to 1973, using data from 1962 to 1969 as historical data.

Table 5.4. Maximum Demand Forecasts for 1970 through 1973.

Year	Maximum Demand (Megawatts)			Per Cent Error
	Peak Annual Forecast	Actual	Error	
1970	2163	2210	- 47	-2.13
1971	2390	2360	30	1.27
1972	2637	2645	- 8	-0.30
1973	2896	2775	121	4.36

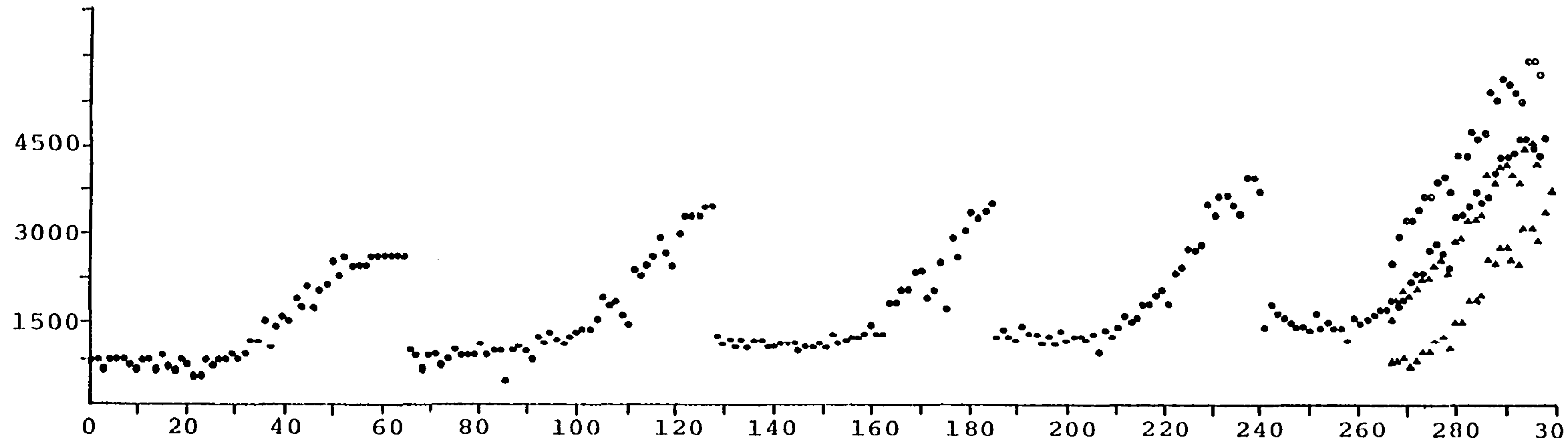


Figure 5.7. Sample Actual Raw Series and Forecast.

• represents actual, △ represents forecast
○ upper 95% confidence limit
▲ lower 95% confidence limit

The forecasts shown in Table 5.4 were plowed back into the historical data set and forecasts were then made for 1974 to 1976, using the same model parameter values, shown in equation 5.32. These forecasts are shown below in Table 5.5.

Table 5.5. Maximum Demand Forecasts for 1970 through 1976.

Year	Maximum Demand (Megawatts)			Per Cent Error
	Peak Annual Forecast	Actual	Error	
1970	2163	2210	- 47	-02.13
1971	2390	2360	30	01.27
1972	2637	2645	- 8	-00.30
1973	2896	2775	121	04.36
1974	3165	3140	25	00.80
1975	3452	3187	265	08.30
1976	3747	3335	416	12.47

The errors for the first 5 years of forecasts, 1970 through 1974 are below 5% of the actual demand. However in 1975 and 1976 the errors were 8.3 and 12.47 per cent, respectively, indicating that the forecast horizon is limited at about 5 years for a given set of model parameter estimates. The parameters were now re-estimated and the forecasts repeated. Table 5.6 shows updated forecasts for 1974 through 1976. The errors are noticeably below 5%. Forecasts for 1977 through 1980 are also shown, using model equation 5.33, derived from historical data up to 1976. The small sizes of the errors and per cent errors show the strength of the modeling technique in forecasting.

Table 5.6. Maximum Demand Forecasts for 1970 through 1980.

Year	Maximum Demand (Megawatts)			Per Cent Error
	Peak Annual Forecast	Actual	Error	
1970	2163	2210	- 47	-2.13
1971	2390	2360	30	1.27
1972	2637	2645	- 8	-0.30
1973	2896	2775	121	4.36
1974	3055	3140	25	0.80
1975	3225	3187	265	8.30
1976	3369	3335	416	12.47
1977	3536	3650	-114	-3.12
1978	3781	3805	- 24	-0.63
1979	4000	--	--	--
1980	4192	--	--	--

CONCLUSION

The forecasts of maximum demand, tabulated in the previous chapter, demonstrated the effectiveness of the methodology developed. The method is a compromise between methods that assume conventional temperature dependence, and time series analysis. The peak load for most utilities is temperature dependent, hence a temperature related characteristic seemed a most likely basis for the extraction of future peak loads. Forecasts were obtained by carrying the whole temperature dependent characteristic into the future. A pseudo series was developed using the basic temperature dependent demand characteristic. This series was also time dependent. A mathematical model was found using Box-Jenkins models and used to forecast demand. Peak annual demand forecasts were within 5% of the actual peak demands that occured, with a 5 year forecast horizon for a given model and parameter estimates.

Forecasts were associated with possible errors, a 95% confidence region, or a probability of occurrence. The temperature distribution was a left-skewed beta density function, and the conditional distribution of load for a given temperature was approximated by a normal density function. These statistical densities were used in the calculation of a probability of load occurrence. The forecasting method and probability calculations may be applied to data collected from geographical areas similar to those studied.

The apparent drawback is the time and effort it takes to analyze the time series results to determine the nature of the model and its parameters. However, a step by step approach of identification, estimation and diagnostic checking will yield a model that may be accepted or rejected.

Areas for Further Research

The model used is based on the total system coincident daily peaks. System load may be broken down into classifications, such as residential, commercial, industrial, and agricultural demand. There are sections of each of these classifications that are dependent on only certain variables, such as weather, the economy, population, and so on. Identification of these variables and the sections of system demand that are affected will greatly enhance forecasting.

An alternative approach will be to find mathematical formulations for each temperature dependent characteristic for each year of historical data. The parameters of these formulations may then be examined for relationships. Forecasts will then be made of the parameters in order to forecast a point on the demand characteristic. The forecast points can be used to generate a basic temperature dependent characteristic for each year. This may also serve as a data preparation stage for the application of time series analysis.

BIBLIOGRAPHY

[1] Asbury, C.E. "Weather Load Model for Electric Demand and Energy Forecasting," IEEE Transactions on PAS, Volume 94, No. 4, July/August 1975, pp. 1111-1116.

[2] Christiaanse, W.R. "Short Term Load Forecasting Using General Exponential Smoothing." IEEE Transactions on PAS, Volume 90, No. 2, March/April 1971, pp. 900-910.

[3] Corpening, S.L., Reppen, N.D., and Ringlee, R.J. "Experience with Weather Sensitive Load Models for Short and Long Term Forecasting," IEEE Transactions on PAS, Volume 92, No. 6, November/December 1973, pp. 1966-1972.

[4] Davey, J., et al. "Practical Application of Weather Sensitive Load Forecasting to System Planning." IEEE Transactions on PAS, Volume 92, No. 3, May/June 1973, pp. 971-977.

[5] Dryar, H.A. "The Effect of Weather on Electric Load," Transaction in American Inst. Engineering, Volume 93, 1944, pp. 1013-1044.

[6] Godard, W.W. "Electric Utility Load Forecasting," Paper No. 55-531, Presented at the AIEE Summer meeting, Swampscott, Mass. June 27-July 1, 1955.

[7] Gupta, P.C. "A Stochastic Approach to Peak Power-Demand Forecasting in Electric Utility Systems," IEEE Transactions on PAS, Volume 91, September/October 1972, pp. 2085-2094.

[8] Gupta, P.C. and Yamada, K. "Adaptive Short Term Forecasting of Hourly Loads Using Weather Information," IEEE Transactions on PAS, Volume 91, September/October 1972, pp. 2085-2094.

[9] Heinemann, G.T., Nordman, D.A. and Plant, E.C. "Relation Between Summer Weather and Summer Load - A Regression Analysis," IEEE Transactions on PAS, Volume 85, November 1966, pp. 1144-1154.

[10] Hoke, T.R. "Temperature-Load Characteristics - Southwest Power Pool," Presented at the Sixth Meeting of the South Central Regional Advisory Committee on July 25, 1967.

[11] A. Keyhani, H. Daniels, "Real Time Load Modeling Technique. An application to short Term Load Prediction," Presented at the Control of Power Systems 1976 Conference and Exposition, pp. 132-136.

[12] Latham, J.H., Jr., et al. "Probability Approach to Electric Utility Load Forecasting," IEEE Transactions on PAS, Volume 87, No. 2, February 1968.

[13] A. Lazzari, "Land - Use Data Improve Load Forecasting," Electrical World, Vol. 158, No. 25, June 1962, pp. 38-40.

[14] Lijesen, D.P. and Rosing, J. "Adaptive Forecasting of Hourly Loads Based on Load Measurements and Weather Information," IEEE Transactions on PAS, Volume 90, No. 4, July/August 1971, pp. 1757-1767.

[15] Marquardt, D.M. "An Algorithm for Least-Squares Estimation of Non-linear Parameters," J. Soc. Indust. Appl Math., 11, 431-441.

[16] Stanton, K.N. "Forecasting by Probability Methods," IEEE Transactions on PAS, Volume 90, May/June 1971, pp. 1183-1189.

[17] Srimvasan, K. and Pronovost, R. "Short Term Load Forecasting Using Multiple Correlation Models," IEEE Transactions on PAS, Volume 94, No. 5, September/October 1975, pp. 1854-1858.

[18] Sachdev, M.S. and Billington, R. "A Review of the Techniques for Forecasting Power and Energy Demands", Trans. of the Canadian Electrical Association, Paper No. 76-SP-141, Vol. 15, Part 2, 1976.

[19] Sachdev, M.S., and Billington, R. "Representative Bibliography on Load Forecasting", IEEE Trans. on PAS, Vol. PAS-96, No. 2, March/April 1977, pp. 697-700.

[20] Scott, A.R., "Electricity Load Forecasting--A Federal Viewpoint of the Canadian Electrical Association", Paper No. 76-SP-142, Vol. 15, Part 2, 1976.

[21] Stanton, K.N., "Medium-Range, Weekly and Seasonal Peak Demand Forecasting by Probability Methods".

[22] Stanton, K.N. and Gupta, P.C., "Forecasting Annual or Season Peak Demand in Electric Utility Systems", Trans. IEEE, Vol. PAS-89, May-June 1970, pp. 951-959.

[23] Thompson, R.P. "Weather Sensitive Electric Demand and Energy Analysis on a Large Geographically Diverse Power System. Application to Short Term Hourly Electric Demand Forecasting," IEEE Transactions on PAS, Volume 89, No. 7, September/October 1970, pp. 1678-1688.

[24] Toyoda, J., Chen, M. and Inoue, Y. "An Application of State Estimation to Short Term Load Forecasting, Part I; Forecasting Modeling," IEEE Transactions on PAS, Volume 89, No. 7, September/October 1970, pp. 1678-1688.

[25] Vemuri, S., Hiu, F. and Balasubramanian, R. "Load Forecasting Using Stochastic Models," Paper N-. TP1-A, Proc. of the 8th CA Conference, Minneapolis, Minnesota, pp. 31-37.

[26] Box, G.E.P., and Jenkins, G.M., "Time Series Analysis, Forecasting and Control", Holden-Day, Inc., San Francisco, 1970.

[27] Kuester, J.L. and Mize, J.H. "Optimization Techniques". McGraw-Hill Book Company, New York, 1975.

[28] Lindgren, "Statistical Theory, "Macmillan Publishing Co., Inc., New York, 1976.

