United States
Environmental Protection
Agency

Office of Monitoring and Technical
Support
Washington DC 20460

EPA-600/4-79-040
June 1979

Research and Development

ENVIRONMENTAL
PROTECTION
AGENCY
DALLAS, TEXAS
LIBRARY

Testing the Validity of the Lognormal Probability Model

Computer Analysis of Carbon Monoxide Data from U.S. Cities

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U S Environmental Protection Agency, have been grouped into nine series These nine broad categories were established to facilitate further development and application of environmental technology Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields The nine series are

1 Environmental Health Effects Research
2 Environmental Protection Technology
3 Ecological Research
4 Environmental Monitoring
5 Socioeconomic Environmental Studies
6 Scientific and Technical Assessment Reports (STAR)
7 Interagency Energy-Environment Research and Development
8 "Special" Reports
9 Miscellaneous Reports

This report has been assigned to the ENVIRONMENTAL MONITORING series. This series describes research conducted to develop new or improved methods and instrumentation for the identification and quantification of environmental pollutants at the lowest conceivably significant concentrations It also includes studies to determine the ambient concentrations of pollutants in the environment and/or the variance of pollutants as a function of time or meteorological factors

This document is available to the public through the National Technical Information Service Springfield Virginia 22161.

EPA-600/4-79-040
June 1979

EPA-600/4-79-040
June 1979

TESTING THE VALIDITY OF THE LOGNORMAL PROBABILITY MODEL: COMPUTER ANALYSIS OF CARBON MONOXIDE DATA FROM U.S. CITIES

by

Wayne R. Ott
David T. Mage
Victor W. Randecker*

OFFICE OF MONITORING AND TECHNICAL SUPPORT
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

*Currently with the Food Safety and Quality Service, U.S. Department of Agriculture, Washington, D.C.

DISCLAIMER

This report has been reviewed by the Office of Research and Development, U.S. Environmental Protection Agency, and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

FOREWORD

The U.S. Environmental Protection Agency (EPA) was created because of increasing public and governmental concern about the dangers of pollution to the health and welfare of the American people. Polluted air, foul water, and spoiled land are tragic testimony to the deterioration of our natural environment. The complexity of that environment and the interplay among its components require a concentrated and integrated attack on the problem.

Physical, chemical, and biological measurements of environmental quality are the most important means available for determining the state-of-the-environment. These monitoring data provide the basis for evaluating the Nation's progress in improving environmental quality and are essential to environmental managers and decision makers. Although great quantities of monitoring data are collected routinely across the Nation, the statistical and mathematical analyses performed upon these data usually are quite rudimentary. In recent years, there has been increased concern over the need to carry out more in-depth analyses of these data. Thus, EPA has increased its emphasis on the development of new data analysis techniques.

This report describes one of a series of projects that EPA's research program is conducting to demonstrate new methods for handling, analyzing, and interpreting environmental data. By providing greater insight into the statistical properties of environmental measurements, we hope that the results of this work will provide information that is useful in the regulatory process for setting policies and standards. We also hope that this work will help stimulate more comprehensive analyses of environmental data by the institutions that collect these data--municipalities, counties, states, federal agencies, private firms, and universities.

Albert C. Trakowski
Deputy Assistant Administrator
for Monitoring and Technical Support

PREFACE

This report is one of several that examines the frequency distributions of observed environmental pollutants. This report focuses on a particular probability model -- the two-parameter lognormal (LN2) probability model -- and it evaluates the suitability of this model for representing U.S. carbon monoxide (CO) air quality frequency distributions. Subsequent reports will examine other probability models, other pollutants, and other approaches (such as graphical techniques) for applying and evaluating various candidate models.

Although the LN2 model frequently has been used by air pollution field personnel, few air quality studies have treated the properties of the LN2 model in detail or examined how well the model fits the raw observations. In this study, a computer program has been developed which can apply the LN2 model to any CO data set, generating summary information about the goodness-of-fit of the model to the data. This computer program has been applied to a carefully selected nationwide sample of CO data sets. To make our work as complete as possible, this report fully documents the properties of the LN2 model, deriving all necessary formulas and equations. Thus, we hope it will serve as a reference document for others in the air pollution field.

This report has required approximately five years for its completion, primarily because the authors have full-time duties that involve other aspects of EPA's research program. A substantial portion of this work has been carried out on the authors' own time, and the present report was typed by the senior author. Despite the slow progress of our work, we hope that the present study is the most exhaustive and complete treatment currently available of the application of the LN2 model to CO frequency distributions. We also hope that, by demonstrating one of the many possible analytical techniques that can be applied to environmental data, we will help increase interest in the development of a formal, fully-funded statistical reasearch and development activity within the EPA.

ABSTRACT

A stratified sample consisting of 11 data sets from an original list of 166 carbon monoxide (CO) air quality data sets in SAROAD was selected as a "national cross section" of observed U.S. CO concentrations, along with a longitudinal group of data sets from a single air monitoring station. The adequacy of the two-parameter longnormal probability model then was evaluated using these data. A special-purpose computer program was developed for calculating the parameters of the LN2 model using four different techniques: (1) direct "Method of Moments," (2) Larsen's "Method of Fractiles," (3) "Maximum Likelihood Estimation" (MLE) for grouped data using an MLE approximation technique, and (4) MLE for grouped data using computer optimization.

The goodness-of-fit of the LN2 model to the data was evaluated using frequency-based approaches (e.g., chi-square, magnitude of the log-likelihood function, Kolmogorov-Smirnov measures) and variate-based approaches (e.g., the difference between the concentration predicted by the model for each interval and the concentration actually observed). The findings show that the method of calculating parameters for the LN2 model exerts a profound influence on the goodness-of-fit of the model to the data, particularly at the higher concentrations of greatest interest for protecting public health. The method of fractiles performs best for the variate-based tests but poorest for the frequency-based tests, suggesting that it is responding more to randomness at the extremes than to the underlying distribution of the process.

ACKNOWLEDGMENTS

Grateful appreciation is expressed to the following individuals, all employees of the Environmental Protection Agency, who were kind enough to review this report and offer many useful suggestions and comments: Ralph Larsen, Lance Wallace, Teri Gardinier, Robert Papetti, William F. Hunt, Gerald Akland, Terence Fitz-Simons, David Holland, and Irene Kiefer (consulting technical editor). We wish to thank them for their excellent technical reviews and the many thoughtful suggestions we received.

CONTENTS

FOREWORD iii
PREFACE iv
ABSTRACT v
ACKNOWLEDGMENTS vi

I. INTRODUCTION 1
STRUCTURE OF REPORT 4
II. LITERATURE REVIEW 6
III. STRUCTURE OF THE NORMAL AND LOGNORMAL PROBABILITY MODELS 14
MOMENTS 16
NORMAL PROBABILITY MODEL 21
LOGNORMAL PROBABILITY MODEL 29
IV. CALCULATION OF PARAMETERS 45
MOMENTS OF THE DATA 45
METHOD OF MOMENTS 49
METHOD OF FRACTILES 52
MAXIMUM LIKELIHOOD ESTIMATION 61
MLE Approximation 64
MLE Optimization 67
V. GOODNESS-OF-FIT 70
FREQUENCY-BASED MEASURES 70
VARIATE-BASED MEASURES 74
VI. DESCRIPTION OF THE COMPUTER PROGRAM 81
MAIN PROGRAM 81
SUBROUTINES 83
VII. SELECTION OF THE DATA 95
GEOGRAPHICAL GROUP 98
LONGITUDINAL GROUP 109
VIII. RESULTS FOR THE GEOGRAPHICAL DATA GROUP 111
FINDINGS FOR INDIVIDUAL CITIES 111
Barstow, CA 113
Norfolk, VA 118
Alexandria, VA 122
Denver, CO 126
Philadelphia, PA 130
Napa, CA 133
Phoenix, AZ 136
Newhall, CA 141
Springfield, MA 144
Pasadena, CA 147
New York City, NY 150
Discussion 153

OVERALL FINDINGS 155
Discussion 170
IX. RESULTS FOR THE LONGITUDINAL DATA GROUP 173
FINDINGS FOR INDIVIDUAL YEARS 173
1966 175
1968 176
1969 182
1970 183
1971 186
1972 187
1973 194
OVERALL FINDINGS 195
Discussion 197
X. CONCLUSIONS 204

REFERENCES 208

APPENDICES

A. LISTING OF THE COMPUTER PROGRAM. 215
B. SAMPLE OUTPUT FOR NEW YORK CITY. 231
C. BASIC STATISTICS, HISTOGRAMS, AND AUTOCORRELATIONS FOR THE GEOGRAPHICAL DATA GROUP. 259
D. BASIC STATISTICS, HISTOGRAMS, AND AUTOCORRELATIONS FOR THE LONGITUDINAL DATA GROUP. 305

I. INTRODUCTION

Our review of the research literature (Chapter II) shows that numerous claims have been made that air pollution concentration data, measured in the ambient air, are "lognormally distributed." Despite the frequent use of the phrase "lognormally distributed," few authors state specifically what they mean by the phrase. When is a distribution judged "lognormal" and how does one make this determination? If some objective criterion were available for lognormality, to what degree would air quality data for cities throughout the U.S. meet this criterion?

A reader carefully reviewing the air pollution literature may be struck even more by the lack of standard statistical nomenclature and a lack of application of well known goodness-of-fit tests. The standard chi-square test, for example, which is useful for examining the hypothesis that the underlying distribution from which a random sample was drawn is lognormal, seldom has been applied in articles on data collected by the air pollution control community. Rather, the usual approach has been to plot the data on logarithmic probability paper. If, upon subjective examination, the resulting line "looks straight," the claim is made that the data are "lognormally distributed."

After careful consideration, we have concluded that two general philosophies of data analysis currently are being

followed: (1) the "statisticians' approach," and (2) the "engineers' approach." Each philosophy addresses a different question. The statistician, generally having a random sample from some larger population, seeks to make inferences about the characteristics of the larger population. The engineer, on the other hand, often possessing the entire population of data, is interested in representing the population graphically or estimating the approximate impact of regulatory pollution control decisions upon this population. The two philosophies, because they are fundamentally different, produce interesting inconsistencies. For example, the engineer may plot environmental data on logarithmic probability paper and decide that the resulting line "looks straight" and that the lognormal model is a good candidate for his application. In contrast, the statistician may apply the chi-square test to the same data and conclude that he should reject the hypothesis that a given data set has a lognormal distribution (i.e., arises from a lognormally distributed population). Very early in this investigation, we observed that data sets which appear as straight lines on logarithmic probability paper can seriously fail the chi-square and other statistical tests.

We began considering the implications of the two approahces in 1975. Although the theoretical basis for the statisticians' approach is well developed, we were troubled that the engineers' approach appears to have slipped into the scientific literature with little fanfare and no formal theoretical explanation.

Except for a recent technical note by Mage and Ott,[1] few papers have discussed the existence of these two fundamentally different approaches. Furthermore, no objective criteria are available to assist the engineer in determining whether a given model is suitable for a given application. Thus, the engineers' approach potentially may be misunderstood by the statistical community, resulting in criticism of a methodology that may be reasonable from an engineering standpoint.

Because of the existence of two disparate data analysis philosophies, there currently is much ambiguity about the claim that a given data set "fits" or "does not fit" a particular distribution, and there is much opportunity for confusion. An important purpose of the present study was to alleviate possible confusion arising from these two philosophies, with particular emphasis on the engineers' approach. By giving additional development and structure to the engineers' approach to air quality data analysis, we hope that we will move one step further to providing engineers with objective, uniform tools by which they can carry out their jobs.

In planning this study, our goal was to conduct the most thorough and extensive evaluation ever undertaken of the suitability of the 2-parameter lognormal (LN2) probability model for representing carbon monoxide (CO) air quality frequency distributions in the United States. Although some past studies have dealt extensively with the LN2 and other distributions, few have applied formal goodness-of-fit tests to the data. In addition,

most past studies were "data limited;" they often used just one or two data sets that were not representative of any particular universe. This study has sought to improve upon previous studies, thereby advancing the state of knowledge about this topic, by incorporating the following features:

- Two data groups are used: (1) a "geographical data group" designed to represent a cross section of U.S. cities in a particular year, and (2) a "longitudinal data group" designed to represent a number of different years in one U.S. city.

- Four different ways to compute the parameters of the 2-parameter lognormal probability model are compared and examined in detail: (1) method of moments, (2) method of fractiles, (3) an approximation of maximum likelihood estimation for grouped data, (4) exact maximum likelihood estimation using computer optimization.

- A number of different goodness-of-fit tests are used to evaluate the adequacy of the model for use with air quality data from U.S. cities.

- An engineering error test (EET) is developed that enables the performance of a model to be evaluated in the context of a criterion familiar to the air pollution professional: namely, the difference between the concentration predicted by the model and the concentration actually observed, expressed in parts-per-million (ppm).

STRUCTURE OF REPORT

The overall structure of this report should be relatively easy for the reader to follow. Chapter II consists of a literature review, including a historical summary of applications of

the lognormal probability model. Chapter III presents the basic structure of the normal and lognormal probability models, and the equations commonly associated with these models are derived and discussed in detail. Chapter IV describes the methodology used in this study to calculate parameters for the lognormal probability model from the CO data sets, and Chapter V describes the methodology used to evaluate the goodness-of-fit of the model to the data. The methodology described in Chapters III, IV, and V is incorporated into a computer program, HIST.IIB, which is described in Chapter VI (and listed in Appendix A). Chapter VII tells how the air quality data used in this investigation were selected. The remaining three chapters cover the results obtained by applying the computer program to the air quality data sets described in Chapter VII. Chapter VIII presents the results for the geographical data group, and Chapter IX presents the results for the longitudinal data group. Finally, the overall conclusions of this research investigation can be found in Chapter X. The appendices include the computer program along with a sample of the full output for one city, and portions of the outputs containing the basic statistics for all cities in the geographical and longitudinal data groups.

A companion report[2] that is planned for publication in 1980 discusses these data sets in the context of a variety of other probability models and presents graphical techniques for selecting and applying the models.

II. LITERATURE REVIEW

As noted by Aitchison and Brown,[3] the two-parameter lognormal distribution has a long history of application in the field of small particle statistics. The size distribution of small particles produced in crushing and grinding operations has been found to be approximately lognormal,[4,5] and a theoretical rationale for the lognormality of particles resulting from breakage processes has been given by Kolmogorov.[6] Analyses of the dust generated by industrial grinding processes have shown the LN2 model to be a good descriptor of the size distribution of these particles.[7]

It was logical to expect that this work would be extended to describe the distributions of particles in the atmosphere. In 1956, Harris and Tabor[8] applied the LN2 model to air pollution particle statistics, and in 1959 Zimmer, Tabor, and Stern[9] described concentrations of total suspended particulate (TSP) in both urban and nonurban ambient atmospheres as approximately LN2. Although they noted that there were some marked deviations from lognormality, they concluded that "...in general, concentrations of particulate matter are lognormally distributed."[9] More recently, deNevers, Lee, and Frank[10] have suggested that the marked deviations from lognormality associated with TSP data may be caused by sampling from two separate lognormal distributions.

Larsen[11] extended the application of the LN2 model from TSP to the gaseous air pollutants, and, in a 1961 paper, he stated that carbon monoxide and oxidant concentrations in the Los Angeles area "...indicate that the variables tend to be logarithmically distributed." In subsequent papers, Larsen[12,13] continued to apply the LN2 model to air quality data, and in 1964 Larsen[14] expressed the suitability of the LN2 model for air quality data in general terms: "Air pollutant concentration data usually fit the bell (Gaussian) shape, if concentration is plotted on a logarithmic scale." Larsen and his associates[15-17] applied the LN2 model to additional air quality data sets, and in a 1967 paper they generalized even further about the suitability of the LN2 model for air quality data: "...concentrations are approximately lognormally distributed for all pollutants in all cities for all averaging times."[18] Larsen[19-23] applied the LN2 model to a variety of pollutants and cities and further refined his "averaging time model" so that the LN2 parameters and maxima could be calculated for any averaging time. In a 1971 report, Larsen[24] presented the averaging time model in considerable detail and concluded: "Pollutant concentrations are lognormally distributed for all averaging times."

In 1972, Lowrimore[25] derived the probability distribution of the sums of lognormally distributed random variables and proposed a compound Poisson-lognormal distribution for some environmental applications. From 1972 to 1975 Larsen[26-32] and others[33] continued to develop the LN2 model and apply it to air quality

data, generally to the exclusion of any other probability models.

During this period and earlier, some authors attempted to explain the apparent lognormality of air quality data in theoretical terms. A 1969 paper by Gifford[34] offered a theoretical basis for the LN2 model, and several papers containing theoretical explanations were presented at a 1972 symposium on statistical aspects of air quality data;[35] others have appeared elsewhere in the literature. The explanations usually are based either on statistical theory[34,36-39] or on properties of the meteorological data.[40,41] The statistical explanations usually involve the central limit theorem and the diffusion law, while the meteorological explanations assume that one of the meteorological variables, such as wind speed, is lognormally distributed.

Although a large number of applications of the LN2 model to air quality data had been undertaken by the mid-1970's, few authors had examined the goodness-of-fit of the model to the data or had considered other candidate models. The first comprehensive comparison of the LN2 model with other models was made by Lynn in 1974.[42] Lynn examined TSP data from Philadelphia and examined the suitability of the normal, LN2, and three-parameter lognormal (LN3) probability models, along with the Pearson Types I and IV distributions and the Gamma distribution. Lynn concluded that "...the two-parameter lognormal does overall slightly better than the three-parameter and in fact does the best of all four distributions (LN2, LN3, Gamma and four-

parameter Pearson)." This result was unexpected, because the LN2 model really is a special case of the LN3 model for which the LN3 third parameter is zero.

The near-unanimous acceptance of the LN2 model continued until 1974 when Mage[43] demonstrated that the LN3 model was superior to the LN2 model for every air quality data set considered, provided that the model is censored when the third parameter is negative. Lack of censorship appearantly explains why Lynn[42] did not find the LN3 model to be superior to LN2 model. In 1975 and 1976, Mage and Ott[44-46] recommended the censored three parameter lognormal (LN3C) model as a general-purpose model which could represent air quality frequency distributions in a large number of circumstances. The LN3 model is a special case of the LN3C model for which the third parameter is positive, and the LN2 model is a special case for which the third parameter is zero. Mage and Ott[44-46] also suggested a possible theoretical rationale for this model: They transformed the second-order partial differential equation for diffusion into a first-order linear differential equation (Langevin equation) and found that, in one special case at least, the LN3C model might naturally arise. They also applied the LN3C model to water quality data.[46]

In the mid-1970's, several investigators began to apply goodness-of-fit criteria to the models, and, upon careful inspection, were finding that the LN2 model did not fit air quality data especially well. In 1976, Kalpasanov and Kurchatova[47] applied the LN2 model to Bulgarian air quality data and found

that it was rejected by the Kolmogorov-Smirnov goodness-of-fit criterion. In some cases, the normal probability model exhibited a goodness-of-fit that was superior to the LN2 model. Using the exponential model as a reference, Curran and Frank[48] analyzed the "tails" of distributions of air quality data. They found that, in general, the cumulative frequencies of the data approached unity more rapidly than the cumulative distribution function (CDF) of the exponential probability model. They called this property "light tailed." They showed that the LN2 was "heavy tailed" since its CDF approached unity more slowly than the exponential model. They reasoned that a light-tailed distribution, such as the Weibull probability model, should be considered for fitting light-tailed data sets.

The Weibull distribution was applied to environmental problems by Mikolaj,[49] who modeled the lengths of oil slicks caused by oil tanker spills. Johnson[50] has reported that ozone data appear better suited to the Weibull model than to the LN2 model. Bencala and Seinfeld[51] have applied the Weibull, Gamma, LN2, and LN3 distributions to air quality data and compared their relative goodness-of-fit. Although they state that the LN3 model gives the best fit of all candidate models, they conclude that the air quality data are approximately LN2 in distributional form.

In 1977, Larsen[52] adopted the LN3 and LN3C models and recommended their use over the standard LN2 model for analysis of air quality data. As noted by Mage and Ott,[1] Larsen's implementation of the LN3C differed from previous authors[44-46] in the manner in which the third parameter was defined, and his initial

paper contained several errors. Larsen[52-56] continued to apply the LN3 and LN3C models to a variety of air quality data sets and generally favored the use of the three-parameter form.

More complicated models than the three-parameter types also have been applied. Mage[57] has extended the lognormal application to the four-parameter lognormal probability (LN4) model, which also is known as the Johnson S_B distribution.[58] He showed that introduction of the fourth parameter allows inclusion of a finite concentration which is approached asymptotically as the CDF approaches unity, thereby giving the light tail that is necessary to fit higher air pollutant concentrations.

In a recent study, Ledolter and his collogues[59] applied a power transform to CO air quality data. This transformation, also known as the Box-Cox transform, makes it possible to test whether various power functions, such as the square root or the cube root, are more suitable than the normal (untransformed) distribution. In the limit, as the nth root is taken, and $|n|$ becomes very large (i.e., $\frac{1}{|n|}$ approaches 0), the distribution approaches the lognormal probability model. For the data they examined, they found that maximum likelihood estimates gave relatively large values of $|n|$, and they concluded that "...the lognormal distribution (or the class of power transforms in general) can provide a good overall description of the frequency distribution." These authors also addressed the problem of fitting the upper tail of the distribution in which the CDF exceeds 0.95. They applied a Pareto distribution to these high frequencies and tested their results using a chi-square test, concluding

that "... the Pareto distribution provides an excellent fit to the upper 5 percent of CO air quality data."[59]

The extreme value distribution also has been proposed for application to air quality data by Singpurwalla[60] and Roberts.[61] Its use seems especially well suited to situations in which the maxima or second highest maxima of air quality data sets are compared with environmental standards. Roberts[62] has provided a number of examples of its application to air quality data.

Although the gamma distribution, when applied by Lynn[42] and Bencala and Seinfeld,[51] was reported to not perform as well as the LN2 and LN3 models, Trajonis[63] analyzed nitrogen dioxide (NO_2) data and observed the "light tail" earlier reported by Curran and Frank.[48] This result suggested that the gamma distribution was better suited to representing these NO_2 data than the heavy-tailed LN2 model.

At the present time, there appears to be a growing consensus that the data analyst should not automatically choose any one probability model (such as the LN2 model) to the exclusion of others, but, rather, that he should carefully evaluate the problem he is addressing and the data at hand before selecting a probability model.[1,64-65] We realize that it may be tempting to use the LN2 model, because many articles are available to facilitate its application. For example, Hunt[66] has developed formulas and tables to assist the analyst in determining the precision associated with random samples from the LN2 model. Although similar aids have not appeared in the air pollution literature for the other models mentioned above, their absence should not dis-

courage the analyst from using a particular model if it fits the data better than the LN2 model. We believe that the choice of the model should be based on the purpose for which the model is intended, the findings of other investigators regarding similar applications, the goodness-of-fit of the model to the data, and the professional judgment of the data analyst. The report by Trajonis[63] contains a good example of a procedure in which different models are tested, and the investigator makes a selection that takes into account the fit of the model to the data.

Although, as we have seen, extensive applications of the LN2 model to air quality data have been made, the use of goodness-of-fit measures is a relatively recent phenomenon in the air pollution literature, and no systematic study of the LN2 model's goodness-of-fit to a large number of data sets has been undertaken. In addition, no investigators have examined the impact of different methods of selecting the model's parameters on the resulting fit. To fill this gap, the present study has been undertaken. It includes a national cross section of air quality data and it considers four different methods of calculating parameters for the LN2 model. We hope that future investigators will consider extending the general approach presented in this study to other pollutants and to additional probability models. The findings of this work are intended to assist data analysts in selecting and applying probability models to meet their particular needs.

III. STRUCTURE OF THE NORMAL AND LOGNORMAL PROBABILITY MODELS

As noted by Akland,[67] probability models have a variety of environmental applications:

- Evaluating environmental standards
- Calculating emission "rollback" levels
- Estimating maximum concentrations
- Approximating threshold concentrations
- Estimating missing observations

Other applications include calculating confidence intervals about certain point estimates, analyzing environmental trends, and estimating environmental health risks. A variety of probability models have been developed for representing the distributions of continuous random variables.

Let X denote a continuous random variable (such as atmospheric concentration levels) and x denote a particular observation (such as 12 ppm measured at a station during 12:00 midnight to 1:00 a.m.). Then a particular probability model designed to represent the distribution of the random variable usually is written in terms of its Probability Density Function (PDF), which is denoted as $f_X(x)$. If $P(x \leq X \leq x + \Delta x)$ denotes the probability that the concentration lies within the range x and $x + \Delta x$, then the relationship between the PDF and probability is as follows:

$$f_X(x) = \lim_{\Delta x \to 0} \left\{ \frac{P(x \leq X \leq x + \Delta x)}{\Delta x} \right\} = \frac{dP}{dx} \qquad (1)$$

Thus, for a continuous random variable (such as pollutant concentration), the PDF is the first derivative of the probability P with respect to the random variable X. Consequently, the probability associated with any point $X = x$ is undefined and cannot be calculated directly from the PDF. Rather, $f_X(x)$ can be interpreted loosely as the probability that X lies within the infinitesimal interval $(x, x + dx)$. The fact that $f_X(x)$ is a limiting function causes no difficulty in environmental applications, because we usually are interested in the probability that X lies above or below a particular value, such as an environmental standard $X = x_s$.

The probability that a given random variable is less than or equal to a specific value $X = x_s$ is obtained by integrating the PDF from negative infinity to $X = x_s$:

$$P(X \leq x_s) = \int_{-\infty}^{x_s} f_X(x)dx = F_X(x_s) \qquad (2)$$

Because of the importance of the integral of the PDF, it is formally defined as the Cumulative Distribution Function (CDF) and is denoted as $F_X(x)$. In the PDF and the CDF, the subscript X is used to show explicitly that a random variable is under consideration. Some authors use the notation $f_X(y)$ and $F_X(y)$ to emphasize that y is a specific member from a sample set denoted by X. If it is understood that a random variable is being considered, the subscript may be deleted, and $f_X(x)$, $F_X(x)$ may be denoted simply by $f(x)$, $F(x)$. For purposes of emphasis, the subscript X will be used throughout this report.

Because $F_X(b)$ gives the probability that $X \leq b$, the

probability that X lies within a specified range between a and b is obtained by subtracting respective CDF's:

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx \qquad (3)$$

Although a strict inequality is given at the left side of X in the statement $P(a < X \leq b)$, with a "less than or equal sign" at the right side of X, the equal signs have little significance. When a approaches b, the area under the curve of a PDF (Figure 1) approaches zero. In the limiting case, if $a = b$, $P(a \leq X \leq b) = 0$. Therefore, the following expressions are equivalent:

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) \qquad (4)$$

In general, the PDF of any probability model has the following properties: (1) it is nonnegative, and (2) its integral from $-\infty$ to $+\infty$ is 1:

$$f_X(x) \geq 0 \quad \text{for all } x \qquad (5)$$

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \qquad (5a)$$

Because the CDF is the integral of the PDF, $F_X(x)$ is a monotonically increasing function that begins at 0 and approaches 1 in the limit as x approaches ∞.

MOMENTS

The "moments" of a probability model are single values

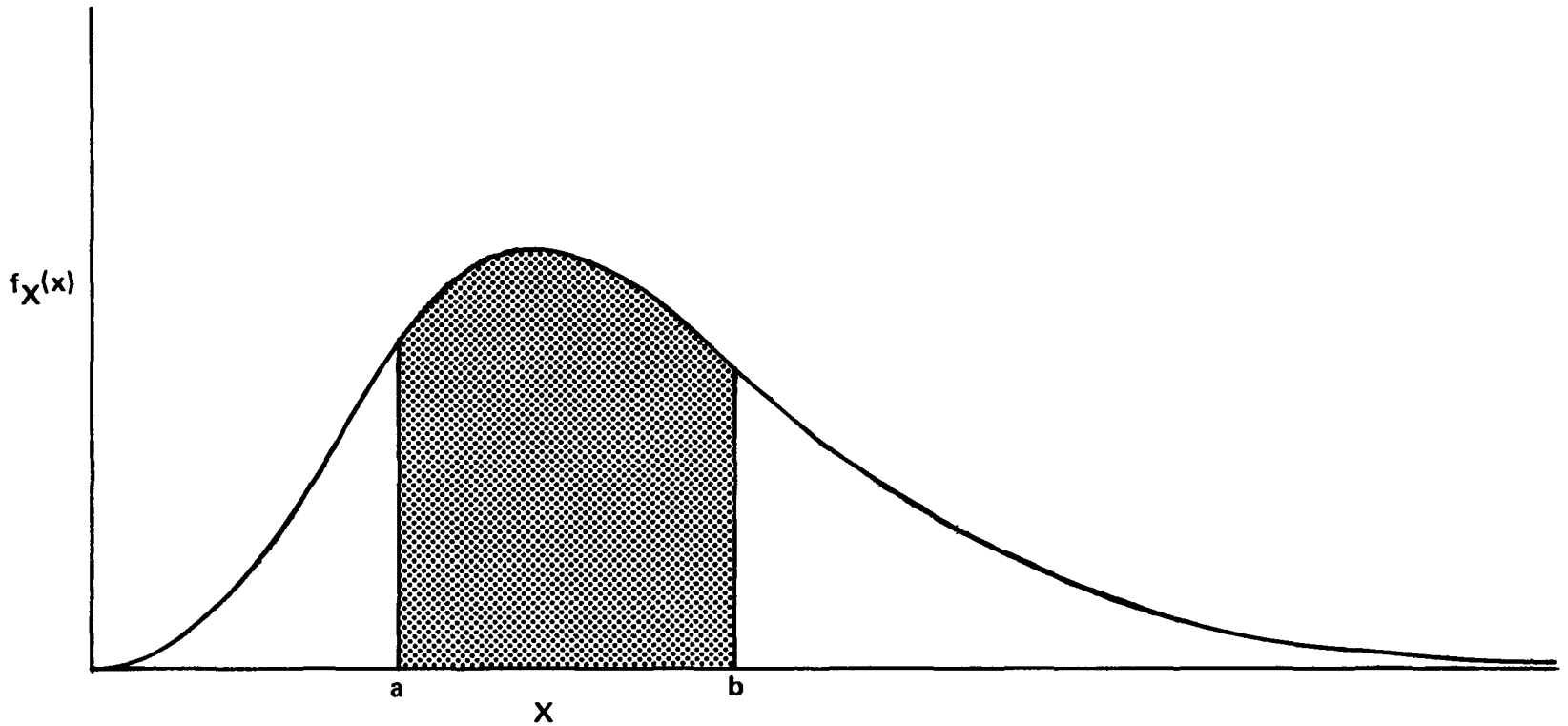

Figure 1. Example of a probability density function (PDF) for a continuous random variable.

which characterize its properties. The first moment about the origin is known as the "expected value," or arithmetic mean, and is an important measure of "central tendency" of the distribution. For random variable X, the expected value $E\{X\}$ is calculated from the PDF as follows:

$$E\{X\} = \int_{-\infty}^{\infty} x f_X(x)\,dx \tag{6}$$

Because of its importance, the notation $\mu = E\{X\}$ often is used to refer to the arithmetic mean of the probability model.

The expectation operator shown above can be applied to other functions of X, say $g(X)$:

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx \tag{7}$$

If $g(X) = X^2$, then Equation 7 gives the second moment of random variable X about the origin. Similarly, if $g(X) = X^3$, the third moment about the origin is obtained. In general, the <u>jth</u> moment about the origin, in which j is a po. i ive integer, is obtained as follows:

$$E\{X^j\} = \int_{-\infty}^{\infty} x^j f_X(x)\,dx \tag{8}$$

If we choose the function $g(X) = (X - E\{X\})^j = (X - \mu)^j$, then

the expectation of $(X - \mu)^j$ is known as the <u>jth</u> moment about the mean of the random variable X:

$$E\{(X - \mu)^j\} = \int_{-\infty}^{\infty} (x - \mu)^j f_X(x)dx \tag{9}$$

Because of its importance in probability theory, the second moment about the mean is given a special name--the "variance" or simply var(X):

$$\mathrm{var}(X) = E\{(X - \mu)^2\} = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x)dx \tag{10}$$

If the square in Equation 10 is expanded,

$$\mathrm{var}(X) = E\{(X - \mu)^2\} = E\{X^2\} - 2\mu E\{X\} + \mu^2 \tag{11}$$

Noting that $\mu = E\{X\}$, we obtain a useful result:

$$\mathrm{var}(X) = E\{X^2\} - \mu^2 \tag{11a}$$

In this report, we shall use the notation μ_j to denote the <u>jth</u> moment about the origin, while m_j denotes the <u>jth</u> moment about the mean. That is,

$$\mu_j = E\{X^j\} \tag{12}$$

$$m_j = E\{(X - \mu)^j\} \tag{13}$$

Thus, μ_1, or simply "μ", denotes the arithmetic mean, and m_2 denotes the variance, var(X). The term σ^2 also commonly is used to denote the variance of a probability distribution; that is, $m_2 = \sigma^2$.

Other measures that are useful for describing a probability distribution are the coefficient of skewness $\sqrt{\beta_1}$ and the coefficient of kurtosis β_2:

$$\sqrt{\beta_1} = \frac{m_3}{m_2^{1.5}} \tag{14}$$

$$\beta_2 = \frac{m_4}{m_2^{2}} \tag{15}$$

These measures may be viewed as indices of the shape of a distribution. The skewness is a measure of the symmetry of the distribution. For symmetrical distributions, $\sqrt{\beta_1} = 0$. For distributions that have tails extending to the right, $\sqrt{\beta_1}$ is positive; for distributions that have tails extending to the left, $\sqrt{\beta_1}$ is negative. The kurtosis β_2 is a measure of the "peakedness" of a distribution--the relationship of its height to its width. For the uniform distribution, $\beta_2 = 1.8$; for the normal distribution, $\beta_2 = 3$; and for the exponential distribution, $\beta_2 = 9$.

As we shall see, the values of β_1 and β_2 for a given distribution can be plotted in the "moment plane" to draw conclusions about the shapes of different probability models in relation to each other. Application of the moment

plane to the data presented in this report is covered in a companion report by Mage and Ott.[2]

If $g(X) = e^{tX}$ in Equation 7, then the expectation operator, if it exists for a given probability model, is called the "moment generating function," $M\{X\}$:

$$M\{X\} = E\{e^{tX}\} = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \tag{16}$$

The moment generating function is a useful tool for obtaining the moments of a probability model. For some probability models, the moments are more easily obtained using the expectation operator, Equation 6, than the moment generating function, Equation 16. For others, the reverse is true. Once an analytical expression for the moment generating function is obtained, the j<u>th</u> moment of the probability model can be found by taking the j<u>th</u> derivative of $M\{X\}$ with respect to t and setting t = 0:

$$E\{X^j\} = \frac{d^j[M\{0\}]}{dt^j} \tag{17}$$

<u>Normal Probability Model</u>

The normal, or Gaussian, probability model is a distribution of particular interest in many fields. If Y is a random variable, the PDF of the normal probability model is symmetrical about the

origin (Figure 2), and its PDF is written as follows:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \qquad \text{for} \quad -\infty < y < +\infty \tag{18}$$

A random variable which has a distribution that can be represented by the normal probability model is said to be normally distributed. Usually, it is of interest to consider normally distributed random variables which are translated such that they are symmetrical about values other than the origin. Thus, we employ the transformation $Y = (X - \mu)/\sigma$, which contains location parameter μ and dispersion parameter σ. With this transformation, $dY/dX = 1/\sigma$, and the PDF of the transformed variate X is written as follows:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2} \tag{19}$$

The moments for this distribution are calculated using the techniques described above. The arithmetic mean is calculated as follows using the expectation operator:

$$E\{X\} = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2} dx \tag{20}$$

Since $y = (x - \mu)/\sigma$, we solve for x and substitute $x = \sigma y + \mu$ into Equation 14. Noting that $dx = \sigma dy$,

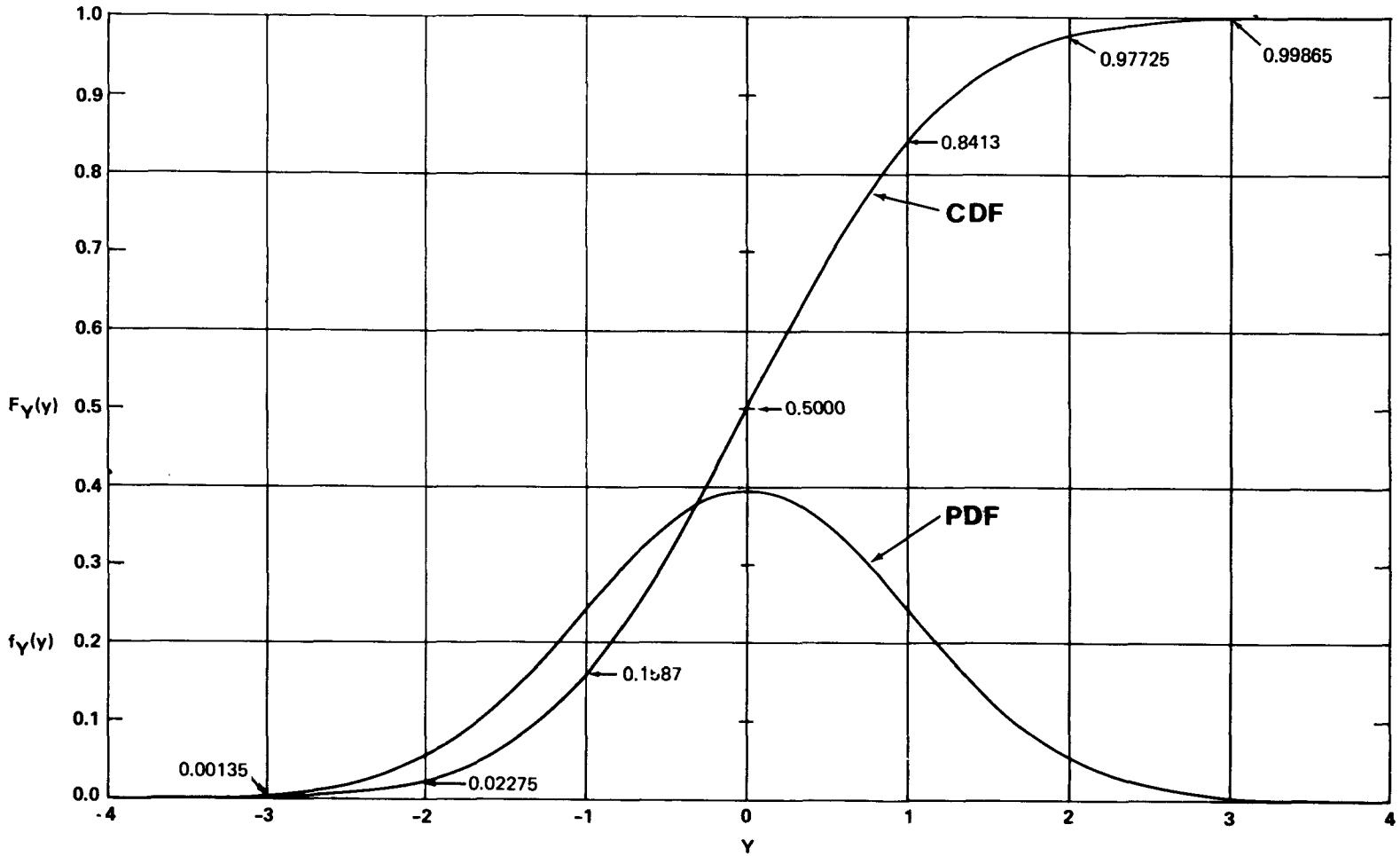

Figure 2. Probability density function (PDF) and cumulative distribution function (CDF) of the normal probability model with μ = 0.0 and σ = 1.0.

$$E\{X\} = \int_{-\infty}^{\infty} (\sigma y + \mu) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} y^2} \sigma dy \tag{20}$$

$$= \sigma \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y^2} dy + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y^2} dy \tag{20a}$$

$$E\{X\} = \mu \tag{20b}$$

The simple result given in Equation 20b is obtained from Equation 20a by the following reasoning. The left-hand integral in Equation 20a is the integral of y multiplied times the PDF of random variable Y, in which Y has mean 0 and standard deviation of unity. The PDF is symmetrical about 0, and the function within the integral sign becomes negative for $-\infty < y < 0$ and positive for $0 < y < +\infty$ when the integration is performed. Because the values obtained by integrating over these two ranges are equal and of opposite sign, the integral over the entire range $-\infty < y < +\infty$ is zero. The second term in Equation 20a is μ times the integral of the PDF of Y over $-\infty < y < +\infty$, which, by definition, is 1.

The variance, or second moment about the mean, is calculated by first calculating $E\{X^2\}$ using Equation 8:

$$E\{X^2\} = \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2} dx \tag{21}$$

Because $x = \sigma y + \mu$, the square of x is $x^2 = \sigma^2 y^2 + 2\mu\sigma y + \mu^2$. Substituting this equation into Equation 21, we obtain an expression for $E\{X^2\}$ that contains three integrals:

$$E\{X^2\} = \int_{-\infty}^{\infty} \sigma^2 y^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy + 2\mu\sigma \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy + \mu^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \quad (22)$$

The first integral can be evaluated by integration by parts. giving σ^2. The second and third terms contain the same integrals as in Equation 20a, which are 0 and 1, respectively. Therefore, $E\{X^2\} = \sigma^2 + \mu^2$. Using Equation 11a to compute the variance, or second moment about the origin, $\mathrm{var}(X) = E\{(X - \mu)^2\} = E\{X^2\} - \mu^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$.

An alternative way to calculate the moments is by use of the moment generating function, $M\{X\}$. The moment generating function for the normal distribution is obtained by integrating the PDF multiplied by e^{tx}, as indicated by Equation 16:

$$M\{X\} = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (23)$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 + tx} dx \quad (23a)$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x^2 - 2(\mu + t\sigma^2)x + \mu^2}{\sigma^2}\right]} dx \quad (23b)$$

To form a perfect square for the numerator of the expression inside the brackets in Equation 23b, we must modify the numerator so that terms which do not involve x or x^2 will be equal to one-half the square of the coefficient of x. That is, the last term must be equal to $(\mu + t\sigma^2)^2 = \mu^2 + 2\mu t\sigma^2 + t^2\sigma^4$. Because μ^2 already is present as the last term in the numerator, we can form the perfect square by adding $2\mu t\sigma^2 + t^2\sigma^4$ to the numerator and then subtracting an equivalent term from the quantity within parentheses:

$$M\{X\} = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x^2 - 2(\mu + t\sigma^2)x + (\mu^2 + 2\mu t\sigma^2 + t^2\sigma^4)}{\sigma^2} - \frac{2\mu t\sigma^2 + t^2\sigma^4}{\sigma^2}\right]} dx \qquad (24)$$

$$= e^{-\frac{1}{2}\left[\frac{2\mu t\sigma^2 + t^2\sigma^4}{\sigma^2}\right]} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x - (\mu + t\sigma^2)}{\sigma}\right]^2} dx \qquad (24a)$$

The integral part of Equation 24a is equivalent to the CDF of a normally distributed random variable with mean $\mu + t\sigma^2$ and variance σ^2. Because this integral is evaluated over the range $-\infty$ to $+\infty$, its value is unity, and the moment generating function for the normal probability model becomes:

$$M\{X\} = e^{-\frac{1}{2}\left[\frac{2\mu t\sigma^2 + t^2\sigma^4}{\sigma^2}\right]} = e^{\mu t + \frac{\sigma^2 t^2}{2}} \qquad (25)$$

To determine the moments of the distribution, we differentiate Equation 25 with respect to t and set t = 0. Thus the first

moment, or the arithmetic mean, is calculated from the moment generating function as follows:

$$\frac{dM}{dt} = (\mu + \sigma^2 t)e^{\mu t + \frac{\sigma^2 t^2}{2}} \tag{26}$$

$$E\{X\} = \frac{dM(0)}{dt} = \mu \tag{27}$$

The second moment about the origin is calculated in a similar manner:

$$\frac{d^2M}{dt^2} = \sigma^2 e^{\mu t + \frac{\sigma^2 t^2}{2}} + (\mu + \sigma^2 t)^2 \, e^{\mu t + \frac{\sigma^2 t^2}{2}} \tag{30}$$

$$E\{X^2\} = \frac{d^2M(0)}{dt^2} = \sigma^2 + \mu^2 \tag{31}$$

Using the same approach as above, the variance is obtained as follows:

$$\mathrm{var}(X) = E\{X^2\} - [E\{X\}]^2 = \sigma^2 \tag{32}$$

For the normal probability model, the moments about the origin are obtained from $M\{X\}$ in the same fashion, giving the following results:

$$E\{X^3\} = \frac{d^3M(0)}{dt^3} = 3\sigma^2\mu + \mu^3 \tag{33}$$

$$E\{X^4\} = \frac{d^4M(0)}{dt^4} = 6\sigma^2\mu^2 + \mu^4 + 3\sigma^4 \tag{34}$$

The notation used in Equations 31, 33, and 34 indicates that we first differentiate $M\{X\}$ with respect to t, and then we substitute $t = 0$ into the result.

General equations for the third and fourth moments about the origin can be derived in a manner analogous to the derivation of the variance given in Equation 11a:

$$E\{(X - \mu)^3\} = E\{X^3\} - 3\mu E\{X^2\} + 2\mu^3 \tag{35}$$

$$E\{(X - \mu)^4\} = E\{X^4\} - 4\mu E\{X^3\} + 6\mu^2 E\{X^2\} - 3\mu^4 \tag{36}$$

Substitution of Equation 31 for $E\{X^2\}$ and Equation 33 for $E\{X^3\}$ into Equation 35 gives the following result:

$$\begin{aligned} E\{(X - \mu)^3\} &= 3\sigma^2\mu + \mu^3 - 3\mu(\sigma^2 + \mu^2) + 2\mu^3 \\ &= 3\sigma^2\mu + \mu^3 - 3\mu\sigma^2 - 3\mu^3 + 2\mu^3 = 0 \end{aligned} \tag{37}$$

Similarly, substitution of Equations 31, 33, and 34 into Equation 36 gives the following result:

$$\begin{aligned} E\{(X - \mu)^4\} &= 6\sigma^2\mu^2 + \mu^4 + 3\sigma^4 - 4\mu(3\sigma^2\mu + \mu^3) + 6\mu^2(\sigma^2 + \mu^2) - 3\mu^4 \\ &= 6\sigma^2\mu^2 + \mu^4 + 3\sigma^4 - 12\mu^2\sigma^2 - 4\mu^4 + 6\mu^2\sigma^2 + 6\mu^4 - 3\mu^4 = 3\sigma^2 \end{aligned} \tag{38}$$

Finally, substitution of the results obtained in Equations 31,

37, and 38 into Equations 14 and 15 gives simple expressions for $\sqrt{\beta_1}$ and β_2 for the normal distribution:

$$\sqrt{\beta_1} = \frac{m_3}{m_2^{1.5}} = \frac{0}{(\sigma^2)^{1.5}} = 0 \tag{39}$$

$$\beta_2 = \frac{m_4}{m_2^{\,2}} = \frac{3\sigma^4}{(\sigma^2)^2} = 3 \tag{40}$$

Lognormal Probability Model

If the logarithmic transformation $Y = (\ln X - \mu)/\sigma$ is employed, then we say that the logarithm of X is normally distributed, or X is "lognormally" distributed. Noting that $dY = dX/(\sigma X)$, the PDF of the transformed variate is written as follows, in which ln denotes the logarithm to the base e:

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} \tag{41}$$

Note that this PDF differs from that for the normal distribution in that x appears in the denominator and ln x is substituted for x. The lognormal PDF is skewed to the right, and X has the range $0 < x < \infty$ (see Figure 4, page 43).

If the quantity in parentheses in Equation 41 is set equal to 0, we obtain an expression for the median of the log-

normal distribution, $\hat{m}$:

$$\frac{\ln \hat{m} - \mu}{\sigma} = 0$$

$$\therefore \quad \ln \hat{m} = \mu, \qquad \text{or} \quad \hat{m} = e^{\mu} \tag{42}$$

In another common form of the lognormal probability model, the PDF does not involve μ; rather, x is expressed as a ratio to the median $\hat{m}$:

$$f_X(x) = \frac{1}{\sigma x \sqrt{2\pi}} \; e^{-\frac{1}{2}\left(\frac{\ln \frac{x}{\hat{m}}}{\sigma}\right)^2} \tag{43}$$

The two forms are equivalent, since the expressions within the parentheses are equivalent.

To calculate the expected value, or first moment about the origin, of the lognormal distribution, we apply the expectation operator, Equation 6:

$$E\{X\} = \int_0^{\infty} x f_X(x) \; dx = \int_0^{\infty} \frac{x}{x\sigma \sqrt{2\pi}} \; e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} dx \tag{44}$$

Note that the integration is performed between 0 and ∞, because the lognormal PDF is defined as 0 from $-\infty$ to 0. To evaluate this

integral, we shall employ the transform $Z = \ln X$, or $X = \exp(Z)$ Then, $dX = e^{Z}dZ$, and Equation 44 can be written as follows:

$$E\{X\} = E\{e^{Z}\} = \int_{0}^{\infty} \frac{e^{z}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^{2}} dz \tag{42}$$

We now observe that, if the substitutions $x = z$ and $t = 1$ are made in the expression for the moment generating function for the normal distribution given by Equation 23, the result will be identical to Equation 42. Thus, we can use the same approach that was used to evaluate Equation 23 to obtain a result for Equation 42. By setting $t = 1$ in Equation 25, we obtain the following expression for the expected value, or arithmetic mean, of the lognormal probability model:

$$E\{X\} = e^{\mu + \frac{\sigma^{2}}{2}} \tag{45}$$

To obtain the higher moments, we apply Equation 8, the expectation operator for the jth moment about the origin:

$$E\{X^{j}\} = \int_{0}^{\infty} \frac{x^{j-1}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}} dx \tag{46}$$

The exponent j-1 appears in the numerator instead of j, because the lognormal PDF contains x in the denominator. Using the same

approach as in Equation 42, we let $Z = \ln X$; then, $X = \exp(Z)$, $dX = e^{Z}dZ$, and $x^{j-1} = e^{(j-1)z}$. Thus,

$$E\{X^j\} = \int_0^\infty \frac{e^{(j-1)z}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} e^z \, dz$$

$$= \int_0^\infty \frac{e^{jz}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz \tag{47}$$

In a similar fashion, we conclude that Equation 47 is identical in form to Equation 23, and we use the result obtained for that situation; mamely, Equation 25 with $t = j$:

$$E\{X^j\} = e^{j\mu + \frac{j^2\sigma^2}{2}} \tag{48}$$

This result implies that the expectation operator for a lognormally distributed random variable X happens to be identical to the moment generating function $M\{Z\}$ for a normally distributed random variable $Z = \ln X$.