[29] Mabert, V., "An Introduction to Short Term Forecasting Using the Box-Jenkins Methodology, Publication #2 in the Monograph series, Production Planning and Control Division, American Institute of Industrial Engineers, Inc., 1975.

[30] Montgomery, D.C., and Johnson, L.A., "Forecasting and Time Series Analysis", McGraw-Hill Book Company, New York, 1976.

[31] Nelson, C.R., "Applied Time Series for Managerial Forecasting", Holden-Day, Inc., San Francisco, 1973.

APPENDIX A

Sample Computer Programs for Calculating
Probabilities of Occurrence of Load and Data Handling.

```
C
C
C     THIS COMPUTER PROGRAM CALCULATES THE PROBABILITY OF LOAD OCCURENCE
C     INPUT REQUIEREMENTS ARE AS FOLLOWS.
C     DEM.....THE AVERAGE DEMAND CHARACTERISTIC IS READ
C     TEMP....STORE TEMPERATURE DATA
C     SD......STANDARD DEVIATION DATA ARE STORED IN DATA BELOW
C     TPROB...CUMULATIVE TEMPERATURE DATA ARE READ FROM AREA FILE
C     TAP.....DENSITY FUNCTION OF TEMP READ FROM AREA FILE
C     BASE....BASE LOAD FOR THE YEAR
C     BIG.....THE PEAK LOAD WHOSE PROBABILITY IS NEEDED.
      DIMENSION DEM(50),TEMP(50),SD(50),TPROB(110),TAP(110)
      DATA BASE,BIG/1300.,2200./
      DATA SD/.1780,.1640,.156,.1490,.143,.136,.131,.124,.112,.106
     1,.099,.094,.088,.084,.08,.076,.072,.069,.066,.063,.06,.058,.056
     2,.055,.053,.052,.051,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,
     3,0.,0.,0.,0.,0.,0.,0.,0.,0.,0.,0.,0./
      J=1
      TPROB(J)=0.
      DO 10 I=2,111
      K=I-1
C
C
C     ALLOCATE THE TEMP DENSITY FILE TO TUBE 'FT14F001'
      READ(14,200)DENS
  200 FORMAT(15X,E15.7)
      TPROB(I)=TPROB(I-1)+DENS
      TAP(K)=DENS
   10 CONTINUE
      WRITE(6,1)TPROB(111)
    1 FORMAT(10X,F10.4)
      WRITE(6,600)
  600 FORMAT(1H1,10X,'LOAD LEVEL',10X,'PROB')
      PROB=0.
      CPROB=0.
      STEP=5.
      DO 5 I=1,50
      IF(I .GE. 29)SD(I)=.05
C
C
C     ALLOCATE THE AVE. DEMAND CHAR. TO 'FT13F001'
      READ(13,300,END=6)DEM(I)
  300 FORMAT(24X,F10.1)
      IF(TEMP(I) .EQ. 9999.)GO TO 6
      K=I
    5 CONTINUE
    6 DO 80 I=1,500
      CO=0.
      X=I
      DO 20 J=1,K
      TEMP(J)=60+K
      JJ=TEMP(J)
      IF(SD(J) .EQ. 0.)GO TO 20
      T=-(BIG-X*STEP-DEM(J))/(BASE*SD(J))
      IF(T .LE. -3. .OR. T .GE. 3.)GO TO 20
      CO=CO+1.
      Y=BIG-X*STEP.
      PROB=ERFC(T)/2.
      PROB=1.-PROB
      PROB=PROB*TAP(JJ)
      CPROB=CPROB+PROB
   20 CONTINUE
      IF(CO .EQ. 0.)GO TO 80
      WRITE(6,400)Y,PROB,CPROB
  400 FORMAT(10X,F10.2,10X,2E15.7)
   80 CONTINUE
      STOP
      END
END OF DATA
```

```
C
C
C     THIS COMPUTER PROGRAM DOES MOST OF THE DATA HANDLING
C     REQUIRED TO ADD AREA DEMANDS ,SETS UP DATA FOR THE BOX-
C     JENKINS LIBRARY RUOTINES.IT MAY ALSO BE EASILY MODIFIED
C     TO SET UP DATA FOR THE PROBABILITY CALCULATION PHASE.
C     BASICALLY IT READS THE LOAD DATA SETS AND WRITES IT IN THE
C     FORM REQUIRED BY FORMAT SETTING.
C
//AGALL JOB NO1NOZIT,'ANNAN',CLASS=J,NOTIFY=LOGONID
/*ACCT PASS=NO1NOZIT
/*JOBPARM D=RMT1
// EXEC FORTGCLG
//FORT.SYSIN DD *
      DIMENSION IYR(366),MON(366),NDA(366),MHOT(366),NCOL(366)
      DIMENSION IDAY(366),LOAD(366),ITER(100),LOAD1(366),LOAD2(366)
      DIMENSION DEM(51,100),IPT(110)
C
C     CONSTANTS INITIALISED
C
      DO 100 IOC=13,13
      DO 100 IYS=76,76
      LIS=1
      NUM=0
      J=0
      DO 6 K=1,100
      ITER(K)=0
      DO 6 KK=1,51
      DEM(KK,K)=0
    6 CONTINUE
   10 J=J+1
C
C     TASK ...TO COMBINE LOADS BY DATA AS EXPLAINED IN CHAP. 4.
C
C
    3 READ(13,1000,END=70)NUM,IYR(J),MON(J),NDA(J),MHOT(J),
     1NCOL(J),IDAY(J),LOAD1(J)
      TOP=LOAD1(J)*MHOT(J)
      TY=LOAD1(J)
      READ(14,1000,END=70)NUM,IYR(J),MON(J),NDA(J),MHOT(J),
     1NCOL(J),IDAY(J),LOAD2(J)
 1000 FORMAT(8(I5,2X))
      TOP=TOP+LOAD2(J)*MHOT(J)
      TY=TY+LOAD2(J)
      READ(15,1000,END=70)NUM,IYR(J),MON(J),NDA(J),MHOT(J),
     1NCOL(J),IDAY(J),LOAD1(J)
      TOP=TOP+LOAD1(J)*MHOT(J)
      TY=TY+LOAD1(J)
      READ(16,1000,END=70)NUM,IYR(J),MON(J),NDA(J),MHOT(J),
     1NCOL(J),IDAY(J),LOAD2(J)
      TOP=TOP+LOAD2(J)*MHOT(J)
      TY=TY+LOAD2(J)
      READ(17,1000,END=70)NUM,IYR(J),MON(J),NDA(J),MHOT(J),
     1NCOL(J),IDAY(J),LOAD1(J)
      TOP=TOP+LOAD1(J)*MHOT(J)
      TY=TY+LOAD1(J)
      MHOT(J)=TOP/TY
      LOAD(J)=TY
      IF(IYR(J) .LT. IYS)GO TO 3
      IF(IYR(J) .GT. IYS)GO TO 70
      IF(J .LT. 2)GO TO 40
      JO=J-1
      IF(NDA(J)-NDA(JO))30,20,30
   30 JP=J-?
```

```
      IF(JP .LE. 0)GO TO 40
      IF(NDA(J)-NDA(JP))40,20,40
   20 J=J-1
      GO TO 10
   40 N=MHOT(J)-59
C
C     SORT LOADS READ BY TEMP.
C
C
      IF(IDAY(J) .GE. 7 .OR. IDAY(J) .EQ. 1)GO TO 10
      IF(N .LE. 0)GO TO 50
      ITER(N)=ITER(N)+1
      L=ITER(N)
      DEM(N,L)=LOAD(J)
   50 IF(MON(J) .EQ. 12 .AND. NDA(J) .EQ. 31)GO TO 70
      GO TO 10
   70 IF(LIS .EQ. 0)GO TO 91
      DO 90 N1=1,51
      NTEMP=N1+59
      TEMP=NTEMP
      L=ITER(N1)
      DO 90 L1=1,L
      LIS=LIS+1
      IF(DEM(N1,L1) .LE. 0.)GO TO 90
      WRITE(9,1100)TEMP,DEM(N1,L1)
 1100 FORMAT(2E15.7,50X)
   90 CONTINUE
      END FILE 9
      STO=9999.
      WRITE(6,1110)LIS
      WRITE(6,1300)
 1300 FORMAT(10X,'IN ORDER OF ...YEAR,TEMP,AVE.LOAD,SIGMA')
 1110 FORMAT(10X,'NUM. OF ENTRIES =',I4)
   91 DO 130 K=1,51
      G=K+59
      SUM=0.
      S=0.
      SU1=0.
      L=ITER(K)
      IF(L .LE. 0) GO TO 130
      DO 120 J=1,L
      SUM=SUM+DEM(K,J)
      SU1=SU1+DEM(K,J)*DEM(K,J)
      S=S+1.
  120 CONTINUE
      DEM(K,100)=SUM/S
      VARR=(SU1/S)-(DEM(K,100)*DEM(K,100))
      SIGMA=SQRT(VARR)
      WRITE(6,1200)IYS,G,DEM(K,100),SIGMA
 1200 FORMAT(10X,I4,3F10.1,36X)
  130 CONTINUE
  100 CONTINUE
      STOP
      END
//GO.FT09F001 DD DSN=ZE0347.AGPS07.DATA,DISP=OLD
//GO.FT13F001 DD DSN=ZE0347.BSAL.DATA,DISP=SHR
//GO.FT14F001 DD DSN=ZE0347.BNEO.DATA,DISP=SHR
//GO.FT15F001 DD DSN=ZE0347.BTPL.DATA,DISP=SHR
//GO.FT16F001 DD DSN=ZE0347.BDPL.DATA,DISP=SHR
//GO.FT17F001 DD DSN=ZE0347.BGOE.DATA,DISP=SHR
//
/*
END OF DATA
```

TEMP. DENSITY DATA

TEMP.(F)	DENSITY
0.1000000E 01	0.0
0.2000000E 01	0.0
0.3000000E 01	0.0
0.4000000E 01	0.0
0.5000000E 01	0.0
0.6000000E 01	0.0
0.7000000E 01	0.0
0.8000000E 01	0.0
0.9000000E 01	0.0
0.1000000E 02	0.0
0.1100000E 02	0.3647637E-03
0.1200000E 02	0.1823819E-03
0.1300000E 02	0.7295276E-03
0.1400000E 02	0.3647637E-03
0.1500000E 02	0.0
0.1600000E 02	0.5471455E-03
0.1700000E 02	0.5471455E-03
0.1800000E 02	0.1823819E-03
0.1900000E 02	0.5471455E-03
0.2000000E 02	0.1276673E-02
0.2100000E 02	0.9119094E-03
0.2200000E 02	0.7295276E-03
0.2300000E 02	0.1094291E-02
0.2400000E 02	0.1459055E-02
0.2500000E 02	0.1641437E-02
0.2600000E 02	0.1823819E-02
0.2700000E 02	0.9119094E-03
0.2800000E 02	0.2918110E-02
0.2900000E 02	0.2006201E-02
0.3000000E 02	0.2188583E-02
0.3100000E 02	0.1823819E-02
0.3200000E 02	0.2918110E-02
0.3300000E 02	0.3830020E-02
0.3400000E 02	0.4741929E-02
0.3500000E 02	0.5106691E-02
0.3600000E 02	0.5289074E-02
0.3700000E 02	0.5836219E-02
0.3800000E 02	0.6200984E-02
0.3900000E 02	0.7295273E-02
0.4000000E 02	0.6930511E-02
0.4100000E 02	0.5653836E-02
0.4200000E 02	0.6748129E-02
0.4300000E 02	0.8571949E-02
0.4400000E 02	0.7660039E-02
0.4500000E 02	0.9483859E-02
0.4600000E 02	0.9301476E-02
0.4700000E 02	0.1076053E-01
0.4800000E 02	0.9301476E-02
0.4900000E 02	0.8936711E-02
0.5000000E 02	0.1331388E-01
0.5100000E 02	0.9301476E-02
0.5200000E 02	0.1404341E-01
0.5300000E 02	0.1349626E-01
0.5400000E 02	0.1294911E-01
0.5500000E 02	0.1221959E-01
0.5600000E 02	0.1532008E-01
0.5700000E 02	0.1258435E-01
0.5800000E 02	0.1240197E-01