The general result given by Equation 48 allows us to calculate all the moments about the origin for the lognormal probability model with relative ease:

$$\mu_1 = e^{\mu + \frac{\sigma^2}{2}} \tag{49}$$

$$\mu_2 = e^{2\mu + 2\sigma^2} \tag{50}$$

$$\mu_3 = e^{3\mu + \frac{9}{2}\sigma^2} \tag{51}$$

$$\mu_4 = e^{4\mu + 8\sigma^2} \tag{52}$$

The variance, or second moment about the mean, is calculated from the above relationships using the moment expansion given in Equation 11a:

$$\mathrm{var}(X) = m_2 = \mu_2 - \mu_1^2 = e^{2\mu + 2\sigma^2} - \left(e^{\mu + \frac{\sigma^2}{2}}\right)^2$$

$$= e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right) \tag{53}$$

Similarly, the third moment about the mean can be obtained using the moment expansion given by Equation 35:

$$m_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3$$

$$= e^{3\mu + \frac{9}{2}\sigma^2} - 3e^{\mu + \frac{\sigma^2}{2}}\left(e^{2\mu + 2\sigma^2}\right) \quad 2\left(e^{\mu + \frac{\sigma^2}{2}}\right)^2$$

$$= e^{3\mu + \frac{9}{2}\sigma^2} - 3e^{3\mu + \frac{5}{2}\sigma^2} + 2e^{3\mu + \frac{3}{2}\sigma^2}$$

$$= e^{3\mu + \frac{3}{2}\sigma^2}\left(e^{3\sigma^2} - 3e^{\sigma^2} + 2\right)$$

$$= e^{3\mu + \frac{3}{2}\sigma^2}\left(e^{\sigma^2} + 2\right)\left(e^{\sigma^2} - 1\right)^2 \tag{54}$$

Using Equations 53 and 54, we can derive an expression for the coefficient of skewness $\sqrt{\beta_1}$:

$$\sqrt{\beta_1} = \frac{m_3}{m_2^{1.5}} = \frac{e^{3\mu + \frac{3}{2}\sigma^2}\left(e^{\sigma^2} + 2\right)\left(e^{\sigma^2} - 1\right)^2}{\left(e^{3\mu + \frac{3}{2}\sigma^2}\right)\left(e^{\sigma^2} - 1\right)^{3/2}}$$

Squaring both sides,

$$\beta_1 = \frac{\left(e^{\sigma^2} + 2\right)^2\left(e^{\sigma^2} - 1\right)^4}{\left(e^{\sigma^2} - 1\right)^3} = \left(e^{\sigma^2} + 2\right)^2\left(e^{\sigma^2} - 1\right)$$

$$\therefore \sqrt{\beta_1} = \left(e^{\sigma^2} + 2\right)\left(e^{\sigma^2} - 1\right)^{1/2} \tag{55}$$

Finally, the fourth moment about the mean can be obtained by using the moment expansion given by Equation 36:

$$m_4 = \mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4$$

$$m_4 = e^{4\mu + 8\sigma^2} - 4e^{\mu + \frac{\sigma^2}{2}\left(e^{3\mu + \frac{9}{2}\sigma^2}\right)} + 6\left(e^{\mu + \frac{\sigma^2}{2}}\right)^2 e^{2\mu + 2\sigma^2} - 3\left(e^{\mu + \frac{\sigma^2}{2}}\right)^4$$

$$= e^{4\mu + 8\sigma^2} - 4e^{4\mu + 5\sigma^2} + 6e^{4\mu + 3\sigma^2} - 3e^{4\mu + 2\sigma^2}$$

$$= e^{4\mu + 2\sigma^2}\left(e^{\sigma^2} - 1\right)^2\left(e^{4\sigma} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3\right) \tag{56}$$

Using Equations 53 and 56, we obtain the following expression for the coefficient of kurtosis β_2:

$$\beta_2 = \frac{m_4}{m_2^2} = \frac{e^{4\mu} + 2\sigma^2\left(e^{\sigma^2} - 1\right)\left(e^{4\sigma} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3\right)}{\left(e^{2\mu} + \sigma^2\right)^2\left(e^{\sigma^2} - 1\right)^2}$$

$$= e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3 \tag{55}$$

If we add 3 to Equation 55 and then subtract 3, another factor can be obtained, and Equation 55 can be written as follows:

$$\beta_2 = 3 + e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 6$$

$$= 3 + \left(e^{\sigma^2} - 1\right)\left(e^{3\sigma^2} + 3e^{2\sigma^2} + 6e^{\sigma^2} + 6\right) \tag{56}$$

Some authors express $\sqrt{\beta_1}$ and β_2 in a simplified form using the substitution $\omega = e^{\sigma^2}$:

$$\sqrt{\beta_1} = (\omega + 2)(\omega - 1)^{1/2} \tag{57}$$

$$\beta_2 = 3 + (\omega - 1)(\omega^3 + 3\omega^2 + 6\omega + 6) \tag{58}$$

When plotting the lognormal distribution, interest also may focus on the mode x_m, which is defined as the value of X for which the PDF has its maximum. To obtain an expression for x_m, we dif-

ferentiate Equation 41, the PDF of the lognormal distribution, and set the result equal to 0:

$$f_X'(x) = -\frac{1}{x_m^2\,\sigma}\left(\frac{\ln x_m - \mu}{\sigma}\right)\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{\ln x_m - \mu}{\sigma}\right)^2}$$

$$-\frac{1}{x_m^2}\,\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{\ln x_m - \mu}{\sigma}\right)^2} = 0 \qquad (59)$$

Then, solving Equation 59 for x_m, we obtain a simple result for the mode of the lognormal distribution:

$$-\frac{1}{\sigma^2}\left(\ln x_m - \mu\right) - 1 = 0$$

$$x_m = e^{\mu-\sigma^2} \qquad (60)$$

Because the expressions for the lognormal distribution usually involve μ and σ, the mean and standard deviation of a normal distribution that describes the logarithm of the variable X of interest, we shall introduce the following notation to avoid confusion:

α = arithmetic mean of X

ξ = arithmetic standard deviation of X

Another statistic of interest is the coefficient of variation ν, which is the ratio of the arithmetic standard deviation to the arithmetic mean:

$$\nu = \frac{\xi}{\alpha} \tag{61}$$

Expressing α in terms of Equation 45 and ξ in terms of Equation 53, the arithmetic mean and variance ξ^2 of a lognormally distributed random variable can be written as follows:

$$\alpha = e^{\mu + \frac{\sigma^2}{2}} \tag{62}$$

$$\xi^2 = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) \tag{63}$$

Thus, the coefficient of variation can be written as the ratio of ξ to α, as obtained from Equations 62 and 63:

$$\nu = \left(e^{\sigma^2} - 1\right)^{1/2} \tag{64}$$

Often the arithmetic mean and standard deviation of the variable of interest are known, and it is necessary to calculate μ and σ. To obtain expressions for μ and σ as a function of α and ξ, we first solve Equation 62 for μ:

$$\mu = \ln \alpha - \frac{1}{2}\sigma^2 \tag{65}$$

Substituting Equation 65 into Equation 63, we can derive an

expression for σ^2 as a function of α and ξ:

$$\xi^2 = e^{2\left[\ln \alpha - \frac{\sigma^2}{2}\right] + \sigma^2}\left(e^{\sigma^2} - 1\right) = e^{2 \ln \alpha}\left(e^{\sigma^2} - 1\right)$$

$$\ln \xi^2 = 2 \ln \alpha + \ln\left(e^{\sigma^2} - 1\right)$$

$$\ln \frac{\xi^2}{\alpha^2} = \ln\left(e^{\sigma^2} - 1\right)$$

$$\frac{\xi^2}{\alpha^2} + 1 = e^{\sigma^2}$$

$$\sigma^2 = \ln\left(\frac{\xi^2}{\alpha^2} + 1\right) \tag{66}$$

Equation 66 also can be written in terms of the coefficient of variation:

$$\sigma = \sqrt{\ln\left(v^2 + 1\right)} \tag{67}$$

If ξ and α are known, we first use Equation 66 to compute σ. Then, we substitute σ and α into Equation 65 to compute μ. Once μ and σ are known, most of the other equations given in this chapter can be applied.

The CDF for the lognormal probability model plots as a straight line on logarithmic probability paper. To determine the vertical position of this line, it is useful to employ the equation for the median of the lognormal distribution,

thus obtaining a value for X which corresponds to the 50 percentile of the CDF. That is, we use Equation 42 to find the median, $\hat{m} = x_{50}$, which is the value for which $F_X(x_{50}) = 0.50$:

$$x_{50} = \hat{m} = e^{\mu} \tag{68}$$

A second point on the straight line can be obtained by finding the value of X which corresponds to plus one standard deviation for the normally distributed random variable Y. This gives the value of X which corresponds to the 84.13 percentile of the CDF (see Figure 2); that is, $F_X(x_{84}) = 0.8413$:

$$y = \frac{\ln x_{84} - \mu}{\sigma} = +1 \tag{69}$$

Solving Equation 69 for x_{84}, we obtain the following simple relationship for the second point on the line:

$$x_{84} = e^{\mu+\sigma} = e^{\mu}e^{\sigma} = e^{\sigma}\hat{m} = e^{\sigma}x_{50} \tag{70}$$

Some authors define the dimensionless factor e^{σ} in Equation 70 as the "standard geometric deviation (SGD)," denoted as σ_g:

$$\sigma_g = e^{\sigma} = \frac{x_{84}}{x_{50}} \tag{71}$$

If Equation 69 is set equal to y = -1, we obtain a value of X

which corresponds to minus one standard deviation for the normally distributed random variable Y. This gives the value of X which corresponds to the 15.87 percentile of the CDF (see Figure 2); that is, $F_X(x_{16}) = 0.1587$. Solving Equation 69 for $y = -1$, we obtain the following result:

$$x_{16} = e^{\mu-\sigma} = e^{\mu}e^{-\sigma} = e^{-\sigma}\hat{m} = \frac{x_{50}}{e^{\sigma}} = \frac{x_{50}}{\sigma_g} \tag{72}$$

Equation 72 actually gives a third point on the straight line, which can be used to check the correctness of the other two points. Notice that the following log-symmetric relationship exists among the three points:

$$\frac{x_{84}}{x_{50}} = \sigma_g = \frac{x_{50}}{x_{16}} \tag{73}$$

We can illustrate the method of plotting the lognormal probability model by selecting a simple example in which the arithmetic mean of X remains constant at $\alpha = 3.0$ ppm CO, and the arithmetic standard deviation takes on four successive values, $\xi = 0.5$, 1.0, 1.5, and 2.0 ppm. Applying the various formulas given in this chapter, the results listed in Table 1 are obtained.

Notice that μ and σ are different for every case, even though the arithmetic mean is constant. The median m is always less than the arithmetic mean α, a general property of the lognormal distribution, although the median becomes increasingly close to the arithmetic mean as the coefficient of variation ν becomes smaller. For values of ν less than about one-sixth

Table 1

Example of Values Calculated for the Lognormal Model

arith. mean α ppm	std. dev. ξ ppm	coef. var. ν	σ	μ	mode x_m ppm	median $\hat{m}$ ppm	SGD σ_g
3.0	0.5	0.16667	0.1655	1.0849	2.8792	2.959	1.1800
3.0	1.0	0.33333	0.3246	1.0459	2.5614	2.846	1.3835
3.0	1.5	0.50000	0.4724	0.9870	2.1466	2.683	1.6038
3.0	2.0	0.66667	0.6064	0.9148	1.7281	2.496	1.8338

(0.16667), the PDF of the lognormal probability model appears almost symmetrical when plotted and resembles the PDF of the normal distribution (see Figure 4).

Figure 3 shows the straight lines plotted on logarithmic probability paper for the four cases given in Table 1. This paper was created by the authors of the present study by cutting and pasting 2-cycle logarithmic probability papers manufactured by Keuffel and Esser Co. This was necessary in order to obtain 3-cycle logarithmic probability paper that could cover the entire range of CO concentrations with a wide range for the cumulative frequencies (that is, from 0.01 to 99.99 percent).

Sometimes it may be necessary to obtain points that are further from the median than those given by Equations 70 and 71 in order to plot the straight line with greater precision. One convenient way to do this is to solve Equation 69 for additional integer multiples of the standard deviation of the normally distributed random variable Y (that is, $y = \pm 2, \pm 3$, etc.). Then

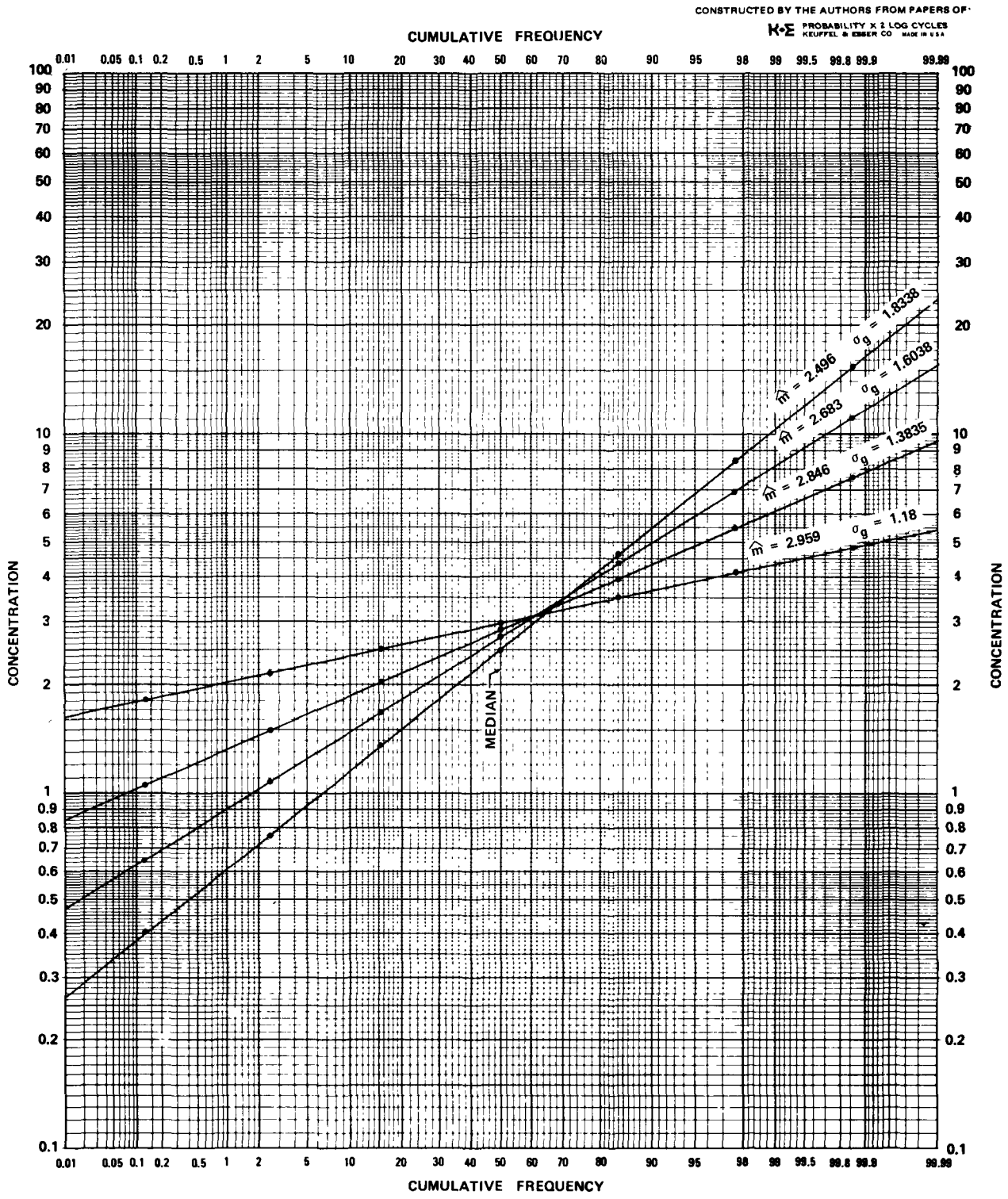

FIGURE 3. EXAMPLE OF LOGARITHMIC PROBABILITY PLOTS FOR FOUR CASES OF THE 2-PARAMETER LOGNORMAL PROBABILITY MODEL.

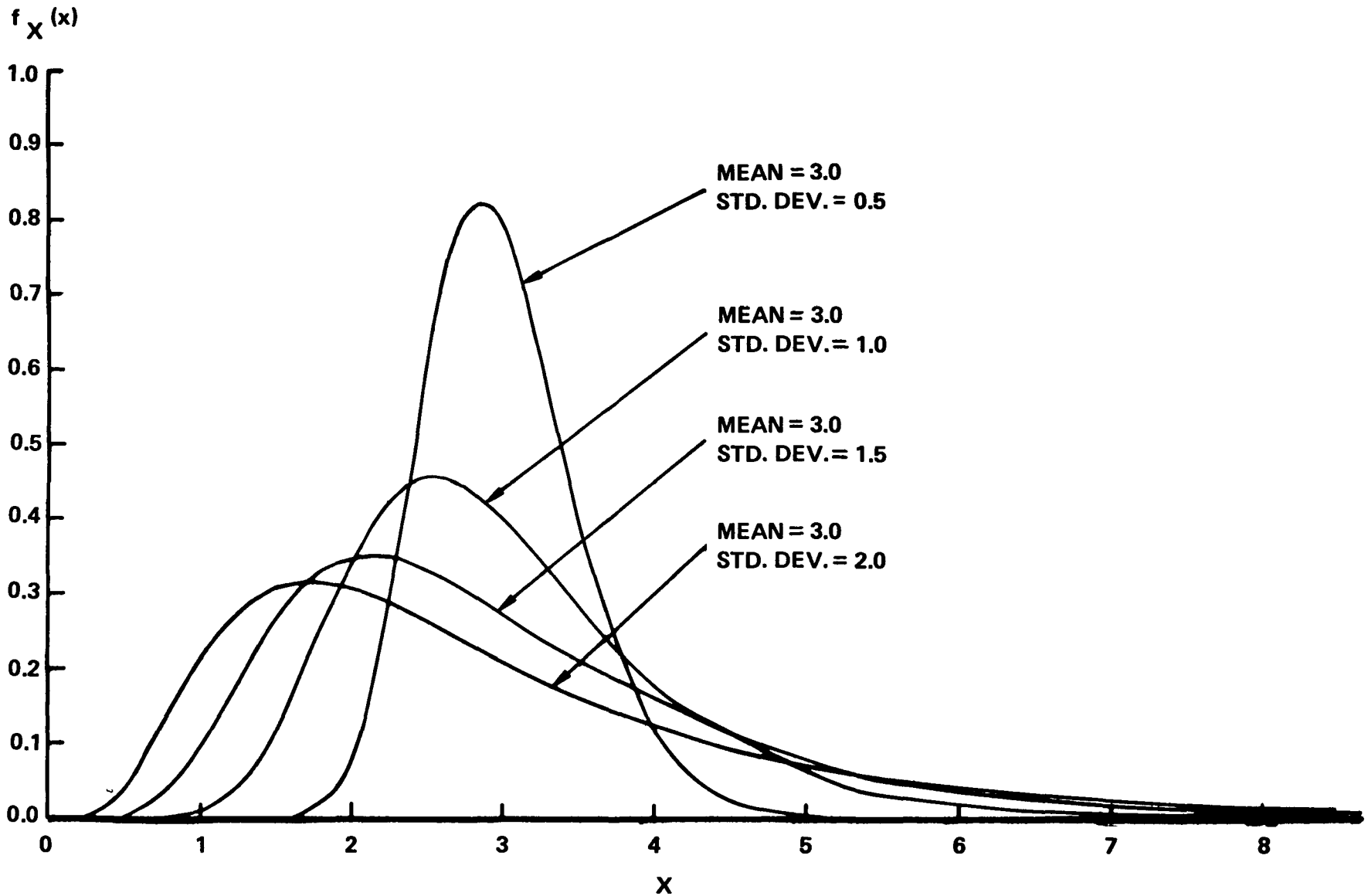

Figure 4. PDF of lognormal probability model showing differences in shape for constant arithmetic mean (3.0), but standard deviations ranging from 0.5 to 2.0.

the points are plotted on logarithmic probability paper for $x = \sigma_g^y x_{50}$ at the appropriate cumulative frequencies as shown in Table 2. In Figure 3, these points are shown as dots along each of the four straight lines for the four cases in the example.

Table 2

Cumulative Frequencies at Various Plotting Positions for the 2-Parameter Lognormal Probability Model

y:	-3	-2	-1	0	+1	+2	+3
x:	x_{50}/σ_g^3	x_{50}/σ_g^2	x_{50}/σ_g	x_{50}	$\sigma_g\ x_{50}$	$\sigma_g^2\ x_{50}$	$\sigma_g^3\ x_{50}$
cum. freq.	0.135%	2.275%	15.87%	50%	84.13%	97.725%	99.865%

Figure 4 shows the PDF's for the four cases given in Table 1. The PDF's are skewed to the right, a general trait of the lognormal probability model, and the skewness increases as the standard deviation increases. From an examination of this figure, it is not at all obvious that the arithmetic mean is the same, $\alpha = 3.0$ ppm, for all four PDF's. As the standard deviation decreases from $\xi = 2.0$ to $\xi = 0.5$, it is apparent that the PDF moves toward a symmetrical distribution resembling the normal distribution. Figures 3 and 4 will be useful in helping the reader understand the results presented in subsequent chapters.

IV. CALCULATION OF PARAMETERS

The two parameters μ and σ fully define the PDF and CDF of the lognormal probability model. If we wish to use this model to represent the frequency distribution of a particular data set, it first is necessary to compute the parameters of the model in some fashion. Although the statistical literature usually describes this process as "estimation of parameters" to emphasize that the parameters of an underlying distribution are being approximated, we shall use the more general phrase "calculation of parameters" to denote any method that is used to obtain the numerical values of the parameters, regardless of whether tests for underlying distributions are involved.

In this research investigation, four general techniques for calculating the parameters were investigated: (1) "method of moments," in which the moments of the model are set equal to the moments of the data; (2) "method of fractiles," in which the CDF of the model is set equal to the cumulative frequencies of the data in the manner suggested by Larsen[21-24], (3) an approximation of the maximum likelihood estimation (MLE) technique for grouped data; and (4) a more exact MLE technique for grouped data using computer optimization.

MOMENTS OF THE DATA

Before describing the four parameter calculation techniques, we shall briefly discuss several statistical calculations which can be made from the raw data in order to describe the data in a

parsimonious form. These calculations were incorporated in the computer program that was developed for this investigation (see Appendix A).

The moments of the observations are computed in a manner similar to that used to calculate the moments of probability models, except that the PDF in the expectation operator is replaced by 1/n, and the actual observed values are used in place of x. The jth moment of the data about the origin μ'_j is computed in a fashion analogous to Equation 8:

$$\mu'_j = \frac{1}{n}\sum_{i=1}^{n} x_i^j \tag{74}$$

where x_i = the ith observation

n = total number of observations

The arithmetic mean $\bar{x}$ of the observations is the first moment about the origin:

$$\bar{x} = \mu'_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{75}$$

Similarly, the second moment about the origin is computed by setting j=2 in Equation 74:

$$\mu'_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \tag{76}$$

The jth moment of the data about the mean, which we shall denote as m'_j, is calculated in a manner analogous to Equation 9:

$$m'_j = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^j \tag{77}$$

Of particular importance is the second moment about the mean, or the variance $\tilde{s}^2$:

$$\tilde{s}^2 = m'_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{78}$$

Because n-1 independent pieces of information are used to calculate Equation 78, $\tilde{s}^2$ is called a "biased" estimate of the variance, and an "unbiased" estimate is obtained by dividing the second moment about the mean by n-1 instead of n:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{79}$$

The square root of the variance is known as the standard deviation of the observations, either $\tilde{s}$ or s. Because $\tilde{s}^2$ and s^2 are related to each other as follows, the biased and unbiased estimates of the variance are approximately the same for large n:

$$s^2 = \tilde{s}^2 \frac{n}{n-1} \tag{80}$$

In some practical applications, computation of the variance is more convenient if the following expansion of Equation 78 is used:

$$\tilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2\bar{x}x_i + \bar{x}^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\bar{x}\,\frac{1}{n}\sum_{i=1}^{n}x_i + \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2 = \mu_2' - \bar{x}^2 \qquad (81)$$

Equation 81 is similar in form to Equation 11a, the expression for the variance of a probability model. If we solve Equation 80 for $\tilde{s}^2$ and substitute the result into Equation 81, the following computational expression for the unbiased estimate of the variance is obtained:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}x_i^2 - \frac{n}{n-1}\,\bar{x}^2 = \frac{1}{n-1}\sum_{i=1}^{n}x_i^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}x_i \qquad (82)$$

Notice that Equations 81 and 82 permit the variance to be calculated by summing the square of the observed values rather than summing the square of the difference between observed values and the mean, as in Equation 79. This expression is very useful when hand calculations are necessary or when certain computer programming efficiencies are required.

Using Equation 77, the third and fourth moments of the

data about the origin are calculated as follows:

$$m_3' = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^3 \tag{83}$$

$$m_4' = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^4 \tag{84}$$

The coefficient of skewness $\sqrt{b_1}$ and coefficient of kurtosis b_2 are calculated in a manner analogous to Equations 14 and 15:

$$\sqrt{b_1} = \frac{m_3'}{m_2'^{1.5}} \tag{85}$$

$$b_2 = \frac{m_4'}{m_2'^{2}} \tag{86}$$

METHOD OF MOMENTS

In the "method of moments," the moments of the model are set equal to the moments computed from the data. To apply this technique to the lognormal probability model, we set Equation 49 equal to Equation 75, $\mu_1 = \mu_1'$, and we set Equation 50 equal to Equation 76, $\mu_2 = \mu_2'$:

$$\mu_1 = e^{\mu + \frac{1}{2}\sigma^2} = \mu_1' = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{87}$$

$$\mu_2 = e^{2\mu + 2\sigma^2} = \mu_2' = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \tag{88}$$

Taking logarithms of Equations 87 and 88,

$$\mu + \frac{1}{2}\sigma^2 = \ln \frac{1}{n}\sum_{i=1}^{n} x_i \tag{89}$$

$$2\mu + 2\sigma^2 = \ln \frac{1}{n}\sum_{i=1}^{n} x_i^2 \tag{90}$$

Multiplying Equation 89 by 2 and subtracting the result from Equation 90, we obtain the following expression for σ^2:

$$\sigma^2 = \ln \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\ln \frac{1}{n}\sum_{i=1}^{n} x_i \tag{91}$$

Multiplying Equation 89 by 4 and subtracting Equation 90 from the result, we obtain the following expression for μ:

$$\mu = 2\ln \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{2}\ln \frac{1}{n}\sum_{i=1}^{n} x_i^2 \tag{92}$$

Rewriting Equations 91 and 92 using the simplified notation for the moments about the origin, we obtain two equations with two unknowns:

$$\sigma^2 = -2\ln \mu_1' + \ln \mu_2' \tag{93}$$

$$\mu = 2\ln \mu_1' - \frac{1}{2}\ln \mu_2' \tag{94}$$

Note that $\mu_1' = \bar{x}$ from Equation 75, and $\tilde{s}^2 = \mu_2' - \bar{x}^2$ from Equation 81; therefore, $\mu_2' = \tilde{s}^2 + \bar{x}^2$. Substituting these values for μ_1' and μ_2' into Equation 93, the following expression for σ^2 is obtained:

$$\begin{aligned}\sigma^2 &= \ln(\tilde{s}^2 + \bar{x}^2) - 2\ln(\bar{x}) \\ &= \ln\left(\frac{\tilde{s}^2 + \bar{x}^2}{\bar{x}^2}\right) = \ln\left(\frac{\tilde{s}^2}{\bar{x}^2} + 1\right)\end{aligned} \tag{95}$$

In a similar manner, substituting these values into Equation 94 gives the following expression for μ:

$$\begin{aligned}\mu &= 2\ln \bar{x} - \frac{1}{2}\ln(\tilde{s}^2 + \bar{x}^2) = \ln \bar{x} + \ln \bar{x} - \frac{1}{2}\ln(\tilde{s}^2 + \bar{x}^2) \\ &= \ln \bar{x} + \ln\left(\frac{\bar{x}^2}{\tilde{s}^2 + \bar{x}^2}\right)^{1/2} = \ln \bar{x} - \ln\left(\frac{\tilde{s}^2 + \bar{x}^2}{\bar{x}^2}\right)^{1/2} \\ &= \ln \bar{x} - \frac{1}{2}\ln\left(\frac{\tilde{s}^2 + \bar{x}^2}{\bar{x}^2}\right) = \ln \bar{x} - \frac{1}{2}\ln\left(\frac{\tilde{s}^2}{\bar{x}^2} + 1\right) = \ln \bar{x} - \frac{1}{2}\sigma^2\end{aligned} \tag{96}$$

If $\tilde{s}$ and $\bar{x}$ in Equation 95 are replaced by ξ and α, respectively, then Equation 95 will be identical to Equation 66 (Chapter III). Similarly, if $\bar{x}$ in Equation 96 is replaced by α, then Equation 96 will be identical to Equation 65. These results imply that the method of moments can be applied to experimental data by either of two approaches. One approach is to calculate the moments of the data about the origin and use Equations 93 and 94 to determine μ and σ. Another approach is to set the arithmetic mean of the model equal to the arithmetic mean of the data, $\alpha = \bar{x}$, and the standard deviation of the model equal to the biased estimate of the standard deviation of the data, $\xi = \tilde{s}$. In the computer program developed for this study (Appendix A), the second approach was used. The arithmetic mean and standard deviation were calculated from the data, and Equations 95 and 96 were used to compute μ and σ. Although the unbiased estimate of the variance s^2 generally should be used in such computations, the difference between the biased and unbiased values will be negligible when n is large, as it usually is for a year of air quality data.

METHOD OF FRACTILES

In the "method of fractiles," the two parameters of the model are chosen in a manner such that the CDF of the model is equal to the cumulative frequencies of the observed data at two points, or fractiles. In this context, a "fractile" is a number between 0 and 1 which represents either the CDF of the model or the percentiles (divided by 100) of the observations. Associated

with each fractile f is a value $X = x_f$ of the CDF, such that $F_X(x_f) = f$. For example, if a fractile is selected as $f = 0.70$, then $F_X(x_{70}) = 0.70$, and we set x_{70} equal to the value occurring at the 70-percentile point of the cumulative frequencies of the observations.

Let f_a and f_b represent the two fractiles selected, with corresponding CDF values of x_a and x_b:

$$F_X(x_a) = f_a \tag{97}$$

$$F_X(x_b) = f_b \tag{98}$$

where $$f_b > f_a$$

If we consider the random variable Y, which is the logarithmic transform $Y = (\ln X - \mu)/\sigma$ (see Chapter III), then we can write Equations 97 and 98 as follows:

$$F_Y\left(\frac{\ln x_a - \mu}{\sigma}\right) = f_a \tag{99}$$

$$F_Y\left(\frac{\ln x_b - \mu}{\sigma}\right) = f_b \tag{100}$$

Solving Equations 99 and 100 for the quantity in parentheses and multiplying by σ, we obtain the following pair of equations, in which $F_Y^{-1}(f_a)$ denotes the inverse CDF of the Gaussian, or normal,

distribution:

$$\ln x_a - \mu = \sigma F_Y^{-1}(f_a) \tag{101}$$

$$\ln x_b - \mu = \sigma F_Y^{-1}(f_b) \tag{102}$$

Subtracting Equation 101 from Equation 102, we obtain an equation which has σ as its only unknown:

$$\ln x_b - \ln x_a = \sigma\left[F_Y^{-1}(f_b) - F_Y^{-1}(f_a)\right] \tag{103}$$

Then, solving for σ,

$$\sigma = \frac{\ln x_b - \ln x_a}{F_Y^{-1}(f_b) - F_Y^{-1}(f_a)} = \frac{\ln\left(\frac{x_b}{x_a}\right)}{F_Y^{-1}(f_b) - F_Y^{-1}(f_a)} \tag{104}$$

Likewise, either Equation 101 or Equation 102 can be solved for μ:

$$\mu = \ln x_a - \sigma F_Y^{-1}(f_a) \tag{105}$$

$$\mu = \ln x_b - \sigma F_Y^{-1}(f_b) \tag{106}$$

Equation 104, when used in conjunction with Equation 105 (or Equation 106), provides a useful approach for calculating the parameters μ and σ for the two-parameter lognormal probability model by the method of fractiles. Any two fractiles f_a and f_b

are selected, and the inverse CDF values $F_Y^{-1}(f_a)$ and $F_Y^{-1}(f_b)$ are obtained from tables of the normal distribution. Then the pollutant concentrations x_a and x_b associated with these fractiles are obtained from the cumulative frequencies of the raw observations. We substitute these four values into Equation 104 in order to calculate σ. Then σ is substituted into Equation 105, and μ is calculated.

Larsen[21-24] has recommended use of the two fractiles $f_a = 0.70$ and $f_b = 0.999$ for analyzing air quality data. Although Aitchison and Brown[3] show that the maximum efficiencies for estimating μ and σ do not occur at these particular fractiles,* we shall use the two Larsen values in this study. By consulting tables of the CDF of the normal distribution, and using interpolation as necessary, we obtain the following inverse CDF values for the two Larsen fractiles:

$$F_Y^{-1}(0.70) = 0.52440 \tag{107}$$

$$F_Y^{-1}(0.999) = 3.09023 \tag{108}$$

Substituting these values into Equation 104, we obtain a relatively simple expression for σ:

---

*Aitchison and Brown[3] show that the maximum efficiency attainable for estimating μ occurs at $f_a = 0.27$ and $f_b = 0.73$, while the maximum efficiency attainable for estimating σ occurs at $f_a = 0.07$ and $f_b = 0.93$ when symmetrical fractiles, or quantiles, are chosen.

$$\sigma = \frac{\ln x_b - \ln x_a}{2.56583} = 0.389737 \ln\left(\frac{x_b}{x_a}\right) \tag{109}$$

Once σ is calculated from Equation 109, we calculate μ using either Equation 105 or 106:

$$\mu = \ln x_b - 3.09023 \tag{110}$$

$$\mu = \ln x_a - 0.5244\sigma \tag{111}$$

Larsen[24] suggests the following equation for determining which observation should be plotted at a particular frequency:

$$r = \frac{fn}{100} + 0.999 \tag{112}$$

where r = the rank order of the observations (where r is truncated to give an integer value)

f = cumulative frequency (percent)

n = number of observations

He states that the constant 0.999 is used "...to give uniform computer results," which select the observation closest to the desired frequency, but no theoretical basis for this approach is presented.

The previous discussion assumes that the 70 percentile of the cumulative frequencies of observed concentrations is readily

available. However, because the data usually are grouped, there may not be a concentration which corresponds exactly to the 70 percentile and 99.9 percentile. Larsen[24] circumvents this problem by approximating the values of x_a and x_b as the concentrations falling closest to the 70 percentile and the 99.9 percentile, respectively. A graphical interpolation approach would be subjective, and too vulnerable to error. Therefore, we seek a way to determine the 70 percentile and 99.9 percentile points that can be adapted to the computer. Obviously, a variety of different ways can be used to approximate these two percentiles.

In our computer program (see Appendix A), the method of fractiles was applied to the data by the subroutine RALPH. The rank order for the 70 and 99.9 percentiles was computed by substituting 70 and 99.9 into Equation 112, along with the number of values n in the data set. This subroutine summed up the number of observations in each interval i, storing the result in $JSUM_i$, and found the interval which met the following condition:

$$JSUM_{i-1} < r \leq JSUM_i \tag{113}$$

Consider, for example, the cumulative frequencies of CO observations obtained for New York City (Figure 5). There are 8,364 hourly observations in this data set. Using Equation 112, the ranking positions, denoted as M1 and M2, are computed by the subroutine RALPH in the following manner (The notation

"Int" denotes the FORTRAN function IFIX(z), which truncates the decimal portion of any floating point number z.):

<u>70 percentile</u>

$$M1 = r = \text{Int}[0.7(8364) + 0.999] = \text{Int}[5855.80] = 5855 \qquad (114)$$

<u>99.9 percentile</u>

$$M2 = r = \text{Int}[0.999(8364) + 0.999] = \text{Int}[8356.64] = 8356 \qquad (115)$$

In the computer program, the number of observations in each interval are stored in the vector JH(I), in which I is an index. As I is incremented, the number of observations in JH are succesively added and stored in JSUM. This subroutine finds the first inteval in which the cumulative number of observations just equals or exceeds r. For New York City, the following values were determined:

$$70\%: \quad 5675 < 5855 \leq 6054 \quad (I = 20,\ x_a = 19 \text{ ppm}) \qquad (116)$$

$$99.9\%: \quad 8354 < 8356 \leq 8356 \quad (I = 49,\ x_b = 48 \text{ ppm}) \qquad (117)$$

The operation of this subroutine is described in greater detail in Chapter V.

Figure 5 shows the cumulative frequencies of the observations for New York City. Notice that the observations are stored in this program at intervals that are offset by 0.5 ppm. Interval Nos. 20 and 49 actually are at 19.5 ppm and

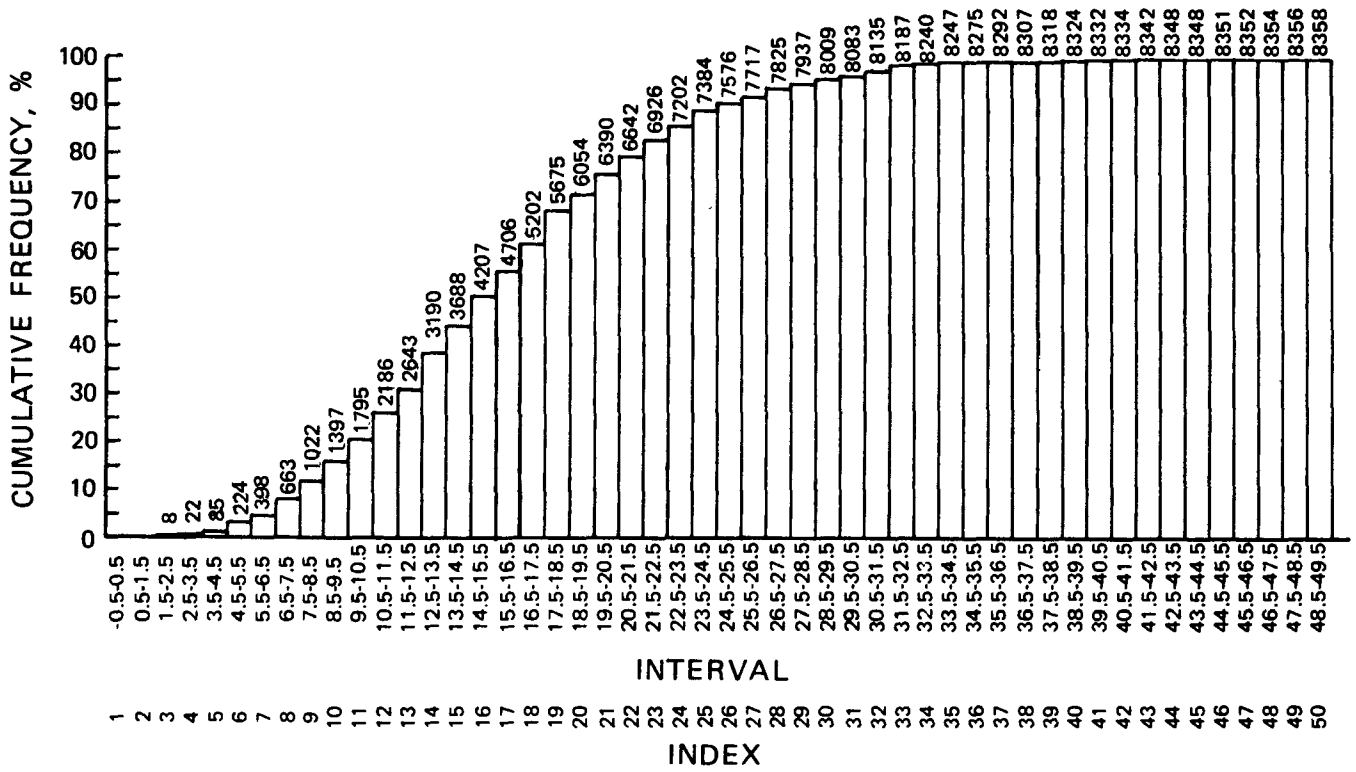

Figure 5. Cumulative frequencies of hourly CO observations in New York City, 1974.

48.5 ppm, respectively, with exact cumulative frequencies of 72.382% and 99.904%. The concentrations used in Equation 109 are the midpoints of the intervals, x_a = 19 ppm and x_b = 48 ppm.

A problem arises if 70 percent or more of the observations are zero. If, for this situation, $x_a = 0$ is substituted into Equation 109, the result is undefined. In the present study, this problem occurred with one data set, Barstow, CA, for which 70.08% of the observations were zero. For this situation, Larsen[68] recommends that the arithmetic mean be substituted for the 70 percentile value; that is, $x_a = \bar{x}$. The subroutine RALPH was designed to make this substitution automatically in these instances.

Although the approach for determining x_a and x_b is patterned after the report by Larsen,[24] the theoretical basis for some of these steps is not entirely clear. Thus, we recommend that other formulas for the ranking position be explored, such as the following simple equation, in which i is the cumulative count of the number of observations:

$$r = \frac{i}{n + 1} \tag{118}$$

Another approach for selecting the fractiles would be to determine the cumulative frequency <u>closest to</u> the 70 percentile or 99.9 percentile. Then the exact cumulative frequencies for these points might be substituted into Equation 104, rather than $f_a = 0.70$ and $f_b = 0.999$ that have been incorporated into Equation 109. Using this approach with New York City's

data, we would substitute $f_a = 0.67842$ (that is, 5675/8365) and $f_b = 0.99892$ (that is, 8356/8365) into Equation 104, with $x_a = 18.5$ ppm and $x_b = 48.5$ ppm. Use of Equation 104 is made computationally easy by the existence of convenient algorithms for computing the CDF and inverse CDF of the normal distribution, such as those incorporated into the computer program used in this study (Appendix A). One advantage of this approach is that the two points selected will coincide exactly with two of the points that are plotted on logarithmic probability paper, because the ranking position given by Equation 118 is the one that is commonly used in graphical work.

Another technique for calculating parameters for the lognormal model that has been suggested by several authors[1,24,63] is a combination of the method of moments and the method of fractiles. Here, the arithmetic mean of the model is set equal to the arithmetic mean of the data (that is, $\alpha = \overline{x}$), and the second parameter is chosen by the method of fractiles in the manner described above. Because the mean of the model then matches the mean of the data, this approach is useful when other averaging times are considered, as they are in the averaging time models. However, this approach needs further development and evaluation before it can be recommended as a preferred technique.

MAXIMUM LIKELIHOOD ESTIMATION (MLE)

Assume that a set of n independent observations $(x_1, x_2, \ldots, x_n)$ is obtained from a distribution whose basic form is known. Using the knowledge of these observations, what can we determine about

most likely values of the parameters of this distribution? For example, if we know that the observations arise from a lognormally distributed random variable X, what are the <u>most probable</u> values of μ and σ, given this particular set of observations? The method of maximum likelihood estimation (MLE) seeks to determine the optimal values of μ and σ that maximize the probability P(A) of the event $A = (x_1, x_2, \ldots, x_n)$.

Assuming that a random variable is lognormally distributed, we form the product of the PDF values for the n observations. We call this product the likelihood function L:

$$L = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x_i - \mu}{\sigma}\right)^2}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \left[\prod_{i=1}^{n} \frac{1}{x_i}\right] e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln x_i - \mu)^2} \qquad (119)$$

The product L usually is a very small number. For computational purposes, it generally is more convenient to work with the logarithm of L. Thus, we form the log-likelihood function L*:

$$L^* = n \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{n} \ln x_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln x_i - \mu)^2 \qquad (120)$$

Maximizing the likelihood function L is equivalent to maximizing

the log-likelihood function L*. To find the values of μ and σ which maximize L*, we take first partial derivatives of L* with respect to μ and σ and then set the resulting functions equal to zero. Differentiating Equation 120 with respect to μ and setting the result equal to zero:

$$\frac{\partial L^*}{\partial \mu} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(\ln x_i - \mu)(-1) = 0$$

Therefore,

$$\sum_{i=1}^{n}(\ln x_i - \mu) = \sum_{i=1}^{n}\ln x_i - n\mu = 0$$

Or,

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\ln x_i \tag{121}$$

Differentiating Equation 120 with respect to σ and setting the result equal to zero:

$$\frac{\partial L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(\ln x_i - \mu)^2 = 0$$

Therefore,

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(\ln x_i - \mu)^2 = n$$

Or,

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i)^2 - \mu^2 \tag{122}$$

This result shows that the maximum likelihood estimators for the lognormal distribution are obtained by first taking logarithms of the observations. Then the mean of the logarithms is calculated using Equation 121, and the (biased) estimate of the variance of the logarithms is calculated using Equation 122.

The underlying theory of maximum likelihood estimation assumes that the observations are independent. In the present investigation, the data consist of hourly observations which are known to be serially correlated, and assumptions about independence are violated. Here, the MLE approach is assumed to be just one of a series of possible candidate techniques for calculating the parameters of the lognormal distribution. Although the approach is satisfactory for our purposes, it should be noted that the resulting estimators, under these circumstances, will not necessarily exhibit the properties associated with the theory.

MLE Approximation

A serious problem arises when we attempt to apply Equations 121 and 122 to data sets containing observations that are zero, because the logarithm of zero is undefined. In many of the data sets used in this investigation, and in most environmental measurements, one frequently encounters observations that are reported as zero. To apply the MLE approach for the lognormal model, it is necessary to develop a means for handling these cases.

In developing our technique, we must note that concentrations reported as zero are not necessarily exactly zero ppm. One reason

is that all observations have been rounded off at the time of the measurement to one or two significant figures; thus, 0 ppm might have been reported even though the true observation was 0.237 ppm. This "automatic rounding" usually is a result of the limited precision of the measuring method, and there is no way to obtain more precise values from today's measuring methods. A second problem is that most measuring instruments have "minimum detectable limits" that cause them to respond poorly, or not at all, at very low concentrations. A final reason that true zeros are unlikely is that even the purest sample of air is likely to contain at least a few molecules of a given pollutant, and, if sufficient precision were possible, we would expect to observe at least several parts-per-billion (ppb) of the pollutant rather than exactly 0.0 ppm.

Nehls and Akland[69] have suggested procedures for the analysis of aerometric data that include ways to handle observations below the minimum detectable limit (MDL) of a given instrument. Although a theoretical basis is not given, they propose that a value that is the mid-point between zero and the MDL of the measurement method be substituted for all observations below the MDL, and they provide a table of the MDL values for various air pollutants. For carbon monoxide, which usually is measured by nondispersive infrared absorption (NDIR) techniques, they recommend substitution of 286 $\mu g/m^3$ (approximately 0.25 ppm) for values below the CO MDL of 573 $\mu g/m^3$ (approximately 0.5 ppm). Kushner[70] has compared this assumption that all measurements below

the MDL are equal to one-half the MDL, which he calls the "L/2 approximation," with a more exact maximum likelihood approach suggested by Hald.[71] He notes that the L/2 approximation is exact if the concentration below the MDL is a uniformly distributed random variable; if it is not, the L/2 approximation introduces slight error. However, for most applications he concludes that this error "...would probably be overshadowed by the error due to finite sample and measurement noise."[71]

In this study, we apply the concept of the L/2 approximation to the first interval (-0.5 ppm to + 0.5 ppm). Since negative concentrations are impossible, the first interval of the histogram actually extends from 0.0 ppm CO to 0.5 ppm CO. Thus, the L/2 approximation is applied by substituting one-half of 0.5 ppm, or X = 0.25 ppm, for any concentration that appears in the first interval. Because all CO concentrations in the first interval are zero values, the result is to substitute 0.25 ppm for any CO concentration reported as 0 ppm. Because of the way in which intervals were chosen, this value happens to coincide with the value suggested by Nehls and Akland.[69]

In the "MLE approximation" technique for calculating the model's parameters, which is implemented in the subroutine AMLE (see Appendix A), logarithms of the raw observations are taken, and Equations 120 and 121 are used to calculate μ and σ. Whenever a zero value is encountered, the subroutine automatically substitutes 0.25 ppm for the observation. It should be noted that the magnitude of the substitution value, if many zeros are present, has a significant effect on μ and σ, because logarithms

are taken. For this reason, this study also examined the effect of using a different substitution value, X = 0.5 ppm, on all data sets.

MLE Optimization

To examine a somewhat different approach for implementing the MLE concept, a computer algorithm was developed that would iteratively find the values of μ and σ that maximize the log-likelihood function L*. In this algorithm, which is contained in subroutine BMLE, L* was calculated by first calculating the probability associated with each interval. Because the intervals are offset by 0.5 ppm, which is one-half the interval width (see Figure 5), the probability that X lies within the first interval is given as follows:

$$P_1 = P(-0.5 \text{ ppm} \leq X < 0.5 \text{ ppm}) = F_X(0.5) - F_X(0.0) \qquad (123)$$

Similarly, the probability that X lies within the second and third intervals is given as follows:

$$P_2 = P(0.5 \text{ ppm} \leq X < 1.5 \text{ ppm}) = F_X(1.5) - F_X(0.5) \qquad (124)$$

$$P_3 = P(1.5 \text{ ppm} \leq X < 2.5 \text{ ppm}) = F_X(2.5) - F_X(1.5) \qquad (125)$$

In general, the probability P_j that X lies within the jth interval is given as follows, in which Δx denotes the interval width, and m denotes the total number of intervals:

$$P\left([j - \tfrac{3}{2}]\Delta x \leq X < [j - \tfrac{1}{2}]\Delta x\right)$$

$$= F_X([j - \tfrac{3}{2}]\Delta x) - F_X([j - \tfrac{1}{2}]\Delta x) \tag{126}$$

$$\text{for } j = 2,3,\ldots,m$$

Using the logarithmic transform $Y = (\ln X - \mu)/\sigma$, P_2 is calculated as follows:

$$P_2 = F_Y\left(\frac{\ln 1.5 - \mu}{\sigma}\right) - F_Y\left(\frac{\ln 0.5 - \mu}{\sigma}\right) \tag{127}$$

For example, if $\mu = 2$ and $\sigma = 1$, P_2 is computed as follows:

$$P_2 = F_Y(\ln 1.5 - 2.0) - F_Y(\ln 0.5 - 2.0)$$

$$= F_Y(0.405446 - 2.0) - F_Y(-0.693147 - 2.0)$$

$$= F_Y(-1.594554) - F_Y(-2.693147)$$

$$= 0.055406 - 0.003539 = 0.051867$$

For each observation in a given interval, we form the likelihood function L by multiplying by the probability associated with the interval. For example, if interval No. 2 contains 12 observations, the factor of L for this interval is P_2^{12} or $(0.051867)^{12}$. The factors for all intervals then are multiplied together:

$$L = \prod_{j=1}^{k} P_j^{m_j} \tag{128}$$

where P_j = probability associated with the j<u>th</u> interval

m_j = number of observations within the j<u>th</u> interval

k = total number of intervals

Because Equation 128 generates numbers that are too small to be stored on modern computers, the subroutine BLME actually works with the logarithms of the probabilities. Thus, the log-likelihood function is computed as follows:

$$L^* = \sum_{j=1}^{k} m_j \log P_j \tag{129}$$

The computer program listed in Appendix A and described in Chapter VI uses an iterative procedure to find the values of μ and σ which maximize L^*.

V. GOODNESS-OF-FIT

Once the model's parameters have been calculated from a particular set of observations, it is important to determine how well the model fits the observations. Such tests may reveal that the parameters have been calculated incorrectly or that the model is unsuited to a particular application. Thus, the investigator requires objective measures of the "goodness-of-fit" of the model to the data.

Two basic types of goodness-of-fit measures are available: (1) frequency-based measures, which depend on the number of observations within intervals, and (2) variate-based measures, which compare concentrations predicted by the model with concentrations observed at each interval. Because of different educational backgrouns, statisticians generally prefer frequency-based measures, while engineers often prefer variate-based measures.

FREQUENCY-BASED MEASURES

One possible frequency-based measure of goodness-of-fit is the difference between the number of observations predicted by the model in each interval of the histogram and the number actually observed. Because this difference sometimes will be negative, it is convenient to form the square of the difference. To obtain a single measure for all k intervals of the histogram, we can add up the squares of the differences, divide by k, and then take the square root of the result, giving the "root-mean-square (RMS)" of predicted minus observed values:

$$\text{RMS} = \left[\frac{1}{k}\sum_{j=1}^{k} (n_p - n_o)_j^2\right]^{1/2} \tag{130}$$

where n_p = number of values predicted by the model in the jth interval

n_o = number of values actually observed in the jth interval

k = number of intervals

The RMS may be viewed as simply an "index" of goodness-of-fit, and no particular probabilistic interpretation is attached to it.

If, instead of dividing the difference squared by k, we divide by the predicted number of observations n_p in each interval and then sum the result, we obtain the chi-square (χ^2) measure of goodness-of-fit:

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_p - n_o)_j^2}{n_p} \tag{131}$$

If the observations are independent, the result computed from Equation 131 will have a chi-square distribution with k - r - 1 degrees of freedom, in which r is the number of parameters in the model. Although the histogram intervals used in this study are of uniform width, it is not necessary that each interval be the same size. Thus, several intervals could be grouped together. Of course, grouping reduces the number of degrees of freedom. The

usual requirement is that at least five expected values should appear in each interval.

Tables of the chi-square CDF are available[72] for various degrees of freedom. By use of the tables, one can make a probability statement such as the following: "We can reject the hypothesis that the observations arise from a lognormal distribution with 95-percent certainty ($P = 0.95$)." Here, for the LN2 model, $r = 2$. Because the CO hourly observations used in this study are serially correlated, the assumption of independence does not apply, making a probabilistic interpretation of χ^2 more difficult. However, χ^2 still provides a useful index for comparing the empirical goodness-of-fit of different parameter calculation techniques.

Another frequency-based measure of goodness-of-fit is the log-likelihood function itself, Equation 129, which always is a negative number, ususally of large magnitude:

$$L^* = \sum_{j=1}^{k} m_j \log P_j \qquad (132)$$

The best fit of the model to the observations occurs when L^* is greatest; that is, when the magnitude of L^* is smallest.

A final type of frequency-based measure of goodness-of-fit is derived from the absolute value of the difference between the CDF and the cumulative frequencies of the observations. In the Kolmogorov-Smirnov test,[73] for example, the ingestigator computes the maximum difference between the CDF and the cumulative fre-

quencies of individual observations. By referring to tables, he can test the hypothesis that the observations arise from the model under study. As discussed earlier, the hourly CO concentrations measured at urban air monitoring stations are intrinsically grouped. Thus, the Kolmogorov-Smirnov computation must be applied to the interval steps rather than to each observation:

$$D = \max_{\text{all } j} \left\{ \left| CDF_j - CUMFREQ_j \right| \right\} \tag{133}$$

where CDF_j = CDF of the model for the $j\underline{th}$ interval

$CUMFREQ_j$ = cumulative frequency of the observations for the $j\underline{th}$ interval

Some authors also recommend computation of the sum of the absolute value of the differences:

$$D_s = \sum_{j=1}^{k} \left| CDF_j - CUMFREQ_j \right| \tag{134}$$

The above discussion covers five different measures of goodness-of-fit, all of which were calculated by the computer program developed for this investigation (See, for example, the program output given in Appendix B). However, not all of these goodness-of-fit measures are discussed in detail in

the summary of findings (Chapters VIII and IX).

VARIATE-BASED MEASURES

The upper boundary of each interval corresponds to a particular cumulative frequency of observations, and we can use the model to calculate the predicted concentration x_p for the interval and compare it with the actual concentration of the interval x_o. For example, if $(x_o)_j$ denotes the observed concentration and $CUMFREQ_j$ denotes the cumulative frequency of observations for the j<u>th</u> interval, then we can use the model to calculate the concentration predicted for that interval by taking the inverse of the CDF: $(x_p)_j = F_X^{-1}(CUMFREQ_j)$. The difference between the predicted and observed concentrations for each interval can be treated as a random variable that is itself a measure of the goodness-of-fit of the model to the observations:

$$d_j = (x_p - x_o)_j \tag{135}$$

By examining the histogram of d, its mean $\overline{d}$, its standard deviation s_d, its range, and other characteristics, the investigator can make statements about the practical suitability of the model for representing a particular data set.

This approach can be illustrated by applying it to the first 10 intervals of the New York City CO data set (See Figure 5, page 59). Although the entire data set consists of 8,364 hourly observations (See Appendix B or C

for the complete histogram), the first 10 intervals contain only 1,397 observations (Table 2).

Table 2

Number of Observations and Cumulative Frequencies for the First 10 Intervals of the New York City CO Data Set

j	Interval (ppm CO)	No. of Observations	Cumulative No. of Obs.	Cumulative Frequency
1	-0.5-0.5	0	0	0.000%
2	0.5-1.5	0	0	0.000%
3	1.5-2.5	8	8	0.096%
4	2.5-3.5	14	22	0.263%
5	3.5-4.5	63	85	1.016%
6	4.5-5.5	139	224	2.678%
7	5.5-6.5	174	398	4.758%
8	6.5-7.5	265	663	7.927%
9	7.5-8.5	359	1022	12.219%
10	8.5-9.5	375	1397	16.703%

For New York City, the arithmetic mean of the hourly CO obser- was $\bar{x}$ = 16.2074 ppm and the arithmetic standard deviation was s = 6.8687 ppm. Calculating the parameters for the LN2 model by the Larsen method of fractiles, we obtain μ = 2.7550 and σ = 0.36119 (See Chapter IV). When the model's parameters are calculated in this fashion, its arithmetic mean is α = 16.7812 ppm and its arithmetic standard deviation is ξ = 6.2644 ppm. If, for example, we wish to calculate x_p for the seventh interval (j = 7), we note that the cumulative frequency of observations for values less than 6.5 ppm (X < 6.5) is 4.758%.

Taking the inverse CDF of the normally distributed random variable Y which describes the logarithms of concentrations, $Y = (\ln X - \mu)/\sigma$, we obtain $F_Y^{-1}(0.04758) = -1.6691$. Then, x_p and y_p for this interval are related as follows:

$$y_p = \frac{\ln x_p - 2.7550}{0.36119} = -1.6691$$

Solving for x_p,

$$\ln x_p = (-1.6691)(0.36119) + 2.7550 = 2.1521$$

$$x_p = e^{2.1521} = 8.603 \text{ ppm}$$

Because $x_o = 6.5$ ppm for this interval, the model overpredicts the concentration by $d_7 = 8.603 - 6.5 = 2.103$ ppm.

The predicted values and the differences d_j for the first 10 intervals are calculated in a similar fashion (Table 3). The results for the entire histogram (the first 50 intervals) can be presented graphically by plotting x_p versus x_o (Figure 6). Here, the 45 degree line, $x_p = x_o$, denotes an exact correspondence between predicted and observed values. The vertical distance of any point above the line is the value of d. For the New York City data set, the model, with parameters calculated by the method of fractiles, tends slightly to overestimate CO concentrations below 15 ppm, with very good correspondence over the range from 15 ppm to 50 ppm. Of course, an exact correspondence

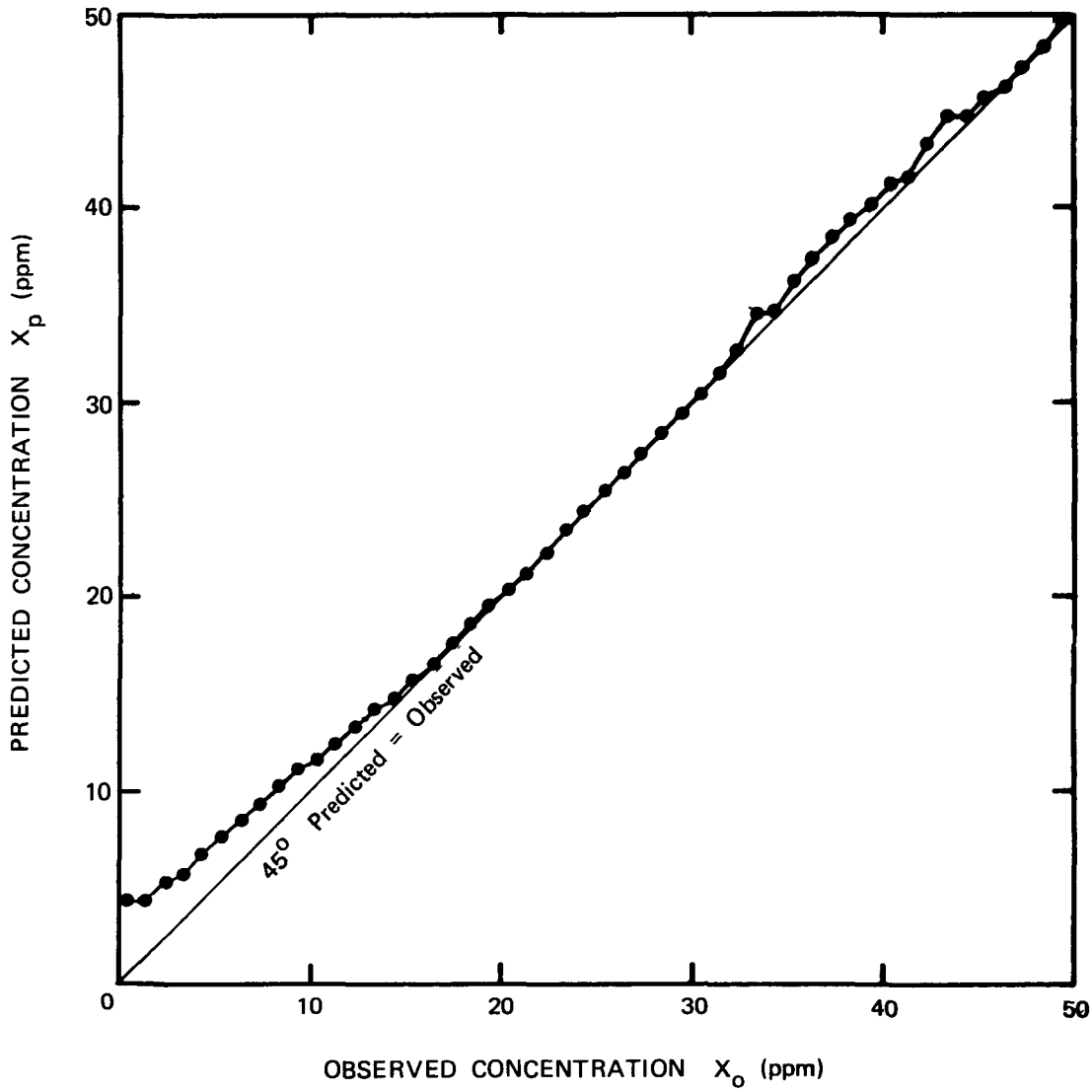

Figure 6. Predicted versus observed CO concentrations for New York City, with parameters of the LN2 model calculated by the method of fractiles.

Table 3

EET Values for the First 10 Intervals of the New York City Data Set, with LN2 Parameters Calculated by the Method of Fractiles

j	x_o (ppm CO)	Cumulative Frequency $CUMFREQ_j$	$F_Y^{-1}(CUMFREQ_j)$	x_p (ppm CO)	$d_j=(x_p-x_o)_j$ (ppm CO)
1	0.5	0.000%	a	a	a
2	1.5	0.000%	a	a	a
3	2.5	0.096%	-3.1037	5.1243	2.6243
4	3.5	0.263%	-2.7910	5.7370	2.2370
5	4.5	1.016%	-2.3207	6.7991	2.2991
6	5.5	2.678%	-1.9308	7.8275	2.3275
7	6.5	4.758%	-1.6691	8.6034	2.1034
8	7.5	7.927%	-1.4103	9.4465	1.9465
9	8.5	12.219%	-1.1642	10.3246	1.8246
10	9.5	16.703%	-0.9659	11.0911	1.5911

[a]Value undefined.

occurs at the intervals corresponding to the 70 percentile (x_o = 19.5 ppm) and the 99.9 percentile (x_o = 48.5 ppm).

The characteristics of the differences can be examined in greater detail by considering the histogram of the d's and the statistical properties of this measure (Figure 7). The average difference ($\bar{d}$ = 0.5233 ppm) was positive, but small. No difference was less than -0.42 ppm, and the maximum positive differences was +2.62 ppm. Of the 50 differences, 58% were in the range from -0.5 ppm to +0.5 ppm, and 80% were in the range from -0.5 ppm to +1.0 ppm. Because the precision of the NDIR measuring instrument is, at best, on the order of $\pm$0.5 ppm, the correspondence of the model to the observations, using this measure of

```
                    *  *  *   DISTRIBUTION OF DATA (INTERVAL WIDTH =  0.5000)   *  *  *

NO. OF VALUES LESS THAN  -5.0000=      0      NO. OF ZERO VALUES =      2

   INTERVAL      NUMBER    PERCENT  +----------------------HISTOGRAM--------------------+ CUM. FREQ.   PLOT POS.
 -5.00- -4.50        0     0.0  %  I                                                  I    0.0  %       0.0  %
 -4.50- -4.00        0     0.0  %  I                                                  I    0.0  %       0.0  %
 -4.00- -3.50        0     0.0  %  I                                                  I    0.0  %       0.0  %
 -3.50- -3.00        0     0.0  %  I                                                  I    0.0  %       0.0  %
 -3.00- -2.50        0     0.0  %  I                                                  I    0.0  %       0.0  %
 -2.50- -2.00        0     0.0  %  I                                                  I    0.0  %       0.0  %
 -2.00- -1.50        0     0.0  %  I                                                  I    0.0  %       0.0  %
 -1.50- -1.00        0     0.0  %  I                                                  I    0.0  %       0.0  %
 -1.00- -0.50        0     0.0  %  I                                                  I    0.0  %       0.0  %
 -0.50-  0.0        16    32.000%  I***************C****************                   I   32.000%     31.373%
  0.0 -  0.50       13    26.000%  I*************************** C                     I   58.000%     56.863%
  0.50-  1.00       11    22.000%  I**********************                 C          I   80.000%     78.431%
  1.00-  1.50        2     4.000%  I****                                     C        I   84.000%     82.353%
  1.50-  2.00        3     6.000%  I******                                       C    I   90.000%     88.235%
  2.00-  2.50        4     8.000%  I********                                       C I   98.000%     96.078%
  2.50-  3.00        1     2.000%  I**                                              CI  100.000%     98.039%
  3.00-  3.50        0     0.0  %  I                                                 CI  100.000%     98.039%
  3.50-  4.00        0     0.0  %  I                                                 CI  100.000%     98.039%
  4.00-  4.50        0     0.0  %  I                                                 CI  100.000%     98.039%
  4.50-  5.00        0     0.0  %  I                                                 CI  100.000%     98.039%
                                  +-----------------------------------------------+

NO. OF VALUES EQUAL TO OR GREATER THAN    5.0000 =     0  ( 0.0  %)

   N =       50

   MEAN =         0.52330536

   2ND MOMENT (VAR.) =          0.694542050              STANDARD DEV. =          0.833391905

        COEFFICIENT OF VARIATION =      1.59255314

   3RD MOMENT =          0.581009209          COEF. OF SKEWNESS =      1.00377178

   4TH MOMENT =          1.40156555           COEF. OF KURTOSIS =      2.90546894

   LOWEST VALUE =        -0.42289734

   HIGHEST VALUE =        2.62427139

        SECOND HIGHEST VALUE =        2.32748318
```

Figure 7. Histogram and summary statistics for the differences obtained from the Engineering Error Test for hourly CO observations from New York City, 1974, with LN2 parameters calculated by the method of fractiles.

goodness-of-fit, appears relatively high.

We shall call this measure of goodness-of-fit the Engineering Error Test (EET). The EET, because it depends on the difference between predicted and observed concentrations at each interval, is a goodness-of-fit measure that has practical significance. When one plots a set of observations on logarithmic probability paper and then draws a straight line to depict the LN2 model, the eye tends visually to weight each point equally and to focus on the concentration difference between the line and the observations. Thus, a line that appears to be a good fit by such graphical approaches will also do well when judged by the EET approach. Unfortunately, not every point plotted on a logarithmic-probability plot represents the same number of observations. Intervals near the center of the distribution contain many observations, while those at the tails, or extremes, contain very few. The EET does not take this fact into account. Thus, as we shall see, the EET tends to provide an excellent measure of how well the model fits a particular empirical data set, which is just one realization of the process under study, but a poor measure of how well the model fits the underlying distribution of the process itself.

VI. DESCRIPTION OF THE COMPUTER PROGRAM

The computer program, which is listed in Appendix A, consists of a main program and seven subroutines, each of which carries out a well-defined function (Table 4). Appendix B contains an example of the output of this program for the data from New York City. This chapter is intended to provide brief documentation of the computer program to assist persons wishing to use the software or the concepts contained in the program.

Table 4

Subroutines Used in the Computer Program and Their Functions

Subroutine	Function
STAT	Computes basic statistics (mean, variance, higher moments, etc.)
HISTOG	Generates and prints a histogram
RALPH	Computes parameters by the method of fractiles
AMLE	Computes parameters by the MLE approximation
BMLE	Computes parameters by the MLE optimization
LN	Examines goodness-of-fit of the model to the data
AUTCOR	Calculates autocorrelation function

MAIN PROGRAM

The main program reads the raw observations into the program

and stores them. Format Statement #110 specifies the format for reading the data. In the present version of the program, 12 values are read per record (one-half day of CO observations). An alphanumeric label identifying the data set also can be read and printed at the beginning of the output. If a label is used, it is inserted immediately before the beginning of the data, and a record containing the integers "11" in columns 1-2 is inserted before the label record. It is assumed that all missing values in the raw data file have been coded with a missing value code, which is denoted in the program as DUMMY. In the present example, DUMMY = 9999.99. Although all observations, including the missing values, are read into the program and counted (L contains the total count), two different vectors are filled. Observations with missing values are saved in the vector STORE(9000), while a second vector, VALUE(9000), contains only the valid observations. Here, missing values have been deleted, and N is the total number of valid observations stored in VALUE.

Once the values have been read and stored, the subroutine STAT is executed in order to compute basic statistics from the raw data. Then the subroutine HISTOG partitions the observations into intervals and prints out a histogram, along with the cumulative frequencies, on the line printer. The statistics generated by STAT and the counts of observations within intervals are used in subsequent steps of the program.

After STAT and HISTOG have been executed, the main program carries out four computation cycles, one for each parameter cal-

culation technique. The computation cycles are executed in the following order: (1) method of moments, (2) method of fractiles, (3) MLE approximation, and (4) MLE optimization. Within each of the computation cycles, the first step is to obtain the values of the parameters, μ and σ; then three subroutines are executed in the following order: LN, HISTOG, and STAT. In the second computation cycle, for example, the subroutime RALPH computes the values of μ and σ (which are represented in the program as U2 and SIG2). Next, LN is executed to determine the goodness-of-fit of a lognormal model with these parameters to the observations. Then, HISTOG is executed to provide a histogram of the differences (d's) for the Engineering Error Test (see Chapter V). Finally, STAT is executed to compute basic statistics of the differences. At the end of the four computation cycles, the subroutine AUTCOR is executed to compute the autocorrelation function.

SUBROUTINES

The following sections describe each subroutine briefly.

STAT(N,VALUE)

The subroutine STAT carries out statistical computations on the N observations stored in the vector VALUE. It computes the arithmetic mean, standard deviation, coefficient of variation (ratio of standard deviation to mean), third and fourth moments about the mean, and coefficients of skewness and kurtosis. It also determines the minimum and maximum observations and the second highest observation. The resulting

values are prited out.

The arithmetic mean (AVE) and the arithmetic standard deviation (STD) are retained in common storage for later use by means of the statement COMMON /F/AVE,STD. Because the method of moments requires a biased estimate of the variance, the value $\tilde{s} = s\sqrt{(n-1)/n}$ is computed in the main program for use in the first computation cycle.

HISTOG(IN,N,VALUE,BASE,DELTA,LIMIT,XI,JH,PV,CUM,HI,IH)

The subroutine HISTOG partitions the observations into intervals of size DELTA and prints out a histogram on the line printer. Like the BASE, which is the lower limit of the observations in the histogram, and LIMIT, which is the number of intervals in the histogram, DELTA is a user option that is set in the main program. In this study, BASE = -0.5 ppm; DELTA = 1.0 ppm; and LIMIT = 60 intervals. Thus, the highest value which could be included on the histogram was (60)(1.0) - 0.5 = 59.5 ppm. Values above this range are stored in the vector HI(1000), and IH contains an integer count of the number of values above the range. If the user wishes to have the values above the range listed after the histogram is printed, he sets IN = 1 in the main program; otherwise the listing will be suppressed. In this study, only one city, Phoenix, had any observations above 59.5 ppm, and this one value was stored in HI in the Phoenix run.

The vector XI(100) contains the concentration values which comprise the abscissa of the histogram. Here, for example,

XI(1) = 0.5 ppm, XI(2) = 1.5 ppm, XI(3) = 2.5 ppm,.... The vector JH(100) contains the number of observations in each interval; the vector PV(100) contains the percentage of the observations in each interval; and the vector CUM(100) contains the cumulative frequency, by interval, expressed as a percentage. The operation of this subroutine is described in greater detail in a report by Ott.[74]

RALPH(U,SIG,AVE1,LIMIT,XI,JH,PV,CUM,N)

The subroutine RALPH computes the two parameters of the lognormal model using the method of fractiles in the manner suggested by Larsen[24] (see Chapter IV). In this computation, the subroutine uses the vectors XI and JH to find the interval that contains the ranked observation corresponding to the 70 percentile and 99.9 percentile of the data. Once the concentrations for the proper intervals have been determined, μ and σ are computed. These floating-point values are represented in the program as U and SIG, respectively.

A problem arises if more than 70 percent of the observations are zero, for Larsen's approach cannot be applied when one of the two concentrations is zero. This situation occurre in the data set for Barstow, CA. In this situation, Larsen[68] recommends that the arithmetic mean of the data set be substituted for the 70-percentile value. Thus, AVE1, the observed arithmetic mean, is transferred to this subroutine, and the substitution is made automatically for this case.

AMLE(U,SIG,N,VALUE)

The subroutine AMLE computes the LN2 parameters by the method of MLE approximation described in Chapter IV. This is accomplished by taking the logarithm of each observation and calculating μ as the mean of the logarithms and σ as the (biased) estimate of the standard deviation of the logarithms. For observations reported as zeros, the "L/2 approximation" is used. That is, zero values are set to the midpoint between 0.0 ppm and 0.5 ppm, or 0.25 ppm, which is the midpoint of the first interval.

BMLE(U,SIG,LIMIT,XI,JH,PV,CUM)

In contrast with AMLE, the subroutine BMLE is relatively complex. It finds the values of μ and σ that maximize L*, the log-likelihood function, by means of an iterative approach. It accomplishes this by systematically incrementing (or decrementing) μ and σ until a maximum is found and the absolute value of the increment (or decrement) reaches an arbitrarily small number. To determine whether to increment or decrement, it tests to see which of the two choices will cause L* to increase. For example, if adding the amount $\Delta\mu$ to μ causes L* to increase (become less negative), the subroutine will continue to "advance" (that is, to increment μ by $\Delta\mu$) on each iteration until L* no longer increases. Then, it will divide $\Delta\mu$ by 2 and repeat the process. Once the maximum for μ has been found, the process is repeated for σ. When the optimum value of σ has been found, it returns to μ, and so on, ultimately

finding the values of μ and σ that correspond to a very small increment (here, the process stops when $\Delta\mu$ or $\Delta\sigma$ is less than 0.000001). Because this subroutine was developed especially for this study as an integral part of the computer program, the comments in the program refer to it as Ott's optimization method.

The steps within the subroutine are reasonably straightforward and should be easy to follow. The value of the log-likelihood function L* is calculated by the DO loop that ends in Statement #20 and is stored in SUM. This DO loop is contained within a second DO loop that ends in Statement #10 and causes the log-likelihood function to be computed three times upon each iteration. To describe this computation, we shall assume that z denotes the value of either μ or σ, depending on which one is being optimized. If L = 1, z = μ, and if L = 2, z = σ. On the first computation (I = 1), the DO loop ending in Statement #10 computes L*(z) and stores the result in SUM1; on the second computation (I = 2), it computes L*(z - d), where d is the increment (represented in the program as D or DELTA), and stores the result in SUM2; on the third computation (I = 3), it computes L*(z + d) and stores the result in SUM3.

Here, SUM1, SUM2, and SUM3 are test values of the log-likelihood function which are used to determine whether z should be incremented ("advance forward") or decremented ("move backward") or modified ("divide by 2.0 and recalculate"). The integer MEM is intended to "remember" the direction. Initially MEM = 0, and if SUM3 is greater than SUM1, indicating that computation should

advance, the program sets MEM = 1. Conversely, if SUM2 is greater than SUM1, the program sets MEM = 2. Depending on the outcome of these tests, the program either increments or decrements z (either U or SIG, depending on the value of the integer L) and returns to Statement #5. Once MEM has been set, the increment (or decrement) cycle is repeated over and over until "overshoot" occurs. Overshoot is determined, if MEM = 1 (advance mode), by testing whether SUM3 is less than SUM1; conversely, if MEM = 2 (backward mode), by testing whether SUM2 is less than SUM1. Overshoot can be interpreted as follows. If the program is in the advance mode, overshoot implies that the log-likelihood function has been successively incremented a number of times when suddenly it is found that, for the next increment, the value of L*(z + d) is less than L*(z). When this happens, control transfers to Statement #25, and the following four steps occur: (1) d is divided by 2.0, (2) MEM is again set to zero, (3) the program tests to see if d is less than its stopping point value (that is, 0.000001), and, if not, (4) control returns to Statement #5.

In order to insure convergence, the algorithm also must handle the situation in which z is very close to its optimum value, and both L*(z - d) and L*(z + d) are less than L*(z). In this case, which we call a "trough," the program tests to see if both SUM2 and SUM3 are less than or equal to SUM1. If so, control transfers to Statement #25, and the four steps listed above are executed. That is, d is divided by 2.0, MEM

is reset, the convergence test is made, and control returns to Statement #25.

This subroutine always begins with SIG as the variable to be optimized (that is, L = 2). Once the convergence test has been satisfied for SIG, the program sets L = 1, and control transfers to Statement #1 so the process can be repeated with U as the variable to be optimized. When the convergence test has been satisfied for U, the program again sets L = 2, control transfers to Statement #1, and the process is repeated on SIG. This process of reversing from U to SIG and back again could be carried on indefinitely, but certain stops have been built into the subroutine. Each time that L is changed, the old value of the log-likelihood function L* is stored in SUMT so that it can be compared with the new value of L*, which is stored in SUM1. If the absolute value of the difference between the old and new values of L*--that is, ABS(SUM1 - SUMT)--is less than 0.0001, then the counter K is incremented. If K is 3 or more, all computations of the subroutine are concluded, and control returns to the main program.

As the steps of this subroutine are executed, considerable output is generated (see Appendix B). On each iteration, the program prints out the following information in 9 columns: the index number of the iteration (the value of ICOUNT); the code indicating whether μ or σ is being optimized (the value of L); the number of times that the probability is zero for an interval computed in the log-likelihood computation (the value

of K1); the value of $L^*(z)$, which is listed as "SUM1"; the value of $L^*(z + d)$, which is listed as "SUM3"; the mean μ, which is listed as "MEAN"; the standard devaition σ, which is listed as "SIGMA"; and the increment d, which is listed as "CHANGE." By examining this last column, we can see that the increment, which begins as d = 0.1, rapidly becomes very small, aad convergence is obtained very quickly.

Although many improvements could be made to this subroutine to make it more efficient and flexible, we found that it worked very well for our purposes. It always converged, and it gave a high degree of precision quite rapidly. Thus, once it was operational, we felt that further refinements would take additional time that could better be spent on other parts of the study. We feel we can recommend this approach for other applications similar to our own, and we feel that other investigators also may wish to refine it more so it can be used for more general-purpose applications.

LN(U,SIG,N,LIMIT,K,X,DIFF,XI,JH,PV,CUM,HI)

The subroutine LN calculates the goodness-of-fit of a LN2 model with parameters μ = U and σ = SIG to the observations contained in the vector JH with intervals specified by the vector XI. Prior to computing these results, it calculates summary information for the LN2 model that is specified by these two parameters (that is, the arithmetic mean, standard deviation, median, mode, geometric standard deviation, and coefficients of

skewness and kurtosis). Then it prints out a table on the line printer that enables one to compare the number of observations predicted for each interval with the number of observations actually observed. In this table, the first column, which is listed as "INTERVAL", gives the abscissa of the histogram (obtained from the vector XI); the next column, which is listed as "LN(X)" is the natural logarithm of the upper bound of the interval. This is followed by "Y", the number of standard deviations from the mean for the normally distributed random variable $Y = (\ln X - \mu)/\sigma$. Then the CDF of Y is given and is listed as "F(Y)". By subtracting values of F(Y) for successive intervals, we obtain the probability associated with each interval, which is listed as "P(Y)". Then the expected number of observations in the interval, listed as "PREDICTED N", is obtained by multiplying the probability for each interval times the total number of observations N. The next column, "OBSERVED N", lists the actual number of observations contained in each interval; this is followed by a column for the difference between predicted and observed values, called "DIFFERENCE". Finally, the last two columns contain the difference squared, listed as "DIFF. SQ.", and the value for chi-square (χ^2), listed as "CHI SQUARE."

The remaining observations are grouped into one last interval, and statistics for this last interval, along with summary information, is printed at the bottom of the table. Information that relates to the Kolmogorov-Smirnov goodness-of-fit statistics also is printed out at the bottom of the table.

After the table giving frequency-based measures of goodness-of-fit has been printed, a second table giving variate-based measures of goodness-of-fit is printed. This table lists the cumulative frequency ("CUM. FREQ."), the number of standard deviations from the mean for the normally distributed random variable Y ("PROBITS"), the concentration predicted by the model for each interval ("PREDICTED X"), the actual concentration observed at the interval ("OBSERVED X"), the difference between predicted and observed concentrations ("DIFFERENCE"), and the square of this difference ("DIFF. SQUARED"). The difference values $d_j = (x_p - x_o)_j$ for each interval j are stored in the vector DIFF(100) so that statistics and other information can be computed from them in subsequent steps of the program. These values are required for the Engineering Error Test (EET) described in Chapter V.

The computations in this subroutine require the ability to calculate the CDF of the normal distribution $F_Y(y)$ and the inverse CDF of the normal distribution $F_Y^{-1}(y)$. These calculations are carried out, with reasonably high precision, using approximations suggested by Hastings.[75] The approximation for $F_Y(y)$ is contained in the function GAUCDF(X), while the approximation for $F_Y^{-1}(y)$ has been programed into the function GCDFI(Q). Both these functions are listed directly after the subroutine LN.

AUTCOR(L,N,STORE,ISTART,JUMP)

The subroutine AUTCOR computes the autocorrelation function

for the L observations stored in the vector STORE, which contains both the valid concentrations and the missing values (which are coded as DUMMY). ISTART specifies the starting point of the autocorrelation, and JUMP is a multiplication factor for the lags. In most applications, ISTART = 1, and JUMP = 1, causing the autocorrelation process to begin at the first hour of the year, with successive lags corresponding to integer multiples of 1 hour. That is, for lag M, there will be L - M pairs of observations used in the correlation computation, and each pair will be separated by M hours.

In the output for New York City (Appendix B), two autocorrelation results are given. In the first case, ISTART = 1, and JUMP = 1; in the second case, ISTART = 6 and JUMP = 6. The latter values cause the pairs of values for which the autocorrelation is computed to be separated by even multiples of 6 hours, beginning at the sixth hour (that is, 6, 12, 18, ..., etc.). The multiplier was included in the subroutine to enable the autocorrelation to be computed for lags that are relatively long in time.

In its present form, this subroutine computes the autocorrelation function for 1 to 75 lags. In the first case, lag 75 corresponds to 75 hours; in the second case, lag 75 corresponds to 6 x 75 = 450 hours. The line printer output shows the lags in hours ("LAG HOURS"), the lags in days ("LAG DAYS"), the number of pairs of values included in the computation of each correlation coefficient ("PAIRS"), the covariance ("COVAR-

IANCE"), and the correlation coefficient r ("CORRELATION COEFFICIENT"). A graph of the correlation coefficients, as a function of the lags, also is printed on the line printer to the right of the columnar output.

Because missing values cannot be included directly in the computation, any pair of values that has at least one missing value (that is, a value coded as DUMMY) is excluded from the computation. As a result, the number of pairs of values (listed under "PAIRS") varies from lag to lag. In this report, the results from the autocorrelation subroutine are not analyzed in detail. The subroutine was included in the program simply to establish whether the observations were serially correlated and the degree of correlation. It was found that most of the data sets exhibited a strong pattern of serial correlation, usually with a major periodicity of 24 hours and a minor periodicity of 12 hours, although the latter was not very strong or regular (see Appendices C and D). The magnitudes of the correlation coefficients appeared to be related to the arithmetic mean of the data set: the higher the mean, the greater the magnitudes of the correlation coefficients.

VII. SELECTION OF THE DATA

When this study was initially planned, we hoped to include as many different pollutants, locations, and years as possible. In particular, we hoped that the data sets ultimately selected would be as representative of the U.S. population as possible. We also hoped that that analytical work would be as thorough as possible, thereby answering the questions addressed in this study in a rigorous and definitive manner. These two competing goals -- the desire for an in-depth analysis and the desire to include as many data sets as possible -- required compromise, for it was not possible to carry out a detailed analysis on every data set in the Nation.

Prior to selecting CO as the pollutant of interest, we examined all candidate air pollutants and data sets available in the U.S. through the Environmental Protection Agency (EPA). The largest source of air quality data on a national basis is the Storage And Retrieval Of Aerometric Data (SAROAD) system maintained within EPA by the Office of Air Quality Planning and Standards (OAQPS) in Research Triangle Park, N.C. In selecting the pollutants, we limited the scope of the study to those pollutants for which National Ambient Air Quality Standards (NAAQS) had been promulgated: total suspended particulates (TSP), sulfur dioxide (SO_2), ozone (O_3), carbon monoxide, nitrogen dioxide (NO_2), and total hydrocarbons. In 1974, TSP was represented by

more stations in the SAROAD data bank than any other pollutant. When we looked into the characteristics of monitoring stations in the data bank, we found considerable variation from station to station. Only 100 stations could be found which measured at least 5 of the 6 NAAQS pollutants. The measurement method, "type" of station, and height of the intake probe above the ground varied. (The "type" denotes whether a station is "urban," "rural," etc.). The measurement methods used for SO_2, O_3, CO, and NO_2 varied from station to station. For example, in 1974, all California stations measured total oxidant instead of ozone. A few stations were mobile while most were stationary. Some were urban and some were rural. Probe heights, for those stations in which the probe height information was available, varied across the Nation from ground level to 100 feet (Figure 8).

Although we sought stations which measured as many NAAQS pollutants as possible, so few stations were available that measured at least 5 of the 6 NAAQS pollutants, and so much variation occurred among them, that we decided to relax the site selection criteria. Thus, we limited our objective to finding all stations which measured just two air pollutants -- O_3 and CO -- at the same location. These pollutants represent both ends of a spectrum with regard to chemical reactivity. Of the 245 stations that measured either O_3 or CO in 1974, only 50 stations could be found in SAROAD that measured both at the same location. In our judgment, this small number of stations could not be considered representative of the U.S. population, and again it was

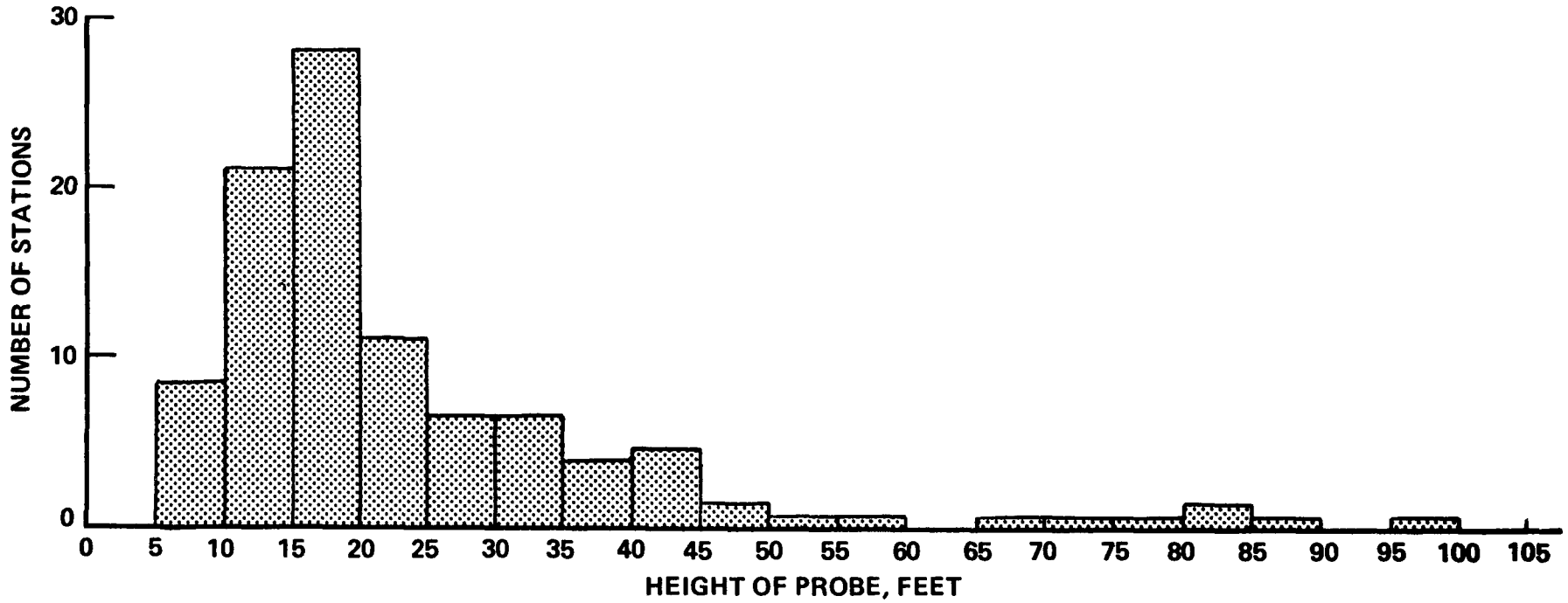

FIGURE 8. HEIGHTS OF CO PROBES AT AIR MONITORING STATIONS IN THE UNITED STATES, 1974.

necessary to limit our objectives and relax the selection criteria. Finally, we decided to select only one pollutant for our analysis: carbon monoxide. Because CO is a primary air pollutant and is relatively inert, it has highly favorable modeling characteristics; other reasons for selecting CO are as follows:

- It has no "surrogates" (i.e., no other pollutants are substituted for CO as an "indicator" of CO levels)
- Almost all air monitoring stations use the same measurement method -- nondispersive infrared absorption (NDIR)
- The measurement method has known validity and reliability
- CO ordinarily is monitored continuously, giving a large number of hourly observations per year at each station
- Considerable literature is available on CO sources, sinks, behavior, and effects
- CO's transport properties, because of its relative nonreactivity, are relatively easy to handle analytically

Once CO was selected as the pollutant of interest in this investigation, two general groups of data sets were taken from SAROAD:

1. Geographical Group - A number of stations in one year that were widely distributed geographically
2. Longitudinal Group - A number of different years at one station

The geographical group consisted of 11 stations, and the longitudinal group consisted of 8 years at one station, Washington, D.C.

GEOGRAPHICAL GROUP

To obtain a geographically diverse group of stations for

1974, we inventoried all CO data sets contained in SAROAD for which a full year (four quarters) of data was available. The result was 167 stations distributed throughout the Nation. We then ranked these 166 stations by their arithmetic means, from lowest to highest (Table 5). The annual arithmetic means ranged from 0.2 mg/m^3 in Yuba City, CA,* to 18.6 mg/m^3 in New York City, NY. Although four quarters were represented in each data set, there were sometimes many missing values due to calibration and instrument malfunction, and the percentage of observations available for the year ranged from 75% to 100%.

Because the 166-station list was too large to carry out detailed analyses of all stations, we sought to select, in a systematic manner, a representative subset of the stations on the list. We decided to use a stratified sample consiting of the lowest station, the highest station, and the deciles in between (that is, the 10-percentile, 20-percentile, and so on, to the 90 percentile).** Because the mean values had been rounded off to whole ppm numbers at the monitoring station, there were many repeated values on the list, and each decile contained more

---

*Examination of this station by the authors revealed that all values were exactly the same - 0.2 mg/m^3. EPA substitutes 0.2 mg/m^3 for all observations in the data bank that are zero values to reflect minimum detectable limits of the instrument. Thus, this data set consisted of nothing but zero values, and it was eliminated from consideration, giving the resulting 166 stations.

**The percentiles (and deciles) are calculated from the ranked data such that the Pth percentile is defined as the P(n+1)/100 order. For example, the 80th percentile (8th decile) is obtained by computing 80(166+1)/100 = 133.6, which then is rounded to 134. Then the 134th observation is obtained from the list, which is 4.3 mg/m^3, the value for Springfield, MA.

TABLE 5

SUMMARY OF CARBON MONOXIDE MONITORING DATA FROM U.S. CITIES IN 1974[a]

CITY, STATE	PROBE HEIGHT (feet)	DATA AVAIL. (%)	ANNUAL ARITH. MEAN (mg/m^3)	STANDARD DEVIATION (mg/m^3)	MAXIMUM (mg/m^3)
YUBA CITY, CA	5	98	0.2	0.00	0.2
BARSTOW, CA	25	80	0.7	0.91	9.2
MILWAUKEE, WI	--	90	0.8	0.90	12.6
CHESAPEAKE, VA	--	86	0.9	0.74	17.8
WEST PALM BEACH, FL	12	77	0.9	1.22	12.1
MILWAUKEE, WI	--	93	1.2	1.32	14.3
MILWAUKEE, WI	--	84	1.2	1.48	31.5
RACINE, WI	15	96	1.3	1.20	25.0
KANSAS CITY, KS	15	99	1.3	0.99	13.0
CAMARILLO, CA	26	91	1.3	1.15	11.5
KANSAS CITY, MO	30	93	1.3	1.42	11.6
MINNEAPOLIS, MN	16	91	1.3	1.23	20.7
TOPEKA, KS	12	97	1.4	1.17	20.0
SEVEN CORNERS, VA	4	96	1.4	1.16	16.1
ROCK HILL, SC	8	76	1.4	2.76	202.0
SCRANTON, PA	12	80	1.5	1.25	20.0
DETROIT, MI	12	86	1.5	1.81	20.9
ALLEN PARK, MI	12	78	1.6	1.40	14.4
NORFOLK, VA	3	94	1.6	1.20	21.8
MEMPHIS, TN	9	97	1.7	1.68	19.5
BAYONNE, NJ	12	88	1.7	1.11	20.0
STOCKTON, CA	42	95	1.7	1.48	18.4
MONTEREY, CA	25	92	1.7	1.05	18.4
LANCASTER, CA	18	98	1.7	1.35	20.7
FRESNO, CA	85	95	1.7	1.43	18.4
MODESTO, CA	88	85	1.8	1.26	20.7
SACRAMENTO, CA	78	98	1.8	1.49	21.8
REDDING, CA	30	99	1.8	0.86	13.8
SAN LUIS OBISPO, CA	40	95	1.8	1.10	17.2
WICHITA, KS	20	94	1.9	2.50	37.0
PITTSFIELD, MA	20	95	1.9	1.36	16.1
SALEM, OR	15	81	1.9	2.10	23.0
HAMPTON, VA	8	91	2.0	1.64	23.6
ALEXANDRIA, VA	30	98	2.0	5.33	461.4
KANSAS CITY, MO	20	92	2.0	1.97	29.9
CHICO, CA	40	99	2.0	1.87	26.4
INDIO, CA	40	98	2.1	1.46	19.5
ST PAUL, MN	32	91	2.1	1.49	18.4
DETROIT, MI	12	81	2.1	1.59	14.7
PENNS GROVE, NJ	15	90	2.1	1.99	24.3
VANCOUVER, WA	15	89	2.1	1.63	26.4

[a]Ranked in order of increasing arithmetic means.

TABLE 5 (Cont'd.)

SUMMARY OF CARBON MONOXIDE MONITORING DATA FROM U.S. CITIES IN 1974

CITY, STATE	PROBE HEIGHT (feet)	DATA AVAIL. (%)	ANNUAL ARITH. MEAN (mg/m^3)	STANDARD DEVIATION (mg/m^3)	MAXIMUM (mg/m^3)
ELIZABETH, NJ	12	89	2.2	1.88	26.2
ST PAUL, MN	13	95	2.2	2.29	28.7
CAMDEN, NJ	12	89	2.2	2.10	38.5
ROCHESTER, NY	--	79	2.2	1.87	23.5
LIVERMORE, CA	5	94	2.2	1.40	17.2
TUCSON, AZ	80	96	2.2	10.85	696.9
DENVER, CO	15	91	2.3	2.67	34.5
SANTA BARBARA, CA	40	96	2.3	2.51	31.0
SALINAS, CA	31	90	2.3	1.19	14.9
SCHENECTADY, NY	30	86	2.3	1.48	24.1
CAMDEN CO., NJ	45	83	2.3	1.13	12.5
NASHUA, NH	30	90	2.3	1.44	17.2
MANCHESTER, NH	35	91	2.4	1.57	18.4
RICHMOND, CA	20	97	2.4	1.36	12.6
FAIRBANKS, AK	10	89	2.4	3.07	35.6
SYRACUSE, NY	--	88	2.4	1.30	16.6
SEATTLE, WA	15	91	2.5	1.80	19.5
RICHMOND, VA	12	86	2.5	1.91	24.7
JACKSONVILLE, FL	11	79	2.5	2.31	30.5
WELBY, CO	15	95	2.5	2.89	27.6
LAS VEGAS, NV	35	85	2.5	2.82	57.5
OMAHA, NB	15	99	2.5	2.28	24.7
LAS CRUCES, NM	50	95	2.6	2.27	23.6
MINNEAPOLIS, MN	40	86	2.6	1.92	35.6
VISALIA, CA	26	97	2.6	1.34	12.6
BAKERSFIELD, CA	56	87	2.6	2.61	29.9
UPLAND, CA	25	97	2.7	1.92	23.0
SUNNYVALE, CA	38	95	2.7	1.88	23.0
PHILADELPHIA, PA	11	78	2.7	2.16	33.3
SYRACUSE, NY	14	89	2.8	1.30	13.2
PHILADELPHIA, PA	17	76	2.8	2.12	29.9
PITTSBURG, CA	25	99	2.8	1.21	13.8
FREMONT, CA	18	98	2.8	1.30	12.6
ANAHEIM, CA	15	94	2.8	2.03	19.5
SHIVLEY, KY	20	97	2.8	2.51	23.0
NEWPORT, KY	15	75	2.9	2.10	23.6
NEW YORK CITY, NY	--	78	2.9	1.31	9.2
NEW YORK CITY, NY	--	85	2.9	1.57	23.0
DES MOINES, IA	20	92	2.9	3.63	33.3
UTICA, NY	--	80	2.9	1.56	21.8
TACOMA, WA	12	94	2.9	2.52	32.2
RENSSELAER, NY	15	92	3.0	1.17	12.4

TABLE 5 (Cont'd.)

SUMMARY OF CARBON MONOXIDE MONITORING DATA FROM U.S. CITIES IN 1974

CITY, STATE	PROBE HEIGHT (feet)	DATA AVAIL. (%)	ANNUAL ARITH. MEAN (mg/m^3)	STANDARD DEVIATION (mg/m^3)	MAXIMUM (mg/m^3)
NEW YORK CITY, NY	65	81	3.0	1.66	14.9
FONTANA, CA	20	90	3.1	1.78	17.2
NAPA, CA	10	98	3.1	1.60	17.2
DENVER, CO	15	88	3.2	3.54	31.0
ARVADA, CO	15	86	3.2	3.42	58.6
SAN FRANCISCO, CA	104	97	3.2	1.95	18.4
TRENTON, NJ	15	88	3.2	2.12	24.6
MEMPHIS, TN	9	93	3.2	2.50	32.2
SAN RAFAEL, CA	26	99	3.3	1.81	20.7
ATLANTA, GA	10	95	3.3	2.14	32.1
REDWOOD CITY, CA	14	100	3.3	1.78	18.4
CONCORD, CA	21	99	3.3	1.50	20.7
LOS ANGELES, CA	25	99	3.3	3.69	40.2
COSTA MESA, CA	10	94	3.4	2.93	26.4
KANSAS CITY, KS	40	98	3.4	1.72	22.0
SAN BERNARDINO, CA	90	93	3.4	2.63	23.0
UPLAND, CA	30	100	3.4	1.99	13.8
PHILLIPSBURG, NJ	18	79	3.4	1.54	18.2
NEW YORK CITY, NY	--	94	3.4	1.60	14.9
CAMBRIDGE, MA	15	90	3.4	2.47	33.3
AZUSA, CA	7	96	3.4	1.73	17.2
PHOENIX, AZ	30	94	3.4	4.67	70.1
BURLINGTON, CA	36	100	3.5	1.79	18.4
SAN JOSE, CA	28	90	3.5	2.57	26.4
LAUDERDALE LAKES, FL	24	92	3.5	1.83	27.6
GREENWICH, CT	45	95	3.6	3.21	46.0
VALLEJO, CA	20	96	3.6	2.11	24.1
MAMARONECK, NY	35	93	3.6	2.12	27.3
BUFFALO, NY	5	92	3.6	1.78	27.6
SOMERVILLE, NJ	15	88	3.7	3.32	32.3
NIAGARA FALLS, NY	12	83	3.7	1.61	22.0
SANTA ROSA, CA	24	99	3.7	1.58	14.9
CHINO, CA	18	98	3.7	2.86	40.2
NORCO, CA	15	89	3.7	1.52	11.5
NEWHALL, CA	18	98	3.8	1.95	16.1
RUBIDOUX, CA	25	83	3.8	2.35	16.1
WHITTIER, CA	15	97	3.8	3.31	29.9
TOMS RIVER, NJ	15	88	3.8	3.25	26.6
OWENSBORO, KY	12	92	3.8	2.27	30.5
HACKENSACK, NJ	15	81	3.9	2.97	30.3
PAULSBORO, NJ	15	88	3.9	2.53	61.5
NEW YORK CITY, NY	--	84	3.9	2.27	20.7

TABLE 5 (Cont'd.)