0.5900000E 02	0.1149006E-01
0.6000000E 02	0.1714390E-01
0.6100000E 02	0.1349626E-01
0.6200000E 02	0.1167244E-01
0.6300000E 02	0.1294911E-01
0.6400000E 02	0.1659675E-01
0.6500000E 02	0.1404341E-01
0.6600000E 02	0.1513769E-01
0.6700000E 02	0.1659675E-01
0.6800000E 02	0.1933248E-01
0.6900000E 02	0.1769104E-01
0.7000000E 02	0.1896772E-01
0.7100000E 02	0.1604960E-01
0.7200000E 02	0.1495532E-01
0.7300000E 02	0.1951486E-01
0.7400000E 02	0.1987962E-01
0.7500000E 02	0.1860295E-01
0.7600000E 02	0.1440817E-01
0.7700000E 02	0.1586723E-01
0.7800000E 02	0.1750866E-01
0.7900000E 02	0.1550246E-01
0.8000000E 02	0.2060916E-01
0.8100000E 02	0.2060916E-01
0.8200000E 02	0.2024439E-01
0.8300000E 02	0.2115630E-01
0.8400000E 02	0.2188583E-01
0.8500000E 02	0.2298012E-01
0.8600000E 02	0.2042677E-01
0.8700000E 02	0.2352726E-01
0.8800000E 02	0.2097392E-01
0.8900000E 02	0.2024439E-01
0.9000000E 02	0.2133868E-01
0.9100000E 02	0.2352726E-01
0.9200000E 02	0.2316250E-01
0.9300000E 02	0.1750866E-01
0.9400000E 02	0.1532008E-01
0.9500000E 02	0.1532008E-01
0.9600000E 02	0.1167244E-01
0.9700000E 02	0.1294911E-01
0.9800000E 02	0.9301476E-02
0.9900000E 02	0.7660039E-02
0.1000000E 03	0.8571949E-02
0.1010000E 03	0.6018601E-02
0.1020000E 03	0.5653836E-02
0.1030000E 03	0.3647638E-02
0.1040000E 03	0.2188583E-02
0.1050000E 03	0.1823819E-02
0.1060000E 03	0.1094291E-02
0.1070000E 03	0.5471455E-03
0.1080000E 03	0.1823819E-03
0.1090000E 03	0.0
0.1100000E 03	0.0

END OF DATA

APPENDIX B

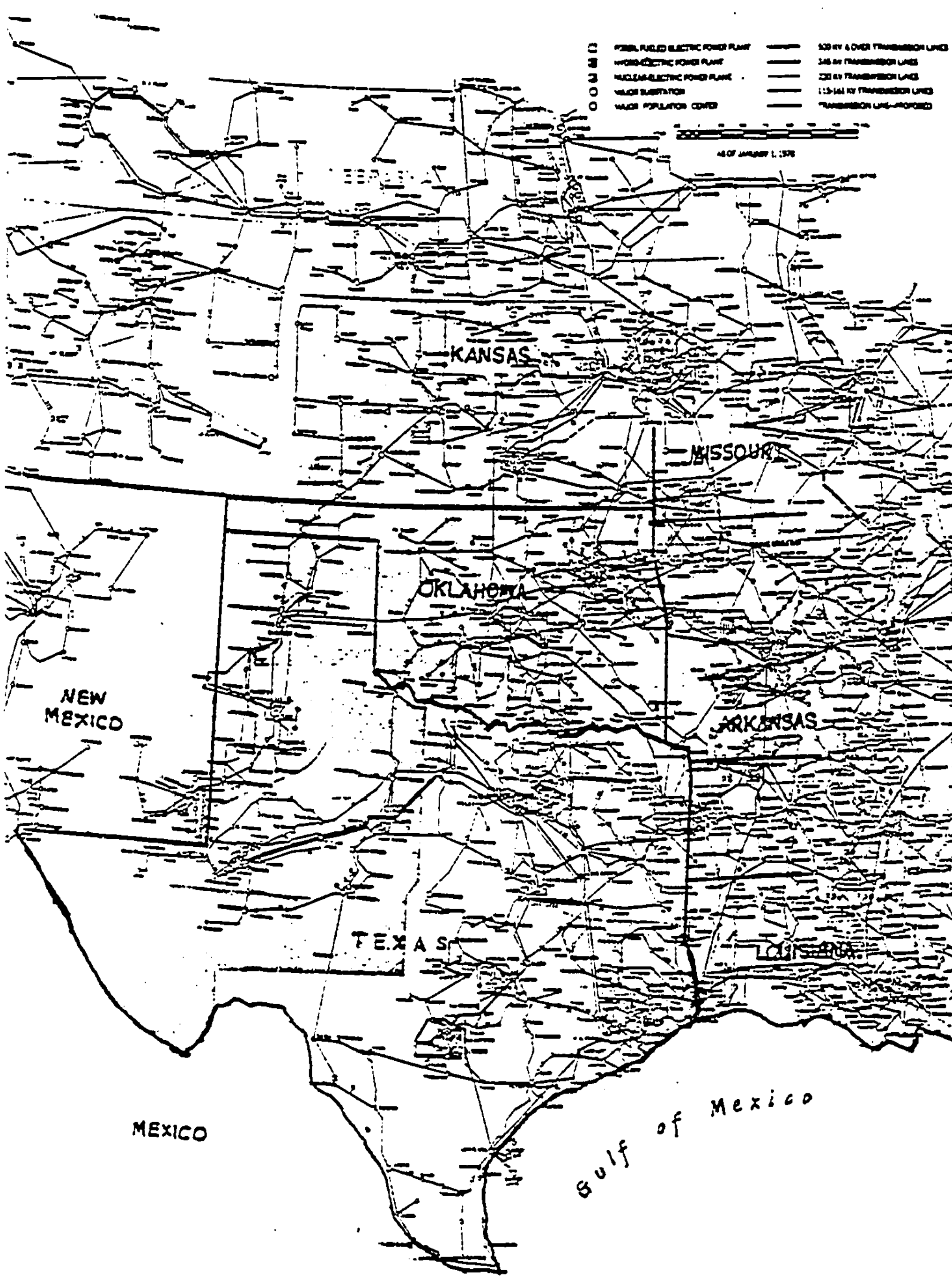
A Map of the Area Studied.

Areas 1 and 2 represent Oklahoma

Areas 3 and 4 represent Texas

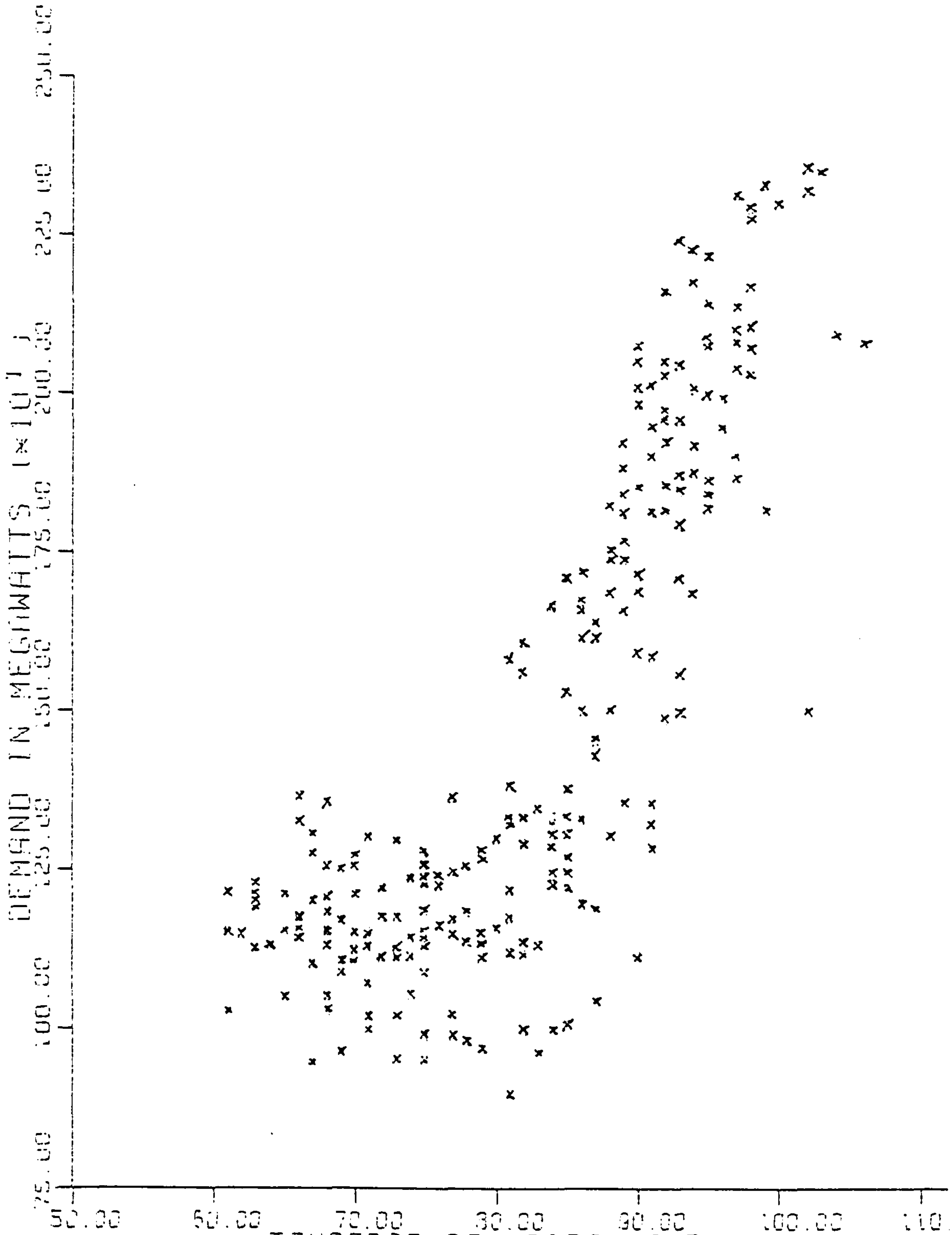
Areas 5 represents Louisiana

Map Showing Oklahoma, Texas and Louisiana.