SUMMARY OF CARBON MONOXIDE MONITORING DATA FROM U.S. CITIES IN 1974

CITY, STATE	PROBE HEIGHT (feet)	DATA AVAIL. (%)	ANNUAL ARITH. MEAN (mg/m^3)	STANDARD DEVIATION (mg/m^3)	MAXIMUM (mg/m^3)
ASBURY PARK, NJ	14	87	3.9	1.88	28.1
POMONA, CA	21	100	3.9	2.25	19.5
LONG BEACH, CA	21	98	3.9	3.00	27.6
LOS ANGELES CO, CA	24	98	4.1	3.80	35.6
NEW YORK CITY, NY	--	89	4.1	1.96	21.8
BEDFORD, MA	15	80	4.1	2.76	29.9
SEATTLE, WA	12	85	4.1	2.75	28.7
PATTERSON, NJ	15	89	4.2	3.55	40.5
PERTH AMBOY, NJ	15	84	4.2	3.25	37.9
SPRINGFIELD, MA	15	95	4.3	2.97	34.5
PORTLAND, OR	18	98	4.3	3.73	27.6
PORTLAND, OR	7	90	4.4	3.89	47.3
CAMDEN, NJ	15	90	4.4	2.20	27.5
LOUISVILLE, KY	35	99	4.5	4.01	120.9
FREEHOLD, NJ	15	89	4.6	3.87	37.2
HEMPSTEAD, NY	15	94	4.6	2.28	30.6
NEWARK, NJ	12	89	4.7	3.33	35.5
WORCHESTER, MA	13	98	4.7	2.82	25.3
BURLINGTON, NJ	15	89	4.7	3.09	28.6
NEW YORK CITY, NY	--	89	4.8	1.97	21.8
RIVERSIDE, CA	12	89	4.8	3.08	24.1
DENVER, CO	30	96	5.0	4.19	67.8
SEATTLE, WA	15	91	5.0	4.48	41.4
PORTLAND, OR	14	85	5.1	3.79	27.5
PASADENA, CA	18	99	5.1	3.99	37.9
LA HABRA, CA	6	97	5.1	3.92	37.9
SAGINAW, MI	10	86	5.2	13.51	276.6
FAIRBANKS, AK	10	78	5.3	4.54	37.4
LOS ANGELES, CA	72	98	5.5	4.19	34.5
ELIZABETH, NJ	15	88	5.6	4.73	45.9
JERSEY CITY, NJ	15	89	5.6	4.64	32.8
LYNWOOD, CA	22	98	5.9	5.23	46.0
DENVER, CO	9	75	6.2	5.24	80.5
LENNOX, CA	20	99	6.3	6.16	52.9
RENO, NV	9	82	6.3	34.05	810.7
MORRISTOWN, NJ	20	88	6.4	5.89	50.5
BURBANK, CA	15	99	6.9	4.22	42.5
BOSTON, MA	10	76	7.0	3.33	36.8
CHICAGO, IL	10	90	7.4	3.67	31.0
NEW YORK CITY, NY	5	88	10.7	5.01	51.7
NEW YORK CITY, NY	5	99	14.6	6.12	49.4
NEW YORK CITY, NY	--	95	18.6	7.90	65.5

than one station (Table 6). For example, at the sixth decile (60 percentile), there were 9 stations with the same arithmetic mean of 3.4 mg/m^3. It was appropriate, in our sampling plan, to select any of the stations that appeared within a particular decile, because they all had the same arithmetic means. This abundance of stations was desirable, because it permitted us to implement a second station selection criterion.

The second criterion sought to emphasize geographical diversity. It was implemented by arranging the 166 data sets in a different manner (Table 7). The left side of Table 7 shows the data sets for four major Air Quality Control Regions (AQCR's): Los Angeles, the New Jersey-New York-Connecticut area, San Francisco, and Philadelphia; the right side shows the remaining stations, listed alphabetically by state. A total of 73 of the 166 stations, or 44%, were in these four AQCR's, with 25 stations, or 15%, in the Los Angeles AQCR, and another 25 stations in the New Jersey-New York-Connecticut AQCR. With so great a proportion of the stations concentrated in just four AQCR's, it was apparent that the 166 data sets were not necessarily representative of the distribution of the U.S. population. Notice, for example, that the Chicago AQCR, which contains a substantial U.S. urban population, does not have a station among the 166 data sets and therefore is not represented on the original list at all.

In selecting the final group of 11 data sets, we decided to include at least one data set from each of the four AQCR's at the left side of Table 7. These four AQCR's (Los Angeles, New Jersey-New York-Connecticut, San Francisco, and Philadelphia)

TABLE 6

CO AIR MONITORING STATIONS IN 1974 AT EACH DECILE

Percentile (%)	City and State	Arithmetic Mean (mg/m^3)	Data Available (%)	Probe Height (feet)	Coefficient of Variation	Station Type*
Lowest	Barstow, CA	0.7	80	25	1.30	CC
10	Allen Park, MI	1.6	78	12	0.88	SM
	Norfolk, VA		94	3	0.75	CR
20	Hampton, VA	2.0	91	8	0.82	SI
	Alexandria, VA		98	30	2.67	CC
	Kansas City, MO		92	20	0.98	SC
	Chico, CA		99	40	0.94	CR
30	Denver, CO	2.3	91	15	1.16	SR
	Santa Barbara, CA		96	40	1.09	CC
	Salinas, CA		90	31	0.52	SC
	Schenectady, NY		86	30	0.64	CR
	Camden Co., NJ		83	45	0.49	RA
	Nashua, NH		90	30	0.63	CC
40	Upland, CA	2.7	97	25	0.71	CC
	Sunnyvale, CA		95	38	0.70	CC
	Philadelphia, PA		78	11	0.80	CM
50	Rensselaer, NY	3.0	92	15	0.39	SC
	New York City, NY	3.0	81	65	0.55	
	Fontana, CA	3.1	90	20	0.57	RI
	Napa, CA	3.1	98	10	0.52	CC
60	Costa Mesa, CA	3.4	94	10	0.86	SR
	Kansas City, KS		98	40	0.51	CC
	San Bernardino, CA		93	90	0.77	CC
	Upland, CA		100	30	0.59	CC
	Phillipsburg, NJ		79	18	0.45	CC
	Cambridge, MA		90	15	0.73	CI
	Azusa, CA		96	7	0.51	SI
	Phoenix, AZ		94	30	1.37	CC
70	Newhall, CA	3.8	98	18	0.51	SC
	Rubidoux, CA		83	25	0.62	RC
	Whittier, CA		97	15	0.86	CC
	Toms River, NJ		88	15	0.86	CC
	Owensboro, KY		92	12	0.60	CC
80	Springfield, MA	4.3	95	15	0.69	CI
	Portland, OR		98	18	0.87	CC
90	Portland, OR	5.1	85	14	0.74	CC
	Pasadena, CA		99	18	0.78	CC
	La Habra, CA		97	6	0.77	SR
Highest	New York City, NY	18.6	95		0.42	

*First Letter: C-Center City, S-Suburban, R-Rural;
Second Letter: C-Commercial, I-Industrial, M-Mobile, R-Residential

Table 7

CO AIR QUALITY DATA SETS IN 1974, GROUPED BY AIR QUALITY CONTROL REGION (AQCR)

CITY	STATE	HEIGHT FEET	PERCENT DATA	MEAN ug/m3	STANDARD DEVIATION	MAXIMUM ug/m3
METROPOLITAN LOS ANGELES AQCR						
AZUSA	CA	7	96	3.4	1.73	17.2
ANAHEIM	CA	15	94	2.8	2.03	19.5
CAMARILLO	CA	26	91	1.3	1.15	11.5
BURBANK	CA	15	99	6.9	4.22	42.5
WHITTIER	CA	15	97	3.8	3.31	29.9
UPLAND	CA	30	100	3.4	1.99	13.8
UPLAND	CA	25	97	2.7	1.92	23.0
SANTA BARBARA	CA	40	96	2.3	2.51	31.0
SAN BERNARDINO	CA	90	93	3.4	2.63	23.0
RUBIDOUX	CA	25	83	3.8	2.35	16.1
PIVERSIDE	CA	12	89	4.8	3.08	24.1
POMONA	CA	21	100	3.9	2.25	19.5
PASADENA	CA	18	99	5.1	3.99	37.9
NORCO	CA	15	89	3.7	1.52	11.5
NEWHALL	CA	18	98	3.8	1.95	16.1
LYNWOOD	CA	22	98	5.9	5.23	46.0
LOS ANGELES CO	CA	24	98	4.1	3.80	35.6
LOS ANGELES	CA	25	99	3.3	3.69	40.2
LOS ANGELES	CA	72	98	5.5	4.19	34.5
LONG BEACH	CA	21	98	3.9	3.00	27.6
LENNOX	CA	20	99	6.3	6.16	52.9
LA HABRA	CA	6	97	5.1	3.92	37.9
FONTANA	CA	20	90	3.1	1.78	17.2
COSTA MESA	CA	10	94	3.4	2.93	26.4
CHINO	CA	18	98	3.7	2.86	40.2
NEW JERSEY NEW YORK CONNECTICUT AQCR						
ELIZABETH	NJ	15	88	5.6	4.73	45.9
BAYONNE	NJ	12	88	1.7	1.11	20.0
ASBURY PARK	NJ	14	87	3.9	1.88	28.1
NEW YORK CITY	NY	5	88	10.7	5.01	51.7
NEW YORK CITY	NY	5	99	14.6	6.12	49.4
NEW YORK CITY	NY		94	3.4	1.60	14.9
NEW YORK CITY	NY		78	2.9	1.31	9.2
NEW YORK CITY	NY		95	18.6	7.90	65.5
NEW YORK CITY	NY		84	3.9	2.27	20.7
NEW YORK CITY	NY		85	2.9	1.57	23.0
NEW YORK CITY	NY		89	4.8	1.97	21.8
NEW YORK CITY	NY		89	4.1	1.96	21.8
NEW YORK CITY	NY	65	81	3.0	1.66	14.9
MAMARONECK	NY	35	93	3.6	2.12	27.3
HEMPSTEAD	NY	15	94	4.6	2.28	30.6
SOMERVILLE	NJ	15	88	3.7	3.32	32.3
PERTH AMBOY	NJ	15	84	4.2	3.25	37.9
PATTERSON	NJ	15	89	4.2	3.55	40.5
NEWARK	NJ	12	89	4.7	3.33	35.5
MORRISTOWN	NJ	20	88	6.4	5.89	50.5
JERSEY CITY	NJ	15	89	5.6	4.64	32.8
HACKENSACK	NJ	15	81	3.9	2.97	30.3
FREEHOLD	NJ	15	89	4.6	3.87	37.2
ELIZABETH	NJ	12	89	2.2	1.88	26.2
GREENWICH	CT	45	95	3.6	3.21	46.0
SAN FRANCISCO BAY AREA AQCR						
SAN JOSE	CA	28	90	3.5	2.57	26.4
SAN FRANCISCO	CA	104	97	3.2	1.95	18.4
SANTA ROSA	CA	24	99	3.7	1.58	14.9
SAN RAFAEL	CA	26	99	3.3	1.81	20.7
VALLEJO	CA	20	96	3.6	2.11	24.1
SUNNYVALE	CA	38	95	2.7	1.88	23.0
CONCORD	CA	21	99	3.3	1.50	20.7
FREMONT	CA	18	98	2.8	1.30	12.6
LIVERMORE	CA	5	94	2.2	1.40	17.2
RICHMOND	CA	20	97	2.4	1.36	12.6
REDWOOD CITY	CA	14	100	3.3	1.78	18.4
NAPA	CA	10	98	3.1	1.60	17.2
PITTSBURG	CA	25	99	2.8	1.21	13.8
BURLINGTON	CA	36	100	3.5	1.79	18.4
METROPOLITAN PHILDELPHIA AQCR						
PAULSBORO	NJ	15	88	3.9	2.53	61.5
PENNS GROVE	NJ	15	90	2.1	1.99	24.3
TRENTON	NJ	15	88	3.2	2.12	24.6
BURLINGTON	NJ	15	89	4.7	3.09	28.6
CAMDEN CO.	NJ	45	83	2.3	1.13	12.5
CAMDEN	NJ	15	90	4.4	2.20	27.5
CAMDEN	NJ	12	89	2.2	2.10	38.5
PHILADELPHIA	PA	17	76	2.8	2.12	29.9
PHILADELPHIA	PA	11	78	2.7	2.16	33.3
OTHER AQCRS						
FAIRBANKS	AK	10	89	2.4	3.07	35.6
FAIRBANKS	AK	10	78	5.3	4.54	37.4
LAS VEGAS	NV	35	85	2.5	2.82	57.5
PHOENIX	AZ	30	94	3.4	4.67	70.1
TUCSON	AZ	80	96	2.2	10.85	696.9
MEMPHIS	TN	9	93	3.2	2.50	32.2
MEMPHIS	TN	9	97	1.7	1.68	19.5
MONTEREY	CA	25	92	1.7	1.05	18.4
SALINAS	CA	31	90	2.3	1.19	14.9
SACRAMENTO	CA	78	98	1.8	1.49	21.8
REDDING	CA	30	99	1.8	0.86	13.8
CHICO	CA	40	99	2.0	1.87	26.4
BAKERSFIELD	CA	56	87	2.6	2.61	29.9
MODESTO	CA	88	85	1.8	1.26	20.7
FRESNO	CA	85	95	1.7	1.43	18.4
STOCKTON	CA	42	95	1.7	1.48	18.4
VISALIA	CA	26	97	2.6	1.34	12.6
SAN LUIS OBISPO	CA	40	95	1.8	1.10	17.2
INDIO	CA	40	98	2.1	1.46	19.5
LANCASTER	CA	18	98	1.7	1.35	20.7
BARSTOW	CA	25	80	0.7	0.91	9.2
ARVADA	CO	15	86	3.2	3.42	58.6
WELBY	CO	15	95	2.5	2.89	27.6
DENVER	CO	15	91	2.3	2.67	34.5
DENVER	CO	30	96	5.0	4.19	67.8
DENVER	CO	15	88	3.2	3.54	31.0
DENVER	CO	9	75	6.2	5.24	80.5
SPRINGFIELD	MA	15	95	4.3	2.97	34.5
ALEXANDRIA	VA	30	98	2.0	5.33	461.4
SEVEN CORNERS	VA	4	96	1.4	1.16	16.1
JACKSONVILLE	FL	11	79	2.5	2.31	30.5
LAUDERDALE LAKES	FL	24	92	3.5	1.83	27.6
WEST PALM BEACH	FL	12	77	0.9	1.22	12.1
ATLANTA	GA	10	95	3.3	2.14	32.1
CHICAGO	IL	10	90	7.4	3.67	31.0
OWENSBORO	KY	12	92	3.8	2.27	30.5
SHIVLEY	KY	20	97	2.8	2.51	23.0
LOUISVILLE	KY	35	99	4.5	4.01	120.9
NEWPORT	KY	15	75	2.9	2.10	23.6
OMAHA	NB	15	99	2.5	2.28	24.7
DES MOINES	IA	20	92	2.9	3.63	33.3
KANSAS CITY	KS	15	99	1.3	0.99	13.0
KANSAS CITY	KS	40	98	3.4	1.72	22.0
KANSAS CITY	MO	20	92	2.0	1.97	29.9
KANSAS CITY	MO	30	93	1.3	1.42	11.6
TOPEKA	KS	12	97	1.4	1.17	20.0
WICHITA	KS	20	94	1.9	2.50	37.0
PITTSFIELD	MA	20	95	1.9	1.36	16.1
WORCHESTER	MA	13	98	4.7	2.82	25.3
BEDFORD	MA	15	80	4.1	2.76	29.9
CAMBRIDGE	MA	15	90	3.4	2.47	33.3
BOSTON	MA	10	76	7.0	3.33	36.8
MANCHESTER	NH	35	91	2.4	1.57	18.4
NASHUA	NH	30	90	2.3	1.44	17.2
SAGINAW	MI	10	86	5.2	13.51	276.6
DETROIT	MI	12	81	2.1	1.59	14.7
DETROIT	MI	12	86	1.5	1.81	20.9
ALLEN PARK	MI	12	78	1.6	1.40	14.4
MINNEAPOLIS	MN	16	91	1.3	1.23	20.7
MINNEAPOLIS	MN	40	86	2.6	1.92	35.6
ST PAUL	MN	13	95	2.2	2.29	28.7
ST PAUL	MN	32	91	2.1	1.49	18.4
RENO	NV	9	82	6.3	34.05	810.7
TOMS RIVER	NJ	15	88	3.8	3.25	26.6
PHILLIPSBURG	NJ	18	79	3.4	1.54	18.2
SCRANTON	PA	12	80	1.5	1.25	20.0
LAS CRUCES	NM	50	95	2.6	2.27	23.6
UTICA	NY		80	2.9	1.56	21.8
SYRACUSE	NY		88	2.4	1.30	16.6
SYRACUSE	NY	14	89	2.8	1.30	13.2
ROCHESTER	NY		79	2.2	1.87	23.5
SCHENECTADY	NY	30	86	2.3	1.48	24.1
RENSSELAER	NY	15	92	3.0	1.17	12.4
NIAGARA FALLS	NY	12	83	3.7	1.61	22.0
BUFFALO	NY	5	92	3.6	1.78	27.6
ROCK HILL	SC	8	76	1.4	2.76	202.0
PORTLAND	OR	18	98	4.3	3.73	27.6
SALEM	OR	15	81	1.9	2.10	23.0
PORTLAND	OR	7	90	4.4	3.89	47.3
PORTLAND	OR	14	85	5.1	3.79	27.5
VANCOUVER	WA	15	89	2.1	1.63	26.4
NORFOLK	VA	3	94	1.6	1.20	21.8
HAMPTON	VA	8	91	2.0	1.64	23.6
CHESAPEAKE	VA		86	0.9	0.74	17.8
RICHMOND	VA	12	86	2.5	1.91	24.7
SEATTLE	WA	15	91	5.0	4.48	41.4
TACOMA	WA	12	94	2.9	2.52	32.2
SEATTLE	WA	12	85	4.1	2.75	28.7
SEATTLE	WA	15	91	2.5	1.80	19.5
RACINE	WI	15	96	1.3	1.20	25.0
MILWAUKEE	WI		90	0.8	0.90	12.6
MILWAUKEE	WI		93	1.2	1.32	14.3
MILWAUKEE	WI		84	1.2	1.48	31.5

represent about 60 million people, or 30% of the U.S. population. It was decided to distribute the remaining 7 stations among the smaller AQCR's. The final listing of the 11 data sets (Table 8) therefore reflects the use of three selection criteria:

- The percentage of observations is as high as possible (that is, as high above the 75% minimum as possible.)
- The stations are a stratified sample (stratified by arithmetic mean) from the universe of 166 CO stations in SAROAD in 1974.
- To the extent possible, the stations are as geographically diverse as possible to avoid clustering in any one AQCR.

By emphasizing geographical diversity, we felt that the group of data sets would better reflect the variety of meteorological conditions and source emissions encountered in different parts of the U.S. than if the stations were concentrated in a few AQCR's.

Careful examination of the original list of 166 stations (Table 5) reveals many pecularities and raises serious quality control questions. For example, Alexandria, VA, one of the data sets included in the geographical data group, has a reported maximum of 461.4 mg/m^3 (hourly average). However, measurements of urban air quality by NDIR seldom gives values above 80 mg/m^3 and almost never gives values above 110 mg/m^3. Thus, we were suspicious about this value, and subsequent discussions with air monitoring field personnel revealed that it was an error, probably caused by keypunching and transfer of the data from the local air pollution control agencies to EPA. Similar problems apparently

TABLE 8

FINAL LIST OF 11 CO DATA SETS SELECTED FROM SAROAD, 1974, TO GIVE GEOGRAPHICAL VARIATION
(Geographical Data Group)

City and State	Street Address	Saroad Site Code	Percentile (%)	Arithmetic Mean (ppm)	Arithmetic Mean (mg/m^3)
Barstow, CA	200 East Bunea Vista	05 0580 001 I01	Lowest	0.44	0.51
Norfolk, VA	P.O. Garage at Brambleton Avenue	48 2140 010 F01	10	1.38	1.58
Alexandria, VA	517 North Asaph Street	48 0080 009 H01	20	1.64	1.88
Denver, CO	2005 South Huron Street	06 0580 011 F01	30	1.91	2.20
Philadelphia, PA	2031 Race Street	39 7140 002 A10	40	2.36	2.71
Napa, CA	2552 Jefferson Avenue	05 5000 003 I01	50	2.73	3.14
Phoenix, AZ	1825 East Roosevelt Street	03 0600 002 G01	60	2.94	3.38
Newhall, CA	24811 San Fernando Road	05 5120 001 I01	70	3.33	3.83
Springfield, MA	East Columbus Avenue	22 2160 005 F01	80	3.71	4.26
Pasadena, CA	1196 East Walnut Street	05 5760 002 A01	90	4.43	5.09
New York City, NY	59th Street and Bridge Pl.	33 4680 023 H01	Highest	16.21	18.64

occurred in the data set from Reno, NV, which contained a maximum value of 810.7 mg/m^3, and the data set for Tucson, AZ, which contained a maximum value of 696.9 mg/m^3. Unusual and doubtful maxima also appeared in the data sets for Saginaw, MI (276.6 mg/m^3) and Rock Hill, SC (202.0 mg/m^3). Unfortunately, for the lower maxima it is difficult to determine with certainty whether the value is a valid outlier or a keypunch error.

Overall, the existence of these peculiar and erroneous values casts doubt on all the values comprising the data sets, and it is evident that EPA should undertake a quality assurance data checking program to compare the observations stored in the data bank with those originally submitted by state and local air pollution control agencies. One simple test, which could be carried out at relatively low cost, would be to compare the minimum, maximum, arithmetic mean, and arithmetic standard deviation for the data set with the same statistics computed by the agency submitting the data. Then inconsistencies could be resolved by changing any incorrect hourly values on an observation-by-observation basis. Hopefully, the offices within EPA that maintain the SAROAD data bank have taken these steps since the preparation of this study. For our purposes, the erroneous maximum of 461.4 mg/m^3 was handled by substituting a missing value code, thereby treating it as a missing value and giving 8606 observations instead of 8607 observations for Alexandria.

LONGITUDINAL GROUP

The longitudinal data group consisted of eight years of

observations at the Continuous Air Monitoring Project (CAMP) station in Washington, DC. This group covered the time period 1966-1974, but data were not available for one year, 1967 (see Table 30 in Chapter IX, page 173). In years 1972 and 1973, the data sets were not complete and contained only 4947 and 4269 observations, respectively. In 1974, only the first quarter of data was available, and the data set consisted of only 1848 observations. Because these observations were not randomly distributed throughout the year, the mean of these data is not a valid estimate of the annual mean, and the procedure described by Hunt[66] for computing the precision of the calculated mean cannot be used. Consequently, the 1974 data set was not included in the analysis, and the results for the 1972 and 1973 data sets are not really comparable with the results for the other data sets in this study.

VIII. RESULTS FOR THE GEOGRAPHICAL DATA GROUP

Previous chapters have discussed the theory behind the methodology used in this investigation, details of the methodology itself, the computer program developed to incorporate this methodology, and the selection of the data used in this investigation. This chapter presents the resulting analyses of the 11 CO hourly data sets which comprise the geographical data group. As indicated in Chapter VII (pages 98-109), these data sets were selected to represent a geographically diverse cross section of U.S. CO data sets for a given year, 1974. In this chapter, we first discuss the results for each of the 11 cities; then the results for all cities grouped together are considered.

FINDINGS FOR INDIVIDUAL CITIES

The 11 cities comprising the geographical data group consisted of stations with the lowest arithmetic mean, highest arithmetic mean, and values at each decile selected from the 166-station list in the manner described in the previous chapter (see Table 8). The basic statistics (Table 9) indicate that the arithmetic mean ranges from 0.442 ppm to 16.207 ppm, and the arithmetic standard deviation ranges from 0.859 ppm to 6.869 ppm. The coefficient of skewness $\sqrt{b_1}$ ranges from 0.805 to 4.977, and the coefficient of kurtosis b_2 ranges from 4.17 to 51.47. The following sections discuss the results for each city in the order of their arithmetic means, from lowest (Barstow) to highest (New

TABLE 9

BASIC STATISTICS FOR THE 11 U. S. CO DATA SETS SELECTED FROM SAROAD, 1974 (GEOGRAPHICAL DATA GROUP)

CITY AND STATE	PERCENTILE (%)	NUMBER OF OBSERVATIONS	MEAN (ppm)	STANDARD DEVIATION (ppm)	COEFFICIENT OF SKEWNESS, $\sqrt{b_1}$	COEFFICIENT OF KURTOSIS, b_2	LOWEST VALUE (ppm)	HIGHEST VALUE (ppm)
Barstow, CA	Lowest	6992	0.442	0.859	2.934	14.71	0.0	8.0
Norfolk, VA	10	8241	1.377	1.061	4.977	51.47	0.0	19.0
Alexandria, VA	20	8606	1.639	1.714	3.332	19.95	0.0	21.0
Denver, CO	30	7952	1.913	2.361	3.187	20.11	0.0	30.0
Philadelphia, PA	40	6836	2.355	1.895	3.736	32.87	0.0	29.0
Napa, CA	50	8558	2.731	1.394	2.100	11.26	0.0	15.0
Phoenix, AZ	60	9246	2.937	4.102	2.859	16.82	0.0	61.0
Newhall, CA	70	8623	3.330	1.699	1.256	6.11	1.0	14.0
Springfield, MA	80	8339	3.706	2.597	2.134	12.95	0.0	30.0
Pasadena, CA	90	8632	4.428	3.466	2.604	12.34	1.0	33.0
New York City, NY	Highest	8364	16.207	6.869	0.805	4.17	2.0	57.0

York City).

Barstow, CA

When the cumulative frequencies of the observations for Barstow are plotted on logarithmic probability paper (Figure 9), a distinct downward curvature is evident. That is, the data points (plotted as small circles) describe the locus of a curve that is concave downward. The data set for Barstow is somewhat unusual in that over 70% (actually, 70.08%; see Appendix C) of the observations are zero. All the zero values have been grouped into the first interval and plotted at the point whose coordinates are 70.08% (plotting position of 70.07%) and 0.5 ppm.

Because the LN2 probability model appears on logarithmic probability paper as a straight line (see Chapter III), each of the four parameter calculation techniques gives a straight line in Figure 9. Each line differs in slope and vertical intercept. Because none of the straight lines passes through all the points, none of the parameter calculation techniques gives a line that unequivocally describes the observations for Barstow. The LN2 lines obtained by the MLE optimization technique and the method of moments both give serious overpredictions at the higher concentrations. For example, at the concentration of 7.5 ppm (interval between 6.5 ppm and 7.5 ppm; plotting position of 99.957%), the model with parameters calculated by MLE optimization predicts 20.5 ppm, an overestimation of 273%, and the model with parameters calculated by the method of moments predicts 19 ppm, an overestimation of 253%.

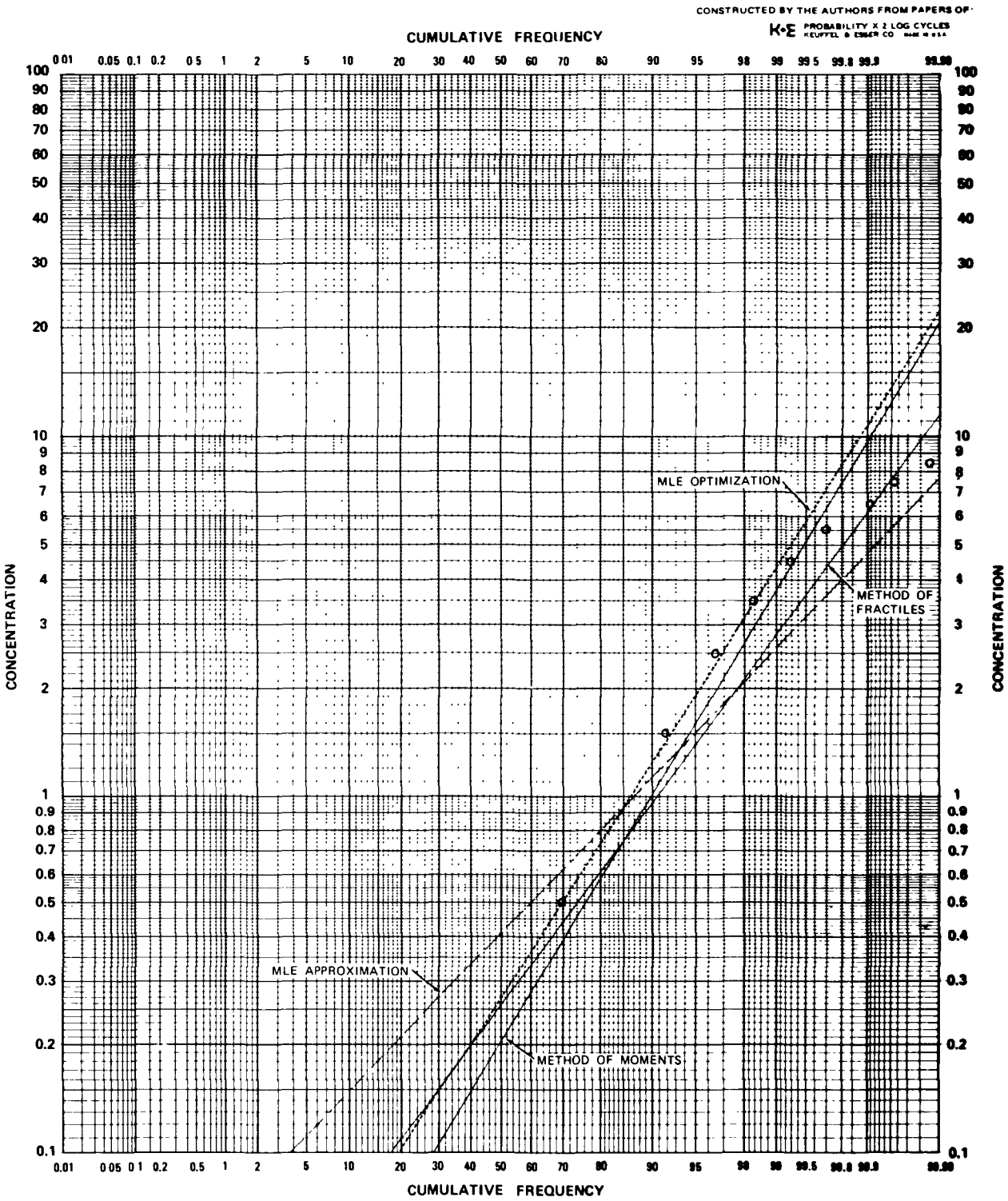

FIGURE 9. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR BARSTOW, CA, 1974 ($n = 8992$).

When the results for the two sets of tests for goodness-of-fit -- the frequency-based tests and the variate-based tests -- are compared, a great contrast is observed. The frequency-based measures indicate that the best fit is obtained with parameters estimated by MLE optimization, while the variate-based Engineering Error Test (EET) does not show that any one method is clearly superior.

Table 10

Frequency-based Measures of Goodness-of-fit for Barstow, CA

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	0.202	3.495	6	181	228	0.065	-6272
Fractiles	0.259	2.765	4	120	370	0.040	-6299
MLE Approx.	0.409	2.210	4	439	912	0.101	-6551
MLE Opt.	0.271	3.296	7	27[1]	36[1]	0.005[1]	-6188[1]

[1]Indicates best result in each column.

For the frequency-based measures (Table 10), the MLE optimization technique gives the best results for all tests. The first two columns list the median ($\hat{m}$) and the geometric standard deviation (s_g), which correspond to the 50-percentile and the slope, respectively, of the straight lines plotted in Figure 10. The third column lists the number of degrees of freedom (d.f.), and the last four columns list the goodness-of-fit tests, beginning with the root-mean-square (RMS) of the difference between the number of observations predicted and observed in each interval, chi-square (χ^2), the maximum difference between the CDF and

the observed cumulative frequency (K-S max), and the log-likelihood function (L*). In each column, a footnote indicates the "best" measure of goodness-of-fit.

Table 11

Variate-based (EET) Measures of Goodness-of-fit for Barstow, CA (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	7	0.546[1]	1.787	-0.56	4.41
Fractiles	7	-0.695	0.581[1]	-1.32	0.13
MLE Approx.	7	-0.991	0.713	-1.79	0.12[1]
MLE Opt.	7	1.158	2.127	-0.15[1]	5.67

[1]Indicates best result in each column.

Despite the fact that MLE optimization causes the model to overestimate concentrations at the higher end of the range, the statistical, or frequency-based, tests of goodness-of-fit indicate that it gives the best fit and greatly surpasses the fit of all the other methods of calculating the LN2 model's parameters for the Barstow data set. By contrast, the results for the EET measures are not so clearcut. Table 11 shows the average difference ($\bar{d}$) between the predicted and observed concentrations, the standard deviation (s_d) of the differences, the minimum difference (d_{min}), and the maximum difference (d_{max}). The smallest values indicate the best fit. Both the method of moments and the MLE optimization approach cause the model to overpredict CO concentrations, with MLE optimization causing the largest overprediction. One reason for the overprediction is the tendency

for the locus of points which comprises the Barstow data to show downward curvature. The method of fractiles gives the smallest standard deviation of differences, largely because it passes directly through the areas where curvature is greatest. It fits the data at the 99.9 cumulative frequency which, for Barstow, corresponds to a relatively high concentration level (6.5 ppm). By contrast, the lines for MLE optimization and the method of moments pass well above the data points at these higher levels (Figure 9), largely because their slopes are determined by the bulk of the data at the lower concentrations.

Overall, we can conclude that much of the differences among the four parameter calculation techniques for Barstow can be attributed to our difficulty in drawing straight lines through points that are better described by a curve. That is, the poor goodness-of-fit that is evident for Barstow may be attributed largely to inadequacies of the LN2 model, which plots as a straight line, when applied to data that show downward curvature. This result suggests that the LN3 or LN4 models, which are covered in other reports,[2,43-46] may be more suitable for this application.

The results for Barstow also illustrate a basic difference in the parameter calculation techniques. The method of MLE optimization selects model parameters in a manner that weights heavily those intervals which contain a large number of observations. By contrast, the method of fractiles, as suggested by Larsen,[24] weights the higher concentrations at the tails more heavily. Because the tails are likely to exhibit considerable random fluctuation, the

method of fractiles appears to fit a particular data set relatively well, when, in fact, it may be responding more to randomness in the data than to underlying distributional properties. If the underlying distribution generating the data is LN2 in form, parameters calculated by the method of fractiles are likely to have the greatest variance of the four parameter calculation techniques, and hence the method of fractiles is a poor approach for estimating parameters of the underlying distribution. It fits the idiosyncracies at the tails of each data set quite well, reflecting the randomness contained in a given realization, but it does not respond to the underlying process generating the data as well as the other approaches. Conversely, the method of MLE optimization will fit the underlying distribution best, giving parameters with the least variance, but it may fit a given realization quite poorly. Of course, we cannot state with certainty that the underlying distribution has an LN2 form. With the Barstow data, the statistical, or frequency-based, measures of goodness-of-fit seem to suggest that the MLE approach is the best one to use for calculating LN2 parameters, while the EET measures do not favor one approach over another. The present authors feel more comfortable using the MLE optimization technique for Barstow than the other approaches, although we realize that strong arguments also can be made for using higher order models.

Norfolk, VA

In the case of Norfolk (Figure 10), we see that the observations do not show consistent downward curvature, but, rather,

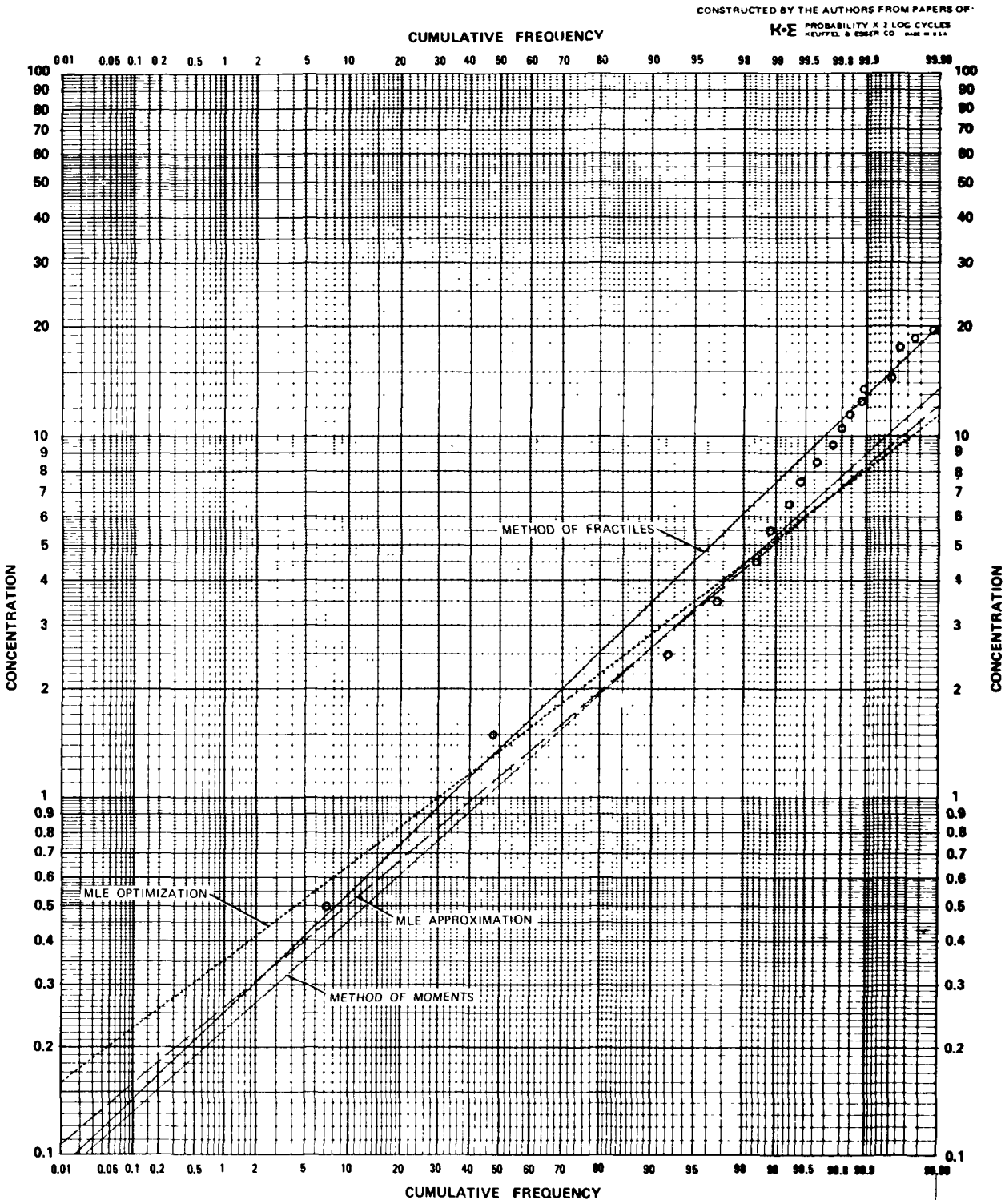

FIGURE 10 LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR NORFOLK, VA, 1974 (n = 8241).

describe a slight zig-zag pattern with reversals of curvature. The four straight lines representing the models are closer to each other than they were for Barstow, and the lines for the MLE approximation and the method of moments are quite close together. In contrast with the Barstow results, these two lines, as well as the one for MLE optimization, cause the model to underestimate CO concentrations at the higher levels. As discussed above, the method of fractiles, by design, gives a line that lies very close to the observed values at the higher concentrations (in this case, 10 to 20 ppm). However, it tends to overpredict concentrations in the range from 2 to 10 ppm for this data set.

When we examine the frequency-based measures of goodness-of-fit for Norfolk (Table 12), we find that the MLE optimization approach gives the best result in three of the four measures of goodness-of-fit, while the method of fractiles gives the best

Table 12

Frequency-based Measures of Goodness-of-fit for Norfolk, VA

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	1.091	1.978	8	682	2706	0.194	-11230
Fractiles	1.364	2.074	10	484[1]	1953	0.126	-11063
MLE Approx.	1.155	1.886	7	681	2183	0.174	-11020
MLE Opt.	1.347	1.797	7	525	1465[1]	0.086[1]	-10785[1]

[1]Indicates best result in each column.

result (i.e., the lowest value) in the root-mean-square (RMS) measure of goodness-of-fit. Unfortunately, the RMS values cannot be compared in this situation, because the RMS value is very sensitive to the number of intervals used in the computation, which varies from method to method, as reflected by the differing number of degrees of freedom. The difference between the predicted and observed number of observations, $(n_p - n_o)_j^2$ for the jth interval, appears in the numerator of both the RMS and chi-square computations (see Chapter V). For the method of fractiles, the sum of the numerators for all intervals gives 3,047,590; for the method of MLE optimization, the sum of the numerators gives 2,752,804. Thus, if the sum of the square of the difference between the predicted and observed number of values is used as a measure of goodness-of-fit instead of the RMS values, the MLE optimization approach again gives the best results. As in the case of Barstow, the frequency-based measures indicate that the MLE optimization approach gives the best performance for Norfolk.

The results for the variate-based EET measures of goodness-of-fit present a very different picture (Table 13). In three of the four categories, the method of fractiles gives the best performance. On the average, all approaches cause the model to underpredict CO concentrations, with the greatest underprediction occurring for the MLE approximation. In Figure 10, it can be seen that the MLE optimization approach fits the lowest five concentrations best, since these values account for the bulk of the observations. By comparison, the line for the

Table 13
Variate-based (EET) Measures of Goodness-of-fit for Norfolk, VA
(ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	14	-1.412	1.585	-4.45	0.38
Fractiles	14	0.896[1]	0.770[1]	-0.42[1]	1.87
MLE Approx.	14	-1.781	1.863	-5.25	0.35[1]
MLE Opt.	14	-1.625	1.929	-5.22	0.60

[1]Indicates best result in each column.

method of fractiles is closest to the two lowest concentrations (0.5 ppm and 1.5 ppm) and also is closest to the highest 10 concentrations which produce the large differences (d_{min}) for the other three approaches. Like Barstow, the LN2 model does not appear to fit the Norfolk data set very well for any of the parameter calculation approaches, and the LN2 model may not be a suitable representation of the underlying distribution.

Alexandria, VA

The CO data set for Alexandria, when plotted on logarithmic-probability paper (Figure 11), resembles the data for Norfolk more than it does the data for Barstow. The zig-zag pattern of the data points is similar to what one might expect as a result of sampling from a lognormal distribution. The lines obtained by the method of moments and the method of fractiles both fit the observations quite well at the higher concentrations. However, the method of moments line diverges from the data points at the lowest concentra-

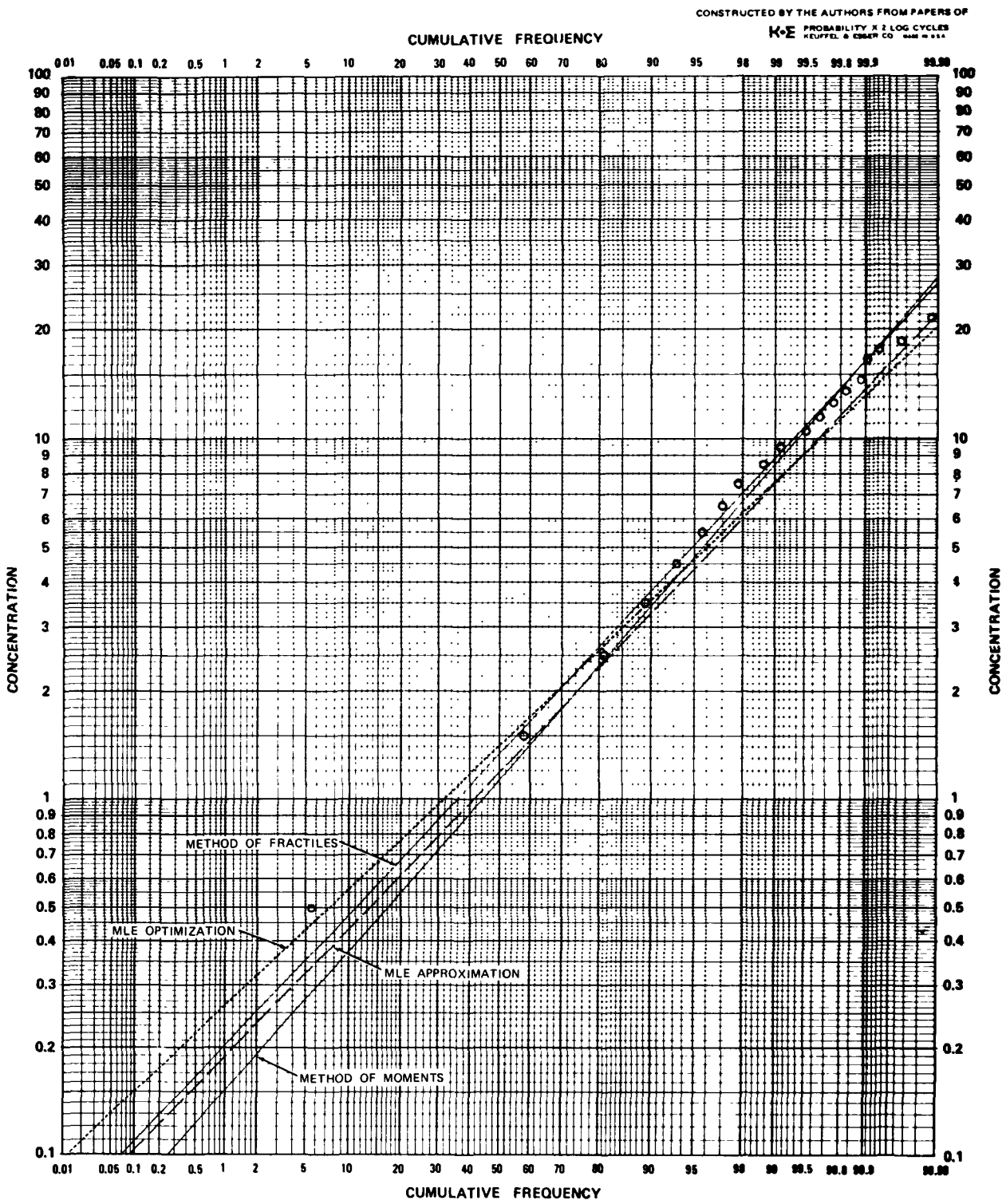

FIGURE 11. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR ALEXANDRIA, VA, 1974 (n = 8606).

Table 14

Frequency-based Measures of Goodness-of-fit for Alexandria, VA

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	1.133	2.362	12	306	872	0.114	-13265
Fractiles	1.308	2.249	12	221	467	0.061	-12986
MLE Approx.	1.184	2.195	11	215	628	0.080	-13072
MLE Opt.	1.387	2.055	10	165[1]	338[1]	0.038[1]	-12879[1]

[1]Indicates best result in each column.

tions while the method of fractiles line appears to fit reasonably well at the lower concentrations. We would expect a good fit to exhibit both positive and negative deviations about the line. By this criterion, the MLE approximation fails completely because it underestimates every data point. Inspection of the MLE optimization line suggests that it gives a relatively good overall fit,

Table 15

Variate-based (EET) Measures of Goodness-of-fit for Alexandria, VA (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	18	-0.016[1]	0.660	-1.01	1.45
Fractiles	18	0.103	0.507[1]	-0.72[1]	1.33
MLE Approx.	18	-1.271	0.740	-2.40	-0.11
MLE Opt.	18	-1.384	1.039	-3.12	+0.13[1]

[1]Indicates best result in each column.

showing a good relationship to the data at the middle range and slight underestimation of the higher concentrations.

The results for the frequency-based measures of goodness-of-fit (Table 14) bear out these findings. The MLE optimization approach gives the best fit by all goodness-of-fit measures. The method of fractiles gives the second-best fit, but it is well behind MLE optimization. The results for the variate-based goodness-of-fit tests (Table 15) present a different picture. For the method of moments, the average deviation ($\bar{d}$) is nearly zero, and the method of fractles also gives an average deviation that is close to zero. Because the method of fractiles also has the smallest standard deviation (s_d) and the smallest range (d_{max} - d_{min}), the variate-based measures indicate that it gives the best fit. The MLE approximation technique always underpredicts the concentration levels; its maximum deviation (d_{max} = -0.11 ppm) is still negative, so the fit is relatively poor, as was noted in the discussion of Figure 11. According to the variate-based goodness-of-fit measures, the MLE optimization approach gives the poorest fit; it has the lowest average deviation, the largest standard deviation, the most undershoot (largest negative d_{min}), and the largest range. Thus, the frequency-based measures suggest that the best fit is obtained by the MLE optimization approach, while the variate-based measures suggest that the best fit is obtained by the method of fractiles. One way to resolve this paradox is to recall that the two sets of goodness-of-fit measures are designed to test for different things. The frequency-based measures are for the purpose of testing whether the hypothe-

sis that the underlying distribution is lognormal can be rejected. The variate-based measures are for determining how well the model represents a particular data set, or realization, in practical terms: the difference between the predicted and observed concentration at each interval. If a hypothesis test for lognormality were applied to the frequency-based measures, the lognormal assumption would be rejected with a high degree of certainty. For example, to reject the lognormal hypothesis at a confidence level of P = 0.995 with 10 degrees of freedom, the chi-square value must exceed $\chi^2 = 25.2$. For the Alexandria data, the MLE optimization approach gives $\chi^2 = 165$, indicating that the underlying distribution is not lognormal. Of course, statistical independence does not necessarily apply to these data, thereby invalidating the hypothesis testing procedure.