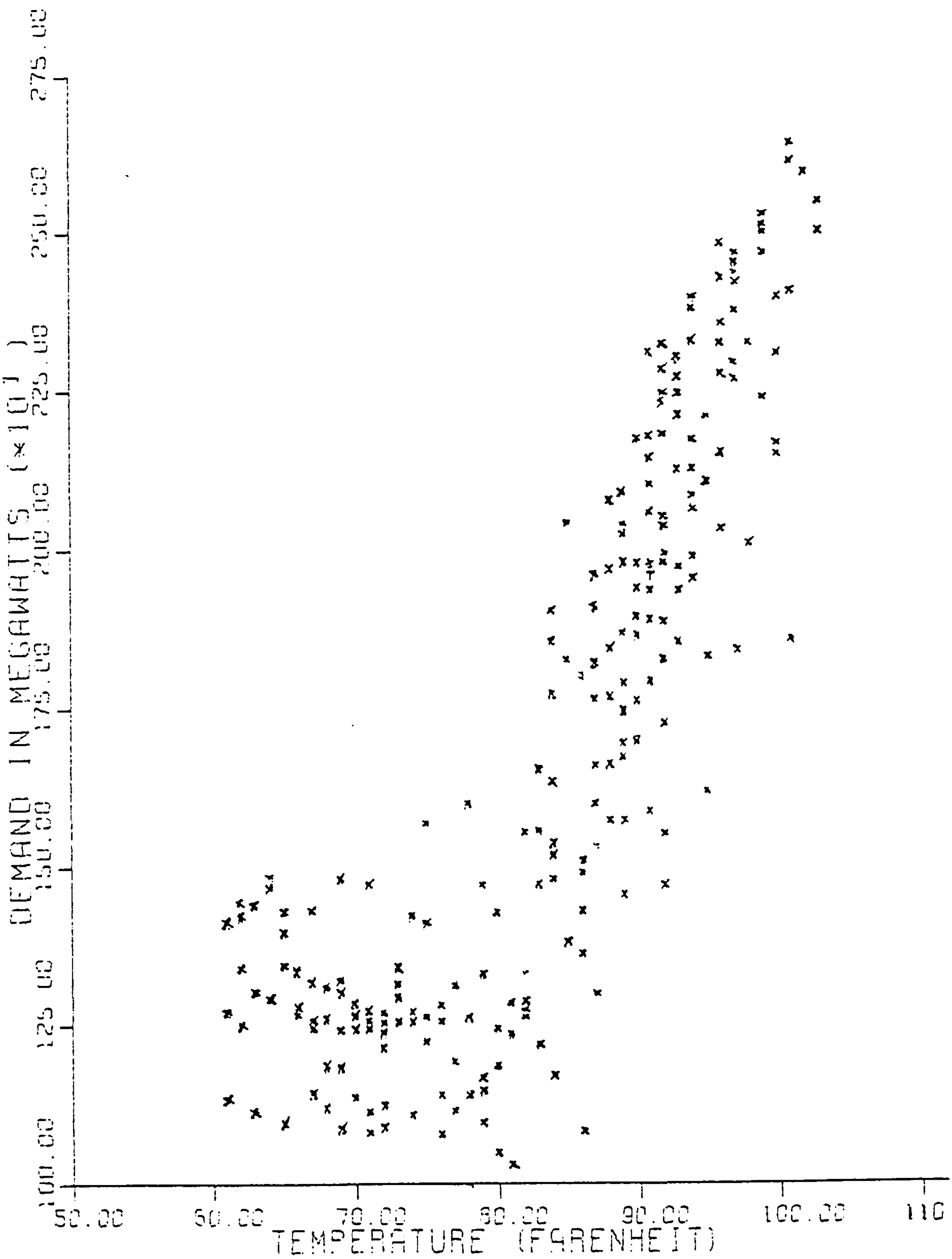


APPENDIX C

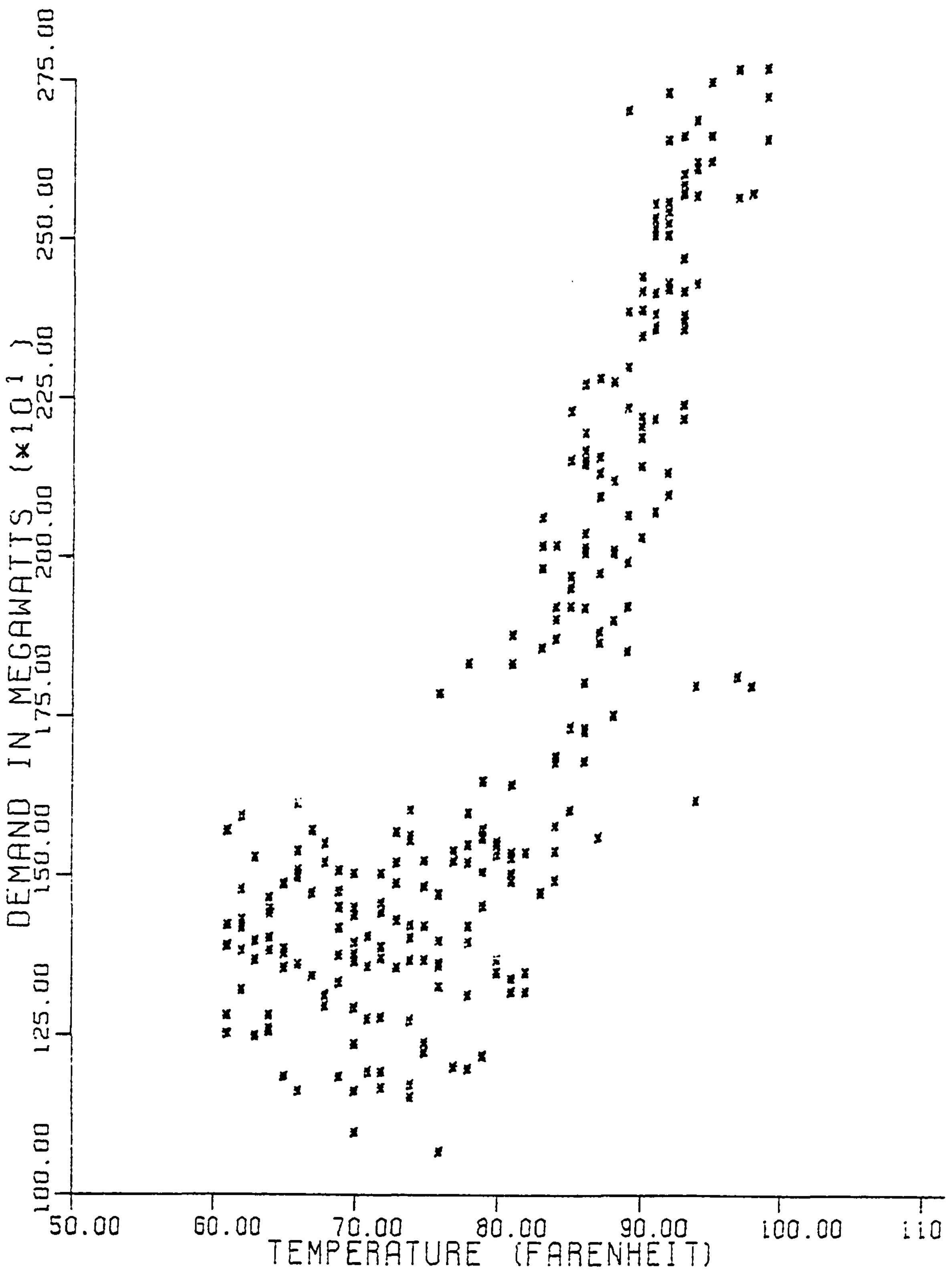
Regional Temperature Dependent Demand Characteristics.



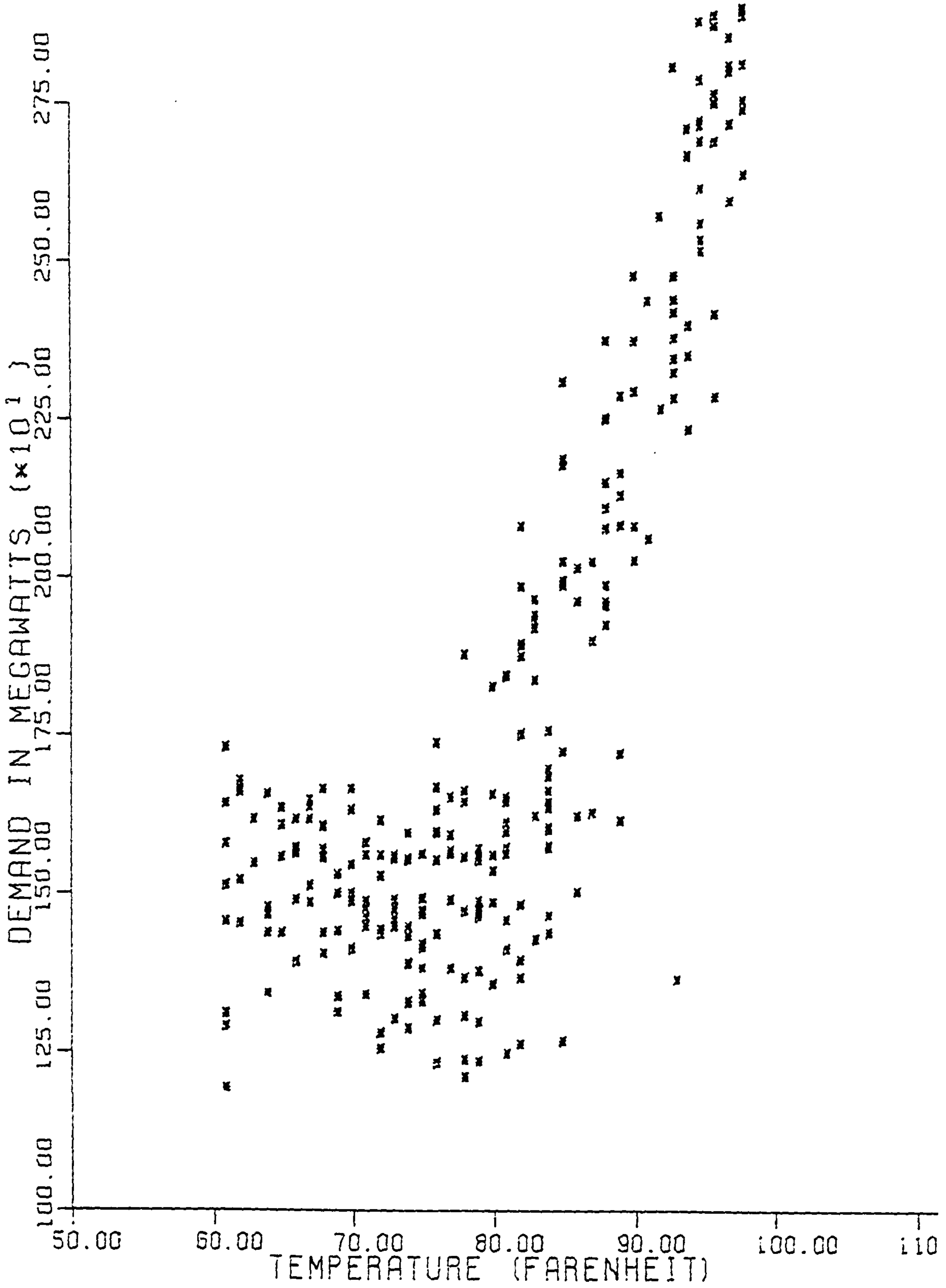
Demand Characteristic for Area 1, 1971.



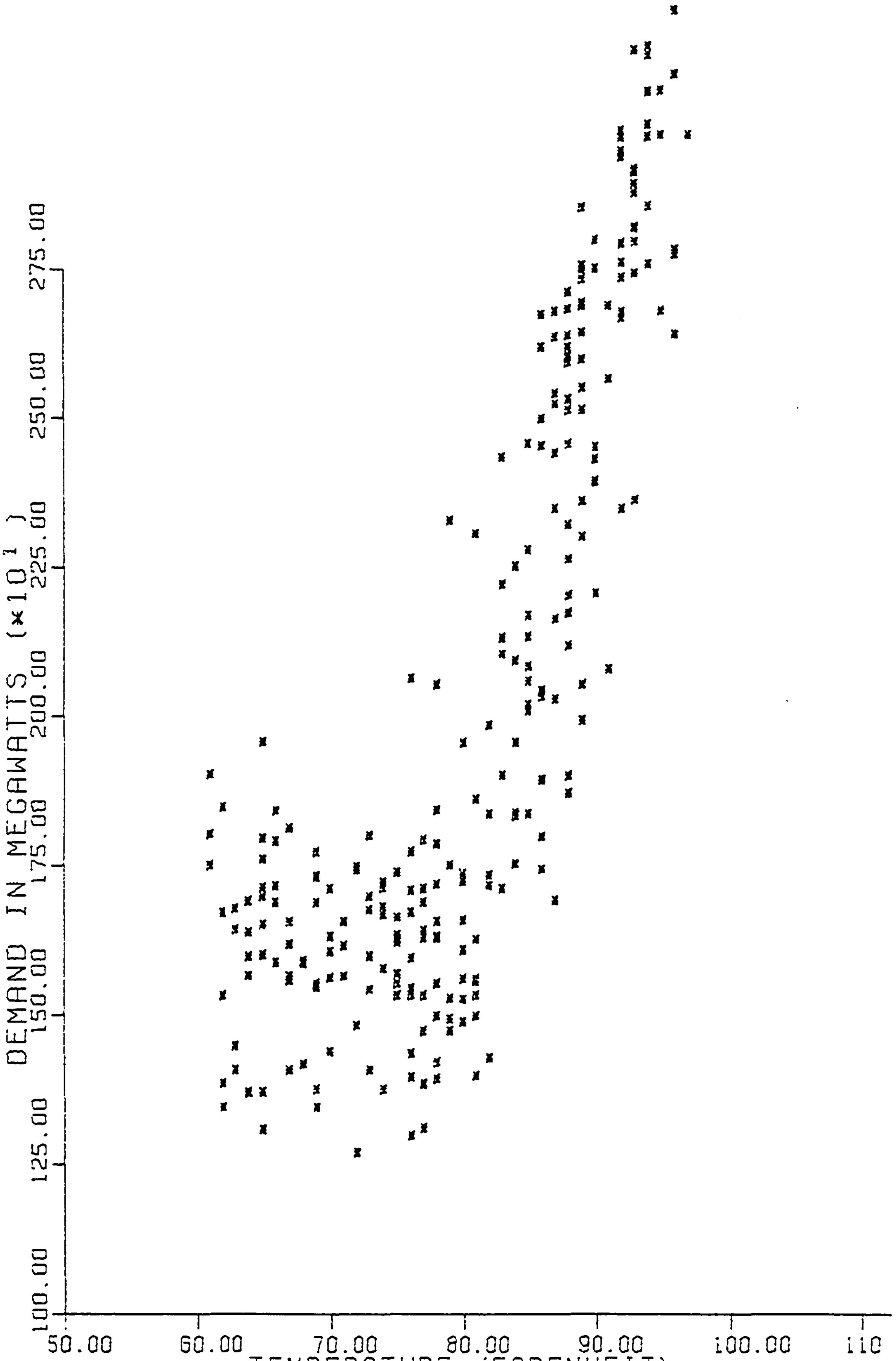
Demand Characteristic for Area 1, 1972.