Denver, CO

The observations for Denver (Figure 12) exhibit a pattern of downward curvature that is similar to that for Barstow. Like the Barstow data, the two MLE techniques (MLE optimization and MLE approximation) fit the observations reasonably well for cumulative frequencies below 90%, but they cause serious overpredictions for the higher levels, sometimes overpredicting the concentrations by a factor of two or more. The lines for the method of moments and the method of fractiles, in contrast, fit the data points quite well at the higher concentrations. Here, the two lines lie relatively close together, which differs from the Barstow results, where the method of moments line differed in

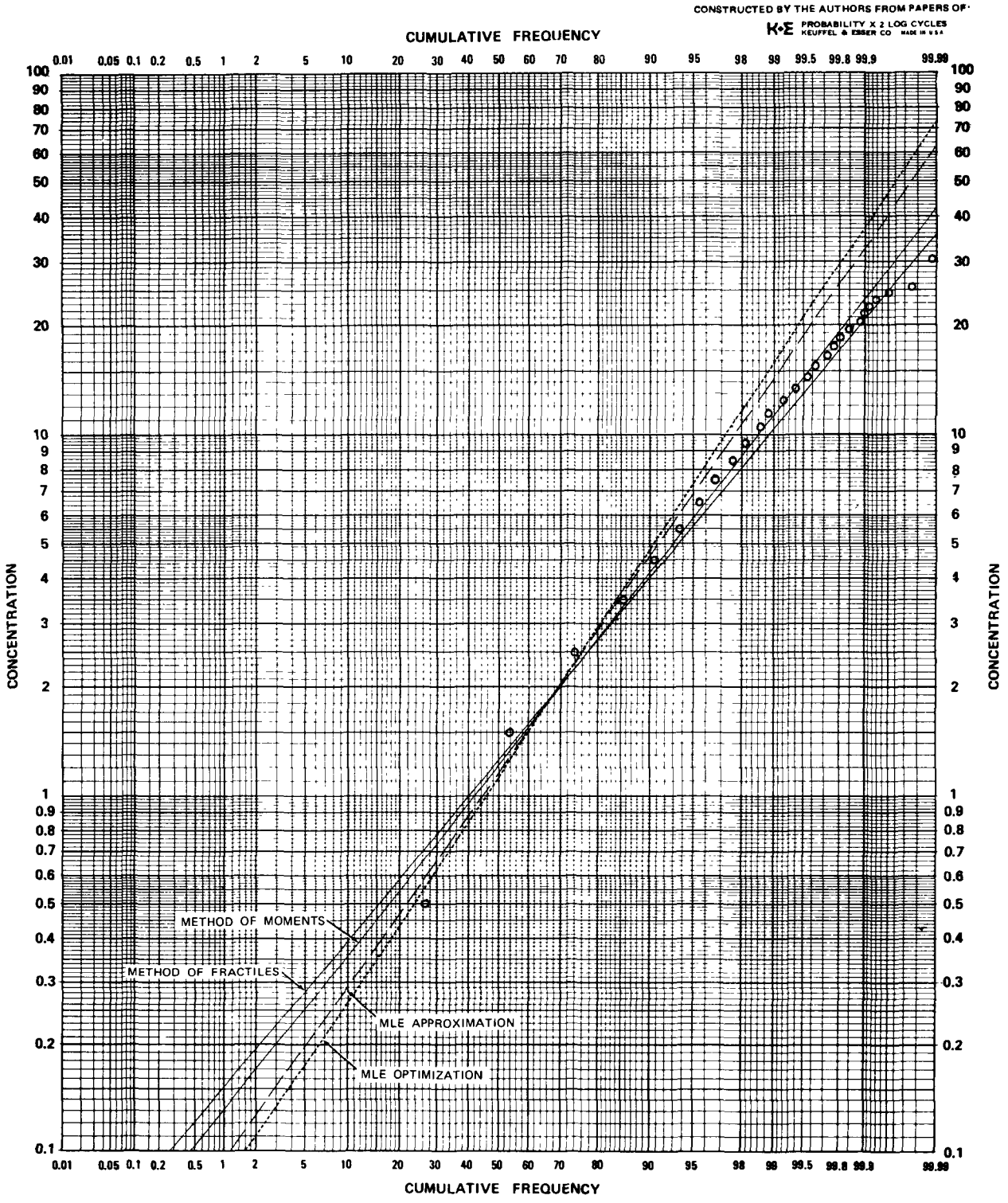

FIGURE 12. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR DENVER, CO, 1974 (n = 7952).

Table 16

Frequency-based Measures of Goodness-of-fit for Denver, CO

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	1.204	2.617	15	310	770	0.088	-15013
Fractiles	1.237	2.500	14	356	1036	0.107	-15127
MLE Approx.	1.152	2.969	19	207	475	0.055[1]	-14875
MLE Opt.	1.110	3.086	19	190[1]	460[1]	0.064	-14866[1]

[1]Indicates best result in each column.

slope from the method of fractiles line, causing overpredictions at the higher concentration levels. According to the frequency-based measures of goodness-of-fit (Table 16), the MLE optimization approach gave the best fit for three of the measures, but the Kolmogorov-Smirnov measure (K-S max) indicated that the MLE approximation gave the best fit. The MLE approaches tend to

Table 17

Variate-based (EET) Measures of Goodness-of-fit for Denver, CO (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	24	0.680	1.376	-0.43	4.99
Fractiles	24	-0.472[1]	0.627[1]	-1.18	1.68[1]
MLE Approx.	24	5.041	5.139	-0.21[1]	17.75
MLE Opt.	24	6.250	6.319	-0.25	21.65

[1]Indicates best result in each column.

give good fits according to the frequency-based measures, because the MLE approaches weight more heavily those intervals which contain large numbers of observations, and the frequency-based measures are sensitive to the number of observations in each intervals. Because most of the observations lie in intervals in the middle concentration ranges, the MLE approaches tend to fit the middle ranges better than the high ranges. When the logarithmic-probability plot of the observations shows downward curvature, as in the case of Barstow and Denver, the MLE optimization approach causes the model to overpredict at the higher concentrations. In Denver, for example, the MLE optimization approach causes serious overpredictions at concentrations above 6 ppm, which corresponds approximately to a cumulative frequency of 95%, but the long tail of the distribution for concentrations above this level contains only 5% of the observations. Because the method of fractiles does not respond to the number of observations in each interval, it fits the tail quite well, and the variate-based measures of goodness-of-fit, which also are not responsive to the number of observations in each interval, indicate that the method of fractiles gives the best fit (Table 17). Although the method of fractiles, on the average, causes a slight underestimation ($\overline{d}$ = -0.472 ppm), the line fits the observations between -1.18 ppm and +1.68 ppm over the entire range of concentrations, which is far superior to the other methods. As in the previous data sets, the frequency-based measures of goodness-of-fit indicate that the MLE approaches give the best fit, while the variate-based measures indicate that the method of fractiles gives

the best fit. The analyst might conclude that the LN2 model, which plots as a straight line, may not be appropriate for representing the curved locus of observations obtained for Barstow and Denver.

Philadelphia, PA

The observations for Philadelphia (Figure 13) exhibit several reversals of curvature that appear similar to the observations for Norfolk (Figure 10). However, the Norfolk data tend to fluctuate around a curve that is concave upward, while the Philadelphia data exhibit the more desirable "snake on a stick" pattern of variation about a line. Because of this tendency of the observations to follow a straight line, the frequency-based measures of goodness-of-fit (RMS, χ^2, and K-S max) in Table 18 all are significantly lower than the corresponding measures for Norfolk given in Table 12. Because the lines are relatively

Table 18
Frequency-based Measures of Goodness-of-fit for Philadelphia, PA

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	1.835	2.027	12	218	638	0.066[1]	-12335
Fractiles	2.015	2.135	15	168[1]	718	0.073	-12390
MLE Approx.	1.832	2.158	14	217	617	0.076	-12350
MLE Opt.	1.844	2.066	13	211	612[1]	0.067	-12331[1]

[1]Indicates best result in each column.

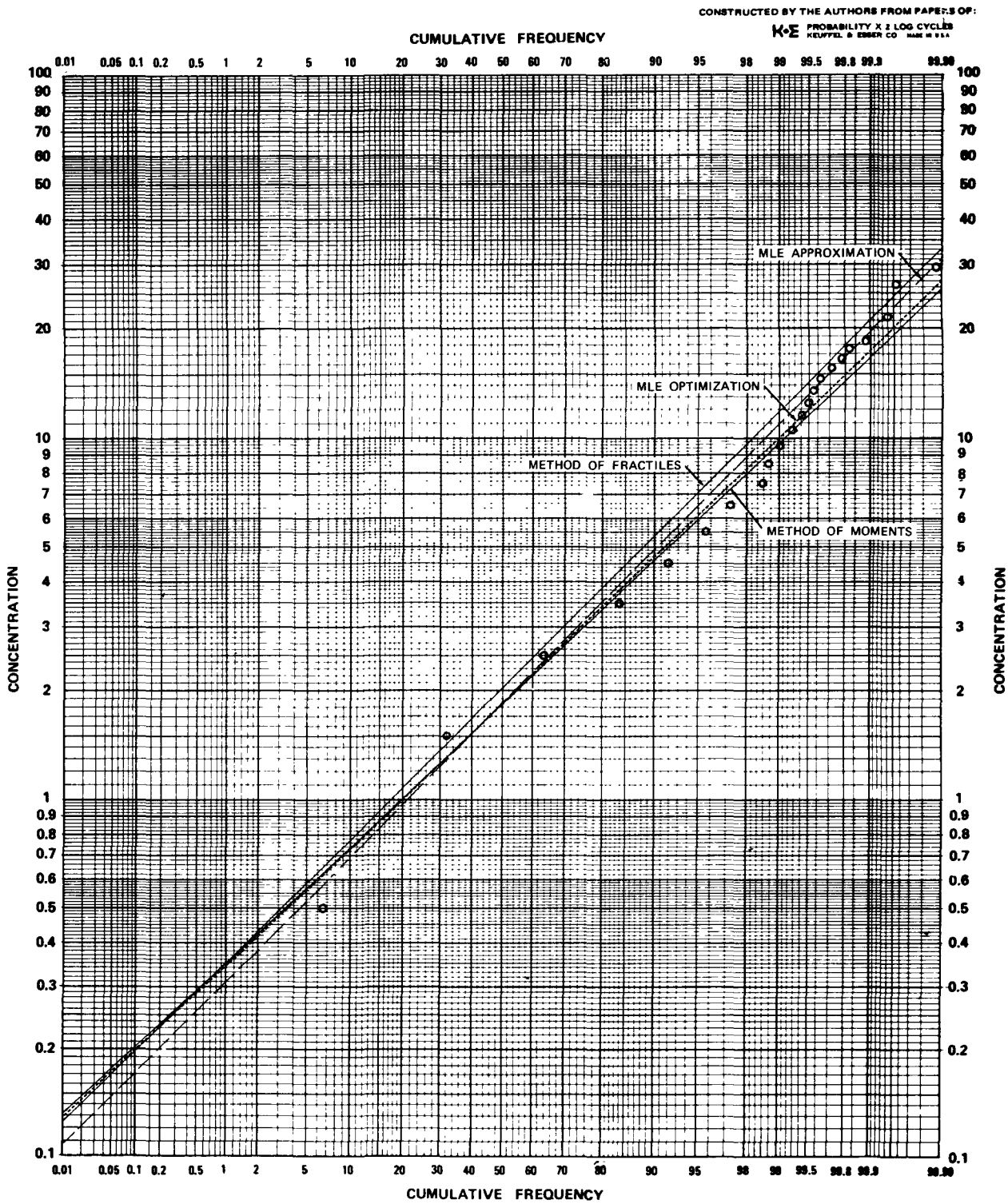

FIGURE 13. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR PHILADELPHIA, PA, 1974 (n = 6836).

close together, the values of the frequency-based measures also are very similar to each other, and no single method of calculating parameters emerges as clearly superior. Actually, each frequency-based measure indicates a different method of calculating parameters is "best." The same ambivalence occurs in the variate-based measures (Table 19), in which each measure selects a different calculation procedure as "best." For example, the method of moments gives minimum overprediction (d_{max} = 0.76 ppm); the method of fractiles gives the least underprediction (d_{min} = -0.08 ppm); the MLE approximation approach gives the smallest standard deviation (s_d = 0.776 ppm); and the MLE optimization technique gives the smallest magnitude of the average difference ($\bar{d}$ = -0.396 ppm). If a hypothesis test were applied to these findings, we would be forced to reject the LN2 model as a representation of the underlying distribution, because of the large magnitudes of the frequency-based measures of goodness-of-fit.

Table 19

Variate-based (EET) Measures of Goodness-of-fit for Philadelphia, PA (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	21	-0.919	1.502	-4.29	0.76[1]
Fractiles	21	1.521	0.861	-0.08[1]	2.67
MLE Approx.	21	0.742	0.776[1]	-0.86	1.92
MLE Opt.	21	-0.396[1]	1.211	-3.25	1.13

[1]Indicates best result in each column.

Rejection of the LN2 model may seem surprising in view of its excellent visual fit to the observations. Rejection means that, if one were to draw the same number of observations, at random, from a real lognormal distribution, frequency-based measures of goodness-of-fit as large as the ones we have obtained would be very rare. Of course, the Philadelphia data set, like the other data sets in this report, is not really a random sample, and therefore application of the theory of hypothesis testing appears questionable.

Napa, CA

The data set for Napa (Figure 14) resembles the data for Philadelphia (Figure 13) in that the observations display the desirable "snake on a stick" pattern about a straight line. The method of fractiles straight line appears visually to fit the observations quite well -- from the lowest value of 0.5 ppm to the highest value of 15.5 ppm.* Because almost 85% of the observations occur in the range from 1.5 ppm to 5.5 ppm, and nearly half of the observations lie in the range from 1.5 ppm to 2.5 ppm, the method of moments and the MLE approaches are greatly affected by concentrations in these ranges. The particular shape of the zig-zag pattern of the observations causes lines which fit well in these ranges to lie below the observations at the higher concentrations, thereby causing underpredictions. Because the method of fractiles ignores the number of observa-

*Note that the actual range for Napa is from 0.0 ppm to 15.0 ppm and that the 0.5 ppm increment results from data grouping.

Table 20

Frequency-based Measures of Goodness-of-fit for Napa, CA

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	2.433	1.618	10	179	334	0.043	-13436
Fractiles	2.260	1.717	11	329	800	0.109	-13712
MLE Approx.	2.442	1.607	9	171	313[1]	0.037	-13422
MLE Opt.	2.466	1.560	9	110[1]	493	0.027[1]	-13393[1]

[1]Indicates best result in each column.

in each interval, the EET measures of goodness-of-fit (Table 21) indicate that the method of fractiles gives the best fit. In contrast, the frequency-based measures (Table 20) indicate that the method of fractiles gives the worst fit, with the two MLE approaches, particularly MLE optimization, performing best. If we examine Figure 14, we note that the minimum underprediction

Table 21

Variate-based (EET) Measures of Goodness-of-fit for Napa, CA (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	13	-0.338	0.477	-1.08	0.26
Fractiles	13	0.065[1]	0.229[1]	-0.32[1]	0.56
MLE Approx.	13	-0.413	0.549	-1.23	0.23
MLE Opt.	13	-0.747	0.861	-2.05	0.19[1]

[1]Indicates best result in each column.

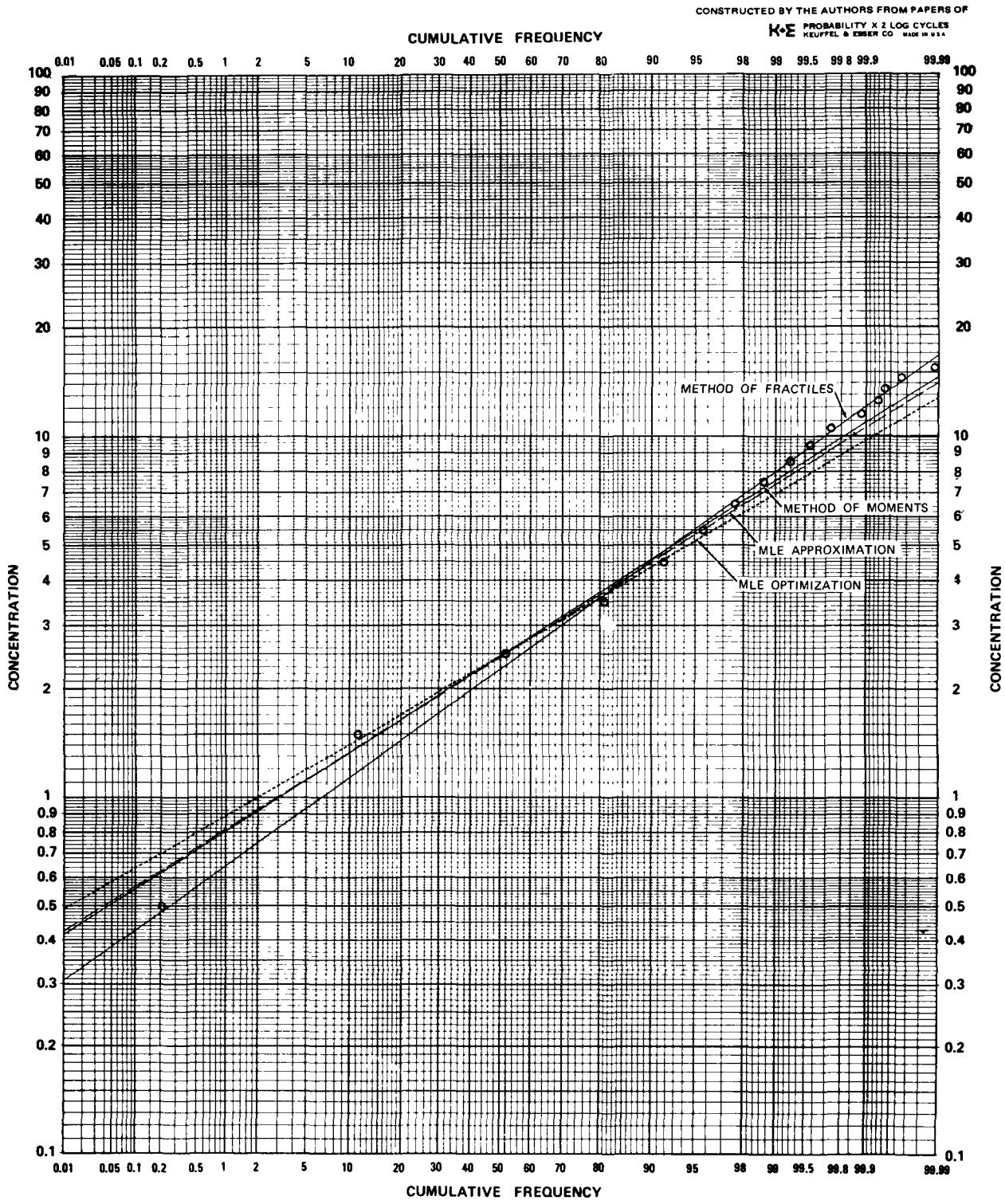

FIGURE 14. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR NAPA, CA, 1974 (n = 8558).

reported in Table 21 (d_{min} = -0.32 ppm) for the method of fractiles occurs at the observed value of 1.5 ppm, which corresponds to a cumulative frequency of 11.5%. If we examine the figure's horizontal fit, the method of fractiles line predicts a cumulative frequency of about 22% at 1.5 ppm, a discrepancy of 9.5%. Notice that a difference in cumulative frequency of 9.5% is quite serious, because it is equivalent to 0.095 x 8558 = 813 observations for the Napa data set. Thus, we have the paradox that a small error (0.32 ppm) according to a variate-based criterion is equivalent to a large error (9.5%) according to a frequency-based criterion. This single data point (1.5 ppm) accounts for the good performance of the method of fractiles when judged by the variate-based measures of goodness-of-fit and the poor performance of the method of fractiles when judged by the frequency-based measures of goodness-of-fit. As with the other data sets, the frequency-based measures would lead to rejection of lognormality if a hypothesis test could be validly applied.

Phoenix, AZ

When the grouped data for Phoenix are plotted on logarithmic-probability paper (Figure 15), they exhibit consistent downward curvature similar to the patterns for the Barstow and Denver data sets, but somewhat more abrupt in the range between cumulative frequencies of 90% and 99.9%. When we first plotted this distribution, we were struck by the two extreme values of 50 ppm and 61 ppm which do not seem to lie on the curve at all.

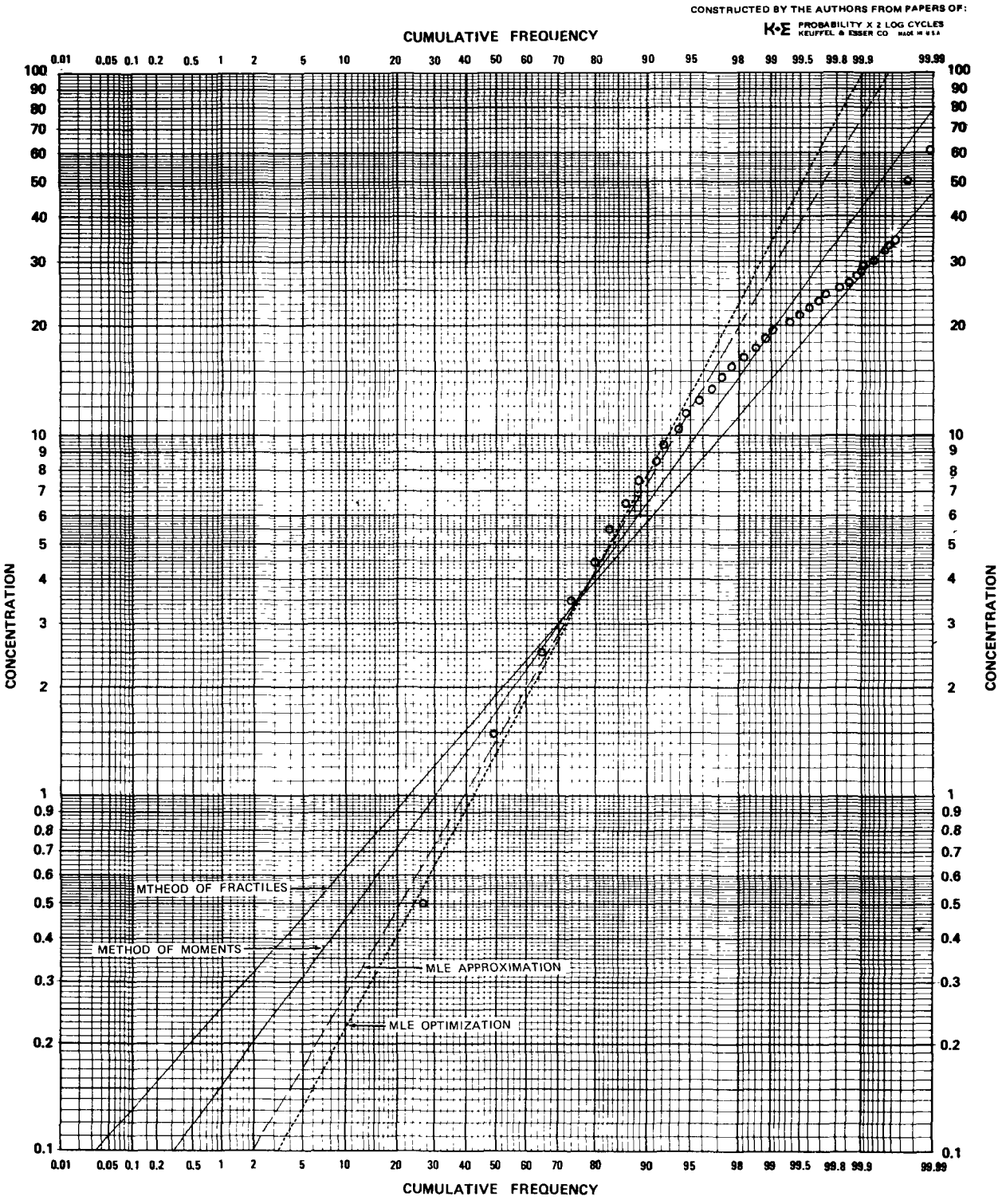

FIGURE 15. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR PHOENIX, AZ, 1974 (n = 8246).

Table 22

Frequency-based Measures of Goodness-of-fit for Phoenix, AZ

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	1.710	2.830	24	299	2224	0.150	-19025
Fractiles	1.900	2.388	19	436	6969	0.206	-20262
MLE Approx.	1.407	3.617	30	145	530	0.058	-18406
MLE Opt.	1.311	4.020	32	103[1]	412[1]	0.042[1]	-18366

[1] Indicates best result in each column.

The next lower value after these two data points, 34 ppm, seems to lie on the curve with the other observations. We were puzzled by these two anomalies, and we initially concluded that these observations must be from a different population. That is perhaps very unusual micrometeorological conditions and sources caused an event that was quite different from the process which generated the other observations. When we obtained the original

Table 23

Variate-based (EET) Measures of Goodness-of-fit for Phoenix, AZ (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	32	3.196	6.582	-2.26	18.97
Fractiles	32	-2.237[1]	1.918[1]	-4.89	1.26[1]
MLE Approx.	32	14.41	19.56	-0.73	59.58
MLE Opt.	32	21.82	28.50	-0.44[1]	87.68

[1]Indicates best result in each column.

data sheets which had been submitted to EPA by Arizona, we found, to our dismay, that these two high values apparently resulted from transcription errors which resulted when the data were keypunched by EPA. In the following list of values, the top line corresponds to the true values that were measured in Phoenix on December 21, 1974, and the bottom line contains the values that actually were stored by EPA in the SAROAD data bank:

Correct: (6,2,1,0,0,1,1,1,1,2,2,1,0,0,0,1,1,1,2,5,6,10,11,10)
SAROAD: (6,2,1,0,0,1,1,1,1,2,2,1,0,0,0,10,10,10,20,50,61,2,2,1)

Notice that erroneous values were substituted in all of the last 9 hours in the 24-hour period, and the reason for such a striking distortion of the correct measurements is unclear. Although these errors were easy to detect because they gave high values which stood out from the rest of the observations, their existence made us worry about other errors that may be present in these data sets but may be more difficult to discover (for example, high values recorded erroneously as low values).

Unfortunately, we had completed all the computer analyses by the time that these errors had been discovered, and, rather than changing the values and repeating the computer runs, we have chosen to present briefly the results for the uncorrected SAROAD data set. These findings should convey a warning to all data analysts to be suspicious of all data that they do not personally collect and process. It is clear that implementation of high levels of quality control at the measuring location does

not prevent the data from being distorted by the time they are stored in the EPA data bank. Again, these findings have convinced us that EPA should institute a quality assurance data checking program to compare the observations stored in its data banks with the values originally submitted by state and local pollution control agencies (see page 109).

Because of the sharp downward curvature of the Phoenix observations, the two MLE approaches, influenced by the number of observations at the lower concentrations, cause serious overpredictions at concentrations above 12 ppm, sometimes overpredicting by 300% or more. The method of moments causes overpredictions at concentrations above 20 ppm, but these are not as serious as for the MLE techniques.

Once again, the frequency-based measures of goodness-of-fit (Table 22) indicate that the MLE optimization approach gives the best fit, and the variate-based measures indicate that the method of fractiles gives the best fit (Table 23). Study of the data points in Figure 15 reveals that none of the parameter calculation techniques gives a very good fit, because the LN2 model attempts to impose a straight line on observations that are better fit by a curve. The poor fit is reflected by the large values obtained for the goodness-of-fit measures. For example, the method of fractiles gives a chi-square value of $\chi^2 = 6988$, and the MLE optimization gives an overprediction of $d_{max} = 87.67$ ppm.

Newhall, CA

The data sets discussed thus far all have contained some observations reported as 0 ppm; that is, observations in the first interval ($-0.5 \leq X < 0.5$). The Newhall data set contains no observations in the first interval, although there are 938 observations in the second interval ($0.5 \leq X < 1.5$). A possible explanation is that vehicular activity 24 hours a day in the vicinity of the monitoring location may prevent concentrations from dropping below 0.5 ppm, even on days with high winds. The overall range of measured concentrations for Newhall was from 1 ppm to 14 ppm, and the data points representing the intervals appear to scatter about a straight line (Figure 16). If we carefully examine the data points at 1.5 ppm and 14.5 ppm, there is some indication that these two end points lie below the others, giving the appearance of slight downward curvature, but one cannot be sure, and the LN2 model would appear to be a reasonable representation of the observations. All four of the straight lines fit the observations reasonably well, but, because the method of fractiles must fit exactly at cumulative frequencies of 70% and 99.9%, it gives larger deviations at the lower frequencies than the other methods. Because the other three methods of calculating parameters are weighted heavily by concentrations near the median (cumulative frequency of 50%) and the arithmetic mean (cumulative frequency of approximately 70%), their goodness-of-fit measures give values that are quite similar to each other (Table 24). For example, the Kolmogorov-Smirnov maximum difference gives 0.036, 0.038, and 0.034, respectively, for the method

Table 24

Frequency-based Measures of Goodness-of-fit for Newhall, CA

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	2.966	1.617	12	152	284	0.036	-15991
Fractiles	3.196	1.534	11	200	1214	0.070	-16280
MLE Approx.	2.918	1.699	13	139[1]	281	0.038	-16002
MLE Opt.	2.953	1.650	12	142	245[1]	0.034[1]	-15979[1]

[1]Indicates best result in each column.

of moments, MLE approximation, and MLE optimization. Similarly, the values of the RMS, chi-square, and log-likelihood functions have very narrow ranges. These results indicate that the MLE optimization approach gives the best fit. As in many of the other data sets, the EET variate-based measures (Table 25) indicate that the method of fractiles gives the best fit. Notice

Table 25

Variate-based (EET) Measures of Goodness-of-fit for Newhall, CA (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	13	0.185	0.376	-0.16	1.30
Fractiles	13	-0.055[1]	0.313[1]	-0.65	0.38[1]
MLE Approx.	13	0.811	1.016	-0.17	3.39
MLE Opt.	13	0.450	0.619	-0.15[1]	2.15

[1]Indicates best result in each column.

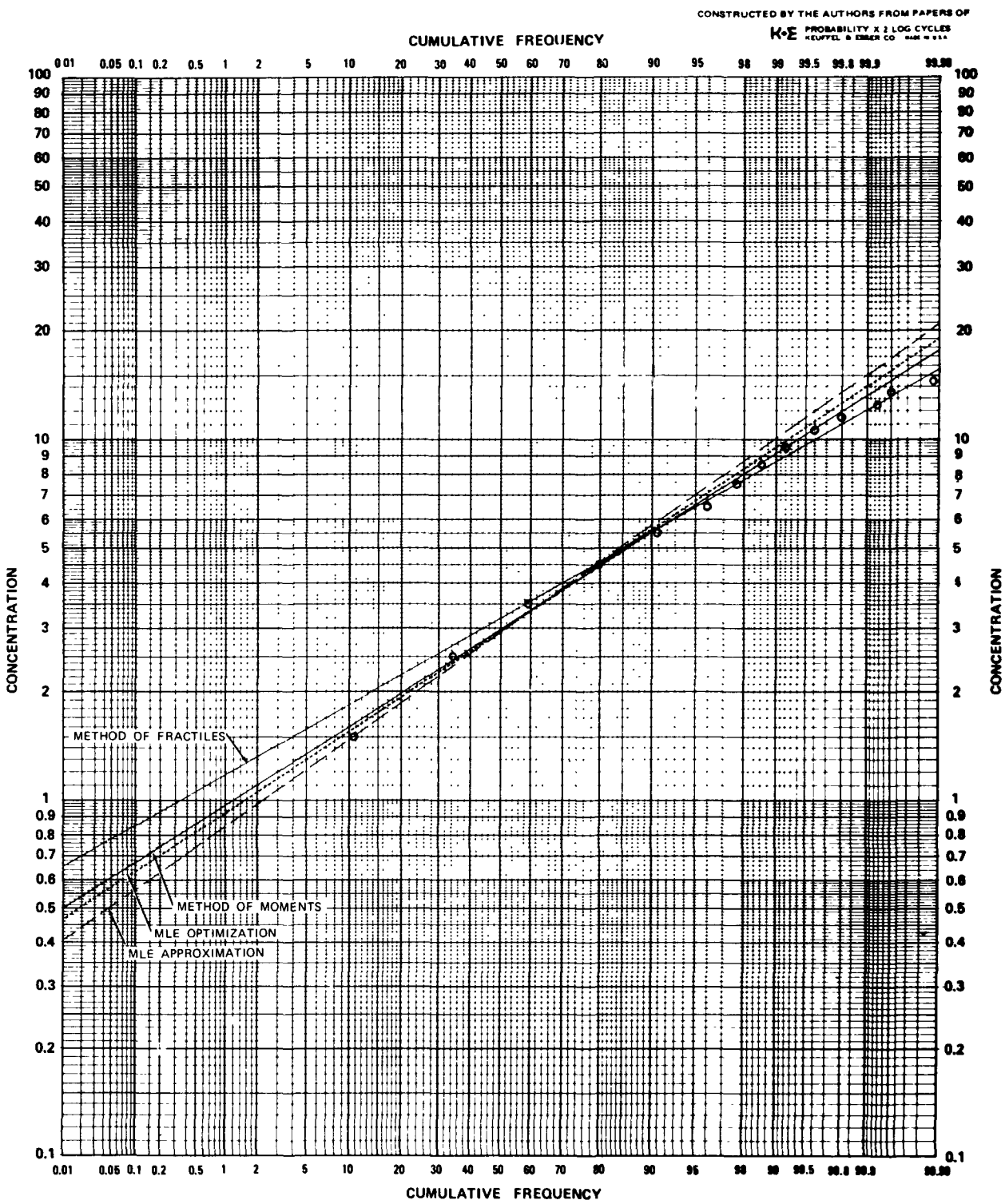

FIGURE 16. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR NEWHALL, CA, 1974 (n = 8623).

that the average deviations ($\bar{d}$) between predicted and observed concentrations all are relatively small in magnitude and that the MLE approximation approach gives the greatest overprediction (d_{max} = 3.39 ppm) due to its larger overshoot at the higher concentrations.

Springfield, MA

The grouped data for Springfield tend to follow the "snake on a stick", or S-shaped, pattern, except perhaps for the lowest value of 0.5 ppm at the cumulative frequency of 3.5% (Figure 17). The lines for the method of fractiles and method of moments seem to fit the upper half of the data reasonably well, although the MLE approaches give lines that appear "pulled down" by the lowest observation of 0.5 ppm. This result is very similar to the pattern which occurred in Denver (Figure 12). In that case, the observations at 0.5 ppm tended to bring the MLE down at the lower values, causing them to tip upward at the higher concentrations

Table 26
Frequency-based Measures of Goodness-of-fit for Springfield, MA

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	3.035	1.881	17	119	4360	0.038[1]	-18828
Fractiles	2.798	1.977	18	171	1736	0.077	-18725
MLE Approx.	2.902	2.165	22	160	1094[1]	0.066	-18693
MLE Opt.	2.942	2.071	21	140[1]	1334	0.054	-18668[1]

[1]Indicates best result in each column.

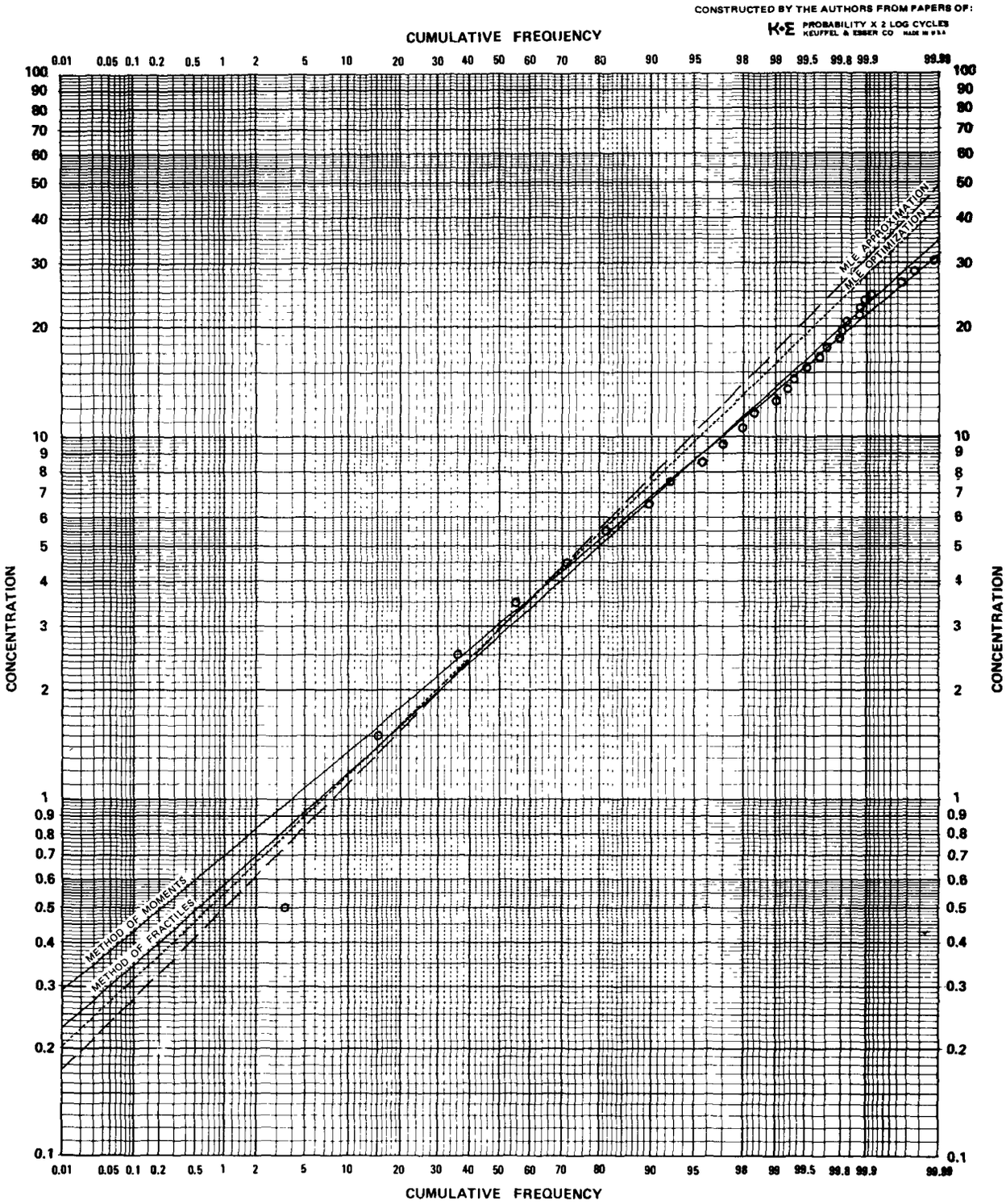

FIGURE 17. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR SPRINGFIELD, MA, 1974 (n = 8339).

Table 27

Variate-based (EET) Measures of Goodness-of-fit for Springfield, MA (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	26	-0.238[1]	0.873	-2.77	0.68
Fractiles	26	0.366	0.604[1]	-0.95	1.32
MLE Approx.	26	4.559	3.450	-0.31	9.52
MLE Opt.	26	2.854	2.057	-0.27[1]	5.45

[1]Indicates best result in each column.

and give large overpredictions. Neither the frequency-based measures of goodness-of-fit (Table 26) nor the variate-based measures (Table 27) show any parameter calculation approach emerging as clearly superior. The goodness-of-fit measures, in general, seem relatively large, and, if hypothesis tests were possible, would strongly reject the LN2 model. For example, to reject the LN2 hypothesis at P = 0.995 with 21 degrees of freedome, chi-square must exceed $\chi^2 = 41.4$; yet the chi-square value for the MLE optimization approach is $\chi^2 = 1,334$. Examination of the chi-square values on an interval-by-interval basis reveals that the first interval makes the largest contribution. For the MLE optimization approach, the first interval contributes $\chi^2 = 853$, which is more than half the total. Thus, elimination of the first interval from the data set would improve the goodness-of-fit measures in all categories and would bring the lines closer together. In the present situation, it is difficult to

claim that any one parameter calculation approach performs best. Despite the poor performance of all four parameter calculation approaches when judged by the frequency-based measures, notice that the variate-based measures do not present as negative a picture. Although no one approach is superior in all categories, the method of fractiles seems to fit reasonably well; the model is within 1.3 ppm (d_{max} = 1.32 ppm) over the entire range of observations, with very small average overprediction ($\bar{d}$ = 0.366 ppm). One reason that the variate-based measures give a more favorable result than the frequency-based measures is that they are less affected by the poor fit that occurs in the first interval.

Pasadena, CA

When plotted on logarithmic-probability paper, the grouped observations for Pasadena give a gently S-shaped curve (Figure 18). Like the Newhall data set, the minimum concentration for

Table 28
Frequency-based Measures of Goodness-of-fit for Pasadena, CA

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	3.487	1.996	22	109	327	0.037	-19561
Fractiles	3.570	1.901	20	83	284	0.037	-19518
MLE Approx.	3.542	1.918	21	85	272[1]	0.034[1]	-19519
MLE Opt.	3.558	1.900	20	82[1]	275	0.035	-19517[1]

[1]Indicates best result in each column.

Table 29

Variate-based (EET) Measures of Goodness-of-fit for Pasadena, CA (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	28	-0.141[1]	1.496	-1.652[1]	5.43
Fractiles	28	-1.304	1.420	-3.39	1.27
MLE Approx.	28	-1.152	1.372[1]	-3.11	1.87
MLE Opt.	28	-1.359	1.438	-3.48	1.14[1]

[1]Indicates best result in each column.

Pasadena was 1 ppm (695 observations), plotted in the 1.5 ppm interval, and there were no observations in the 0.5 ppm interval. Of course, if more observations were taken from the same station (for example, several years of data), we would expect more values to appear in the first interval. The results for the four methods give lines that are extremely close together; the lines for the method of fractiles, MLE approximation, and MLE optimization are almost coincident. For these three approaches, $\hat{m}$ and s_g are almost identical (Table 28), and the values for the frequency-based measures of goodness-of-fit are approximately the same. The variate-based measures of goodness-of-fit (Table 29) also yield similar values for these three approaches, and no one approach stands out as clearly superior. Overall, the negative values of $\bar{d}$ suggest that the model tends to underpredict concentrations for all four approaches, although the underpredictions

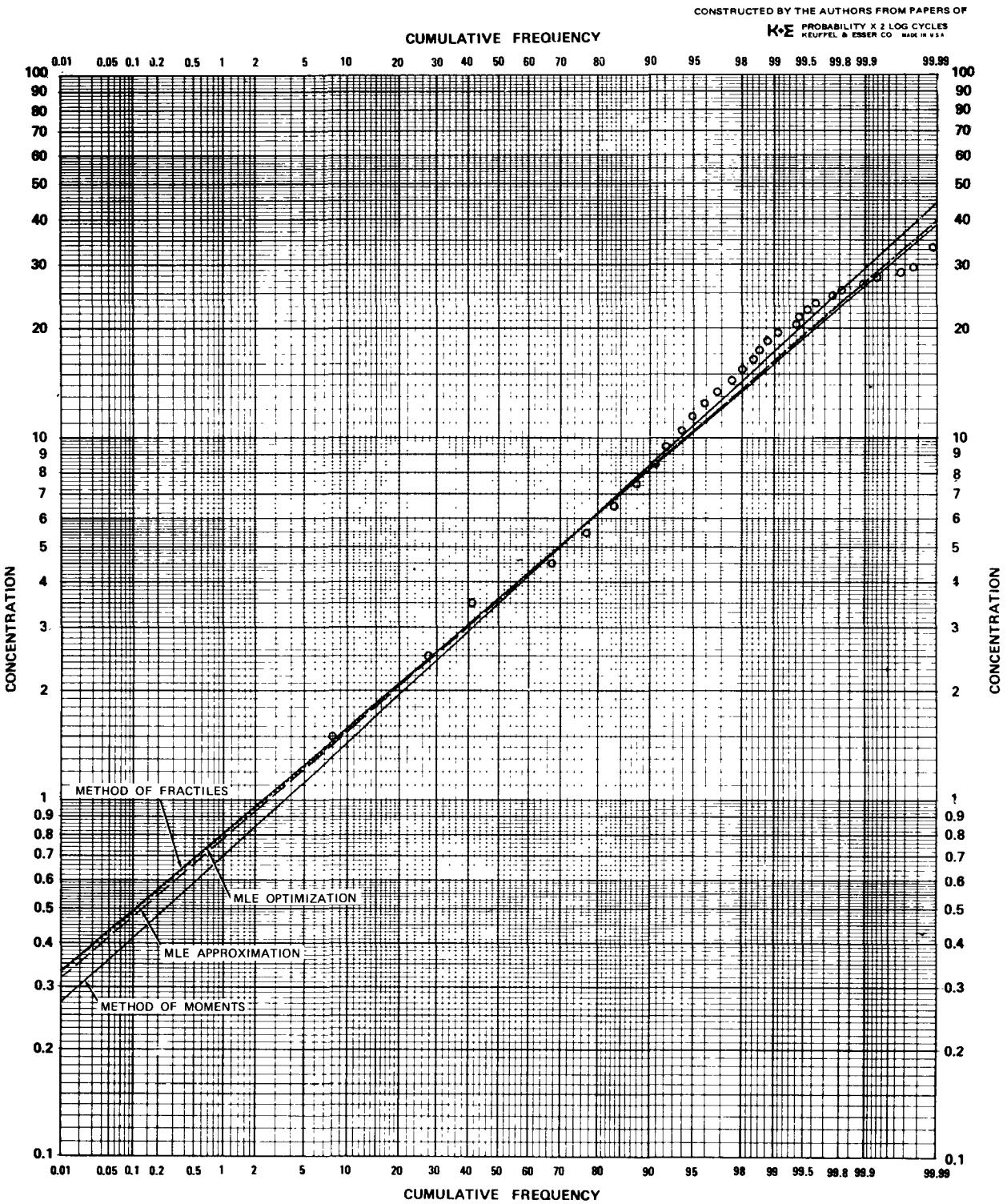

FIGURE 18. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR PASADENA, CA, 1974 (n = 8632).

Table 30

Frequency-based Measures of Goodness-of-fit for New York City, NY

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	14.92	1.501	44	48	893	0.039[1]	-27798
Fractiles	15.72	1.435	41	62	4010	0.085	-28284
MLE Approx.	14.72	1.581	48	43	416[1]	0.046	-27710
MLE Opt.	14.75	1.567	47	43[1]	430	0.044	-27707[1]

[1]Indicates best result in each column.

are not great. The slight tendency toward underprediction occurs because, although the observations lie close to the straight lines, more data points lie above (upper left) than lie below (lower right) the lines.

New York City

The New York City data set was the highest of any CO data

Table 31

Variate-based (EET) Measures of Goodness-of-fit for New York City, NY (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	50	1.357	1.731	-0.67	4.99
Fractiles	50	0.523[1]	0.833[1]	-0.42[1]	2.62[1]
MLE Approx.	50	3.730	4.623	-0.74	13.83
MLE Opt.	50	3.284	4.071	-0.70	12.18

[1]Indicates best result in each column.

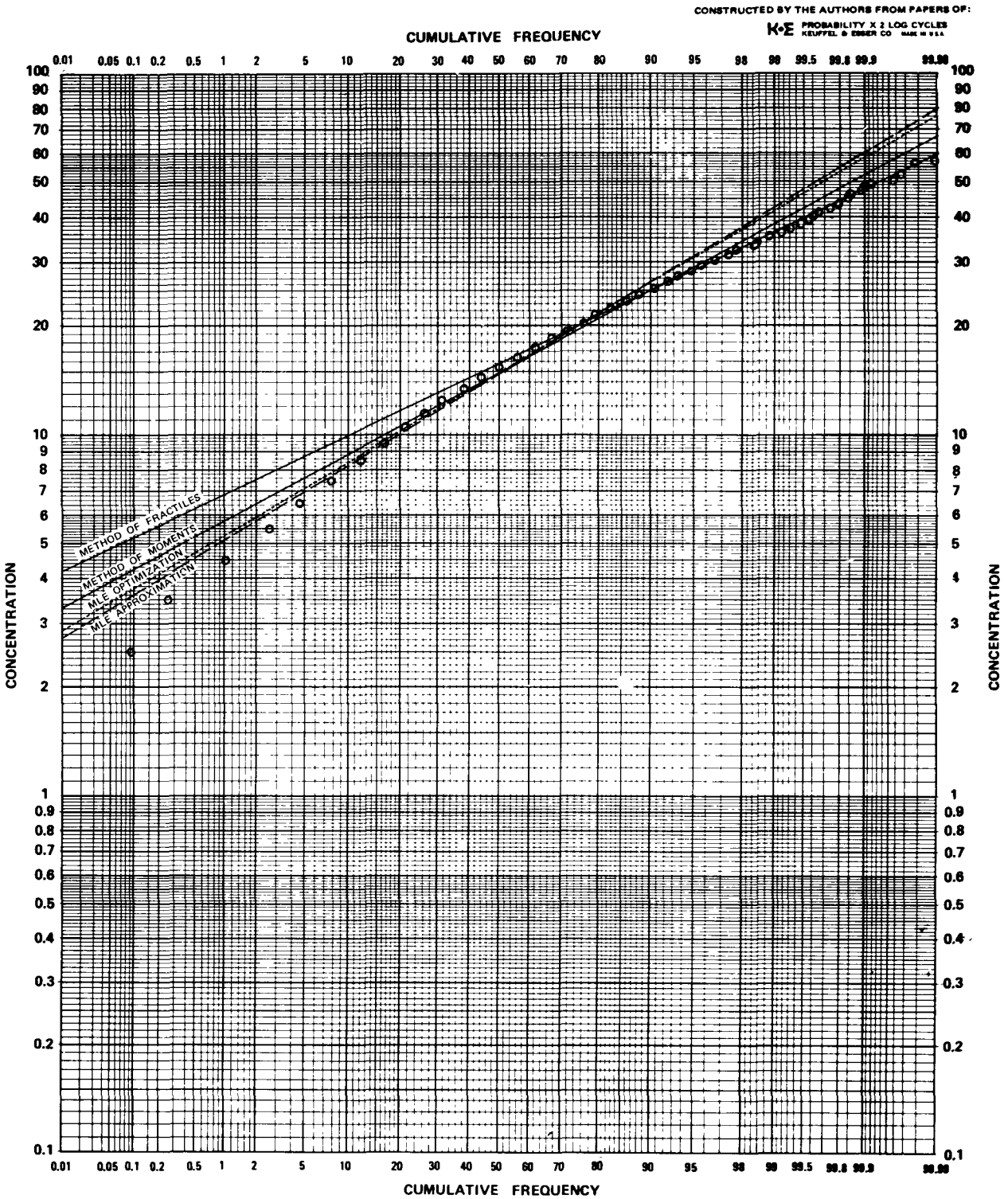

FIGURE 19. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR NEW YORK CITY, NY, 1974 (n = 8364).

set in the SAROAD data bank in 1974. When plotted on logarithmic probability paper (Figure 19), these observations, which range from 2 ppm to 56 ppm, showed consistent downward curvature. Unlike the high (erroneous) outliers for Phoenix, all of the New York city observations seem to fall on a smooth, though gradual, curve. Like the other data sets that exhibited downward curvature (Phoenix, Denver, and Barstow), the line for the method of fractiles fits the observations quite well at cumulative frequencies between 70% and 99.99%, but seriously overpredicts concentrations at the lower levels. The two MLE approaches fit best in the middle ranges (10-20 ppm) between cumulative frequencies of 20% and 80%, where most of the observations are located. The MLE optimization and MLE approximation lines almost coincide, and the method of moments line seems somewhat intermediate between the line for the method of moments and the two MLE lines. The frequency-based measures of goodness-of-fit indicate that the best fit is obtained from the MLE optimization approach, although K-S max measure selects the method of moments as best (Table 30). According to the frequency-based measures, the method of fractiles gives the worst fit. The variate-based measures indicate the opposite: the method of fractiles gives the best fit (Table 31). Inspection of Figure 19 suggests that none of the parameter calculation approaches can be very satisfactory over the full range of the observations because of the basic difficulty of drawing a straight line through observations that are better described by a curve.