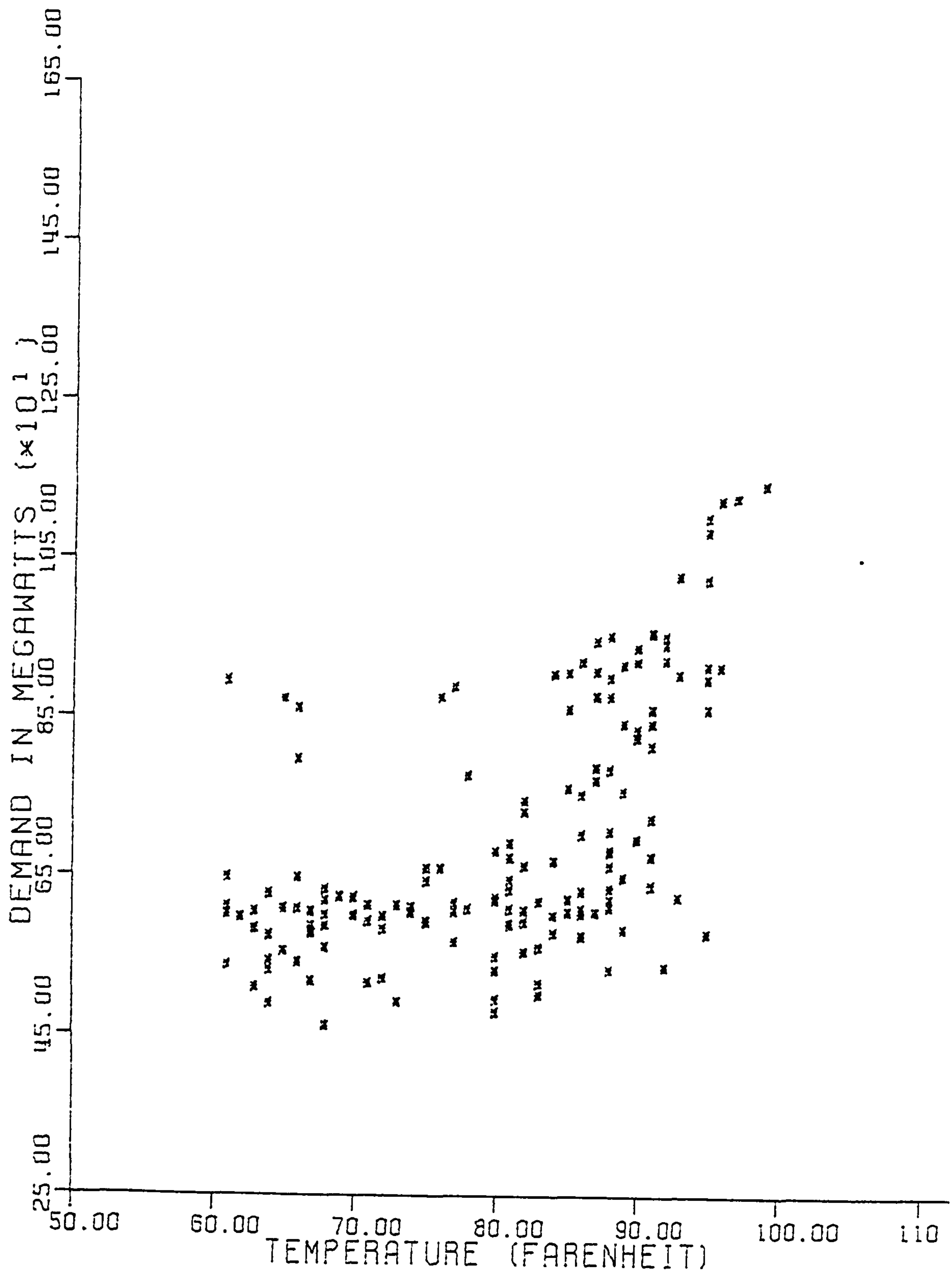
Demand Characteristic for Area 1, 1973.



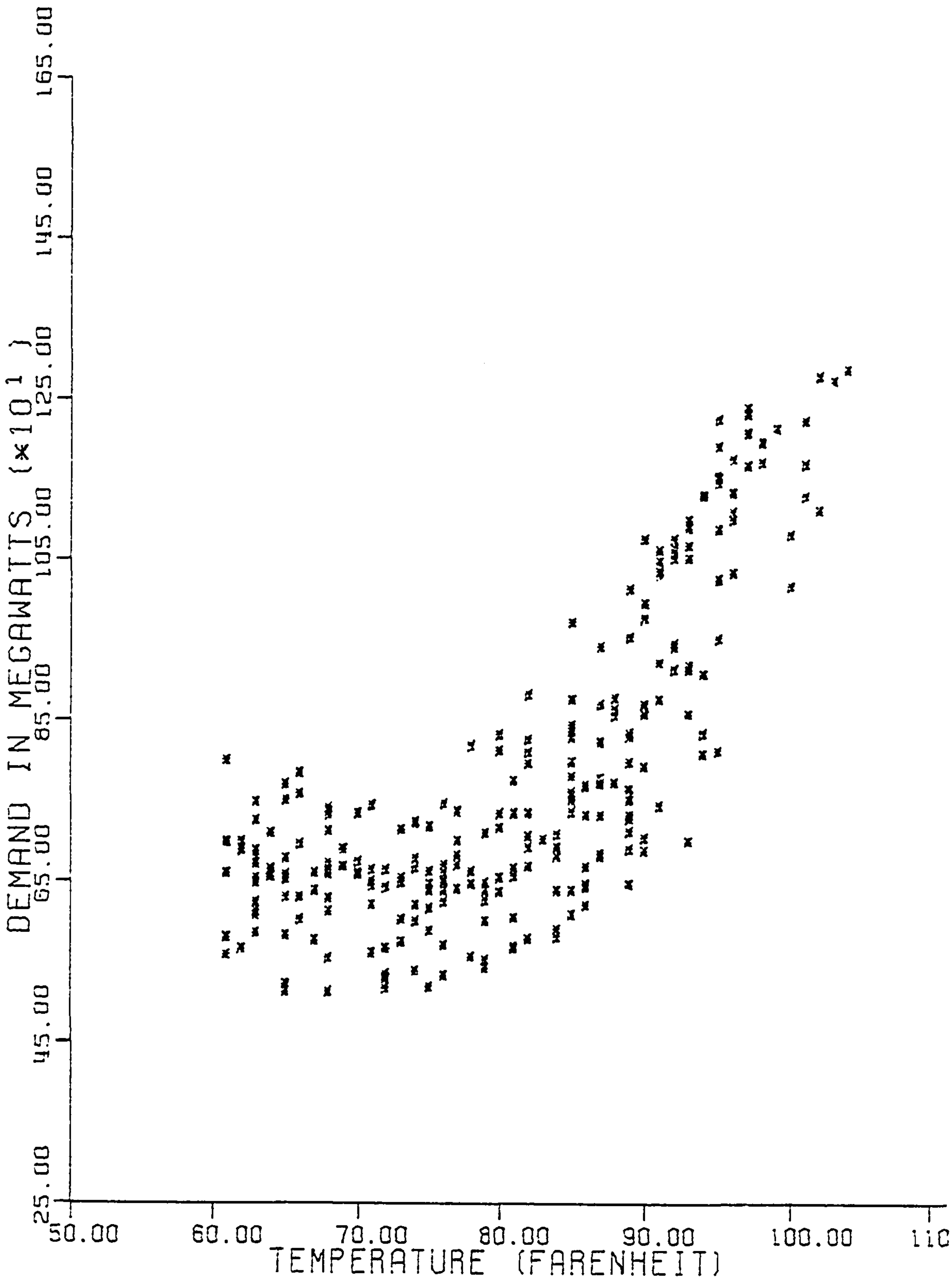
Demand Characteristic for Area 1, 1974.



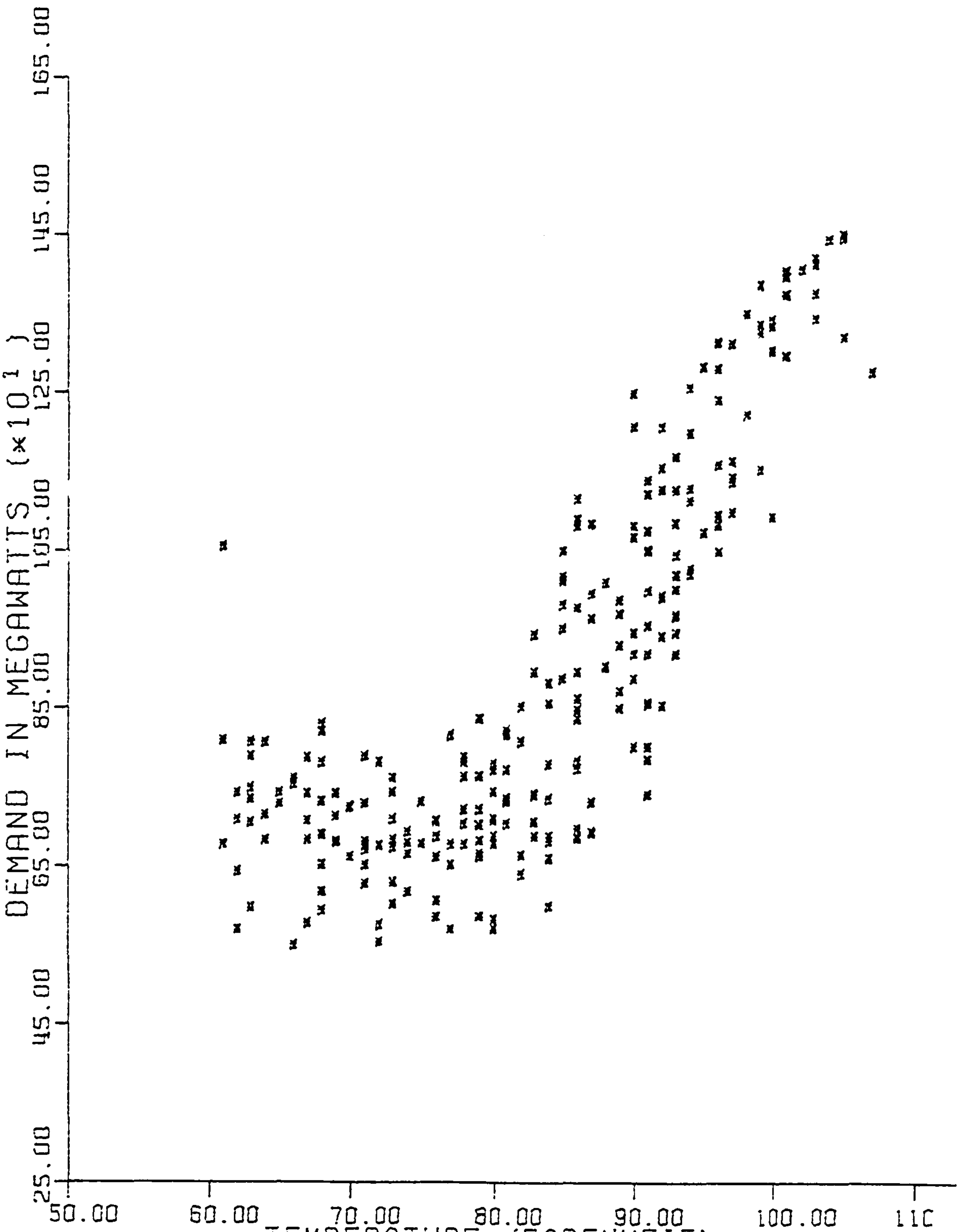
Demand Characteristic for Area 1, 1975.



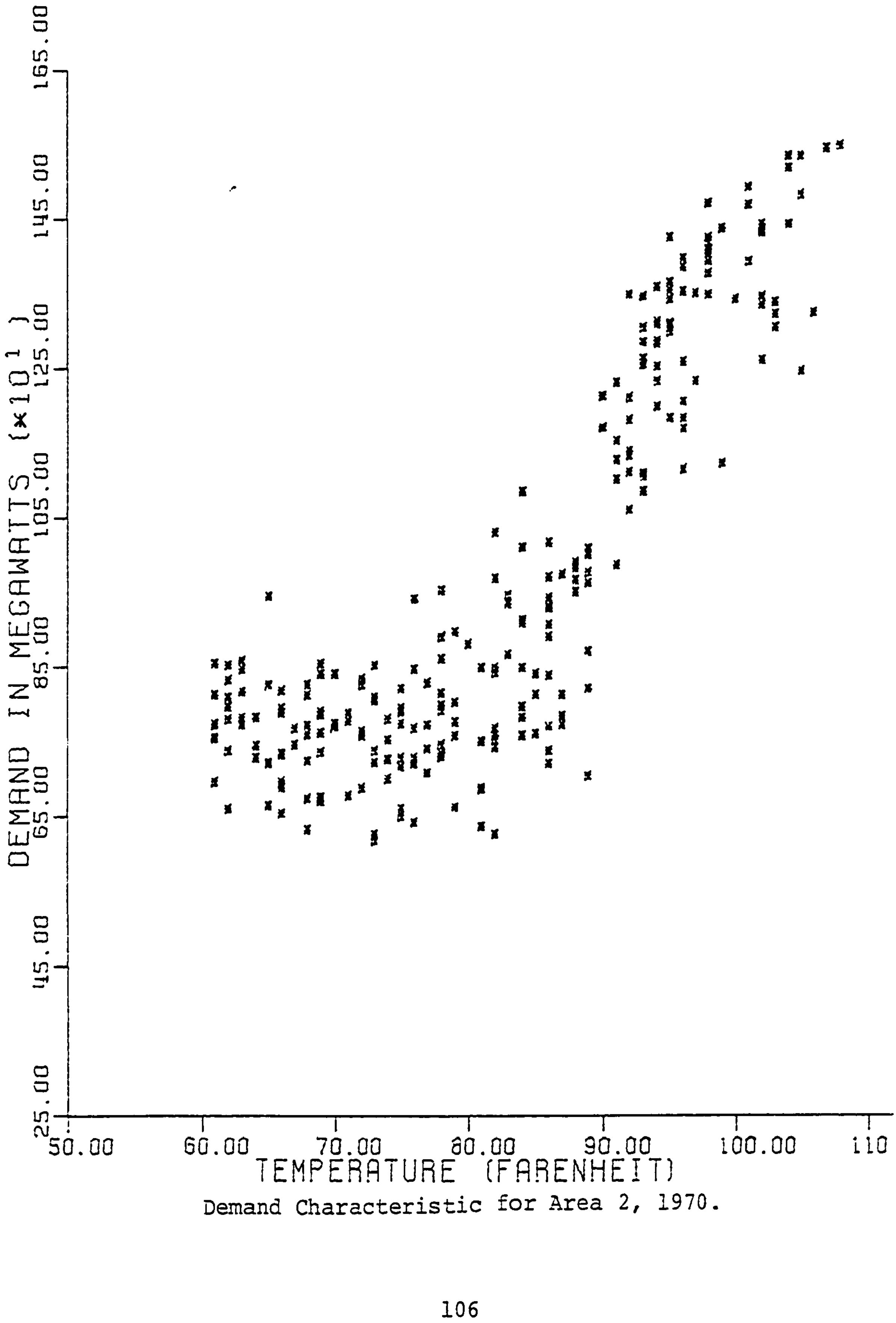
Demand Characteristic for Area 2, 1967.



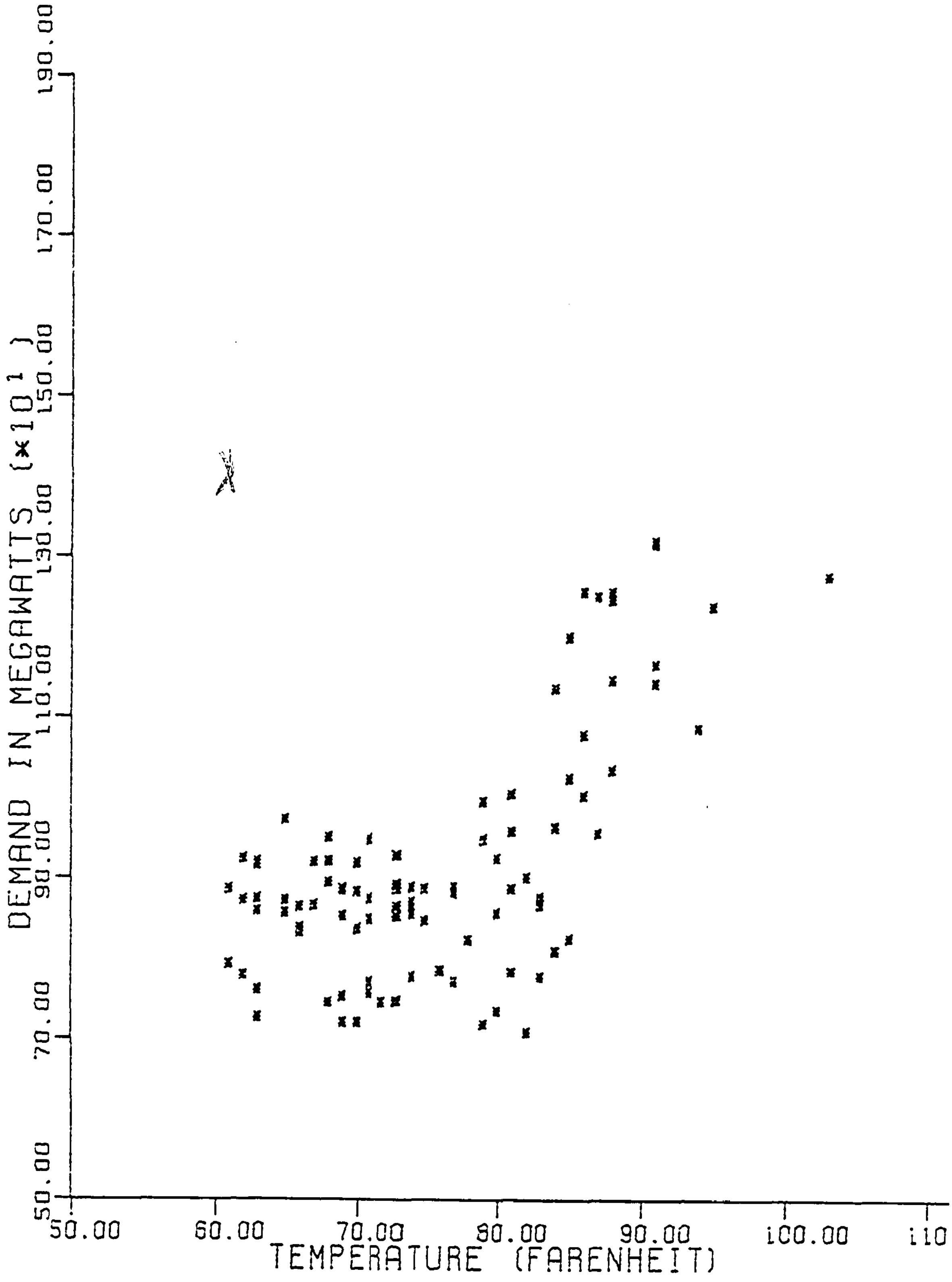
Demand Characteristic for Area 2, 1968.



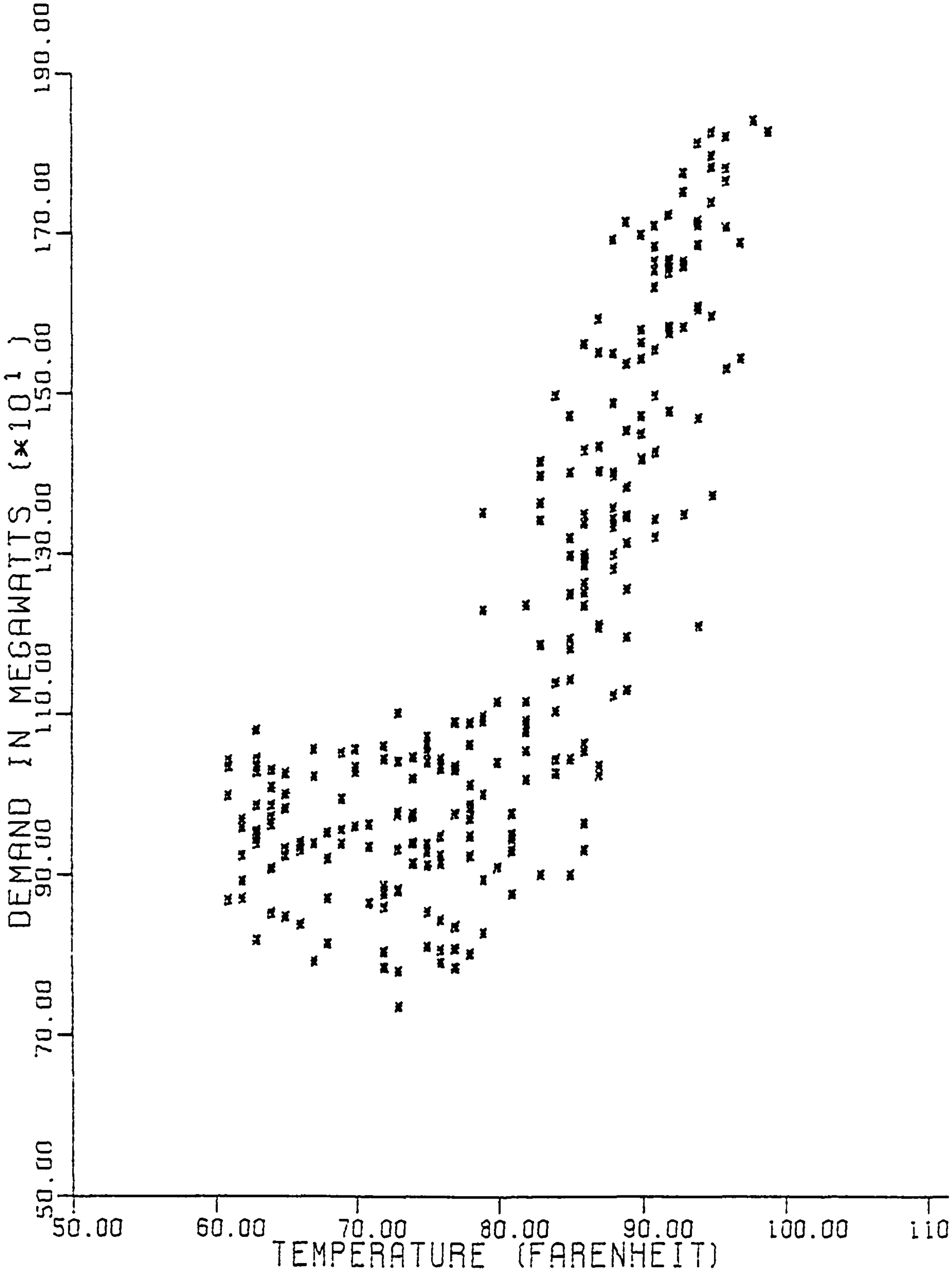
Demand Characteristic for Area 2, 1969.