Discussion

If the results for the 11 data sets comprising the geographical data group are compared (Table 32), it is seen that no single parameter calculation approach emerges as superior in all situations. According to the frequency-based measures of goodness-of-fit, the MLE optimization approach is best in a majority of the data sets, although the MLE approximation performs almost as well in two cases. According to the variate-based measures of goodness-of-fit, the method of fractiles was the best approach in a majority of the data sets, although no single approach emerged as superior in four of these. The method of moments was never judged to be the best approach for calculating parameters, although it offers the practical advantage that the mean of the model matches the mean of the data. Four of the 11 data sets (36%) exhibited concave downward curvature, with the remainder showing a zig-zag pattern that varied from slight reversals of curvature to more distinct S-shaped patterns.

Thus, the method of fractiles seems best when judged by variate-based criteria, and the MLE optimization approach seems best when judged by frequency-based criteria. This dichotomy occurs because the two sets of criteria measure different things. Frequency-based measures, like chi-square, are sensitive to the number of values within each interval, and variate-based measures, like the EET tests, are sensitive to the predicted and observed concentrations at each interval. When an observer examines a plot of the observations on logarithmic-probability paper, the

Table 32

SUMMARY OF FINDINGS FOR THE GEOGRAPHICAL DATA GROUP

City	"Best" by the Frequency-based Measures	"Best" by the Variate-based Measures	Graphical Pattern
Barstow, CA	MLE Optimization	None	Downward Curvature
Norfolk, CA	MLE Optimization	Method of Fractiles	Zig-zag
Alexandria, VA	MLE Optimization	Method of Fractiles	Zig-zag
Denver, CO	MLE Optimization	Method of Fractiles	Downward Curvature
Philadelphia, PA	None	None	Zig-zag
Napa, CA	MLE Approaches[a]	Method of Fractiles	Zig-zag
Phoenix, AZ	MLE Optimization	Method of Fractiles	Downward Curvature
Newhall, CA	MLE Optimization	Method of Fractiles	Zig-zag
Springfield, MA	None	None	Zig-zag
Pasadena, CA	MLE Approaches[a]	None	Zig-zag
New York City, NY	MLE Optimization	Method of Fractiles	Downward Curvature

[a]Both the MLE approximation and MLE optimization approaches perform well.

eye tends to emphasize the variate-based approach, because the paper displays the logarithm of the concentration on its variate (vertical) scale, but it greatly compresses the (horizontal) frequency scale in those areas (e.g., 20% to 80% cumulative frequencies) where large numbers of observations occur. As a result, the user of logarithmic-probability paper tends to give considerable emphasis to the tail of the distribution, which is represented by a large number of intervals but relatively few observations. Because the tail of the distribution is subject to considerable randomness, resulting variate-based fits, such as the method of fractiles, tend also to be heavily influenced by randomness in the data. Thus they fit the specific data set generated by a process rather than the underlying distribution of the process itself. In this investigation, we unfortunately cannot be certain about the nature of this underlying distribution, so it is not possible to attach any particular significance to our findings for frequency-based fits such as the MLE optimization approach.

OVERALL FINDINGS

Tables 33-36 summarize the overall findings for the 11 cities and the four parameter calculation approaches. In general, the values for the frequency-based measures of goodness-of fit are quite large, indicating a poor overall fit of the LN2 model to the 11 data sets, regardless of the approach used to calculate its parameters. For example, all chi-square values

TABLE 33

SUMMARY OF RESULTS FOR THE METHOD OF MOMENTS

CITY AND STATE	MODEL PARAMETERS		STATISTICAL TESTS				ENGINEERING ERROR TEST			
	Arithmetic Mean (ppm)	Arithmetic Standard Deviation (ppm)	Root Mean Square Of Difference	Chi Square	Kolmogorov-Smirnov Maximum Difference	Log Likelihood Function	Arithmetic Mean (ppm)	Arithmetic Standard Deviation (ppm)	Lowest Value (ppm)	Highest Value (ppm)
Barstow, CA	0.44	0.86	181	228	0.0650	- 6272	0.55	1.79	-0.56	4.41
Norfolk, VA	1.38	1.06	682	2706	0.1937	-11230	-1.41	1.59	-4.45	0.38
Alexandria, VA	1.64	1.71	306	872	0.1140	-13265	-0.02	0.66	-1.01	1.45
Denver, CO	1.91	2.36	310	770	0.0888	-15013	0.68	1.38	-0.43	4.99
Philadelphia, PA	2.36	1.90	218	638	0.0665	-12335	-0.92	1.50	-4.29	0.76
Napa, CA	2.73	1.39	179	334	0.0426	-13436	-0.34	0.48	-1.08	0.26
Phoenix, AZ	2.94	4.10	299	2224	0.1504	-19025	3.20	6.58	-2.26	18.97
Newhall, CA	3.33	1.70	152	284	0.0362	-15991	0.18	0.38	-0.16	1.30
Springfield, MA	3.71	2.60	119	4360	0.0377	-18828	-0.24	0.87	-2.77	0.68
Pasadena, CA	4.43	3.47	109	327	0.0372	-19561	-0.14	1.50	-1.65	5.43
New York City, NY	16.21	6.87	48	893	0.0391	-27798	1.36	1.73	-0.67	4.99
ARITHMETIC MEAN			237	1240	0.0792	-15705	0.26	1.68	-4.45[a]	18.97[b]

[a]Minimum of values listed in column.

[b]Maximum of values listed in column.

TABLE 34

SUMMARY OF RESULTS FOR THE METHOD OF FRACTILES

CITY AND STATE	MODEL PARAMETERS		STATISTICAL TESTS				ENGINEERING ERROR TEST			
	Arithmetic Mean (ppm)	Arithmetic Standard Deviation (ppm)	Root Mean Square Of Difference	Chi Square	Kolmogorov-Smirnov Maximum Difference	Log Likelihood Function	Arithmetic Mean (ppm)	Arithmetic Standard Deviation (ppm)	Lowest Value (ppm)	Highest Value (ppm)
Barstow, CA	0.43	0.58	120	370	0.0403	- 6299	-0.70	0.58	-1.32	0.13
Norfolk, VA	1.78	1.49	484	1953	0.1258	-11063	0.90	0.77	-0.42	1.87
Alexandria, VA	1.82	1.75	221	467	0.0611	-12986	0.10	0.51	-0.72	1.33
Denver, CO	1.88	2.16	356	1036	0.1067	-15127	-0.47	0.63	-1.18	1.68
Philadelphia, PA	2.69	2.37	168	718	0.0734	-12390	1.52	0.86	-0.08	2.67
Napa, CA	2.61	1.52	329	800	0.1092	-13712	0.06	0.23	-0.32	0.56
Phoenix, AZ	2.78	2.96	436	6969	0.2064	-20262	-2.24	1.92	-4.89	1.26
Newhall, CA	3.50	1.57	200	1214	0.0701	-16280	-0.06	0.31	-0.65	0.38
Springfield, MA	3.53	2.71	171	1736	0.0772	-18725	0.37	0.60	-0.95	1.32
Pasadena, CA	4.39	3.14	83	284	0.0371	-19518	-1.30	1.42	-3.39	1.27
New York City, NY	16.78	6.26	62	4010	0.0854	-28284	0.52	0.83	-0.42	2.62
ARITHMETIC MEAN			239	1778	0.0902	-15877	-0.12	0.79	-4.89[a]	2.67[b]

[a]Minimum of values listed in column.

[b]Maximum of values listed in column.

TABLE 35

SUMMARY OF RESULTS FOR THE METHOD OF MLE APPROXIMATION

CITY AND STATE	MODEL PARAMETERS		STATISTICAL TESTS				ENGINEERING ERROR TEST			
	Arithmetic Mean (ppm)	Arithmetic Standard Deviation (ppm)	Root Mean Square Of Difference	Chi Square	Kolmogorov-Smirnov Maximum Difference	Log Likelihood Function	Arithmetic Mean (ppm)	Arithmetic Standard Deviation (ppm)	Lowest Value (ppm)	Highest Value (ppm)
Barstow, CA	0.56	0.52	439	912	0.1011	- 6551	- 0.99	0.71	-1.79	0.12
Norfolk, VA	1.41	0.99	681	2183	0.1743	-11020	- 1.78	1.86	-5.25	0.35
Alexandria, VA	1.61	1.49	215	628	0.0797	-13072	- 1.27	0.74	-2.40	- 0.11
Denver, CO	2.08	3.14	207	475	0.0550	-14875	5.04	5.14	-0.21	17.75
Philadelphia, PA	2.46	2.21	217	617	0.0762	-12350	0.74	0.78	-0.86	1.92
Napa, CA	2.73	1.37	171	313	0.0372	-13422	- 0.41	0.55	-1.23	0.23
Phoenix, AZ	3.22	6.61	145	530	0.0585	-18406	14.41	19.56	-0.73	59.58
Newhall, CA	3.36	1.91	139	281	0.0384	-16002	0.81	1.02	-0.17	3.39
Springfield, MA	3.91	3.53	160	1094	0.0657	-18693	4.56	3.45	-0.31	9.52
Pasadena, CA	4.38	3.18	85	272	0.0344	-19519	-1.15	1.37	-3.11	1.87
New York City, NY	16.35	7.90	43	416	0.0458	-27710	3.73	4.62	-0.74	13.84
ARITHMETIC MEAN			227	702	0.0697	-15602	2.15	3.62	-5.25[a]	59.58[b]

[a]Minimum of values listed in column.

[b]Maximum of values listed in column.

TABLE 36

SUMMARY OF RESULTS FOR THE METHOD OF MLE OPTIMIZATION

CITY AND STATE	MODEL PARAMETERS		STATISTICAL TESTS				ENGINEERING ERROR TEST			
	Arithmetic Mean (ppm)	Arithmetic Standard Deviation (ppm)	Root Mean Square Of Difference	Chi Square	Kolmogorov-Smirnov Maximum Difference	Log Likelihood Function	Arithmetic Mean (ppm)	Arithmetic Standard Deviation (ppm)	Lowest Value (ppm)	Highest Value (ppm)
Barstow, CA	0.55	0.98	27	36	0.0051	- 6188	1.16	2.13	-0.15	5.67
Norfolk, VA	1.60	1.02	525	1465	0.0863	-10785	- 1.63	1.93	-5.22	0.60
Alexandria, VA	1.80	1.48	165	338	0.0377	-12879	- 1.38	1.04	-3.12	0.13
Denver, CO	2.09	3.35	190	460	0.0644	-14866	6.25	6.32	-0.25	21.65
Philadelphia, PA	2.40	2.00	211	612	0.0668	-12331	- 0.40	1.21	-3.24	1.13
Napa, CA	2.72	1.27	110	493	0.0271	-13393	- 0.75	0.86	-2.05	0.19
Phoenix, AZ	3.45	8.41	103	412	0.0416	-18366	21.82	28.50	-0.44	87.68
Newhall, CA	3.35	1.79	142	245	0.0339	-15979	0.45	0.62	-0.15	2.15
Springfield, MA	3.84	3.21	140	1334	0.0538	-18668	2.85	2.06	-0.27	5.45
Pasadena, CA	4.37	3.12	82	275	0.0351	-19517	- 1.36	1.44	-3.48	1.14
New York City, NY	16.32	7.71	43	430	0.0437	-27707	3.28	4.07	-0.70	12.18
ARITHMETIC MEAN			158	555	0.0450	-15516	2.75	4.56	-5.22[a]	87.68[b]

[a]Minimum of values listed in column.

[b]Maximum of values listed in column.

but one are greater than $\chi^2 = 100$, and, if hypothesis testing were valid (and taking into account the number of degrees of freedom), the LN2 model would be rejected at a high level of confidence ($P > 0.995$) for all cities and all parameter calculation procedures. The one data set for which chi-square is less than 100 is Barstow, CA ($\chi^2 = 36$ for the MLE optimization approach with 7 degrees of freedom), and, to avoid rejecting the LN2 hypothesis for Barstow at $P = 0.995$, chi-square would have to be less than $\chi^2 = 20.3$. Although chi-square gives values that are too large for all data sets, the "best" (that is, lowest) values of this frequency-based measure of goodness-of-fit always are obtained with the MLE optimization approach, and this approach clearly would be the preferred way to calculate (that is, "estimate") parameters if hypothesis tests were to be undertaken. Of course, hypothesis testing is questionable in view of the high serial correlation, and consequent lack of statististical independence, of the observations.

Tables 33-36 also briefly summarize the results obtained from the Engineering Error Test. The method of fractiles gave the smallest average difference between predicted and observed values ($\bar{d} = -0.12$ ppm), while the method of moments, MLE approximation, and MLE optimization approaches gave average overpredictions that were increasingly large ($\bar{d} = 0.26$ ppm, 2.15 ppm, and 2.75 ppm, respectively). The method of fractiles had a minimum difference of -4.89 ppm and a maximum difference of 2.67 ppm for all data sets (Table 34), indicating that this parameter calculation approach fits the observations quite well, according to

the variate-based concepts. The other three approaches had maxima of 18.97, 59.95, and 87.68 ppm, although their minima were similar to the minimum for the method of fractiles.

Another view of the EET results can be obtained by developing histograms of the differences for each parameter calculation approach, with all 11 cities grouped together. To do this, the differences were first listed in tabular form (Tables 37-40). Each column gives the count for the number of occurrences of differences within intervals of size 0.5 ppm; for example, the first interval in the center of the table represents the number of differences in the range (0 ppm $\leq$ d < 0.5 ppm). For each table, there was a total of 246 differences, and the numerical values of differences of 5 ppm or more were listed below the table.

The histograms generated from these tables (Figure 20) allow the EET results for the four approaches to be compared visually. It is seen that the histograms for all approaches but one, the method of fractiles, have long tails extending to the right, with very striking right tails for the two MLE approaches. For example, the MLE approximation approach has 17 differences that overpredict the concentration by 12 ppm or more, and the MLE optimization approach has 19 differences that overpredict the concentration by 12 ppm or more. In contrast, all 246 differences for the method of fractiles are less than 3.0 ppm, and this is the only approach showing a slight tendency to underpredict, reflected by the leftward skewness of the histogram. The method of fractiles also is unique for the overall "narrowness" of the

Table 37

NUMBER OF VALUES IN INTERVALS FOR ENGINEERING ERROR TEST WITH PARAMETERS ESTIMATED BY METHOD OF MOMENTS

	Xp-Xo (ppm)																						
CITY AND STATE	<-5	-5 -4.5	-4.5 -4	-4 -3.5	-3.5 -3	-3 -2.5	-2.5 -2	-2 -1.5	-1.5 -1	-1 -0.5	-0.5 0	0 0.5	0.5 1	1 1.5	1.5 2	2 2.5	2.5 3	3 3.5	3.5 4	4 4.5	4.5 5	>5	TOTAL
Barstow, CA	0	0	0	0	0	0	0	0	0	2	3	0	0	1	0	0	0	0	0	1	0	0	7
Norfolk, VA	0	0	1	1	1	1	1	1	1	1	3	3	0	0	0	0	0	0	0	0	0	0	14
Alexandria, VA	0	0	0	0	0	0	0	0	1	4	5	5	2	1	0	0	0	0	0	0	0	0	18
Denver, CO	0	0	0	0	0	0	0	0	0	0	12	4	0	4	0	1	2	0	0	0	1	0	24
Philadelphia, PA	0	0	1	0	1	1	4	1	0	2	3	4	4	0	0	0	0	0	0	0	0	0	21
Napa, CA	0	0	0	0	0	0	0	0	1	4	4	4	0	0	0	0	0	0	0	0	0	0	13
Phoenix, AZ	0	0	0	0	0	0	3	5	3	2	3	4	0	0	0	1	0	1	0	1	0	9	32
Newhall, CA	0	0	0	0	0	0	0	0	0	0	3	9	0	1	0	0	0	0	0	0	0	0	13
Springfield, MA	0	0	0	0	0	1	0	2	1	3	5	12	2	0	0	0	0	0	0	0	0	0	26
Pasadena, CA	0	0	0	0	0	0	0	3	3	7	4	8	1	0	0	0	0	1	0	0	0	1	28
New York City, NY	0	0	0	0	0	0	0	0	0	7	7	6	6	5	3	0	3	4	5	2	2	0	50
TOTAL	0	0	2	1	2	3	8	12	10	32	52	59	15	12	3	2	5	6	5	4	3	10	246
CUMULATIVE	0	0	2	3	5	8	16	28	38	70	122	181	196	208	211	213	218	224	229	233	236	246	

VALUES GREATER THAN ±5 ppm

Phoenix	
5.8	14.4
6.5	15.2
10.5	18.0
11.3	19.0
12.6	

Pasadena
5.4

Table 38

NUMBER OF VALUES IN INTERVALS FOR ENGINEERING ERROR TEST WITH PARAMETERS ESTIMATED BY THE METHODS OF FRACTILES

	Xp-Xo (ppm)																						
CITY AND STATE	<-5	-5 -4.5	-4.5 -4	-4 -3.5	-3.5 -3	-3 -2.5	-2.5 -2	-2 -1.5	-1.5 -1	-1 -0.5	-0.5 0	0 0.5	0.5 1	1 1.5	1.5 2	2 2.5	2.5 3	3 3.5	3.5 4	4 4.5	4.5 5	>5	TOTAL
Barstow, CA	0	0	0	0	0	0	0	0	3	1	2	1	0	0	0	0	0	0	0	0	0	0	7
Norfolk, VA	0	0	0	0	0	0	0	0	0	0	3	2	1	4	4	0	0	0	0	0	0	0	14
Alexandria, VA	0	0	0	0	0	0	0	0	0	2	5	8	2	1	0	0	0	0	0	0	0	0	18
Denver, CO	0	0	0	0	0	0	0	0	4	10	6	3	0	0	1	0	0	0	0	0	0	0	24
Philadelphia, PA	0	0	0	0	0	0	0	0	0	0	1	3	1	6	3	5	2	0	0	0	0	0	21
Napa, CA	0	0	0	0	0	0	0	0	0	0	5	7	1	0	0	0	0	0	0	0	0	0	13
Phoenix, AZ	0	6	1	4	4	1	1	2	3	2	3	3	1	1	0	0	0	0	0	0	0	0	32
Newhall, CA	0	0	0	0	0	0	0	0	0	1	4	8	0	0	0	0	0	0	0	0	0	0	13
Springfield, MA	0	0	0	0	0	0	0	0	0	1	6	8	6	5	0	0	0	0	0	0	0	0	26
Pasadena, CA	0	0	0	0	6	1	3	4	1	2	4	6	0	1	0	0	0	0	0	0	0	0	28
New York City, NY	0	0	0	0	0	0	0	0	0	0	16	13	11	2	3	4	1	0	0	0	0	0	50
TOTAL	0	6	1	4	10	2	4	6	11	19	55	62	23	20	11	9	3	0	0	0	0	0	246
CUMULATIVE	0	6	7	11	21	23	27	33	44	63	118	180	203	223	234	243	246	246	246	246	246	246	

VALUES GREATER THAN ±5 ppm

Table 39

NUMBER OF VALUES IN INTERVALS FOR ENGINEERING ERROR TEST WITH PARAMETERS ESTIMATED BY THE METHOD OF MLE APPROXIMATION

	Xp-Xo (ppm)																						
CITY AND STATE	<-5	-5 -4.5	-4.5 -4	-4 -3.5	-3.5 -3	-3 -2.5	-2.5 -2	-2 -1.5	-1.5 -1	-1 -0.5	-0.5 0	0 0.5	0.5 1	1 1.5	1.5 2	2 2.5	2.5 3	3 3.5	3.5 4	4 4.5	4.5 5	>5	TOTAL
Barstow, CA	0	0	0	0	0	0	0	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0	7
Norfolk, VA	1	0	1	1	1	1	1	1	0	2	3	2	0	0	0	0	0	0	0	0	0	0	14
Alexandria, VA	0	0	0	0	0	0	3	5	4	2	4	0	0	0	0	0	0	0	0	0	0	0	18
Denver, CO	0	0	0	0	0	0	0	0	0	0	2	3	2	2	1	0	2	0	1	1	0	10	24
Philadelphia, PA	0	0	0	0	0	0	0	0	0	1	2	7	2	5	4	0	0	0	0	0	0	0	21
Napa, CA	0	0	0	0	0	0	0	0	3	2	5	3	0	0	0	0	0	0	0	0	0	0	13
Phoenix, AZ	0	0	0	0	0	0	0	0	0	5	6	2	0	1	1	0	1	0	0	1	0	15	32
Newhall, CA	0	0	0	0	0	0	0	0	0	0	2	4	3	2	0	1	0	1	0	0	0	0	13
Springfield, MA	0	0	0	0	0	0	0	0	0	0	3	3	0	1	1	1	1	1	1	0	1	13	26
Pasadena, CA	0	0	0	0	2	5	2	4	2	1	4	7	0	0	1	0	0	0	0	0	0	0	28
New York City, NY	0	0	0	0	0	0	0	0	0	7	6	6	5	2	1	1	2	0	2	0	1	17	50
TOTAL	1	0	1	1	3	6	6	12	11	21	38	38	12	13	9	3	6	2	4	2	2	55	246
CUMULATIVE	1	1	2	3	6	12	18	30	41	62	100	138	150	163	172	175	181	183	187	189	191	246	

VALUES GREATER THAN ±5ppm

Norfolk	Denver		Phoenix			Springfield			New York		
-5.3	5.4	9.5	6.2	20.1	41.9	5.4	7.6	8.5	5.8	8.7	11.5
	5.9	11.4	7.8	23.7	47.1	5.5	8.0	8.5	6.4	8.9	11.6
	7.5	13.1	9.1	26.1	50.0	6.3	8.2	8.8	6.8	9.4	11.8
	8.3	13.9	14.3	35.3	58.6	7.0	8.4	9.5	7.3	10.3	12.4
	8.4	17.8	16.8	38.1	59.6	7.1			8.1	10.8	13.8
									8.7	11.1	

Table 40

NUMBER OF VALUES IN INTERVALS FOR ENGINEERING ERROR TEST WITH PARAMETERS ESTIMATED BY THE METHOD OF MLE OPTIMIZATION

	Xp-Xo (ppm)																						
CITY AND STATE	<-5	-5 -4.5	-4.5 -4	-4 -3.5	-3.5 -3	-3 -2.5	-2.5 -2	-2 -1.5	-1.5 -1	-1 -0.5	-0.5 0	0 0.5	0.5 1	1 1.5	1.5 2	2 2.5	2.5 3	3 3.5	3.5 4	4 4.5	4.5 5	>5	TOTAL
Barstow, CA	0	0	0	0	0	0	0	0	0	0	2	2	1	0	1	0	0	0	0	0	0	1	7
Norfolk, VA	1	0	1	1	1	1	1	1	0	1	2	3	1	0	0	0	0	0	0	0	0	0	14
Alexandria, VA	0	0	0	0	2	1	0	8	1	1	3	2	0	0	0	0	0	0	0	0	0	0	18
Denver, CO	0	0	0	0	0	0	0	0	0	0	2	3	1	2	1	1	0	1	1	0	1	11	24
Philadelphia, PA	0	0	0	0	1	0	1	1	4	1	3	4	5	1	0	0	0	0	0	0	0	0	21
Napa, CA	0	0	0	0	0	0	1	3	1	2	2	4	0	0	0	0	0	0	0	0	0	0	13
Phoenix, AZ	0	0	0	0	0	0	0	0	0	0	7	3	1	1	0	1	0	1	1	0	0	17	32
Newhall, CA	0	0	0	0	0	0	0	0	0	0	3	6	2	1	0	1	0	0	0	0	0	0	13
Springfield, MA	0	0	0	0	0	0	0	0	0	0	4	2	1	1	2	1	1	1	2	3	5	3	26
Pasadena, CA	0	0	0	0	6	2	3	3	1	2	4	6	0	1	0	0	0	0	0	0	0	0	28
New York City, NY	0	0	0	0	0	0	0	0	0	6	6	7	5	3	1	2	0	2	0	1	0	17	50
TOTAL	1	0	1	1	10	4	6	16	7	13	38	42	17	10	5	6	1	5	4	4	6	49	246
CUMULATIVE	1	0	2	3	13	17	23	39	46	59	97	139	156	166	171	177	178	183	187	191	197	246	

VALUES GREATER THAN ±5ppm

Barstow	Norfolk	Denver		Phoenix			Springfield	New York		
5.7	-5.2	5.6	11.9	5.5	25.3	61.6	5.3	5.1	7.7	10.1
		6.8	14.1	7.6	30.0	69.3	5.3	5.7	7.8	10.1
		7.3	16.2	10.3	35.2	73.8	5.4	6.0	8.3	10.5
		9.2	17.1	12.7	38.8	86.8		6.5	9.1	10.9
		10.3	21.7	14.6	51.7	87.8		7.1	9.5	12.2
		10.5		21.7	56.0			7.7	9.7	

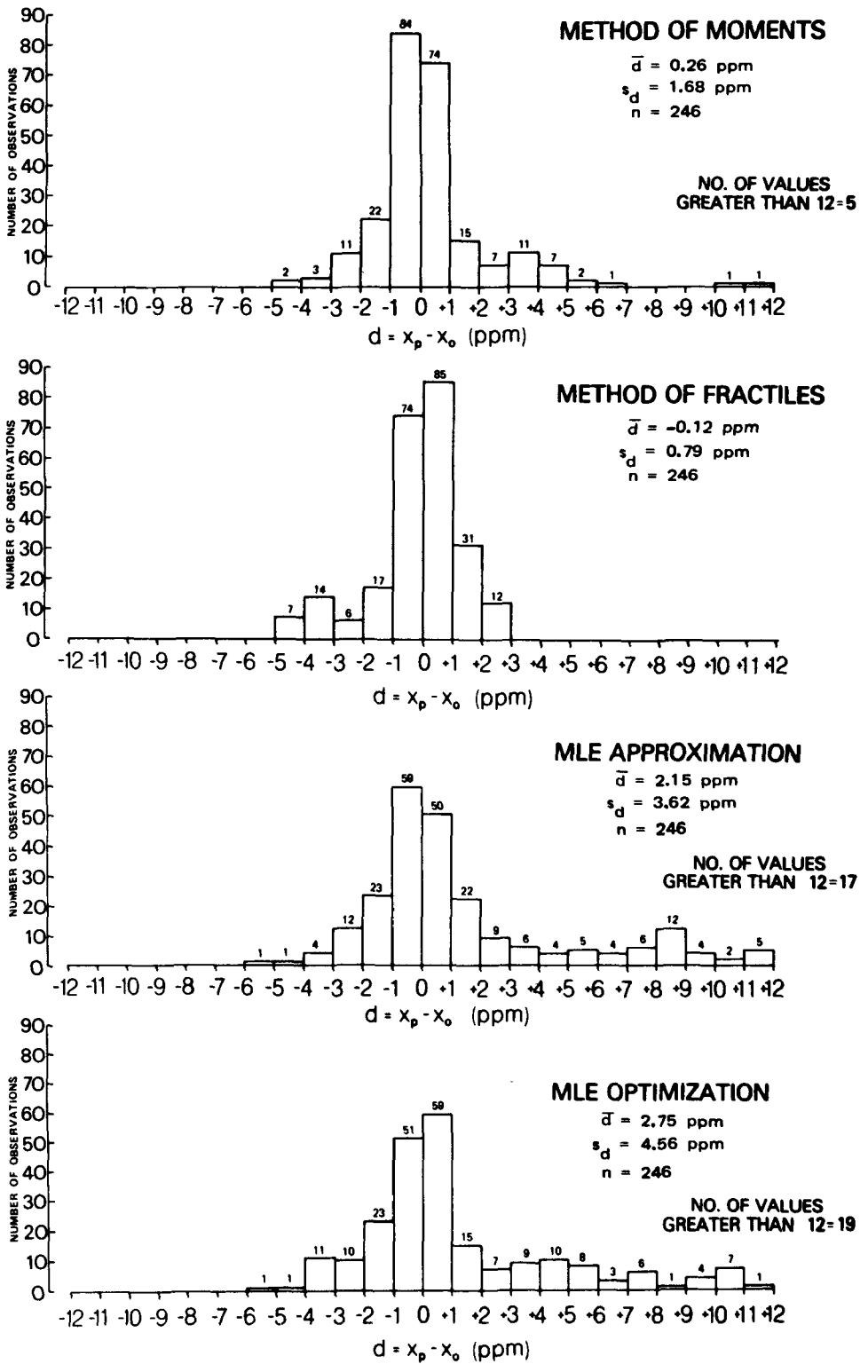

Figure 20. Histograms of the deviations between predicted and observed concentrations for four different methods of calculating LN2 parameters for the geographical data group.

histogram.

These results also can be compared graphically by plotting the cumulative counts of the EET differences (Figure 21). For ease of plotting, curves have been drawn through the points. All curves lie close together for differences to the left of d = 0.0 ppm, but the curves show distinct variations for positive differences. The method of fractiles has a very small projection to the right, but the two MLE approaches have striking rightward tails and almost coincide. The curve for the method of moments is intermediate between the other curves but lies fairly close to the method of fractiles.

To determine the percentage of the differences which lie within stated error ranges, we examine the bottom row marked "CUMULATIVE" in each of Tables 37-40. For example, to find the percentage of the differences that fall in the range $\pm$ 2.0 ppm for the Larsen method of fractiles, we read from Table 38 that 27 differences were less than d = -2.0 ppm and 234 differences were less than d = +2.0 ppm. By subtracting the two values, we determine that 234-27 = 207 differences were within the range d = $\pm$ 2.0 ppm. Because the table contains 246 differences, we conclude that 207/246 = 84% of the differences were between $\pm$ 2.0 ppm for the method of fractiles.

It is possible, in the same fashion, to construct a table for various "error ranges" which summarizes the percentage of the differences falling into various ranges (Table 41). It is seen that that the first two techniques -- the method of moments and the method of fractiles -- have similar error

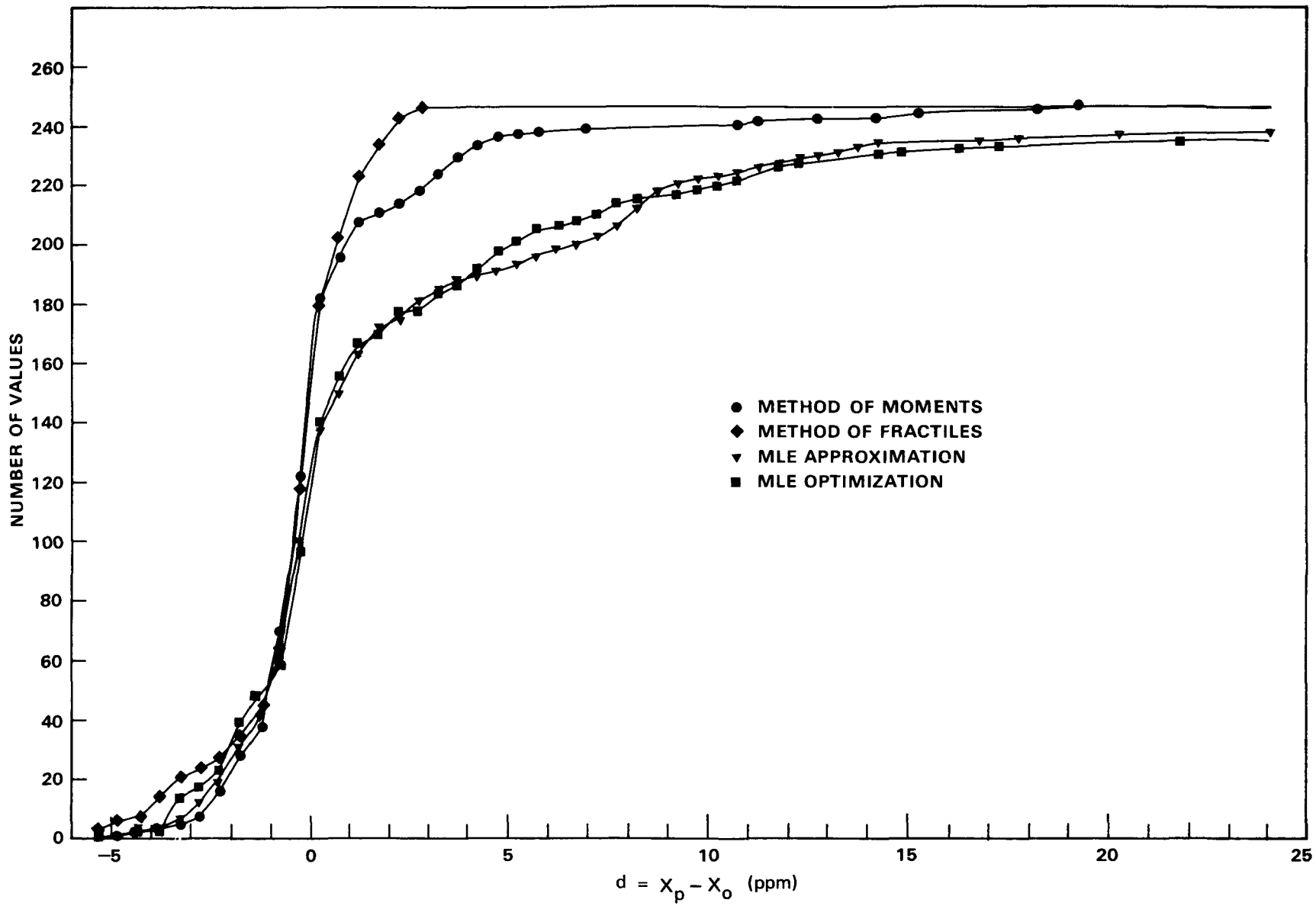

Figure 21. Cumulative counts of the number of EET differences for four methods of calculating the LN2 parameters on all 11 cities comprising the geographical data group.

Table 41

Percent of LN2 Predictions Within Stated Error Ranges

	CO Concentration (ppm)					
	±0.5	±1.0	±1.5	±2.0	±3.0	±4.0
Method of Moments	45.1%	66.3%	72.3%	79.3%	86.6%	92.3%
Method of Fractiles	47.6%	64.6%	77.2%	84.1%	91.5%	97.2%
MLE Approximation	30.9%	44.3%	54.1%	62.6%	71.1%	75.2%
MLE Optimization	32.5%	44.7%	51.6%	60.2%	67.1%	76.8%

although the method of fractiles performs slightly better than the method of moments. The results for the last two approaches are quite similar to each other and are distinctly different from the first two approaches. If the analyst intends to represent the frequency distribution of a particular data set by the LN2 model, he can gain some insight from this table. The NDIR instrument used to make CO measurements customarily is assumed to have a measurement error on the order of ± 1 ppm (hourly averages). If the analyst wishes the model's predictions to be within ± 1.0 ppm over the entire range of concentrations, it can be seen from Table 41 that all methods perform rather poorly. Only two-thirds of the predictions lie within this range for the Larsen method of fractiles, and fewer than half of the predictions are within ± 1.0 of the observed values for the two MLE approaches. If the allowable error range is expanded to ± 1.5 ppm, then 77.2% of the predictions using Larsen's approach will be considered successful. Finally, with a ± 2.0 ppm error band, 84.1% of the predictions from the method of fractiles are successful, and 60.2% of the predictions from MLE optimization are successful. It is not until the error band is expanded to ± 3.0 ppm that we can state that at least 90% of the predictions are successful, and, even with this relatively wide error band, only 71% of the predictions from the MLE optimization technique are successful.

Discussion

Whether the above error bands are considered "too wide" or

"acceptable" is a matter of opinion and should be determined by the purpose for which the model is intended. It should be emphasized that the word "prediction," as used here, does not mean predicting future values and should be interpreted in the following narrow sense.

Suppose that the analyst obtains a CO data set containing 8760 hourly observations (ignoring missing values), and he wishes to represent the year of data as parsimoniously as possible. If he partitions the observations into 50 intervals and makes a histogram, he can represent the frequency distribution of the data set by just 50 numerical values: the counts of the number of observations in each interval. In effect, 8760 numerical values have been reduced to just 50 numerical values, a significant reduction. Although all information on the time series of the observations has been removed, the histogram retains information that is very useful for enforcement purposes. If the analyst now wishes to apply the LN2 model to the data set, he can reduce the information he must store to just two numerical values: the two parameters of the LN2 model. By doing so, he has reduced 8760 numerical values to just 2 numerical values. With the model and the two parameters, he can reconstruct the entire histogram. Great parsimony has been attained, and the cost is paid in terms of the error that may result. As we have seen, the LN2 model can be used to predict the concentrations for 90% of the intervals within $\pm$ 3 ppm and for 84% of the intervals within $\pm$ 2 ppm for the method of fractiles. Because the geographical data group was selected as a

cross section of U.S. CO data sets in 1974, it is likely that the LN2 model, with parameters calculated by the method of fractiles, will attain the same performance for other data sets in the population.

Notice that the data analyst, because he is "predicting" concentrations for a data set that is available to him, can determine immediately how well his model fits the observations. Thus, the goodness-of-fit tests described in this report can immediately be used to see how well the LN2 model predicts the true concentrations. It is important to stress that a model applied in this manner should not be used to predict the future concentrations for the following year, nor should it be used to assess the impact of control strategies. Variate-based tests are useful techniques for assessing how well a model describes the concentrations of a particular data set, which is one realization of some underlying process. In order to use a model for predictions, we must be certain that the basic distribution of the process has been modeled correctly, not just one realization. Thus, it is possible, even likely, that the MLE optimization approach gives the best estimate of the underlying distribution of the process, even though it may not perform well on a given realization when judged by variate-based tests. Of course, an even more likely result is that the LN2 model is the wrong distribution for modeling air quality data, and more complex models are needed to describe the underlying process. At the present time, we are inclined toward this view.

IX. RESULTS FOR THE LONGITUDINAL DATA GROUP

The longitudinal data group was included in this study to enable us to examine the distributions of hourly CO observations at a single air monitoring station in successive years. The data were obtained from SAROAD for the Continuous Air Monitoring Project (CAMP) station in Washington, D.C., and covered the period from 1966 to 1974 (Table 42). However, data for 1967 were not available from SAROAD, and the data set for 1974 included only several months of observations and therefore was not analyzed in detail. Although the data sets for 1972 and 1973 were incomplete, they were included in the analysis. Over the eight years of data, the arithmetic means were relatively similar, ranging from 3.31 ppm to 4.35 ppm. The arithmetic standard deviations also were close to each other, ranging from 2.47 ppm to 3.80 ppm.

This chapter first presents the findings for the individual years; then the overall findings are discussed briefly. The numerical analysis for the longitudinal data group is not quite as detailed as that for the geographical data group (Chapter VIII).

FINDINGS FOR INDIVIDUAL YEARS

The results for the individual years, from 1966 to 1973, are discussed in the following sections.

TABLE 42

CO DATA SETS FROM THE CAMP STATION IN WASHINGTON, DC (LONGITUDINAL DATA GROUP)

YEAR	NO. OF HOURLY OBSERVATIONS	ARITHMETIC MEAN (ppm)	STANDARD DEVIATION (ppm)	COEFFICIENT OF SKEWNESS, $\sqrt{b_1}$	COEFFICIENT OF KURTOSIS, b_2	MINIMUM (ppm)	MAXIMUM (ppm)
1966	7822	3.3108	2.4727	3.2670	23.7206	0.0	38.0
1968	7266	3.4140	2.6232	2.4363	13.7693	0.0	27.0
1969	6056	3.5416	2.3929	2.3007	12.9541	0.0	25.0
1970	7748	3.4907	2.5859	3.8101	37.5566	0.0	45.0
1971	7254	3.4888	2.5706	3.6183	31.1403	0.0	46.0
1972	4947*	4.3499	3.7744	3.0540	20.7536	0.0	55.0
1973	4269*	3.6798	3.1338	3.5097	30.6102	0.0	50.0
1974	1848*	3.8825	3.8006	1.7992	8.0855	0.0	26.0

*Incomplete Data Sets

Table 43
Frequency-based Measures of Goodness-of-fit for Washington, D.C. - 1966

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	2.653	1.946	16	124	520	0.040	-15881
Fractiles	2.798	1.977	18	108	575	0.048	-15903
MLE Approx.	2.664	1.987	17	134	497[1]	0.046	-15893
MLE Opt.	2.689	1.922	16	107[1]	589	0.030[1]	-15877[1]

[1]Indicates best result in each column.

1966

The observations for 1966, when plotted on logarithmic probability paper (Figure 22), appear to scatter narrowly about a straight line. Because the parameters that were calculated by the four methods were nearly identical, the resulting LN2 lines all had similar slopes and vertical intercepts, and, when

Table 44
Variate-based (EET) Measures of Goodness-of-fit for Washington, D.C. - 1966
(ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	25	-0.786	1.110	-2.79	0.61
Fractiles	25	0.356	0.721[1]	-0.95[1]	1.74
MLE Approx.	25	-0.110[1]	0.783	-1.62	0.96
MLE Opt.	25	-0.985	1.267	-3.19	0.55[1]

[1]Indicates best result in each column.

plotted, almost coincided. The medians ($\hat{m}$) were very similar to each other, as were the standard geometric deviations (s_g), and the resulting frequency-based goodness-of-fit measures also were very close in value (Table 43). Although the differences among the measures were small, three of the four frequency-based measures indicated that the MLE optimization approach gave the best fit. If hypothesis tests were undertaken, the horizontal scatter about the lines in Figure 22, although small to the eye, would be too large to be caused by chance alone, assuming that we really were sampling from a lognormal distribution. This can be seen from the relatively high values of chi-square (χ^2) and the Kolmogorov-Smirnov maximum difference (K-S max). The variate-based measures of goodness-of-fit also are very close to each other, with the method of fractiles having a slight, though debatable, advantage over the other three methods. According to these results, the LN2 model, with parameters calculated by the method of fractiles, can represent the entire distribution with a minimum underprediction of 0.95 ppm and a maximum overprediction of 1.74 ppm, and an average overprediction of only 0.36 ppm. It appears that the LN2 model could be used to represent this particular data set adequately for many practical applications.

1968

The Washington, D.C., data for 1968, when plotted on logarithmic-probability paper (Figure 23), appear very similar to the 1966 data (Figure 22), and the similarity suggests that they may arise from the same underlying distribution. The medians for the

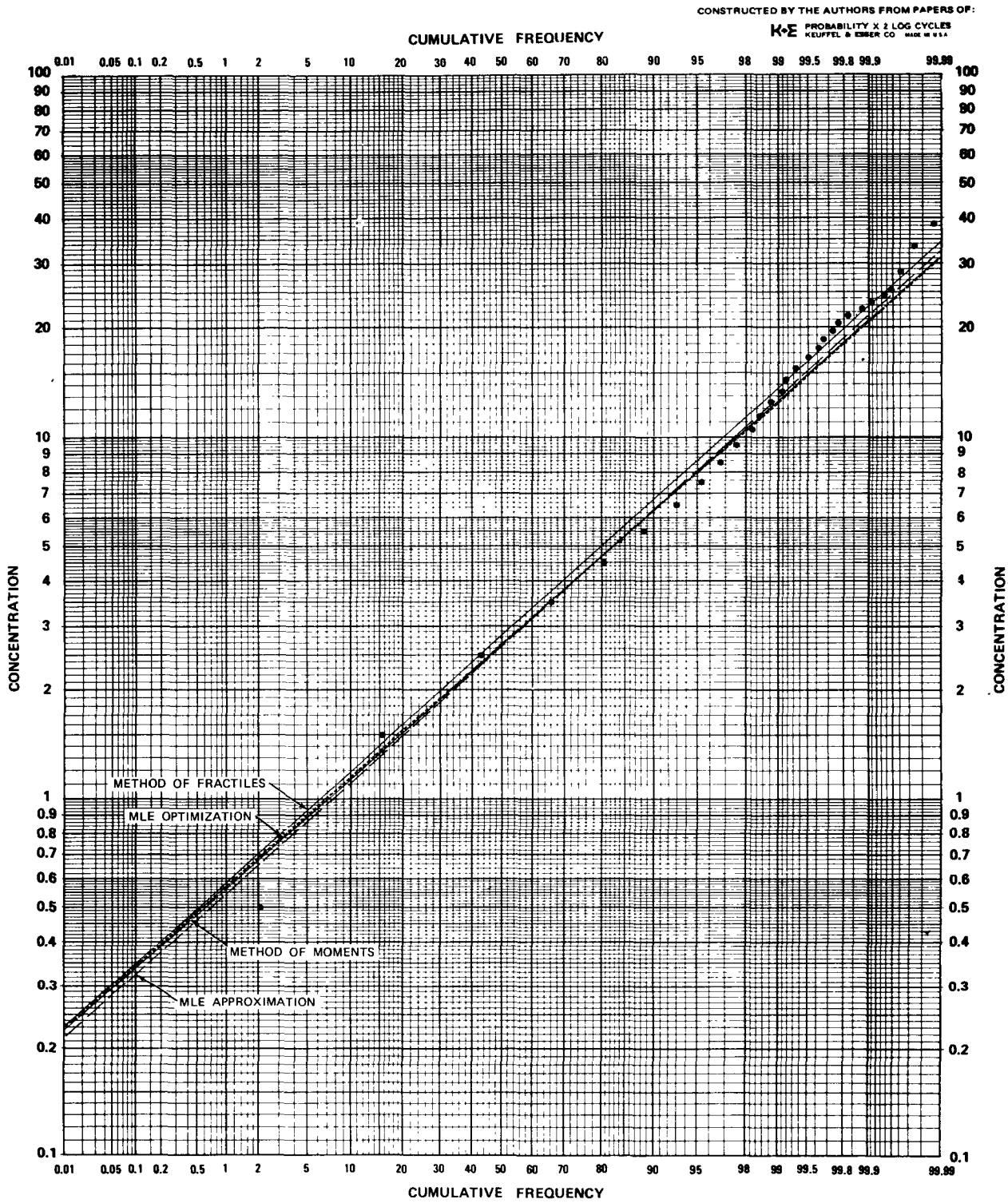

FIGURE 22. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR WASHINGTON DC, 1966 (n = 7822).

Table 45

Frequency-based Measures of Goodness-of-fit for Washington, D.C. - 1968

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	2.707	1.976	17	67	541	0.023	-15499
Fractiles	2.823	1.943	17	65[1]	853	0.028	-15547
MLE Approx.	2.640	2.132	19	91	309[1]	0.035	-15476
MLE Opt.	2.662	2.072	18	80	327	0.023[1]	-15468[1]

[1]Indicates best result in each column.

two data sets are very similar ($\hat{m}$ between 2.6 and 2.8 ppm), and their standard geometric deviations are quite close (s_g between 1.9 and 2.1 ppm). However, the 1968 data show some tendency toward concave downward curvature while the 1966 data do not. The frequency-based goodness-of-fit measures (Table 45) give much

Table 46

Variate-based (EET) Measures of Goodness-of-fit for Washington, D.C. - 1968 (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	25	-0.101	0.682	-1.36	1.48
Fractiles	25	-0.101[1]	0.678[1]	-1.45	1.17[1]
MLE Approx.	25	2.147	2.144	-0.13[1]	7.69
MLE Opt.	25	1.237	1.393	-0.16	5.16

[1]Indicates best result in each column.

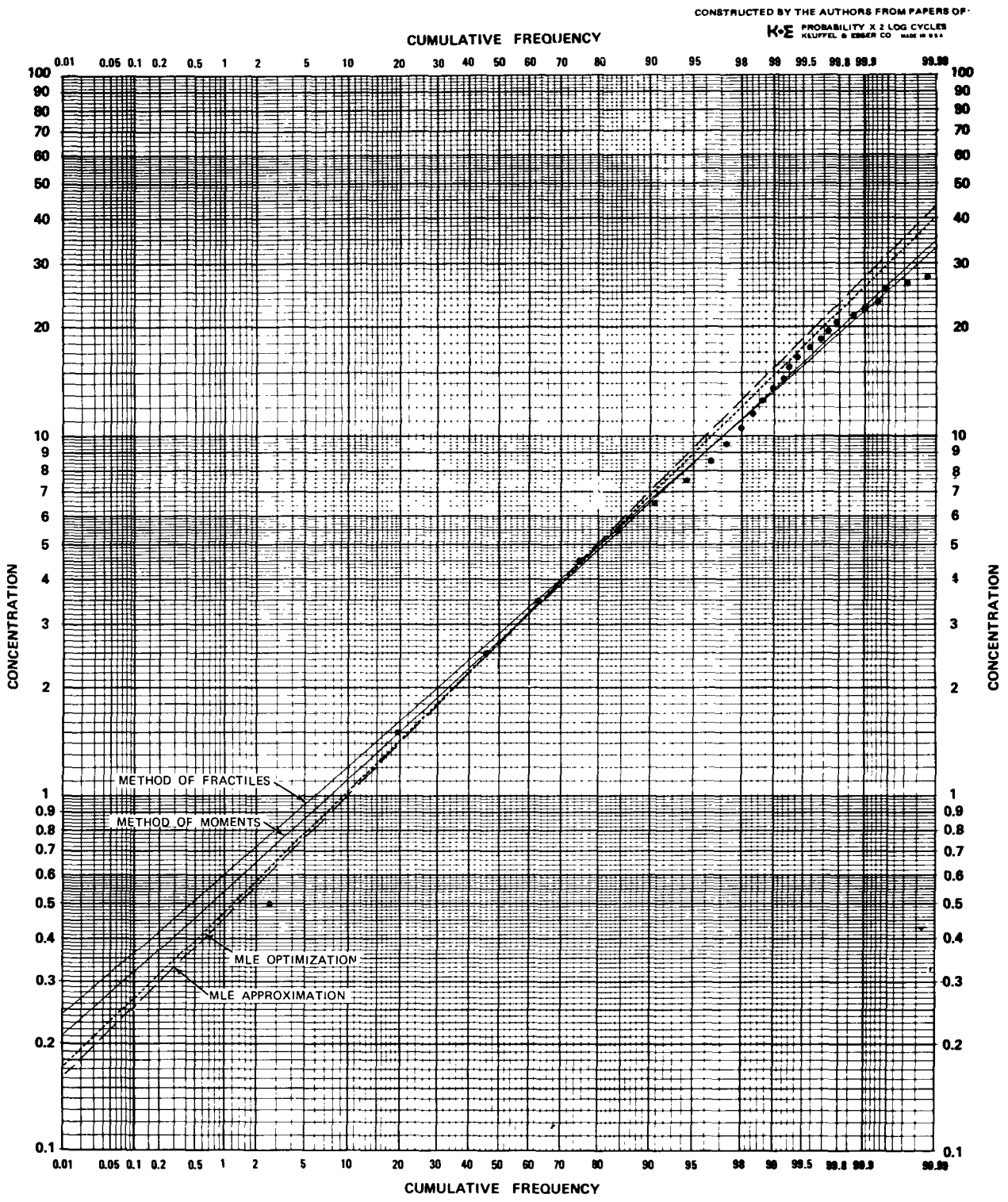

FIGURE 23. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR WASHINGTON, DC, 1968 (n = 7266).