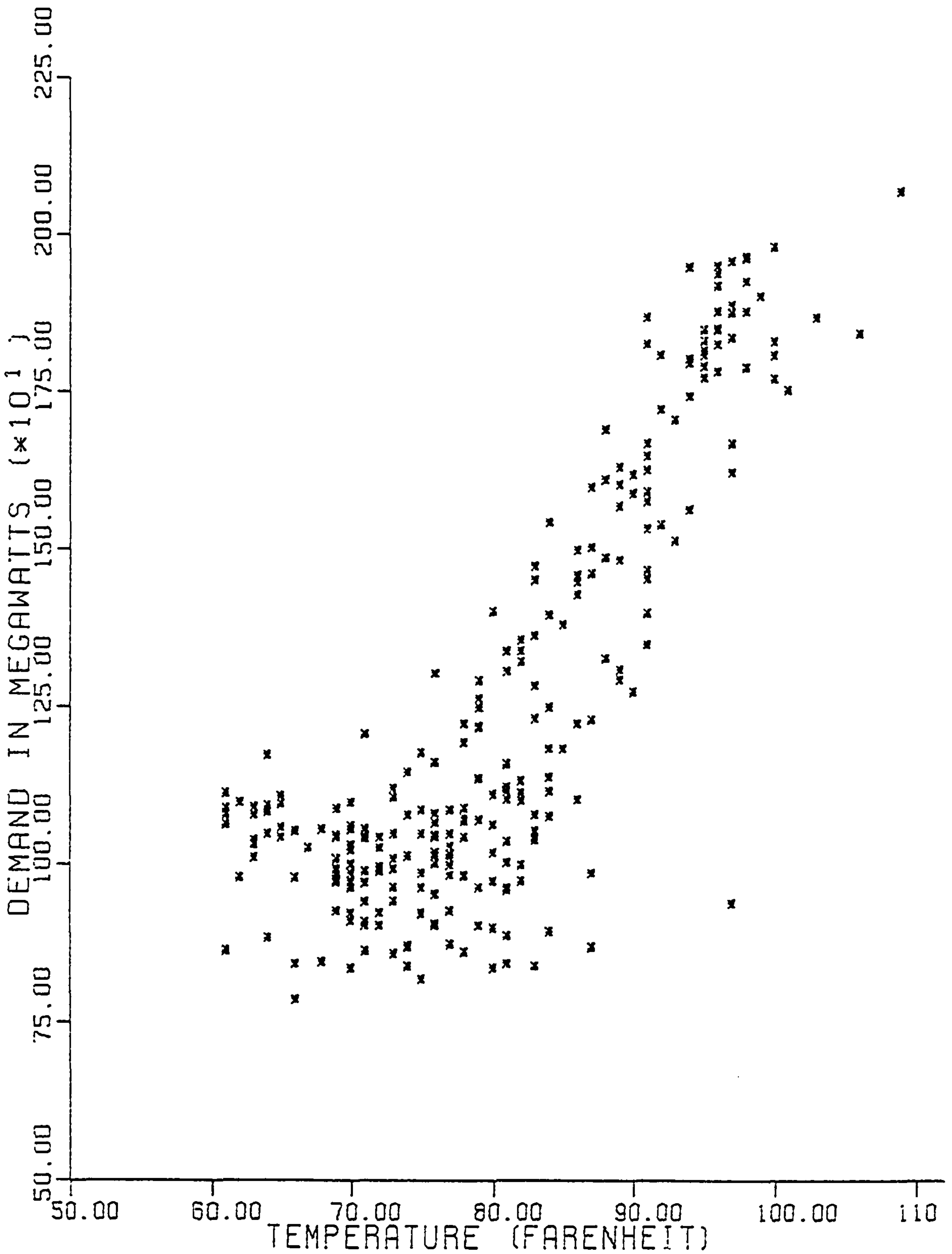
Demand Characteristic for Area 2, 1970.



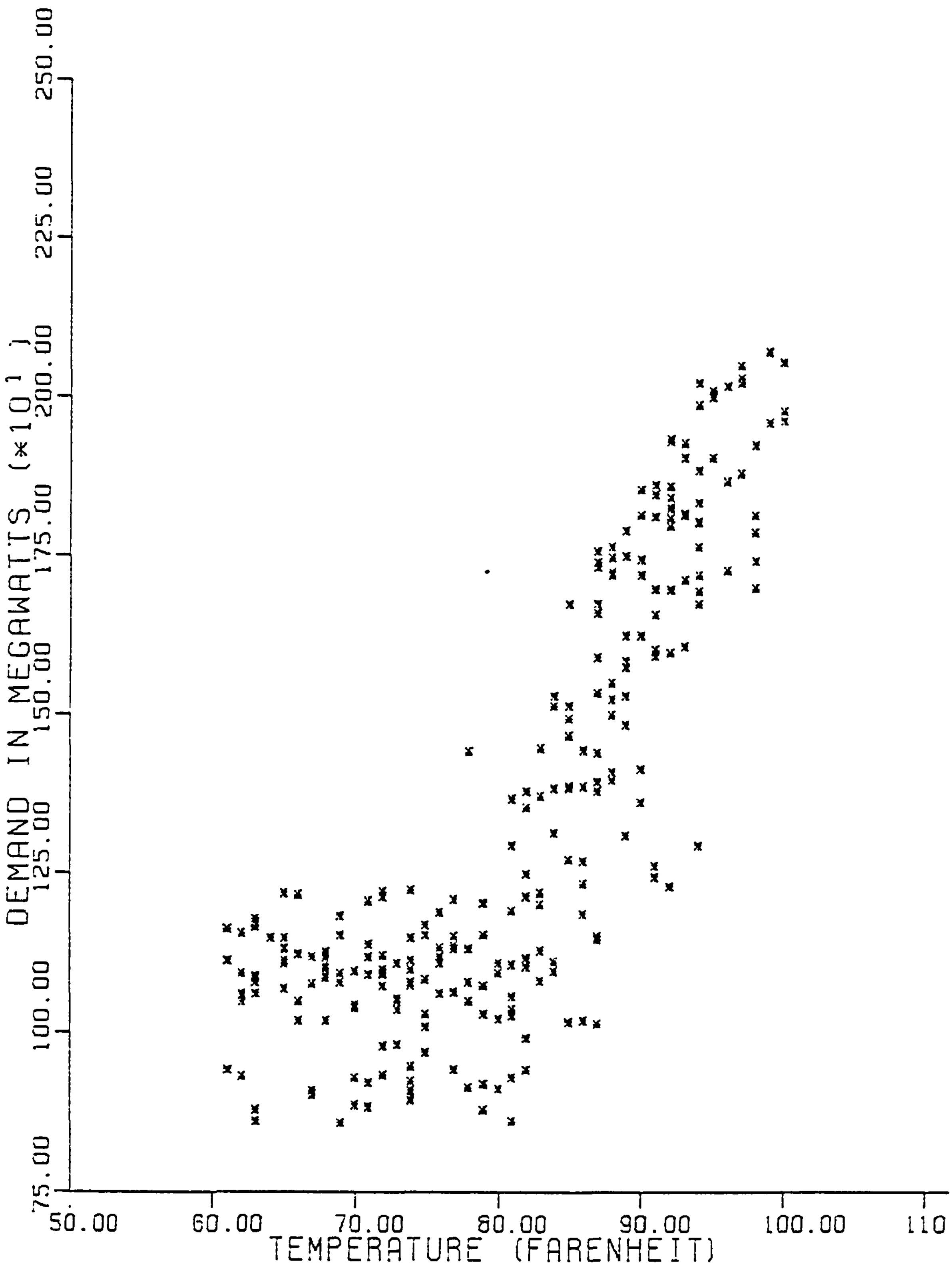
Demand Characteristic for Area 2, 1972.



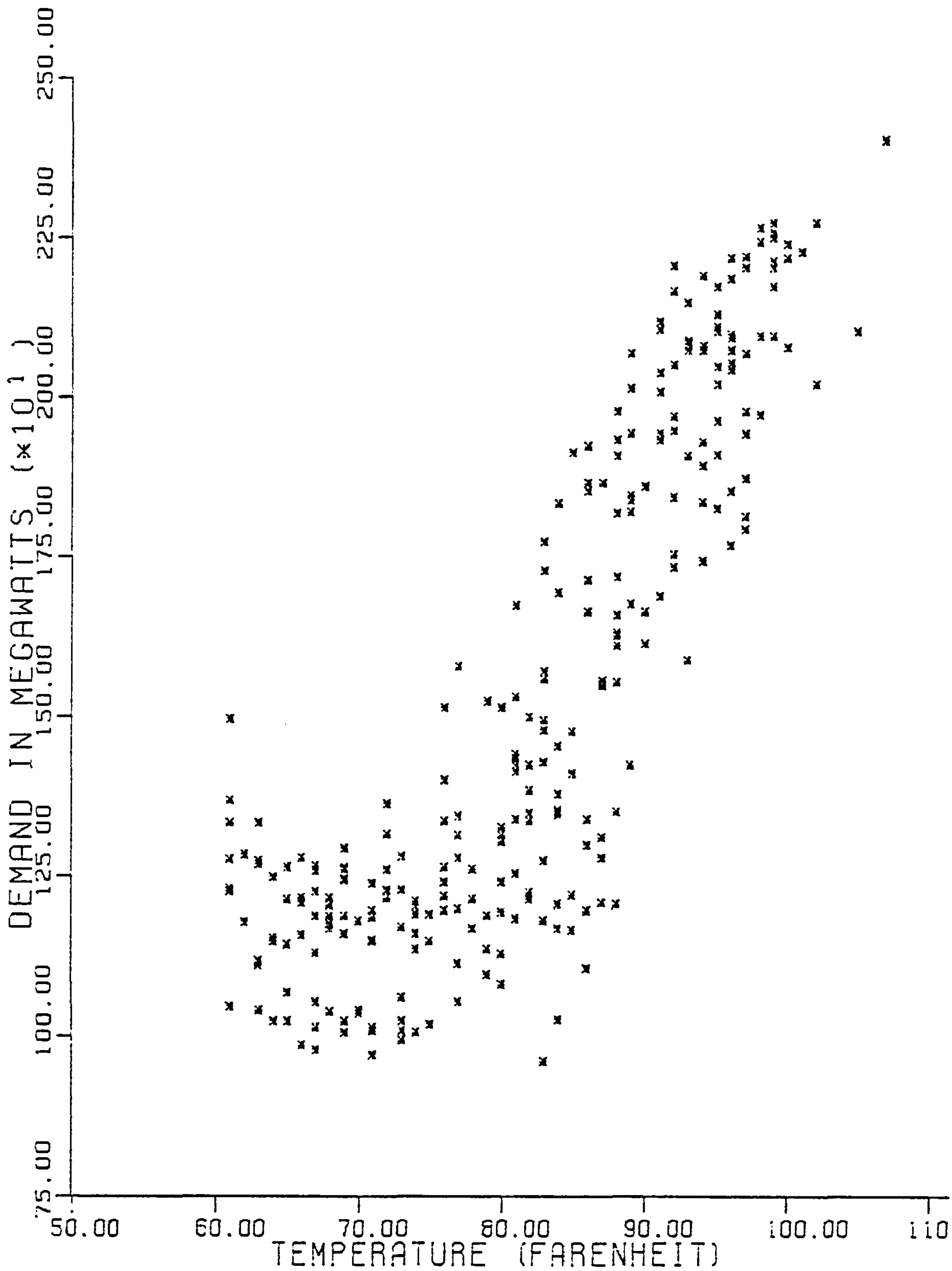
Demand Characteristic for Area 2, 1973.