Table 47

Frequency-based Measures of Goodness-of-fit for Washington, D.C. - 1969

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	2.935	1.846	15	103[1]	2547	0.046[1]	-12969
Fractiles	2.879	1.872	15	118	1933	0.060	-12951
MLE Approx.	2.862	2.041	18	147	959[1]	0.074	-12955
MLE Opt.	2.899	1.947	17	124	1308	0.061	-12931[1]

[1]Indicates best result in each column.

lower values for the two MLE approaches in the 1968 data set than they do in the 1966 data set. This occurs because the observations in the lower and central ranges of the 1968 data (cumulative frequencies of 2% to 80%) happen to fit the model quite well for the MLE approaches, giving correspondingly lower values of chi-square and the Kolmogorov-Smirnov maximum differ-

Table 48

Variate-based (EET) Measures of Goodness-of-fit for Washington, D.C. - 1969 (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	22	-0.306	0.636	-1.42	0.52[1]
Fractiles	22	-0.148[1]	0.563[1]	-1.19	1.07
MLE Approx.	22	2.316	2.118	-0.32	7.80
MLE Opt.	22	1.014	1.010	-0.25[1]	4.13

[1]Indicates best result in each column.

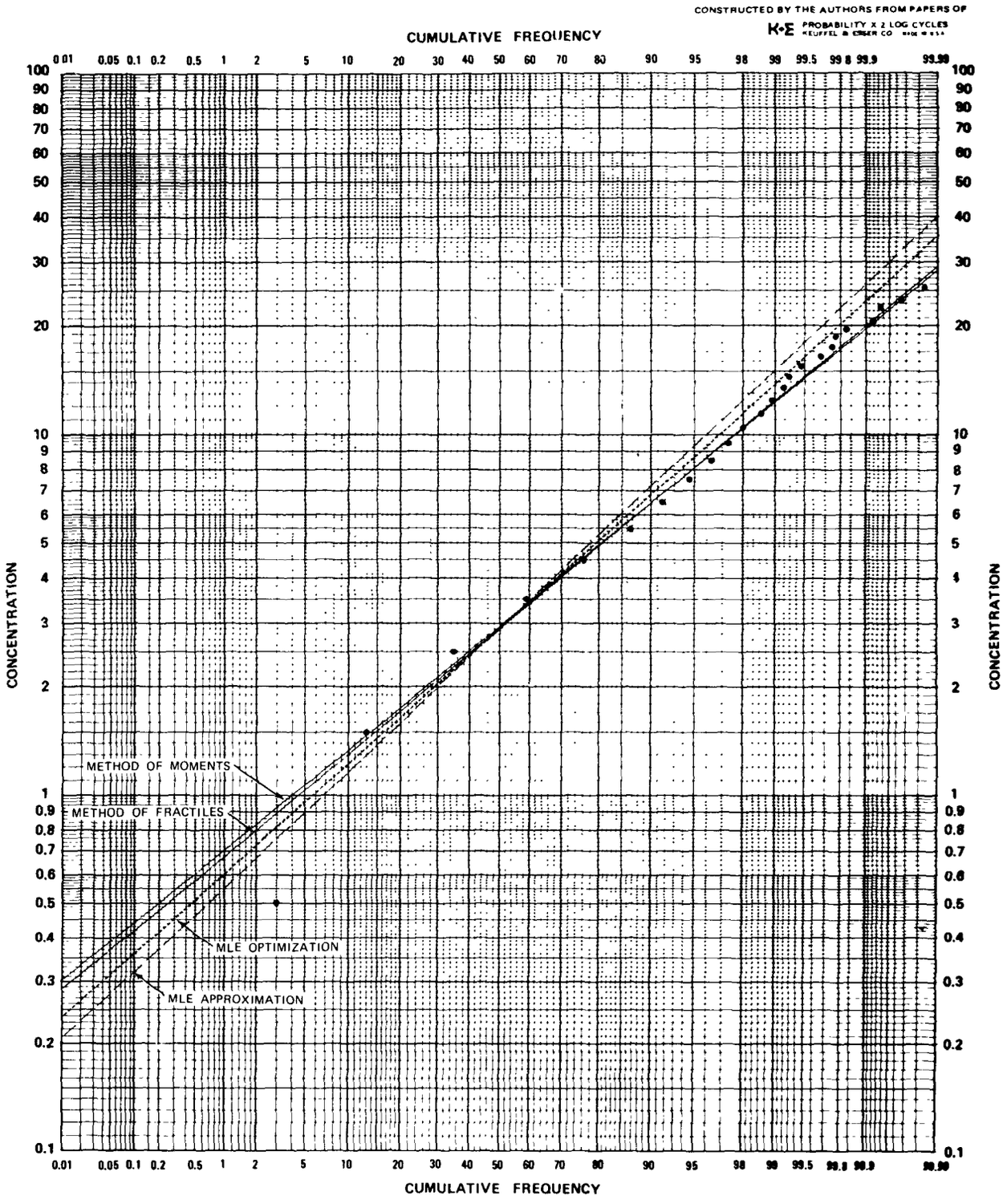

FIGURE 24. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR WASHINGTON, DC, 1969 (n = 6056).

ence. According to the variate-based EET measures (Table 46), the method of fractiles gives the best results, but the values are so similar to those for the method of moments that it would be difficult to choose the method of fractiles over the method of moments. The relatively large overpredictions for the MLE approximation and MLE optimization approaches (7.69 ppm and 5.16 ppm, respectively) occur because of the downward curvature at the higher concentrations in 1968. As discussed in the previous chapter, downward curvature at the extremes has little influence on the computed values of the MLE parameters, causing a characteristic overprediction when evaluated by the EET tests.

1969

The observed CO values for Washington, D.C., CAMP in 1969, when plotted on logarithmic-probability paper (Figure 24), appear quite similar to the 1966 and 1968 data sets. As given in Table 47, the medians (approximately $\hat{m}$ = 2.9 ppm) are slightly higher than those for the 1966 and 1968 data, and the standard geometric deviations (approximately s_g = 1.9 ppm) are almost identical to those for the 1966 and 1968 data. As with the 1966 data set, the observations do not appear to have definite curvature in either direction, although the scatter of the points about the four straight lines is larger. As a result of the scatter, the values of χ^2 and K-S max for 1969 are larger than those for 1966 and 1968. According to the frequency-based measures of goodness-of-fit (Table 47), no one approach emerges as clearly superior, partly because the two parameters for the LN2 model are nearly

the same for every approach. Surprisingly, the method of moments gives the best fit in two of the categories (RMS and K-S max). According to the EET variate-based measures (Table 48), no single approach gives the best fit, although the method of moments and the method of fractiles give fairly good fits. For the method of fractiles, the LN2 model predicts the observed hourly CO concentration over the full range of the observations within 2.3 ppm ($d_{max} - d_{min} = 1.07 - [-1.19] = 2.26$ ppm), with an average underprediction of less than 0.2 ppm ($\bar{d} = -0.148$ ppm) and a standard deviation of only 0.6 ppm ($s_d = 0.563$ ppm).

1970

The 1970 data from Washington, D.C., CAMP, when plotted on logarithmic-probability paper (Figure 25), also resemble the data for the previous years (1966, 1968, and 1969). As reported in Table 49, the value of the model's median (approximately $\hat{m}$ = 2.8 ppm) and standard geometric deviation (approximately s_g = 2.0 ppm) are nearly identical to those in the previous years, and the plot of the four straight lines resembles the previous plots very closely. If we consider the characteristics that all four of these data sets (1966, 1968, 1969, and 1970) share in common, we note that, in all cases, 98% of the data are 10 ppm or less, and all medians lie in the narrow range of 2.64 ppm $\leq \hat{m} \leq$ 2.94 ppm. Thus, all four of these distributions at the CAMP station appear quite similar.

The frequency-based measures of goodness-of-fit for the 1970 data (Table 49) do not identify any one approach as clearly

Table 49

Frequency-based Measures of Goodness-of-fit for Washington, D.C. - 1970

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	2.805	1.937	17	143[1]	1897	0.050[1]	-16623
Fractiles	2.668	2.164	20	216	1086[1]	0.088	-16699
MLE Approx.	2.773	2.093	20	181	1120	0.063	-16638
MLE Opt.	2.811	1.996	18	155	1450	0.051	-16609[1]

[1]Indicates best result in each column.

superior. The two lines for the method of moments and the method of MLE optimization are very close together and therefore produce similar values of K-S max and the log-likelihood function (L*). Although the method of fractiles performs poorly with three of the four frequency-based measures, surprisingly it gives the best (lowest) value of chi-square (χ^2 = 1086). Of course, this chi-

Table 50

Variate-based (EET) Measures of Goodness-of-fit for Washington, D.C. - 1970 (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	34	-2.235	3.331	-8.95	1.00[1]
Fractiles	34	2.038	1.808[1]	-0.83[1]	4.51
MLE Approx.	34	1.048	1.931	-2.77	3.54
MLE Opt.	34	-0.955[1]	2.687	-6.54	2.01

[1]Indicates best result in each column.

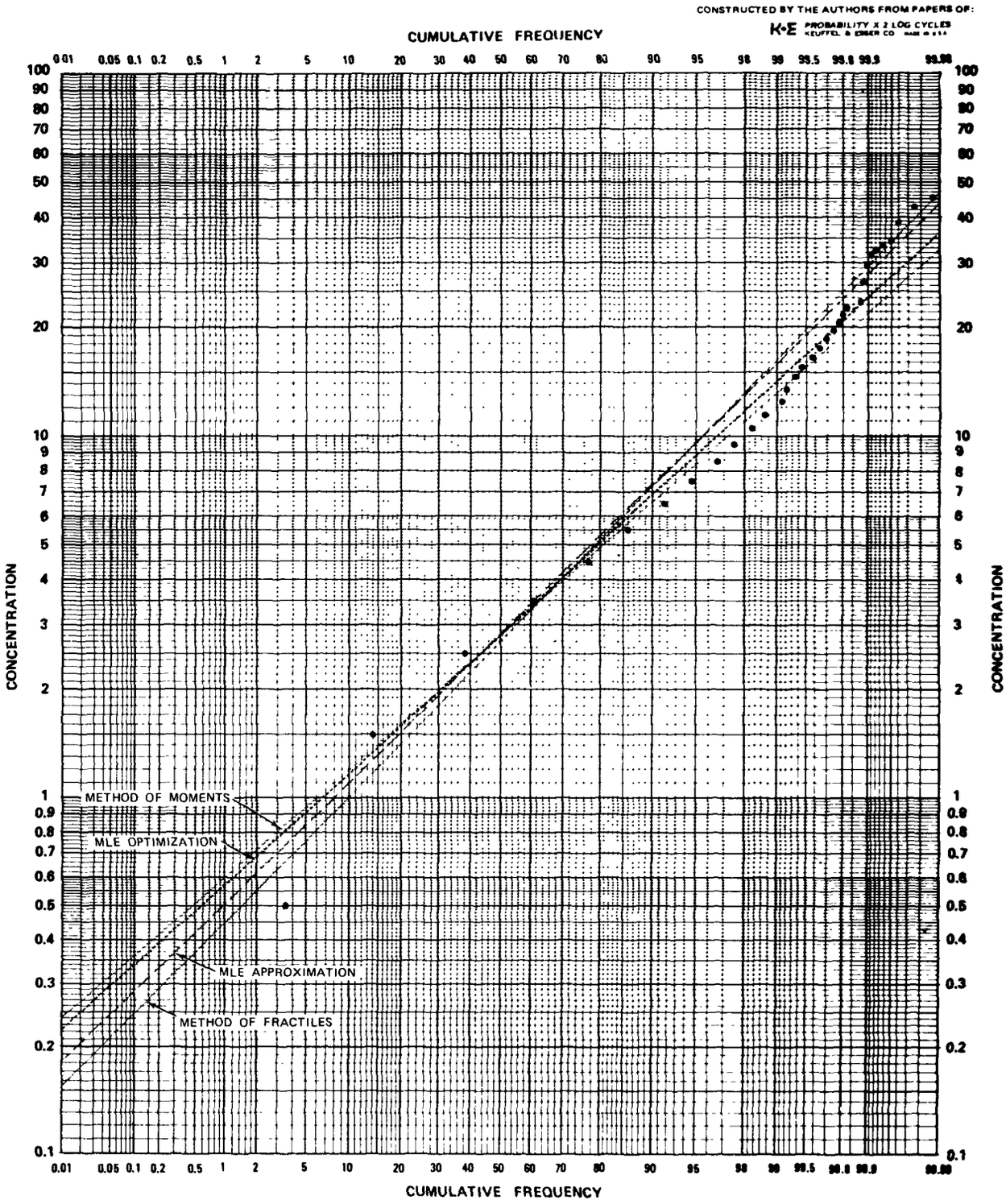

FIGURE 25. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR WASHINGTON, DC, 1970 (n = 7748).

square value, like the others in Table 49, is quite large, and, if hypothesis test were undertaken, would strongly lead to rejection of the lognormal hypothesis. The EET variate-based measures (Table 50) tend to indicate that the method of fractiles has the smallest standard deviation (s_d = 1.808 ppm) and the least serious underprediction (d_{min} = -0.83 ppm), but it has the most serious overprediction (d_{max} = 4.51 ppm) and the second highest average overprediction ($\bar{d}$ = 2.038 ppm). Overall, the goodness-of-fit measures suggest that the observations exhibit scatter about a straight line that is somewhat greater than might be expected by chance alone and that no one parameter calculation approach is clearly superior for this data set.

1971

The data for 1971, when plotted on logarithmic-probability paper (Figure 26), once again resemble the plots for the previous years at this station (1966, 1968, 1970, and 1971). As seen in Table 51, the median of the LN2 model (approximately $\hat{m}$ = 2.8 ppm) and standard deviation of the model (approximately s_g = 1.9 ppm) lie in the same range as those for the other four years. According to the frequency-based measures of goodness-of-fit (Table 51), the MLE optimization approach gives the best performance in all four categories (RMS, χ^2, K-S max, and L*). Conversely, the variate-based measures of goodness-of-fit (Table 52) indicate that the method of fractiles gives the best fit. The superior performance of the method of fractiles with this data set occurs because the data points above 10 ppm describe an upward curve, and

the method of fractiles fits this upward curve at a cumulative frequency of 99.9%. Because the EET measures are more strongly influenced by the tail of the distribution than the body of the distribution, they indicate that the method of fractiles performs quite well. The other three approaches are influenced more by the body of the distribution and therefore do not perform as well when judged by the EET measures.

1972

When the CAMP data for 1972 are plotted on logarithmic-probability paper (Figure 27), the resulting plot stands in sharp contrast with the plots for the period 1966-1971. The lines for the four parameter calculation approaches have considerable spread, and each line has a different slope. In addition, a sharp discontinuity, or "bump", appears at the higher concentrations. Finally, as indicated by Table 53, the concentrations for this data set are higher than those for the previous five years; the medians for the four approaches range from 3.0 ppm to 3.5 ppm, with an average value of 3.25 ppm, which is significantly higher than the medians for the period 1966-1971 (approximately $\hat{m} = 2.8$ ppm). Similarly, the standard deviations for the four approaches in 1972 have an average value of 2.31 ppm, which is significantly higher than the standard geometric deviations for the period 1966-1971 (approximately $s_g = 1.9$ ppm).

The anomaly, or bump, which occurs at the higher concentrations appears to be an abrupt deviation from the remaining pattern of observations. Examination of the histogram (Appendix

Table 51

Frequency-based Measures of Goodness-of-fit for Washington, D.C. - 1971

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	2.809	1.932	17	87	171	0.037	-14691
Fractiles	2.728	2.074	19	144	449	0.073	-14838
MLE Approx.	2.856	1.885	16	63	126	0.022	-14664
MLE Opt.	2.883	1.850	15	46[1]	123[1]	0.018[1]	-14658[1]

[1]Indicates best result in each column.

D) indicates that there are 2 observations of 24 ppm, followed by 14 observations of 25 ppm, followed by 21 observations of 26 ppm, and then followed by only 1 observation of 27 ppm. From our analysis of the geographical data group (Chapter VIII), it is clear that this behavior is highly unusual, and one suspects a possible error in the measurement or data transfer process, or

Table 52

Variate-based (EET) Measures of Goodness-of-fit for Washington, D.C. - 1971 (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	30	-1.425	1.947	-5.44	0.55
Fractiles	30	0.919[1]	0.800[1]	-0.43[1]	1.96
MLE Approx.	30	-2.059	2.408	-6.91	0.37
MLE Opt.	30	-2.563	2.768	-8.05	0.23[1]

[1]Indicates best result in each column.

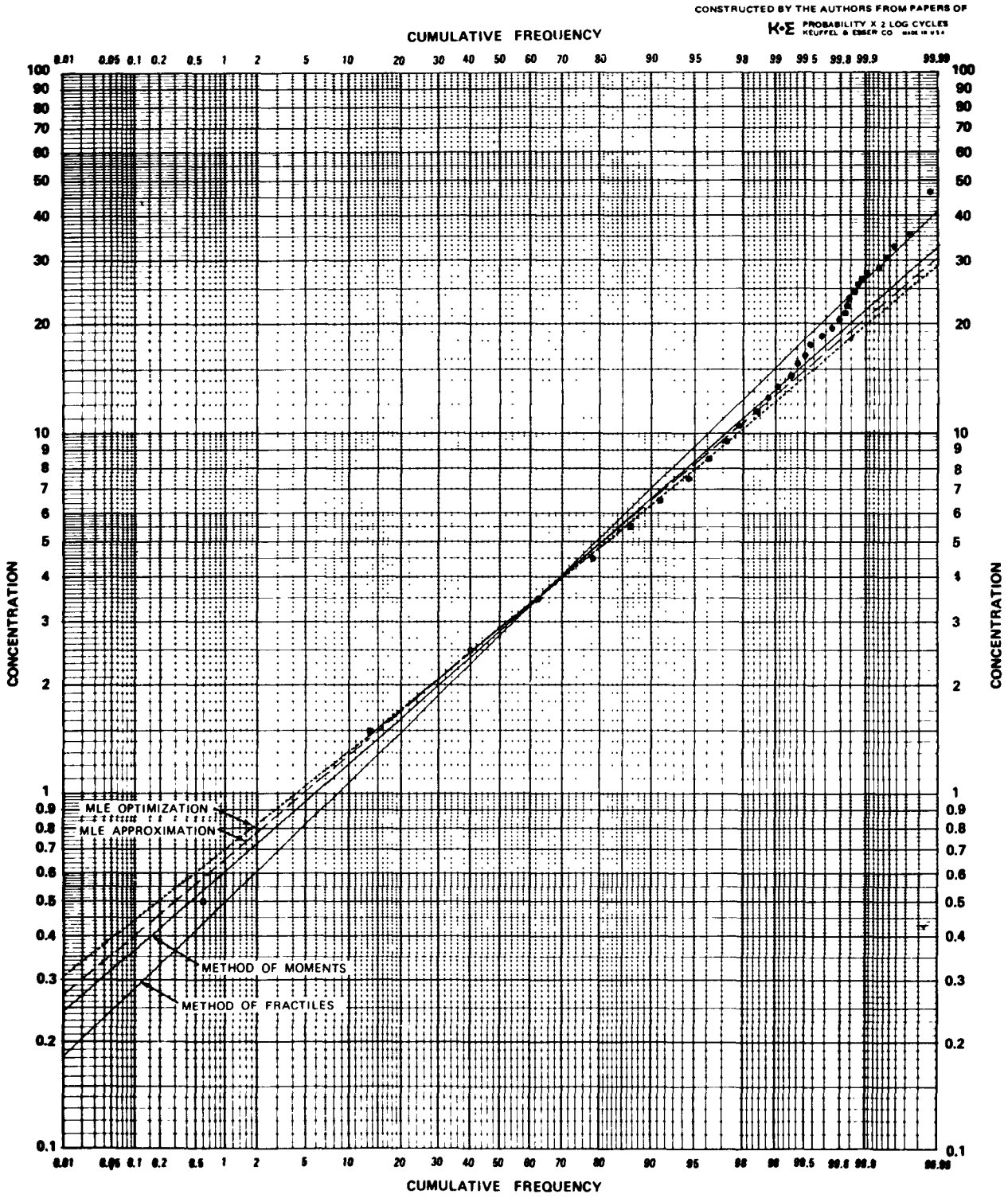

FIGURE 26. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR WASHINGTON, DC, 1971 (n = 7254).

Table 53

Frequency-based Measures of Goodness-of-fit for Washington, D.C. - 1972

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	3.286	2.115	21	155	5874	0.088	-12909
Fractiles	3.542	1.929	19	140[1]	25122	0.084[1]	-13350
MLE Approx.	3.043	2.669	28	169	2009[1]	0.111	-12720
MLE Opt.	3.112	2.508	26	167	2148	0.105	-12703[1]

[1]Indicates best result in each column.

else a very unusual event at the air monitoring station.

Because of this anomaly in the 1972 data, the frequency-based measures of goodness-of-fit (Table 53) all give very large values, indicating a poor fit for all four approaches. Notice that the chi-square values for all four approaches exceed χ^2 = 2000, with the method of fractiles giving the extremely large value of χ^2 = 25122. According to the EET variate-based mea-

Table 54

Variate-based (EET) Measures of Goodness-of-fit for Washington, D.C. - 1972 (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	35	0.808[1]	3.021	-4.58	7.35
Fractiles	35	-1.893	2.355[1]	-7.29	0.94[1]
MLE Approx.	35	13.848	15.253	-0.84	41.39
MLE Opt.	35	9.631	10.942	-0.75[1]	29.45

[1]Indicates best result in each column.

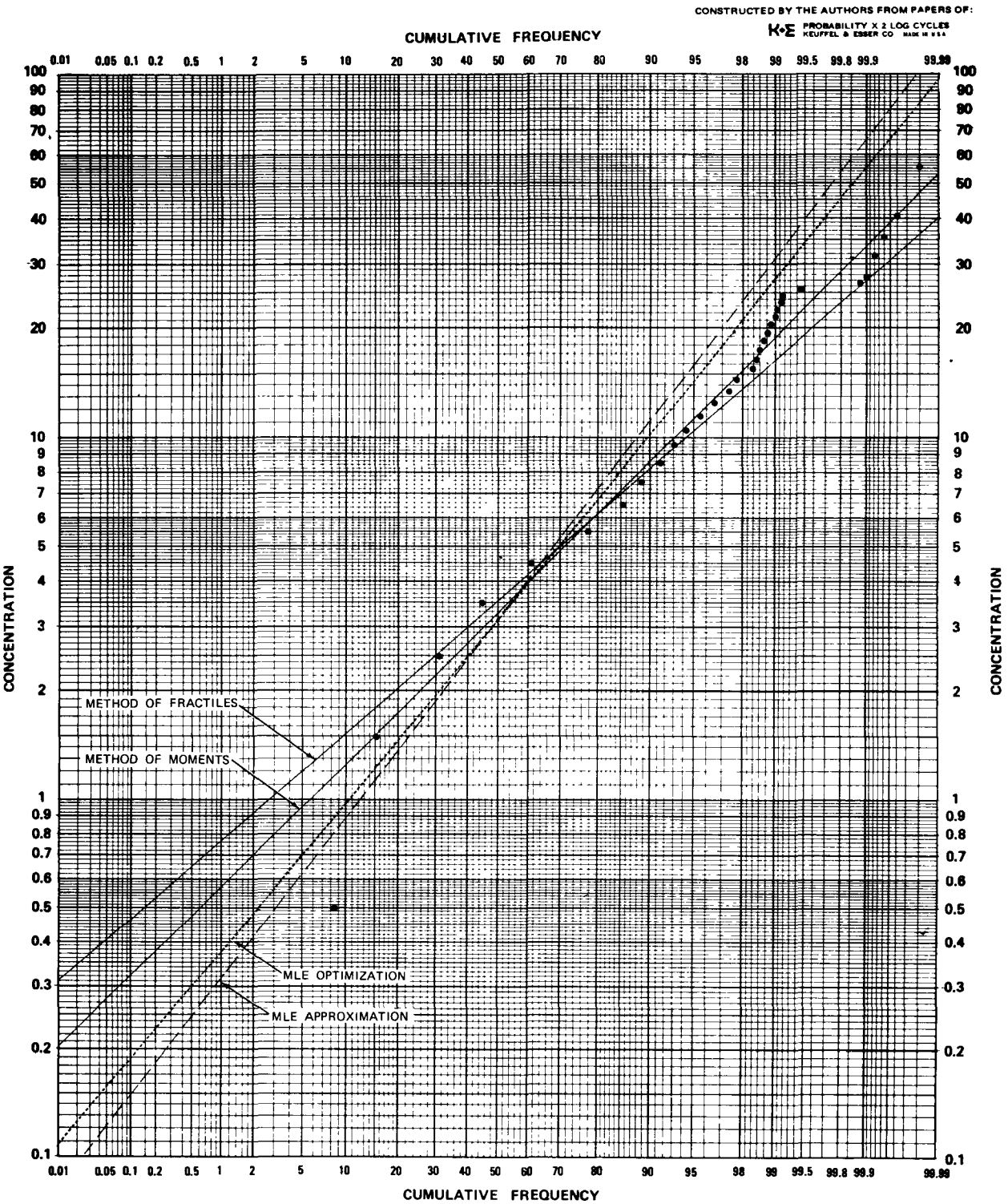

FIGURE 27. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR WASHINGTON, DC, 1972 (n = 4947).

Table 55

Frequency-based Measures of Goodness-of-fit for Washington, D.C. - 1973

Method	$\hat{m}$	s_g	d.f.	RMS	χ^2	K-S max	L*
Moments	2.802	2.093	17	55[1]	373	0.030[1]	-9591
Fractiles	2.728	2.074	17	63	381	0.044	-9598
MLE Approx.	2.751	2.251	19	68	265[1]	0.044	-9586
MLE Opt.	2.774	2.182	18	62	274	0.038	-9580[1]

[1]Indicates best result in each column.

sures (Table 54), the method of fractiles gives a better fit than the method of moments, and both methods perform much better than the two MLE approaches, which give serious overpredictions. The MLE approaches perform poorly because they are greatly affected by the middle ranges of the data; with the downward curvature at the low end and the anomaly at the high end, the MLE lines do not fit the overall distribution very well. Of course, it

Table 56

Variate-based (EET) Measures of Goodness-of-fit for Washington, D.C. - 1973 (ppm)

Method	n	$\bar{d}$	s_d	d_{min}	d_{max}
Moments	38	-0.514[1]	2.398	-7.89	3.11
Fractiles	38	-1.439	2.652	-9.47	1.53[1]
MLE Approx.	38	3.288	2.974	-0.78[1]	10.22
MLE Opt.	38	1.580	2.344[1]	-3.99	7.01

[1]Indicates best result in each column.

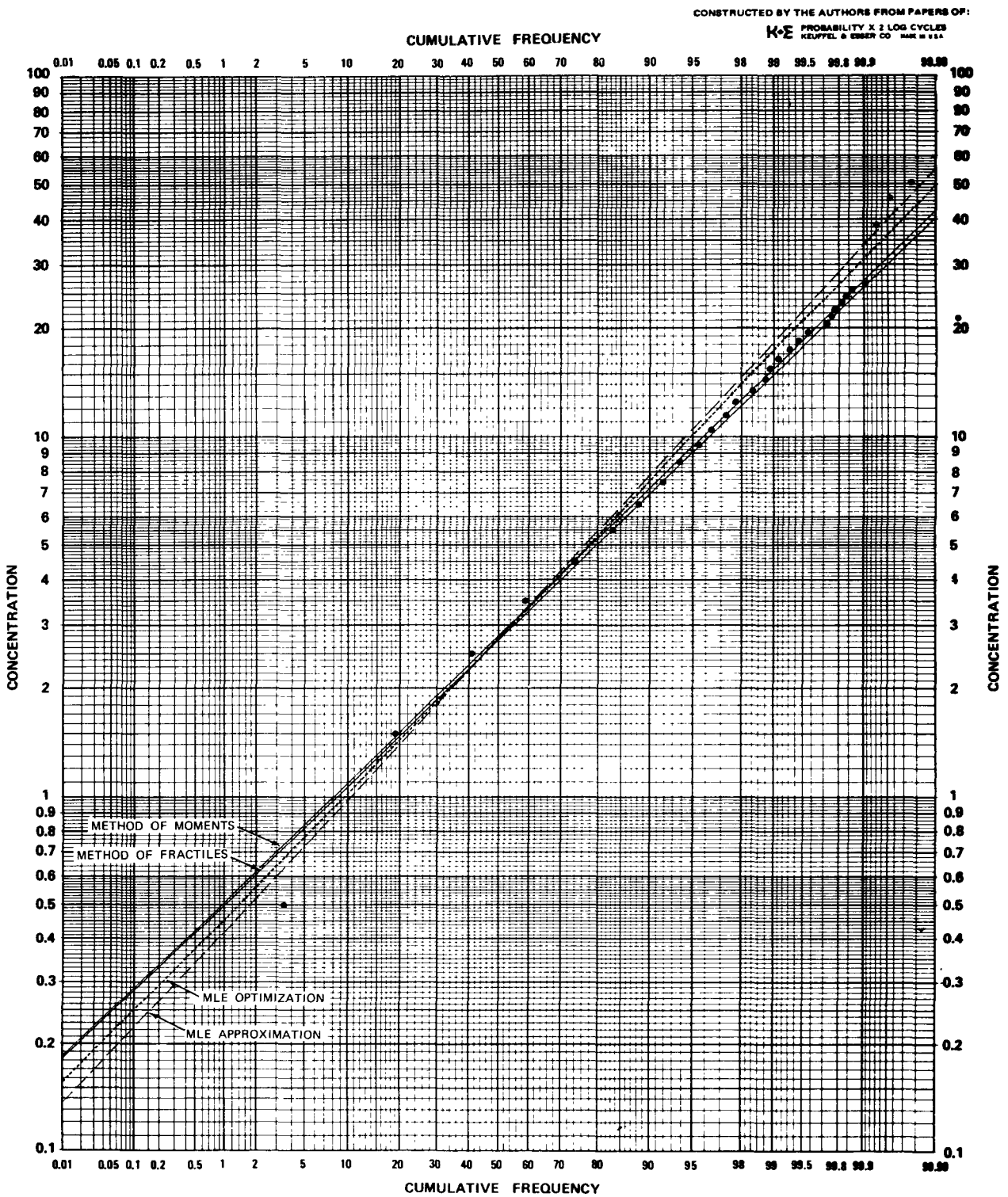

FIGURE 28. LOGARITHMIC PROBABILITY PLOT OF CO HOURLY OBSERVATIONS FOR WASHINGTON, DC, 1973 (n = 4269).

should be borne in mind that the 1972 data set is incomplete, with only 4947 observations out of a possible 8760 hourly observations for the year.

1973

The CAMP 1973 data set, like the 1972 data set, also is incomplete, but it exhibits no serious anomalies. Like the 1972 data, its levels are significantly higher than those for the earlier five years, but the medians of the LN2 model for the four approaches are not quite as high as those for the 1972 data. Unlike the 1972 data, the straight lines for the four approaches, when plotted on logarithmic-probability paper (Figure 28), have the same close proximity to each other that was characteristic of the first five years of data (1966-1971). The data points tend to scatter about a straight line, and the lines for the method of moments and the method of fractiles tend, subjectively at least, to give the best visual fit. The frequency-based measures (Table 55) do not indicate that any one approach is superior, although the method of moments performs surprisingly well. One reason that no single approach emerges as superior is that the lines are sufficiently close together that the resulting goodness-of-fit measures give very similar values. For example, the log-likelihood function varies from L* = -9580 to L* = -9598, only an 18 point spread. The variate-based measures (Table 56) also do not indicate any of the four approaches clearly to be superior, and each approach is identified as "best" by a different variate-based measure of goodness-of-fit.

OVERALL FINDINGS

If we combine the EET differences for all data sets at the CAMP station over the period 1966-1973, it is seen that the resulting histograms for the longitudinal data group (Figure 29, page 196) resemble those for the geographical data group (Figure 20, page 166), although some differences also are evident. Unlike the findings for the geographical data group, the histogram for the method of moments (top of Figure 29) does not show any overpredictions greater than 12 ppm, and the distribution is slightly skewed to the left. The histogram for the method of fractiles, like the corresponding histogram for the geographical data group, does not show any tendency for overprediction (that is, it does not have any noticeable tail to the right), and its average deviation ($\bar{d}$ = -0.04 ppm) is the smallest of the four methods. In the geographical data group, the method of fractiles also produces the smallest average deviation ($\bar{d}$ = -0.12 ppm). In both data groups, the method of moments gives the next best average deviation ($\bar{d}$ = -0.50 ppm for the longitudinal data group and $\bar{d}$ = 0.26 ppm for the geographical data group), with similar standard deviations (1.88 ppm and 1.68 ppm, respectively). The overall EET statistics for the MLE approximation approach are similar in the two data groups ($\bar{d}$ = 2.93 ppm and s_d = 3.94 ppm for the longitudinal data group and $\bar{d}$ = 2.15 ppm and s_d = 3.71 ppm for the geographical data group), and the histograms are similar to each other. For the MLE optimization approach, both data groups show a similar tendency for large overpredictions above 12 ppm, although the average overprediction is much smaller in the longi-

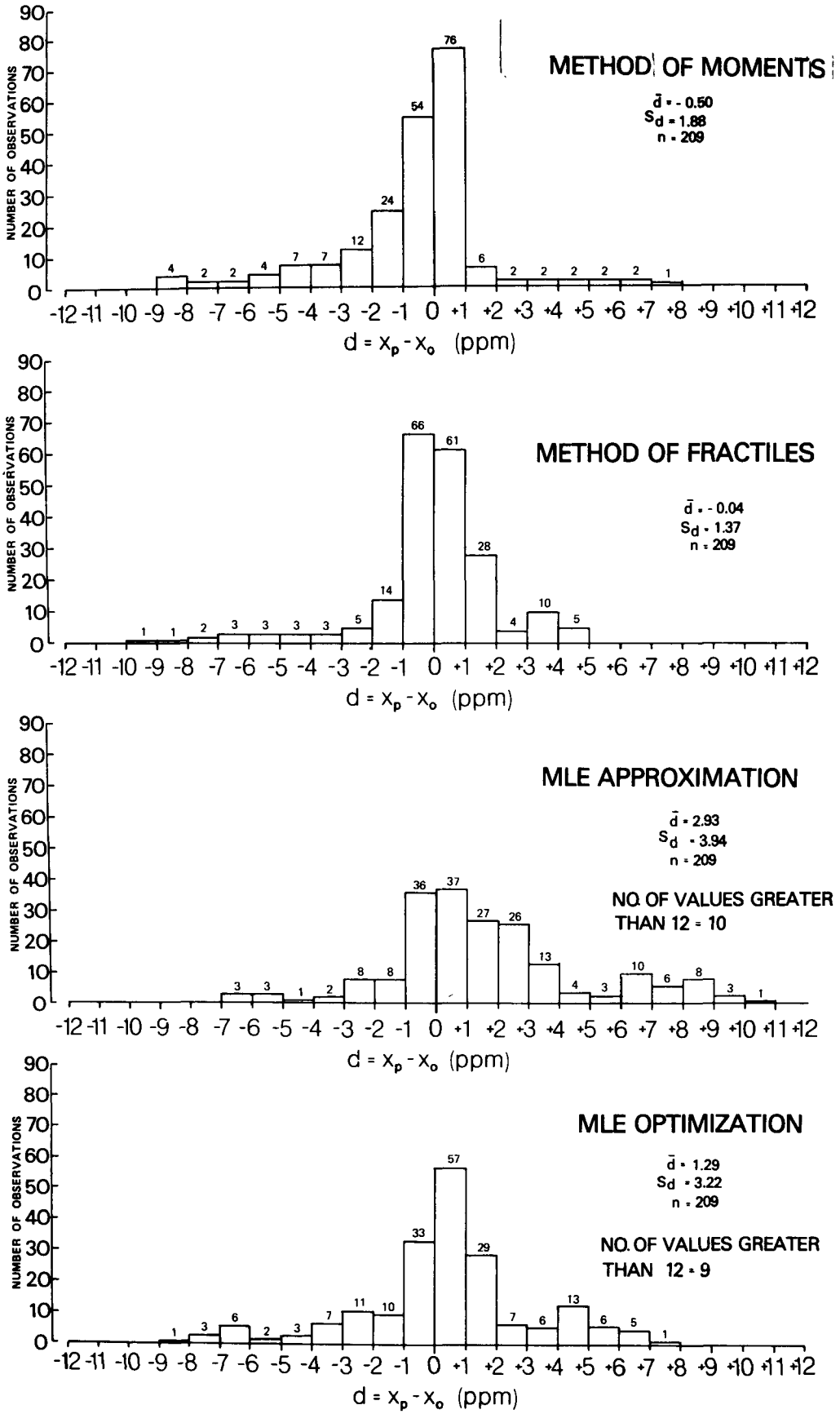

Figure 29. Histograms of the deviations between predicted and observed CO concentrations for four different methods of calculating parameters, Washington, D.C., CAMP, 1966-1973.

tudinal data group ($\bar{d}$ = 1.29 ppm) than in the geographical data group ($\bar{d}$ = 2.75 ppm). The standard deviation also is somewhat less (s_d = 3.22 ppm for the longitudinal data group and s_d = 4.56 ppm for the geographical data group).

According to the EET measures of goodness-of-fit reflected in Figure 29, the method of fractiles gives the best overall performance. How good is this? For the longitudinal data group, 61% of the predictions from the LN2 model, with parameters calculated by the method of fractiles, are within ± 1 ppm of the true value; 81% are within ± 2 ppm; 85% are within ± 3 ppm; and 91% are within ± 4 ppm. These findings are similar to those for the geographical data group (see Table 41, page 169) in which 65% of the observations are within ± 1 ppm of the true value; 84% are within ± 2 ppm; 92% are within ± 3 ppm; and 97% are within ± 4 ppm. Using the method of fractiles to represent a data set within either data group, therefore, we can be 80% certain that the LN2 model will fit the entire range of observations within ± 2 ppm. We can be at least 90% certain that the model will fit the entire distribution within ± 4 ppm, and it usually will do better than this for the geographical data group. Whether this result is satisfactory for representing a data set will depend on the particular application to be made by the data analyst (see related discussion on pages 170-172).

Discussion

An important application of probability models in air pollution control is to estimate the value of the maximum concen-

tration, or, sometimes, the second highest maximum concentration for the year. In such applications, we usually are interested not in the maximum for a particular data set, which can be viewed as one realization from some underlying distribution, but in the expected maximum of the underlying distribution. Assume, for example, that the LN2 model correctly describes the underlying distribution. Assume, also, that the air pollution problem (the magnitude of source emissions and their spatial configuration) remains unchanged from year to year. Then, meteorological events alone (ignoring any error introduced by the measurement system) would be responsible for changes in the observed concentrations from year to year. If we treat these events as random phenomena, we can assume that each year is a realization from the same underlying LN2 distribution. For air pollution control purposes, we are interested in the average of the maxima that would result from this process in a number of successive years. That is, we are interested in the expected value of the maxima for a number of years and not in the maximum caused by the particular meteorological conditions occuring in any one year.

If we actually were sampling, at random, from an LN2 distribution, the method of MLE optimization would be the best approach for estimating parameters, for the resulting model would give estimates of the expected value of the maxima that have the lowest possible variance. Because the CAMP data for the period 1966-1971 seemed to arise from similar distributions, and because emissions may not have changed significantly over this brief period, it is of interest to examine the maxima pre-

dicted by the model using two different approaches for calculating parameters: the method of MLE optimization and the method of fractiles.

Using the notation for the LN2 model introduced in Chapter III (see pages 29 to 44), X denotes concentration, and Y is a normally distributed random variable that describes the logarithm of X; that is, $Y = (\ln X - \mu)/\sigma$. Then the maximum concentration predicted by the model for each data set is the value $X = x_{max}$ which corresponds to the probability associated with the highest value in the data set. In a data set containing n observations (assuming that missing values are randomly distributed over the year), we find the inverse CDF corresponding to the probability (or cumulative frequency) of $f_{max} = \frac{n}{n+1}$:

$$x_{max} = F_X^{-1}\left(\frac{n}{n+1}\right) \tag{136}$$

Because tables are readily available to evaluate the CDF of the normal distribution, we first determine the inverse CDF of Y:

$$\frac{\ln x_{max} - \mu}{\sigma} = y_{max} = F_Y^{-1}\left(\frac{n}{n+1}\right) \tag{137}$$

Solving Equation 137 for x_{max},

$$x_{max} = e^{(\sigma y_{max} + \mu)} = e^{\sigma y_{max}} e^{\mu} \tag{138}$$

Noting, from Equation 42, that $\hat{m} = e^{\mu}$ and, from Equation 71, that

$\sigma_g = e^{\sigma}$, we can write Equation 138 very simply in terms of the model's median and standard geometric deviation:

$$x_{max} = \hat{m} [\sigma_g]^{y_{max}} \tag{139}$$

To apply this relationship to the five similar CAMP data sets, we calculate y_{max} from the inverse normal CDF for each data set using Equation 136. Then we use the values of $\hat{m}$ and s_g given in Tables 43-52 of this chapter to calculate x_{max} using Equation 139 and setting $\sigma_g = s_g$. The results are given in Table 57.

The resulting maxima calculated by the two approaches are quite different from each other. The average of observed maxima for the five years of data is 36.2 ppm (last column of Table 57). The average value of the maxima for the method of fractiles is 35.3 ppm, which comes relatively close to the observed average of 36.2 ppm. The average value of the maxima for the MLE optimization approach is 32.2 ppm, which underestimates the observed average value. Notice, however, that the standard deviation of the maxima for the MLE optimization approach is smaller than the standard deviations of maxima for either the method of fractiles or the actual observations. Thus, the actual observations and the fractiles result show considerably more variation from year to year than does the MLE optimization approach. The observed maximum for any one year is a relatively poor estimate of the expected value of the maximum, since it may be considerably higher or lower than the true value. Application of the LN2 model in any year tends to reduce the variance,

Table 57

Maximum Hourly CO Concentrations for CAMP in Washington, D.C., Predicted by the LN2 Model with Parameters Calculated by Two Approaches

Year	n	f_{max}	y_{max}	Maximum by Fractiles (ppm)	Maximum by MLE Optimization (ppm)	Actual Maximum (ppm)
1966	7822	0.9998722	3.65662	33.8	29.3	38
1968	7266	0.9998624	3.63767	31.6	37.7	27
1969	6056	0.9998349	3.59048	27.3	31.7	25
1970	7748	0.9998709	3.65418	44.8	35.1	45
1971	7254	0.9998621	3.63725	38.7	27.0	46
			Arithmetic Mean:	35.3	32.2	36.2
			Arith. Std. Dev:	6.7	4.3	9.8

thereby increasing our chances of accurately estimating the true expected value of the maximum concentration. At first glance, it might appear from these findings that the method of fractiles is the best approach for making this estimation.

The confidence interval about the average value of the maxima can be calculated as $\overline{x_{max}} \pm \frac{ts}{\sqrt{n}}$, in which t is obtained from (two-sided) tables of the Student's t-distribution and s is the standard deviation of maxima. For the observed maxima and n = 5, the 95% confidence interval about the mean is calculated as $36.2 \pm (2.776)(4.4) = 36.2 \pm 12.2$ ppm. Thus, we can say that the true expected value of the maximum CO concentration lies, with 95% assurance, between 24.0 ppm and 48.4 ppm. Although the average value obtained from the method of fractiles is closer to the observed average value, this finding is not conclusive because of the wide confidence interval that exists around the observed average value. Notice that, in three of the five years of data (1968, 1969, and 1970), the MLE estimate of the maximum actually is closer to the average of observed maxima than is the method of fractiles estimate, and the MLE probably is the better approach.

However, this five-year time series appears to be too brief to compare the performance of the two parameter calculation procedures. Furthermore, our assumption that the underlying distribution of these observations is LN2 in form may be false, and our assumption that the air pollution problem remained constant over this period may not be valid. Also, we have disregared the slight difference in the number of observations available for

each year of data. Nevertheless, the findings do suggest that the LN2 model, when applied in this manner, can be useful in estimating the expected value of the maximum concentration and therefore can be used for predictive purposes. A chief advantage of using the model (and various other candidate probability models as well) is that it reduces the variance of the estimate. In this example, the LN2 model, with parameters calculated by the method of fractiles, reduces the variance, or square of the standard deviation, from $(9.8)^2 = 96.0$ to $(6.7)^2 = 44.9$, a reduction of 53%. The MLE optimization approach for calculating parameters performs even better, reducing the variance to $(4.3)^2 = 18.5$, a reduction of 81%.

Thus, if only one year of observations is available, the predicted maximum concentration should lie closer to the average value of maxima if we use the model than if we use the actual maximum for that year. The observed maximum, because it is subject to considerable variability, is not as good an estimate of the average of maxima as is the value predicted by the model. Because of the importance attached to maximum concentrations by regulatory officials, these results deserve greater attention. The advantages of the MLE optimization approach, and the need to find other models that are superior to the LN2 probability model also should receive great attention. It would be useful, in future research, to examine a time series covering a larger span than five years and to investigate, in greater detail, the significance of the assumptions used in this example.

X. CONCLUSIONS

The authors' conclusions about the results of this investigation are summarized briefly as follows:

* * *

When the lognormal (LN2) probability model is applied to real air quality data, the method chosen for calculating its parameters exerts considerable influence on the predictions and hence on the goodness-of-fit of the model to the data.

* * *

Air pollution concentration data for CO (hourly observations), when plotted on logarithmic-probability paper, show either (1) a zig-zag pattern about a straight line, or (2) curvature that is concave downward. Concave upward curvature is very uncommon.

* * *

For data sets that show concave downward curvature, the method of fractiles approach for calculating parameters gives overpredictions at the lower concentrations, and the MLE optimization approach gives overpredictions at the higher concentrations.

* * *

According to frequency-based measures of goodness-of-fit (for example, tests based on the difference between the number of observations predicted and the number observed in each interval), the MLE approach gives the best fit.

* * *

According to variate-based measures of goodness-of-fit (that is, statistics based on the difference between the predicted and observed concentrations at each interval), the method of fractiles -- with parameters calculated at the 70% and 99.9% cumulative frequencies in the manner suggested by Larsen -- gives the best fit.

* * *

According to all goodness-of-fit measures, the method of moments and the method of MLE approximation seldom give the best fit.

* * *

Graphical plots on logarithmic-probability paper, when examined subjectively by an observer, tend to emphasize variate-based goodness-of-fit measures and underemphasize frequency-based goodness-of-fit measures.

* * *

Variate-based measures of goodness-of-fit, because they weight each interval equally regardless of the number of observations it contains, are unduly influenced by randomness, errors, and fluctuation which occurs at the tails of the distribution.

* * *

Frequency-based measures of goodness-of-fit, because they weight the number of observations in each interval, are less sensitive to randomness at the tails of the distribution.

* * *

If hypothesis tests are undertaken, the MLE optimization

approach gives the best values for chi-square and other frequency-based measures of goodness-of-fit.

* * *

Using the LN2 model with the method of fractiles to represent an hourly CO air quality data set, one can be 80% certain that the model will fit the entire range of observations within $\pm$ 2 ppm. One can be at least 90% certain that the model will fit the entire distribution within $\pm$ 4 ppm.

* * *

If hypothesis testing is undertaken for urban air quality data, and if problems of statistical independence are ignored, then the LN2 model would be rejected with a high degree of confidence ($P > 0.995$) for all data sets and all cities considered.

* * *

Although a particular data set may not be "lognormal" in the statistical sense (that is, it does not satisfy frequency-based measures of goodness-of-fit), it may be lognormal in the engineering sense (that is, it satisfies variate-based EET measures sufficiently for its intended application); we shall call such data sets "psuedo-lognormal."

* * *

Few U.S. air quality data sets appear to be lognormal in the conventional statistical sense; whether they are considered psuedo-lognormal will depend on the particular application facing the data analyst and the criteria he establishes for that application.

* * *

The analyst should not automatically choose the LN2 model for use with air quality data. Rather, he should evaluate candidate models and consider each data set on an individual basis. Other probability models for possible consideration include the gamma, beta, Weibull, LN3C, and Johnson distributions.

* * *

Many of the data sets in this investigation exhibited their largest chi-square values for intervals that were close to the origin, suggesting that probability models which fit well at the lower concentrations (such as the LN3 and LN3C models) would significantly improve the frequency-based measures of goodness-of-fit.

* * *

Application of a proability model (such as the LN2 model) to air quality data appears to be helpful in predicting the expected value of the maximum concentration. If only one year of data is available, for example, application of the LN2 model can reduce the variance of the prediction, thereby reducing the chance of error in predicting the (long-term) expected value of the maximum hourly CO concentration.

* * *

A quality assurance data checking program should be implemented by EPA to compare data contained in the SAROAD data bank with the data originally submitted by state and local air pollution control agencies (see pages 107-109 and 138-140). All discrepancies should be investigated and corrected.