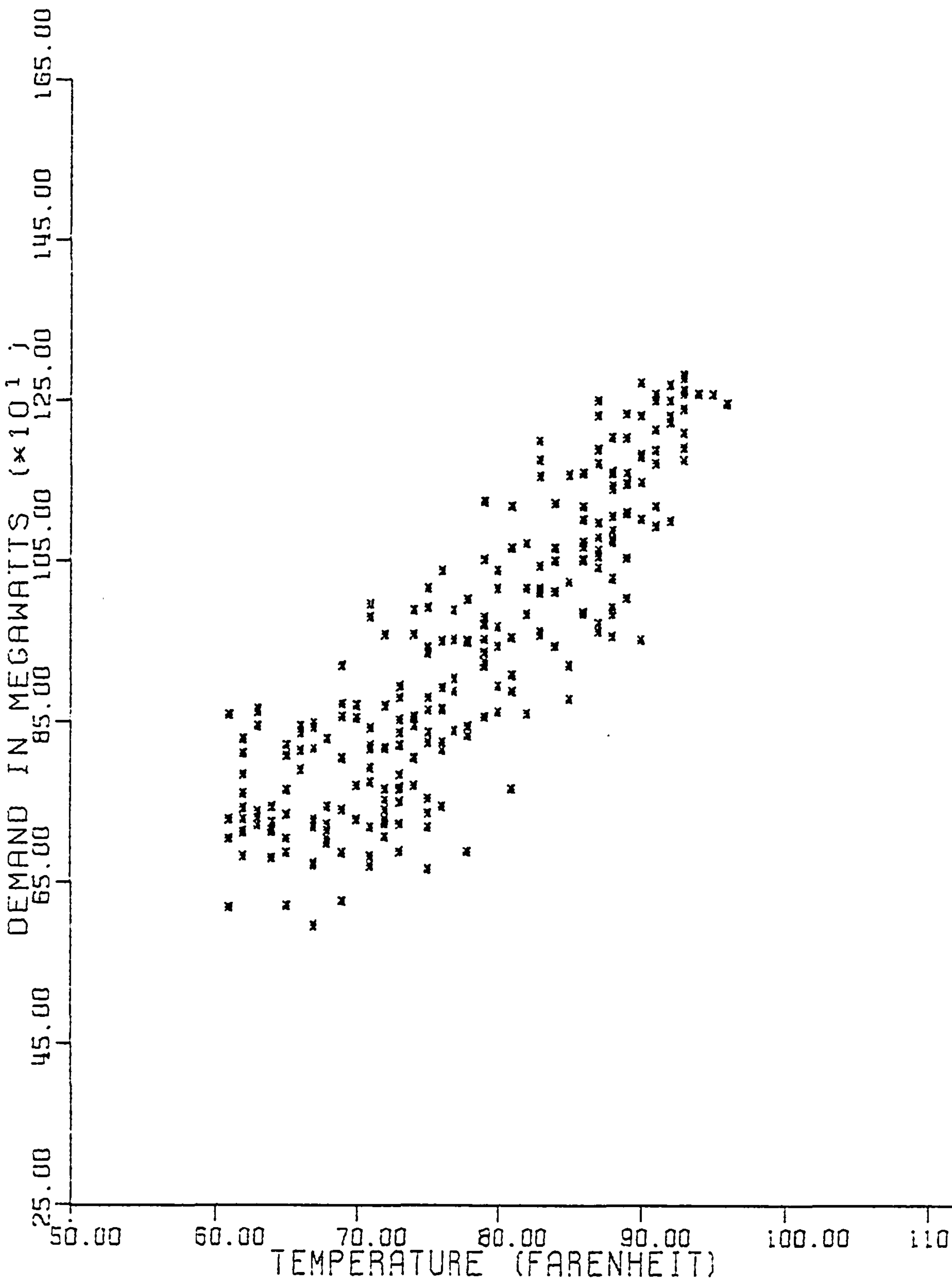
Demand Characteristic for Area 2, 1974.



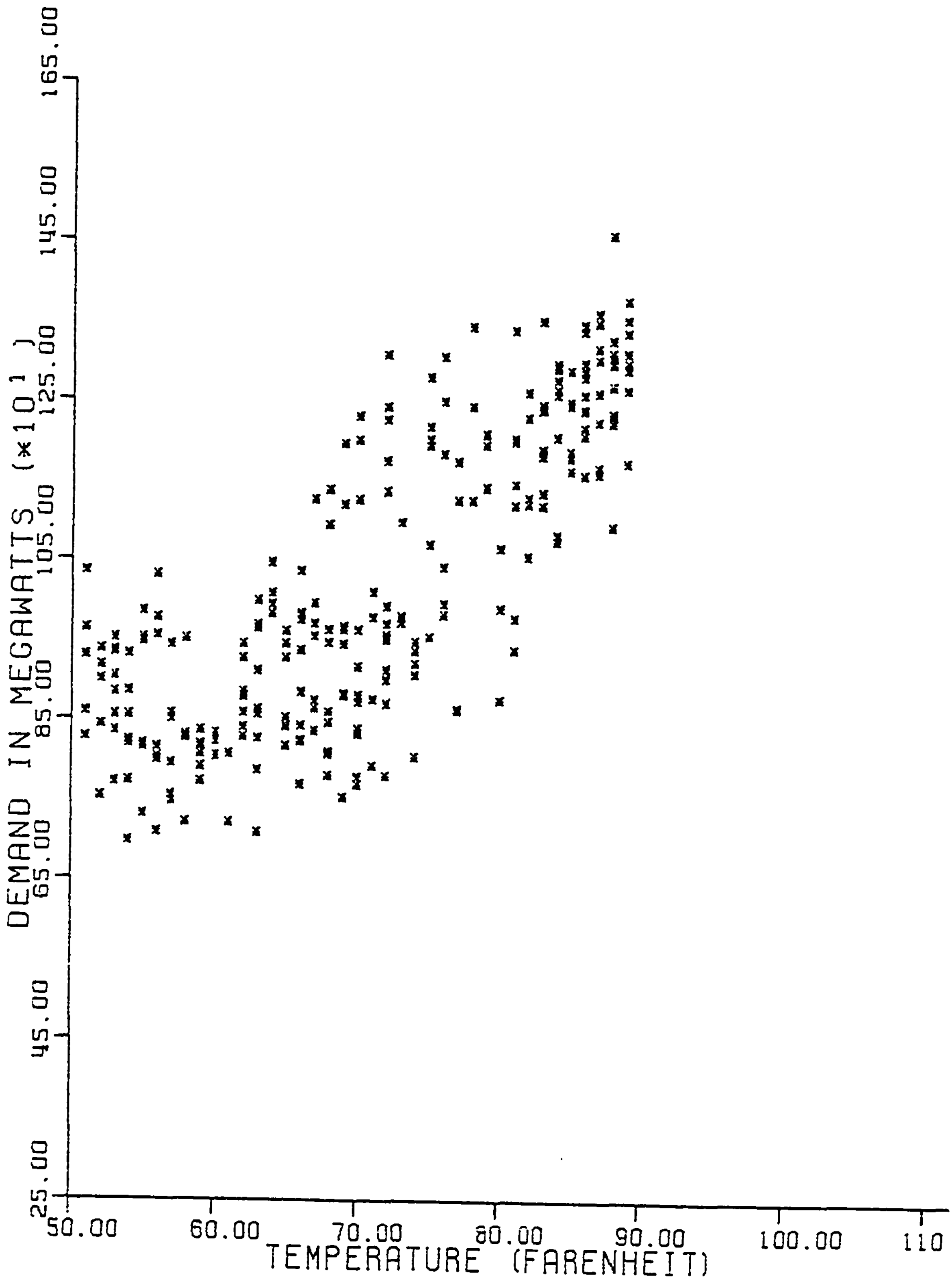
Demand Characteristic for Area 2, 1975.



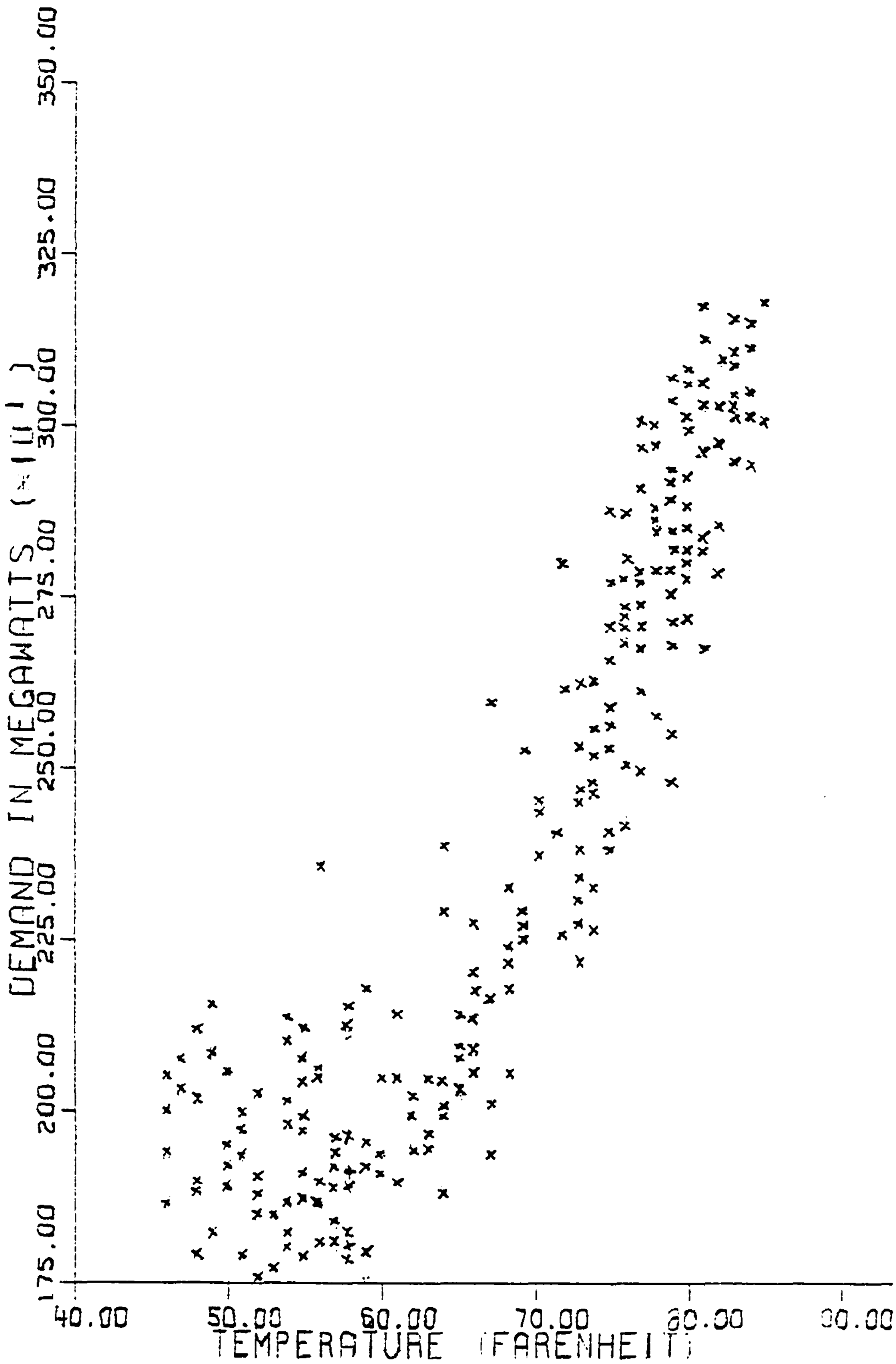
Demand Characteristic for Area 2, 1977.



Demand Characteristic for Area 5, 1967.



Demand Characteristic for Area 5, 1968.



Demand Characteristic for Area 5, 1976.