REFERENCES

1. Mage, D. T., and W. R. Ott, "Refinements of the Lognormal Probability Model for Analysis of Aerometric Data," J. Air Poll. Control Assoc. 28, p. 796 (1978).

2. Mage, D.T., and W.R. Ott, "Testing the Validity of Selected Probability Models: Graphical Analysis of Carbon Monoxide Data from U.S. Cities," Environmental Protection Agency, Research Triangle Park, N.C., to be published in 1980.

3. Aitchison, J. and J.A.C. Brown, The Lognormal Distribution, Cambridge University Press, New York (1966).

4. Hatch, T. and S. P. Choate, "Statistical Description of the Size Properties of Non-uniform Particulate Substances," J. Franklin Inst. 207, p. 369 (1929).

5. Krumbein, W.C., "Application of Logarithmic Moments to Size Frequency Distributions of Sediments," J. Sediment. Petrol. 6, p. 35 (1936).

6. Kolmogorov, A. N., "Uber das logarithmisch normale Verteilungsgesetz der Dimensionen der Teilchen bei Zerstuckelung," C.R. Acad. Sci. U.R.S.S. 31, p. 99 (1941).

7. Drinker, P. and T. Hatch, Industrial Dust, 2nd Ed., Mc Graw-Hill, New York, p. 149 (1954).

8. Harris, E.D., and E.C. Tabor, "Statistical Considerations Related to the Planning and Operation of a National Air Sampling Network," Proceedings of the 49th Annual Meeting of the Air Pollution Control Association, Buffalo, N.Y. (1956).

9. Zimmer, C.E., E.C. Tabor, and A.C. Stern, "Particulate Pollutants in the Air of the United States," J. Air. Poll. Control Assoc., 9, p. 136 (1959).

10. deNevers, N., K.W. Lee, and N.H. Frank, "Patterns in TSP Distribution Functions," J. Air Poll. Control Assoc., 29, p. 32 (1979).

11. Larsen, R.I., "A Method for Determining Source Reduction Required to Meet Air Quality Standards," J. Air Poll. Control Assoc., 11, p. 71 (1961).

12. Larsen, R.I., "Parameters of Aerometric Measurements for Air Pollution Research," Am. Ind. Hyg. Assoc. J., 22, p. 97 (1961).

13. Larsen, R.I., "United States Air Quality," Arch. Environ. Health, 8, p. 325 (1964).

14. Larsen, R.I., "Determining Basic Relationships Between Variables," Symposium on Environmental Measurements, PHS Publication No. 999-AP-15, Sanitary Engineering Center, Cincinnati, Ohio, pp. 251-263 (1964).

15. Zimmer, C.E., and R.T. Larsen, "Calculating Air Quality and Its Control," J. Air Poll. Control Assoc., 15, p. 565 (1965).

16. Larsen, R.I., "Air Pollution from Motor Vehicles," Annals of the New York Academy of Sciences, 136(12), p. 275 (1966).

17. Larsen, R.I., "Determining Source Reduction Needed to Meet Air Quality Standards," Proceeding, Part I, of the International Clean Air Congress, pp. 60-74, London (1966).

18. Larsen, R.I., C.E. Zimmer, D.A. Lynn, and K.G. Blemel, "Analyzing Air Pollutant Concentration and Dosage Data," J. Air Poll. Control Assoc., 17, p. 85 (1967).

19. Larsen, R.I., "Determining Reduced-Emission Goals Needed to Achieve Air Quality Goals--A Hypothetical Case," J. Air Poll. Control Assoc., 17, p. 823 (1967).

20. Larsen, R.I., "Future Air Quality Standards and Industrial Control Requirements," Proceedings: The Third National Conference on Air Pollution, PHS Publication No. 1649, U.S. Government Printing Office, Washington, D.C. (1967).

21. Larsen, R.I., "A New Mathematical Model of Air Pollutant Concentration, Averaging Time, and Frequency," J. Air Poll. Control Assoc., 19, p. 24 (1969).

22. Larsen, R.I. and H.W. Burke, "Ambient Carbon Monoxide Exposures," APCA Paper No. 69-167 presented at the 62nd Annual Meeting of the Air Pollution Control Association (1969).

23. Larsen, R.I., "Relating Air Pollutant Effects to Concentration and Control," J. Air Poll. Control Assoc., 20, p. 214 (1970).

24. Larsen, R.I., "A Mathematical Model for Relating Air Quality Measurements to Air Quality Standards," Publication No. AP-89, U.S. Environmental Protection Agency, Research Triangle Park, N.C. (1971).

25. Lowrimore, G.R., "The Probability Distribution of the Logarithm of the Sum of Two or More Lognormal Random Variables,

and a Compound Poisson-lognormal Probability Distributior with Applications to the Environmental Sciences," Ph.D. Thesis, Virginia Polytechnic University, University Micro films, Ann Arbor, Michigan, No. 74-29,135 (1972).

26. Larsen, R.I., "Discussion of Important Factors for the Sulfur Dioxide Concentration in Central Stockholm," Atmospheric Environment, 6, p. 423 (1972).

27. Larsen, R.I., "Air Quality Frequency Distributions and Meteorology," Proceedings of the Fourth Meeting of the Expert Panel on Air Pollution Modeling, NATO Committee on the Challenges of Modern Society, May 28-30, 1973.

28. Larsen, R.I., "An Air Quality Data Analysis System for Interrelating Effects, Standards, and Needed Source Reductions," J. Air Poll. Control Assoc., 23, p. 933 (1973).

29. Larsen, R.I., "An Air Quality Data Analysis System for Interrelating Effects, Standards, and Needed Source Reductions -- Part 2," J Air Poll. Control Assoc., 24, p. 551 (1974).

30. Larsen, R.I., "Relating Air Pollutant Effects to Concentration and Control," in Detection and Control of Air Pollution, pp. 94-122, edited by J.C. Webb, J. Leroux, et. al., MSS Information Corp., 655 Madison Avenue, New York, N.Y. 10021 (1974).

31. Larsen, R.I., "Relating Air Quality Data to Effects, Standards, and Needed Source Reductions," Proceedings of the Conference on Ambient Air Quality Measurements, pp. IV: 40-55, edited by W.B Chadick and G.F. Hoffnagle, Southwest Section of the Air Pollution Control Association, March 10-11 (1975).

32. Chapman, L.D., G.G. Akland, J.F. Finklea, R.I. Larsen, T.D. Mount, W.C. Nelson, D.C. Quigley, and W.C. Wilson, "Electricity Demand: Project Independence and the Clean Air Act," Publication No. ORNL-NSF-EP-89, Oak Ridge National Laboratory, Oak Ridge, Tennessee (1975).

33. Phinney, D.E., and J.E. Newman, "The Precision Associated with the Sampling Frequencies of Total Particulate at Indianapolis, Indiana," J. Air Poll. Control Assoc., 22, p. 692 (1972).

34. Gifford, F.A., Jr., "The Lognormal Distribution of Air Pollution Concentrations," Air Resources Atmospheric Turbulence and Diffusion Laboratory, ESSA, Oak Ridge, Tennessee (preprint, 3p.)(1969).

35. Proceedings of the Symposium on Statistical Aspects of Air Quality Data, edited by L.D. Kornreich, U.S. Environmental Protection Agency, Research Triangle Park, N.C., No. EPA-650/4-74-038 (1974).

36. Gifford, F.A., Jr., "The Form of the Frequency Distribution of Air Pollution Concentrations," Proceedings of the Symposium on Statistical Aspects of Air Quality Data, U.S. Environmental Protection Agency, Research Triangle Park, N.C., No. EPA-650/4-74-038, pp. 3-1 to 3-6 (1974).

37. Knox, J.B., and R.I. Pollack, "An Investigation of the Frequency Distribution of Surface Air Pollutant Concentrations," Proceedings of the Symposium on Statistical Aspects of Air Quality Data, U.S. Environmental Protection Agency, Research Triangle Park, N.C., No. EPA-650/4-74-038, pp. 9-1 to 9-17 (1974).

38. Kahn, H.D., "Note on the Distribution of Air Pollutants," J. Air Poll. Control Assoc. 23, p. 973 (1973).

39. Pollack, R.I., "Studies of Pollutant Concentration Frequency Distributions," U.S. Environmental Protection Agency, Research Triangle Park, N.C. No. EPA-650/4-74-004 (1975).

40. Benarie, M., "Sur la Validite de la Distribution Logarithmico-normale des Concentrations de Pollutant," Proceedings of the 2nd International Clean Air Congress, Washington, D.C., 1970, Academic Press, New York, N.Y., pp. 68-70 (1971).

41. Benarie, M., "The Use of the Relationship Between Wind Velocity and Ambient Pollutant Concentration Distributions for the Estimation of Average Concentrations from Gross Meteorological Data," Proceedings of the Symposium on Statistical Aspects of Air Quality Data, U.S. Environmental Protection Agency. Research Triangle Park, N.C., No. EPA-650/4-74-038, pp. 5-1 to 5-17 (1974).

42. Lynn, D.A., "Fitting Curves to Urban Suspended Particulate Data," Proceedings of the Symposium on Statistical Aspects of Air Quality Data, U.S. Environmental Protection Agency, Research Triangle Park, N.C., No. EPA-650/4-74-038, pp. 13-1 to 13-27 (1974).

43. Mage, D.T., "On the Lognormal Distribution of Air Pollutants," Proceedings of the Fifth Meeting of the Expert Panel on Air Pollution Modeling, NATO/CCMS N.35, Roskilde, Denmark, June 4-6 (1974).

44. Mage, D.T. and W.R. Ott, "An Improved Statistical Model for Analyzing Air Pollution Concentration Data," Paper No.

75-51.4 presented at the 68th Annual Meeting of the Air Pollution Control Association, Boston, MA (1975).

45. Mage, D.T., and W.R. Ott, "An Improved Model for Analysis of Air and Water Pollution Data," International Conference on Environmental Sensing and Assessment, Volume 2, IEEE No. #75-CH 1004-1 ICESA, p. 20-5, September (1975).

46. Ott, W.R., and D.T. Mage, "A General Purpose Univariate Probability Model for Environmental Data Analysis," Comput. and Ops. Res., 3, p. 209 (1976).

47. Kalpasanov, W., and G. Kurchatova, "A Study of the Statistical Distributions of Chemical Pollutants in Air," J. Air Poll. Control Assoc. 26, p. 981 (1976).

48. Curran, T.C., and N.H. Frank, "Assessing the Validity of the Lognormal Model When Predicting Maximum Air Pollutant Concentrations," Paper No. 75-51.3, presented at the 68th Annual Meeting of the Air Pollution Control Association, Boston, MA (1975).

49. Milokaj, P.G., "Environmental Applications of the Weibull Distribution Function: Oil Pollution," Science, 176, p. 1019 (1972).

50. Johnson, T., "A Comparison of the Two-parameter Weibull and Lognormal Distributions Fitted to Ambient Ozone Data," Proceedings, Quality Assurance in Air Pollution Measurements, New Orleans, LA, edited by the Air Pollution Control Association, March 11-14, pp. 312-321 (1979).

51. Bencala, K.E., and J.H. Seinfeld, "On Frequency Distributions of Air Pollutant Concentrations," Atmos. Environ. 10, p. 941 (1976).

52. Larsen, R.I., "An Air Quality Data Analysis System for Interrelating Effects, Standards, and Needed Source Reductions: Part 4, A Three-parameter Averaging-time Model," J. Air Poll. Control Assoc., 27, p. 454 (1977).

53. Larsen, R.I., "An Air Quality Data Analysis System for Interrelating Effects, Standards, and Needed Source Reductions -- A Summary," Proceedings of the Fourth International Clean Air Congress, edited by S. Kasuga et. al., The Japanese Union of Air Pollution Prevention Associations, Tokyo, Japan, pp. 323-325 (1977).

54. Finklea, J.F., C.M. Shy, J.B. Moran, W.C. Nelson, R.I. Larsen, and G.G. Akland, "The Role of Environmental Health Assessment in the Control of Pollution," in Advances in Environmental Science and Technology, Vol. 7, edited by

J.N. Pitts, Jr., and R.L. Metcalf, John Wiley and Sons, New York, N.Y., pp. 315-389 (1977).

55. Larsen, R.I., "Relating Data to Effects, Standards, and Source Reductions," in Quality Assurance Practices for Health Laboratories, edited by S.L. Inhorn, American Public Health Association, Washington, D.C., pp. 329-338 (1978).

56. Larsen, R.I., "An Air Quality Data Analysis System for Interrelating Effects, Standards, and Needed Source Reductions -- Part 6, Calculating Needed Source Reductions from an Analysis of Several Years of Air Quality Data," APCA Paper No. 78-63.6, presented at the 71st Annual Meeting of the Air Pollution Control Association, June (1978).

57. Mage, D.T., "Data Analysis by Use of Univariate Probability Models," Paper No. 1-G/75, presented at the Seminar on Air Pollution Control, Ministry of Health and Social Affairs, Seoul, Korea (1975).

58. Johnson, N.L., "Systems of Frequency Curves Generated by Methods of Translation," Biometrika 36, p. 149 (1949).

59. Ledolter, J., G.C. Tiao, G.B. Hudak, J.T.Hsieh, and S.B. Graves, "Statistical Analysis of Multiple Time Series Associated with Air Quality Data: New Jersey CO Data," Technical Report No. 529, Department of Statistics, University of Wisconsin, Madison (1978).

60. Singpurwalla, N.D., "Extreme Values from a Lognormal Law with Applications to Air Pollution Problems," Technometrics, 14, p. 703 (1972).

61. Roberts, E. M., "Review of Statistics of Extreme Values with Applications to Air Quality Data, Part I, Review," J. Air Poll. Control Assoc., 29, p. 632 (1979).

62. Roberts, E. M., "Review of Statistics of Extreme Values with Applications to Air Quality Data, Part II, Applications," J. Air Poll. Control Assoc., 29, p. 733 (1979).

63. Trajonis, John, "Empirical Relationships between Atmospheric Nitrogen Dioxide and its Precursors," U.S. Environmental Protection Agency, Research Triangle Park, N.C., No. EPA-600/3-78-018 (1978).

64. "Guidelines for the Interpretation of Air Quality Standards," U.S. Environmental Protection Agency, Research Triangle Park, N.C., OAQPS No. 1.2-008, revised February 1977 (1977).

65. Mage, D.T., T.R. Fitz-Simons, D.M. Holland, and W.R. Ott, "Techniques for Fitting Probability Models to Experimental Data," Proceedings, Quality Assurance in Air Pollution Measurement, New Orleans, Louisiana, edited by the Air Pollution Control Association, March 11-14, pp. 304-311 (1979).

66. Hunt, W.F., Jr., "The Precision Associated with the Sampling Frequency of Log-Normally Distributed Air Pollutant Measurement," J. Air Poll. Control Assoc., 22, p. 687 (1972).

67. Akland, Gerald, "Applications of Probabilistic Models to Air Quality Data," internal memorandum, U.S. Environmental Protection Agency, Environmental Monitoring and Systems Laboratory, Research Triangle Park, N.C., September 15, 1978.

68. Larsen, R.I., U.S. Environmental Protection Agency, Research Triangle Park, N.C., personal communication (1978).

69. Nehls, Gerald J., and Gerald G. Akland, "Procedures for Handling Aerometric Data," J. Air Poll. Control Assoc., 23, p. 180 (1973).

70. Kushner, E.J., "On Determining the Statistical Parameters for Pollution Concentration from a Truncated Data Set," Atmos. Environ., 10, p. 975 (1976).

71. Hald, A., "Maximum Likelihood Estimation of the Parameters of a Normal Distribution which is Truncated at a Known Point," Skandinavisk Aktuarietidskrift, 32, pp. 119-134 (1949).

72. Hahn, Gerald J., and Samuel S. Shapiro, Statistical Models in Engineering, John Wiley & Sons, Inc., New York, N.Y. (1967).

73. Green, J.R., and U.A.S. Hegazy, "Powerful Modified-EDF Goodness-of-Fit Tests," J. American Statistical Assoc., 71, pp. 204-209 (1976).

74. Ott, Wayne R., "A Fortran Program for Computing the Pollutant Standards Index (PSI)," U.S. Environmental Protection Agency, Washington, D.C., No. EPA-600/4-78-001, May (1978).

75. Hastings, C., Jr., assisted by J.T. Hayward and J.P. Wong, Jr., Approximations for Digital Computers, Princeton University Press, Princeton, N.J. (1955).

APPENDIX A

LISTING OF THE COMPUTER PROGRAM

```
C  * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
C        PROGRAM NAME = HISTOGRAM.II    VERSION B    MAY 20, 1979
C        (VERSION B READS DATA IN INDEX.PLOT FORMAT)
C PROGRAMMER IS WAYNE R. OTT, U. S. ENVIRONMENTAL PROTECTION AGENCY
C
C PROGRAM ANALYZES ENVIRONMENTAL DATA SETS:
C    1.  COMPUTES BASIC STATISTICS
C    2.  GENERATES AND PLOTS A HISTOGRAM
C    3.  EXAMINES GOODNESS OF FIT FOR LOGNORMAL MODEL
C    4.  CALCULATES AND PLOTS THE ATOCORRELATION FUNCTION
C
C
C  * * * USER INSTRUCTIONS * * *
C
C MODIFY INPUT FORMAT #110 AS APPROPRIATE FOR ENTERING DATA
C IF LABEL IS DESIRED, INSERT 2 EXTRA LINES AT BEGINNING OF DATA SET:
C     FIRST LINE MUST HAVE "11" IN COL. 1-2
C     SECOND LINE MUST HAVE LABEL IN COL. 1-80
C SET 'DUMMY' TO THE CODE FOR MISSING VALUES
      DUMMY=9999.99
C SET 'BASE' AS LOWEST VALUE FOR CALCULATION
C SET 'DELTA' AS INTERVAL SIZE AND LIMIT AS NUMBER OF INTERVALS
C     (NOT MORE THAN 100 INTERVALS ARE POSSIBLE)
      BASE=-0.50
      DELTA=1.0
      LIMIT=60
C IF VALUES ABOVE RANGE ARE TO BE PRINTED, SET INSTR1=1
      INSTR1=1
C SET DIMENSION VALUE(...) LARGE ENOUGH TO HOLD ALL DATA
C SET DIMENSION STORE(...) LARGE ENOUGH TO HOLD ALL DATA
C
C
C  * * * PROGRAM HOUSEKEEPING * * *
C
      DIMENSION STORE(9000)
      DIMENSION VALUE(9000)
      DIMENSION LABEL(80)
      DIMENSION TEMP(12)
      DIMENSION XI(100),JH(100),PV(100),CUM(100),HI(1000)
      DIMENSION XB(100),JB(100),PB(100),CUB(100),HB(50)
      DIMENSION X(100),DIFF(100)
      COMMON /B/MARK1,MARK2,IBLANK,LINE(50)
      COMMON /F/AVE,STD
      COMMON /G/DUMMY
C INITIALIZE VALUES (L IS NO. OF VALUES READ; N IS NO. OF VALID VALUES)
    1 L=0
      N=0
C CLEAR ARRAYS
      DO 10 I=1,80
      LABEL(I)=IBLANK
   10 CONTINUE
C
C
C  * * * DATA INPUT SECTION * * *
C
C READ INPUT VALUE ACCORDING TO SAROAD CARBON MONOXIDE FORMAT
    5 READ(5,110) ITEST,TEMP
  110 FORMAT(I2,6X,12F6.2)
C  FINISH READING DATA IF ITEST=88 OR ITEST=99
      IF(ITEST.EQ.88)   GO TO 30
      IF(ITEST.EQ.99) GO TO 30
```

```
C READ AND WRITE LABEL IF ITEST=11
      IF (ITEST.NE.11) GO TO 20
      READ(5,120) LABEL
  120 FORMAT(80A1)
      WRITE(6,220) LABEL
  220 FORMAT(1H0,80A1)
C LABEL HAS BEEN READ; RETURN FOR ANOTHER RECORD
      GO TO 5
   20 CONTINUE
C ALL 12 VALUES READ ARE STORED IN "STORE"
      DO 24 I=1,12
      L=L+1
      STORE(L)=TEMP(I)
   24 CONTINUE
C ONLY THE VALID VALUES READ ARE STORED IN "VALUE"
C DELETE MISSING VALUES CODED AS DUMMY
      DO 25 I=1,12
      IF(TEMP(I).EQ.DUMMY)   GO TO 25
      N=N+1
      VALUE(N)=TEMP(I)
   25 CONTINUE
C ONE RECORD BEEN STORED; RETURN FOR ANOTHER RECORD
      GO TO 5
C LAST VALUE HAS BEEN DETECTED
   30 CONTINUE
C LIST SUMMARY INFORMATION
      WRITE(6,230) L,N
  230 FORMAT(1H0,'NO. OF VALUES READ =',I6,7X,
     1'NO OF VALID VALUES STORED =',I6)
C
C
C * * * MAIN COMPUTATION * * *
C
      WRITE(6,301)
  301 FORMAT(///,1X,'BASIC STATISTICS-----',///)
      CALL STAT(N,VALUE)
      WRITE(6,302)
  302 FORMAT(1H1,1X,'HISTOGRAM OF RAW DATA-----')
      CALL HISTOG(INSTR1,N,VALUE,BASE,DELTA,LIMIT,XI,JH,PV,CUM,HI,IH)
      WRITE(6,303)
  303 FORMAT(1H1,1X,'PARAMETERS EST. BY THE METHOD OF MOMENTS-----')
C...USE BIASED ESTIMATE OF STANDARD DEVIATION
      STD1=(SQRT(FLOAT(N-1)/FLOAT(N)))*STD
      AVE1=AVE
      T=ALOG((STD1/AVE1)*(STD1/AVE1)+1.0)
      SIG=SQRT(T)
      U=ALOG(AVE1)-0.5*T
      CALL LN(U,SIG,N,LIMIT,K,X,DIFF,XI,JH,PV,CUM,IH)
      CALL HISTOG(0,K,DIFF,-5.0,0.5,20,XB,JB,PB,CUB,HB,IHD)
      CALL STAT(K,DIFF)
      WRITE(6,304)
  304 FORMAT(1H1,1X,'PARAMETERS EST. BY THE METHOD OF LARSEN-----')
      CALL RALPH(U2,SIG2,AVE1,LIMIT,XI,JH,PV,CUM,N)
      CALL LN(U2,SIG2,N,LIMIT,K,X,DIFF,XI,JH,PV,CUM,IH)
      CALL HISTOG(0,K,DIFF,-5.0,0.5,20,XB,JB,PB,CUB,HB,IHD)
      CALL STAT(K,DIFF)
      WRITE(6,305)
  305 FORMAT(1H1,1X,'PARAMETERS EST. BY APPROXIMATE M. L. E.-----')
      CALL AMLE(U3,SIG3,N,VALUE)
      CALL LN(U3,SIG3,N,LIMIT,K,X,DIFF,XI,JH,PV,CUM,IH)
      CALL HISTOG(0,K,DIFF,-5.0,0.5,20,XB,JB,PB,CUB,HB,IHD)
```

```
      CALL STAT(K,DIFF)
      WRITE (6,306)
  306 FORMAT(1H1,1X,'PARAMETERS EST. BY OPTIMIZATION OF M. L. E.-----')
      CALL BMLE(U3,SIG3,LIMIT,XI,JH,PV,CUM)
      CALL LN(U3,SIG3,N,LIMIT,K,X,DIFF,XI,JH,PV,CUM,IH)
      CALL HISTOG(0,K,DIFF,-5.0,0.5,20,XB,JB,PB,CUB,HB,IHD)
      CALL STAT(K,DIFF)
C COMPUTE AUTOCORRELATION FUNCTIUN
      CALL AUTCOR(L,N,STORE,1,1)
      CALL AUTCOR(L,N,STORE,6,6)
C
C IF ITEST=88, BEGIN COMPUTATIUN OVER AGAIN
      IF(ITEST.EQ.88)   GO TO 1
C
C  * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
      END


      BLOCK DATA
      COMMON /B/MARK1,MARK2,IBLANK,LINE(50)
      DATA MARK1/'*   '/,MARK2/'U   '/,IBLANK/'
      END


      SUBROUTINE STAT(N,VALUE)
C SUBROUTINE TO COMPUTE STATISTICAL MOMENTS
      COMMON /F/AVE,STD
      DIMENSION VALUE(9000)
      IF (N.EQ.1)   GO TO 50
C...COMPUTE MEAN
      T=FLOAT(N)
      SUM=0.0
      DO 10 I=1,N
      SUM=SUM+VALUE(I)
   10 CONTINUE
      AVE=SUM/T
      WRITE(6,200) N
      WRITE(6,210) AVE
C...COMPUTE OTHER MOMENTS
      SUM2=0.0
      SUM3=0.0
      SUM4=0.0
      DO 20 I=1,N
      D=VALUE(I)-AVE
      Z2=D*D
      Z3=D*Z2
      Z4=D*Z3
      SUM2=SUM2+Z2
      SUM3=SUM3+Z3
      SUM4=SUM4+Z4
   20 CONTINUE
C...COMPUTE VARIANCE,STANDARD DEVIATION,AND COEFFICIENT OF VARIATION
      S2=SUM2/(T-1.0)
```

```
      STD=SQRT(S2)
      COF=STD/AVE
      WRITE(6,220) S2,STD
      WRITE(6,225) COF
C...COMPUTE COEFFICIENT OF SKEWNESS
      IF(S2.EQ.0.0)   GO TO 50
      S3=SUM3/T
      SKEW=S3/(S2**1.5)
      WRITE(6,230) S3,SKEW
C...COMPUTE COEFFICIENT OF KURTOSIS
      S4=SUM4/T
      BETA2=S4/(S2*S2)
      WRITE(6,240) S4,BETA2
C...FIND LOWEST AND HIGHEST VALUES
      VMAX=-99999999999.0
      VMAX2=-99999999999.0
      VMIN=99999999999.0
      DO 30 I=1,N
      IF(VALUE(I).LT.VMIN)   VMIN=VALUE(I)
      IF(VALUE(I).GT.VMAX)   IMAX=I
      IF(VALUE(I).GT.VMAX)   VMAX=VALUE(I)
   30 CONTINUE
C...FIND SECOND HIGHEST MAXIMUM VALUE
      DO 40 I=1,N
      IF(I.EQ.IMAX)   GO TO 40
      IF(VALUE(I).GT.VMAX2)   VMAX2=VALUE(I)
   40 CONTINUE
      WRITE(6,245) VMIN
      WRITE(6,246) VMAX
      WRITE(6,247) VMAX2
      GO TO 60
   50 CONTINUE
      WRITE(6,250)
  250 FORMAT(1H ,5X,'NO ADDITIONAL STATISTICS POSSIBLE')
   60 CONTINUE
  200 FORMAT(1H0,5X,'N =',I8)
  210 FORMAT(1H0,5X,'MEAN =',F18.8)
  220 FORMAT(1H0,5X,'2ND MOMENT (VAR.) =',F22.10,12X,'STANDARD DEV. =',
     1F20.9)
  225 FORMAT(1H0,10X,'COEFFICIENT OF VARIATION =',F15.8)
  230 FORMAT(1H0,5X,'3RD MOMENT =',F20.9,9X,'COEF. OF SKEWNESS =',F15.8)
  240 FORMAT(1H0,5X,'4TH MOMENT =',F20.9,9X,'COEF. OF KURTOSIS =',F15.8)
  245 FORMAT(1H0,5X,'LOWEST VALUE =',F18.8)
  246 FORMAT(1H0,5X,'HIGHEST VALUE =',F18.9)
  247 FORMAT(1H0,10X,'SECOND HIGHEST VALUE =',F18.9)
      RETURN
      END
```

```
      SUBROUTINE HISTOG(IN,N,VALUE,BASE,DELTA,LIMIT,XI,JH,PV,CUM,HI,IH)
C THIS SUBROUTINE SORTS EACH OF N VALUES INTO INTERVALS OF SIZE DELTA
C IF IN=1, THE VALUES ABOVE THE RANGE ARE LISTED
C...XI CONTAINS THE HISTOGRAM ABSCISSA
C...PV CONTAINS THE INDIVIDUAL FREQUENCIES
C...CUM CONTAINS THE CUMULATIVE FREQUENCIES
C...ILOW IS THE NUMBER OF VALUES BELOW THE RANGE
C...IZERO IS THE NUMBER OF ZERO VALUES
C...IEXCES IS THE NUMBER OF VALUES ABOVE THE RANGE
C...HI CONTAINS THE INDIVIDUAL VALUES ABOVE THE RANGE
      DIMENSION VALUE(9000)
      COMMON /B/MARK1,MARK2,IBLANK,LINE(50)
      DIMENSION XI(100),JH(100),PV(100),CUM(100)
      DIMENSION HI(1000)
C CLEAR ARRAYS
      DO 1 I=1,100
      JH(I)=0
      PV(I)=0.0
      XI(I)=0.0
      CUM(I)=0.0
    1 CONTINUE
C...CLEAR VALUES
      ILOW=0
      IZERO=0
      IEXCES=0
C...SET TOP OF RANGE
      TOP=FLOAT(LIMIT)*DELTA + BASE
C BEGIN SORTING ROUTINE
      DO 10 I=1,N
C...DETERMINE IF POINT IS A ZERO VALUE
      IF (VALUE(I).EQ.0.0)  IZERO=IZERO+1
C...MODIFY VALUE TO ADJUST FOR MACHINE DECIMAL CONVERSION ERROR
      PVALUE=VALUE(I) + 0.000001
C...DETERMINE IF POINT IS BELOW RANGE
      IF(PVALUE.GE.BASE)   GO TO 15
      ILOW=ILOW+1
      GO TO 10
   15 CONTINUE
C...DETERMINE IF POINT IS ABOVE RANGE
      IF(PVALUE.LT.TOP)   GO TO 20
      IEXCES=IEXCES+1
      IF(IEXCES.GT.1000)  GO TO 28
      IF(IN.EQ.1)   HI(IEXCES)=VALUE(I)
      GO TO 10
   20 CONTINUE
C...POINT IS WITHIN RANGE; CARRY OUT SORTING
      SDELTA=0.0
      DO 25 J=1,100
      SDELTA=DELTA+SDELTA
      TDELTA=SDELTA+BASE
      IF(PVALUE.LT.TDELTA)   GO TO 27
   25 CONTINUE
   27 CONTINUE
C PVALUE IS GE THAN VALUE(I); THEREFORE INTERVAL HAS BEEN FOUND:
      JH(J)=JH(J)+1
      GO TO 10
   28 CONTINUE
C ERROR: NUMBER OF POINTS ABOVE RANGE EXCEEDS ARRAY SIZE
      IF(IEXCES.EQ.1001)   WRITE(6,2099)
 2099 FORMAT(1H1,2X,'TOO MANY POINTS ARE OUTSIDE RANGE; INCREASE DELTA')
   10 CONTINUE
```

```
C TOTAL HISTOGRAM POINTS (THOSE ABOVE BASE)
      NUM=0
      DO 30 I=1,LIMIT
      NUM=NUM+JH(I)
   30 CONTINUE
      NUM=NUM+IEXCES
      S=FLOAT(NUM)
C FILL ABSCISSA, INDIVIDUAL, AND CUMULATIVE FREQUENCY VECTORS
      SUM=0.0
      DO 40 I=1,LIMIT
      XI(I)=(FLOAT(I))*DELTA+BASE
      PV(I)=(FLOAT(JH(I))*100.0)/S
      SUM=SUM+PV(I)
      CUM(I)=SUM
   40 CONTINUE
C CLEAR LINE ARRAY
      DO 50 K=1,50
      LINE(K)=IBLANK
   50 CONTINUE
C PRINT RESULTS
      WRITE(6,2000) DELTA
      WRITE(6,2001) BASE,ILOW,IZERO
      WRITE(6,2002)
      DO 60 I=1,LIMIT
      Z=PV(I)+0.5
      M=IFIX(Z)
      IF(M.EQ.0)   GO TO 80
      IF(M.GT.50)  M=50
C...FILL LINE ARRAY WITH INDIVIDUAL FREQUENCIES
      DO 70 J=1,M
      LINE(J)=MARK1
   70 CONTINUE
   80 CONTINUE
C...FILL LINE ARRAY WITH CUMULATIVE FREQUENCIES
      Q=(CUM(I)/2.)+0.5
      L=IFIX(Q)
      IF(L.NE.0)   LINE(L)=MARK2
C...PRINT ONE LINE
      A=XI(I)-DELTA
      B=XI(I)
      IC=JH(I)
      D=PV(I)
      E=CUM(I)
      F=E*(S/(S+1.0))
      IF(DELTA.LT.0.1) WRITE(6,2003) A,B,IC,D,(LINE(K),K=1,50),E,F
      IF(DELTA.GE.0.1) WRITE(6,2004) A,B,IC,D,(LINE(K),K=1,50),E,F
C...CLEAR LINE ARRAY
      IF(M.EQ.0)   GO TO 90
      DO 90 K=1,M
      LINE(K)=IBLANK
   90 CONTINUE
      IF(L.NE.0)   LINE(L)=IBLANK
C...CHANGE PAGE
      IF(I.NE.51)   GO TO 60
      WRITE(6,2005)
      WRITE(6,1999)
      WRITE(6,2002)
   60 CONTINUE
C...PRINT BOTTOM OF GRAPH
      WRITE(6,2005)
```

```
C...PRINT NO. OF VALUES ABOVE RANGE
      PEXCES=(FLOAT(IEXCES)/S)*100.0
      WRITE(6,2006) TOP,IEXCES,PEXCES
      IH=IEXCES
 1999 FORMAT(1H1)
 2000 FORMAT(1H0,22X,        '*  *  *   DISTRIBUTION OF DATA (INTERVAL WI
     1DTH =',F8.4,')    *  *  *')
C...IF IN=1, PRINT VALUES ABOVE THE RANGE
      IF(IN.NE.1)   GO TO 110
      IF(IEXCES.EQ.0)  GO TO 110
      WRITE(6,2007)
      DO 100 I=1,IEXCES
      WRITE(6,2008) I,HI(I)
  100 CONTINUE
  110 CONTINUE
 2001 FORMAT(1H0,2X,'NO. OF VALUES LESS THAN ',F8.4,'=',I6,5X,
     1'NO. OF ZERO VALUES =',I6)
 2002 FORMAT(1H0,5X,'INTERVAL',5X,'NUMBER',3X,'PERCENT',
     12X,'+---------------------HISTOGRAM--------------------+',
     21X,'CUM. FREQ.',2X,'PLOT PUS.')
 2003 FORMAT(1H ,1X,F7.4,'-',F7.4,1X,I7,2X,F7.3,
     1'%',2X,'I',50A1,'I',1X,F7.3,'%',4X,F7.3,'%')
 2004 FORMAT(1H ,1X,F7.2,'-',F6.2,1X,I8,2X,F7.3,
     1'%',2X,'I',50A1,'I',1X,F7.3,'%',4X,F7.3,'%')
 2005 FORMAT(1H ,36X,'+-------------------------------------------------
     1-+')
 2006 FORMAT(1H0,2X,'NO. OF VALUES EQUAL TO OR GREATER THAN ',F9.4,
     1' =',I5,2X,'(',F6.3,'%)')
 2007 FORMAT(1H1,///,8X,'LISTING OF VALUES ABOVE RANGE:',//)
 2008 FORMAT(1H0,12X,I4,5X,F14.6)
      RETURN
      END




      SUBROUTINE RALPH(U,SIG,AVE1,LIMIT,XI,JH,PV,CUM,N)
C SUBROUTINE TO DETERMINE LN PARAMTERS BY LARSEN METHOD OF FRACTILES
      DIMENSION XI(100),JH(100),PV(100),CUM(100)
C FIND 70 AND 99.9 PERCENTILES
      IL1=0
      IL2=0
      JSUM=0
      JSPREV=0
      T=FLOAT(N)
      R=0.7*T + 0.999
      M1=IFIX(R)
      R=0.999*T + 0.999
      M2=IFIX(R)
      DO 10 I=1,LIMIT
      JSUM=JSUM+JH(I)
      IF (M1.GT.JSPREV.AND.M1.LE.JSUM)   IL1=I
      IF (M2.GT.JSPREV.AND.M2.LE.JSUM)   IL2=I
      JSPREV=JSUM
   10 CONTINUE
      Y1=0.52440
      Y2=3.09023
      IF(IL2.EQ.0)   GO TO 20
      X1=XI(IL1)-0.5
      IF(X1.NE.0.0)   GO TO 16
```

```
      X1=AVE1
      WRITE(6,200)
   16 X2=XI(IL2)-0.5
      WRITE(6,202) X1,Y1,X2,Y2
      SIG=(ALOG(X2/X1))/(Y2-Y1)
      U=ALOG(X1)-SIG*Y1
      GO TO 30
  200 FORMAT(1H0,'70 % ARE ZEROS; MEAN IS USED FOR 70 PERCENTILE')
   20 WRITE(6,201)
  201 FORMAT(1H0,'LARSEN METHOD NOT POSSIBLE: NO 99.9 PERCENTILE')
   30 CONTINUE
  202 FORMAT(1H0,5X,'FOR THE 70 PERCENTILE, X =',F7.3,' AND Y =',
     1F9.5,';  FOR THE 99.9 PERCENTILE, X =',F7.3,' AND Y =',F9.5)
      RETURN
      END
```

```
      SUBROUTINE AMLE(U,SIG,N,VALUE)
C SUBROUTINE TO COMPUTE PARAMETERS BY THE APPROXIMATE M.L.E. METHOD
C IGNORES GROUPING; ASSUMES ZEROS ARE AT MINIMUM DETECTABLE LIMT (.50)
C TAKES LOGARITHMS OF RAW DATA; CALCULATES MOMENTS FROM RESULTING LOGS
      DIMENSION VALUE(9000)
      SUM=0.0
      SUM2=0.0
      DO 10 I=1,N
      P=VALUE(I)
      IF(P.EQ.0.0)   P=0.25
      Y=ALOG(P)
      SUM=SUM+Y
      SUM2=SUM2+Y*Y
   10 CONTINUE
      Q=FLOAT(N)
      U=SUM/Q
      VAR=SUM2/Q-U*U
      SIG=SQRT(VAR)
      RETURN
      END
```

```
      SUBROUTINE BMLE(U,SIG,LIMIT,XI,JH,PV,CUM)
C SUBROUTINE TO COMPUTE PARAMETERS BY M.L.E. OPTIMIZATION METHOD
C THE LIKELIHOOD FUNCTION IS COMPUTED FOR THE INTERVALS
      DIMENSION XI(100),JH(100),PV(100),CUM(100)
      ICOUNT=0
      C=1.0/SQRT(3.141592654)
      SUM=0.0
      WRITE(6,200)
C SET ARBITRARY STARTING POINTS
C     U=1.0
C     SIG=0.5
C SET L=1 IF U IS TO BE INCREMENTED
C SET L=2 IF SIG IS TO BE INCREMENTED
      L=2
      K=0
```

```
      SUMT=0.0
    1 CONTINUE
      MEM=0
      D=0.1
    5 CONTINUE
C COMPUTE LOG-LIKELIHOOD FUNCTION FOR U, SIG
      TU=U
      TSIG=SIG
      K1=0
      DO 10 I=1,3
      TEMP=0.0
      DO 20 J=1,LIMIT
      XL=ALOG(XI(J))
      Y=(XL-TU)/TSIG
      CDF=GAUCDF(Y)
      PROB=CDF-TEMP
      TEMP=CDF
C......DETERMINE OCCURRENCE OF ZEROS
      IF(PROB.EQ.0.0)   K1=K1+1
      IF(PROB.EQ.0.0)   GO TO 20
      IF(JH(J).EQ.0)    GO TO 20
C......CALCULATE LOG-LIKELIHOOD FUNCTION
      Z=ALOG(PROB)*FLOAT(JH(J))
      SUM=SUM+Z
   20 CONTINUE
C SUM1 IS SUM(X); SUM2 IS SUM(X-D); SUM3 IS SUM(X+D)
      IF(I.EQ.1)   SUM1=SUM
      IF(I.EQ.2)   SUM2=SUM
      IF(I.EQ.3)   SUM3=SUM
      SUM=0.0
      IF(I.EQ.1.AND.L.EQ.1)  TU=U-D
      IF(I.EQ.1.AND.L.EQ.2)  TSIG=SIG-D
      IF(I.EQ.2.AND.L.EQ.1)  TU=U+D
      IF(I.EQ.2.AND.L.EQ.2)  TSIG=SIG+D
   10 CONTINUE
      IF(MEM.NE.2)   DELTA=D
      IF(MEM.EQ.2)   DELTA=-1.0*D
C PRINT RESULTS OF EACH ITERATION
      WRITE(6,201) ICOUNT,L,K1,MEM,SUM1,SUM3,U,SIG,DELTA
      ICOUNT=ICOUNT+1
C LIMIT THE MAXIMUM NUMBER OF RUNS
      IF(ICOUNT.GT.200)   GO TO 50
C OTT'S OPTIMIZATION METHOD:
C   THE GOAL IS TO FIND THE MAXIMUM VALUE OF SUM
C   MODIFY U,SIG BY COMPARING THE MAGNITUDE OF SUM2 AND SUM3 WITH SUM1
C   IF SUM3 IS GREATER THAN SUM1, ADVANCE FORWARD (SET MEM=1)
C   IF SUM2 IS GREATER THAN SUM1, MOVE BACKWARD (SET MEM=2)
C   IF YOU HAVE OVERSHOT, DIVIDE BY 2.0 AND RECALCULATE
C   IF NEITHER SUM2 NOR SUM3 EXCEEDS SUM1, DIVIDE BY 2.0 AND RECALCULATE
C......SET DIRECTION
      IF(MEM.EQ.0.AND.SUM3.GT.SUM1)   MEM=1
      IF(MEM.EQ.0.AND.SUM2.GT.SUM1)   MEM=2
C......TEST FOR OVERSHOOT
      IF(MEM.EQ.1.AND.SUM3.LT.SUM1)   GO TO 25
      IF(MEM.EQ.2.AND.SUM2.LT.SUM1)   GO TO 25
C......TEST FOR TROUGH
      IF(SUM3.LE.SUM1.AND.SUM2.LE.SUM1)   GO TO 25
C......INCREMENT VARIABLE
      IF(MEM.EQ.1.AND.L.EQ.1)   U=U+D
      IF(MEM.EQ.1.AND.L.EQ.2)   SIG=SIG+D
      IF(MEM.EQ.2.AND.L.EQ.1)   U=U-D
```

```
      IF(MEM.EQ.2.AND.L.EQ.2)   SIG=SIG-D
      GO TO 29
   25 D=D/2.0
      MEM=0
   29 CONTINUE
C RETURN IF CONVERGENCE IS NOT YET OBTAINED
      TEST=ABS(DELTA)
      IF(TEST.GE.0.000001)   GO TO 5
C ALLOW FOR ONLY 3 REVERSALS OF L AFTER CONVERGENCE
      T=ABS(SUM1-SUMT)
      SUMT=SUM1
      IF(T.LT.0.0001)  K=K+1
      IF(K.GE.3)   GO TO 50
C CONVERGENE IS OBTAINED; REVERSE L
      GO TO (30,40), L
   30 L=2
      GO TO 1
   40 L=1
      GO TO 1
   50 CONTINUE
C CHANGE PAGE
      WRITE(6,203)
  203 FORMAT(1H1)
  200 FORMAT(1H0,1X,'NUMBER',3X,'L',2X,'K1',1X,'MEM',9X,'SUM1',15X,
     1'SUM3',17X,'MEAN',15X,'SIGMA',13X,'CHANGE')
  201 FORMAT(1H ,2X,I3,5X,I1,2X,I2,2X,I1,1X,2(1X,F18.8),3(1X,E18.6))
      RETURN
      END
```

```
      SUBROUTINE LN(U,SIG,N,LIMIT,K,X,DIFF,XI,JH,PV,CUM,IH)
C SUBROUTINE TO CALCULATE LOGNORMAL FREQUENCIES FOR PARAMETERS AVE,STD
      DIMENSION XI(100),JH(100),PV(100),CUM(100)
      DIMENSION X(100),DIFF(100)
C...INITIALIZE VALUES
      TEMP=0.0
      Q=FLOAT(N)
      SUMCHI=0.0
      SUMSQ=0.0
      SUMS=0.0
      SUMZ=0.0
      JSUM=0
      JOSUM=0
      NSUM=0
      ISUM=0
      ISUMD=0
      ISTOP=LIMIT
      DSUM=0.0
      CUMAX=0.0
      M=0
      S=0.0
      Z=0.0
C...COMPUTE PARAMTERS
      VAR=SIG*SIG
      AVE=EXP(U+0.5*VAR)
      STD=SQRT(EXP(2.0*U+VAR)*(EXP(VAR)-1.0))
      GMED=EXP(U)
      GMODE=EXP(U-VAR)
```

```
      SGD=EXP(SIG)
      W=EXP(VAR)
      B1SQ=(W+2.0)*SQRT(W-1.0)
      B2=3.0+(W-1.0)*(W*W*W+3.0*W*W+6.0*W+6.0)
      WRITE(6,200) AVE,STD
      WRITE(6,201) AVE,STD
      WRITE(6,202) GMED,GMODE,SGD
      WRITE(6,203) U,SIG
      WRITE(6,207) B1SQ,B2
      WRITE(6,199)
      WRITE(6,204)
      WRITE(6,205)
      WRITE(6,206)
      WRITE(6,204)
C...MAIN COMPUTATION
      DO 10 I=1,LIMIT
      M=M+1
      XL=ALOG(XI(I))
      Y=(XL-U)/SIG
      CDF=GAUCDF(Y)
      PROB=CDF-TEMP
C...J IS THE PREDICTED NO. OF VALUES IN THE INTERVAL
C...JO IS THE OBSERVED NO. OF VALUES IN THE INTERVAL
      J=IFIX(PROB*Q+0.5)
      JO=JH(I)
C...STOP COMPUTATION IF NO. OF PREDICTED VALUES IS LESS THAN 5 AFTER MEDIAN
      VAL=XI(I)
      IF(J.LT.5.AND.VAL.GT.GMED)  ISTOP=I
      IF(J.LT.5.AND.VAL.GT.GMED)  GO TO 12
      JSUM=JSUM+J
      JOSUM=JOSUM+JO
      NDIF=J-JO
      NSUM=NSUM+NDIF
      NDIF2=NDIF*NDIF
      ISUMD=ISUMD+NDIF2
      E=FLOAT(J)
      O=FLOAT(JO)
      SQ=(E-O)*(E-O)
      IF(E.EQ.0.0)   CHISQ=0.0
      IF(E.EQ.0.0)   GO TO 15
      CHISQ=SQ/E
C...COMPUTE SUM OF SQUARES OF DIFF. BETWEEN PREDICTED AND OBSERVED NO.
   15 SUMSQ=SUMSQ+SQ
C...COMPUTE SUM OF CHI-SQUARE
      SUMCHI=SUMCHI+CHISQ
      WRITE(6,210) XI(I),XL,Y,CDF,PROB,J,JH(I),NDIF,NDIF2,CHISQ
C......FIND KOLMOGOROV-SMIRNOV RELATED STATISTICS
      CUMDIF=ABS(CDF-FLOAT(JOSUM)/Q)
      IF(CUMDIF.GT.CUMAX)   INT=I
      IF(CUMDIF.GT.CUMAX)  CUMAX=CUMDIF
      DSUM=DSUM+CUMDIF
C......CALCULATE LOG-LIKELIHOOD FUNCTION
      TEMP=CDF
      IF(PROB.EQ.0.0)   GO TO 10
      IF(JH(I).EQ.0)    GO TO 10
      Z=ALOG(PROB)*FLOAT(JH(I))
      SUMZ=SUMZ+Z
   10 CONTINUE
   12 CONTINUE
      WRITE(6,204)
      WRITE(6,209) JSUM,JOSUM,NSUM,ISUMD,SUMCHI
```

```
C SUM UP REMAINING VALUES
C...COMPUTE SUM OF OBSERVED NO. OF VALUES FOR LAST INTERVAL
      DO 20 I=ISTOP,LIMIT
      ISUM=ISUM+JH(I)
C......CALCULATE LOG-LIKELIHOOD FUNCTION FOR LAST INTERVALS
      XL=ALOG(XI(I))
      Y=(XL-U)/SIG
      CDF=GAUCDF(Y)
      PROB=CDF-TEMP
      TEMP=CDF
      IF(PROB.EQ.0.0)   GO TO 20
      IF(JH(I).EQ.0)    GO TO 20
      Z=ALOG(PROB)*FLOAT(JH(I))
      SUMZ=SUMZ+Z
   20 CONTINUE
      ISUM=ISUM+IH
      J=N-JSUM
      PROB=FLOAT(J)/Q
      NDIF=J-ISUM
      NDIF2=NDIF*NDIF
      E=FLOAT(J)
      O=FLOAT(ISUM)
      SQ=(E-O)*(E-O)
      IF(E.EQ.0.0)   GO TO 25
      CHISQ=SQ/E
   25 SUMSQ=SUMSQ+SQ
      SUMCHI=SUMCHI+CHISQ
      WRITE(6,212) PROB,J,ISUM,NDIF,NDIF2,CHISQ
      RMS1=SQRT(FLOAT(ISUMD)/FLOAT(M))
      WRITE(6,211) SUMSQ,RMS1,SUMCHI
      NDF=ISTOP-3
      WRITE(6,213) NDF
      WRITE(6,214)
      WRITE(6,215) CUMAX,INT
      WRITE(6,216) DSUM
C COMPARISON OF X FOR OBSERVED FREQUENCIES WITH X FOR MODEL
      WRITE(6,225) SUMZ
      WRITE(6,198)
      WRITE(6,220) AVE,STD
      WRITE(6,199)
C...DETERMINE LAST INTERVAL WITH NUMERICAL VALUES
      DO 30 I=1,LIMIT
      IF(CUM(I).LT.99.95)   K=I
   30 CONTINUE
C...FIND CORRESPONDING X VALUE FOR EACH CUMULATIVE FREQUENCY
      WRITE(6,221)
      WRITE(6,222)
      DO 40 I=1,K
      CDF=CUM(I)/100.0
      IF(CDF.EQ.0.0)   X(I)=0.0
      IF(CDF.EQ.0.0)   DIFF(I)=0.0
      IF(CDF.EQ.0.0)   Y=0.0
      IF(CDF.EQ.0.0)   GO TO 35
      Y=GCDFI(CDF)
      X(I)=EXP(SIG*Y+U)
      DIFF(I)=X(I)-XI(I)
      S=DIFF(I)*DIFF(I)
      SUMS=SUMS+S
   35 WRITE(6,223) CUM(I),Y,X(I),XI(I),DIFF(I),S
   40 CONTINUE
      WRITE(6,222)
```

```
      RMS2=SQRT(SUMS/FLOAT(K))
      WRITE(6,224)  RMS2
      WRITE(6,198)
198 FORMAT(1H1)
199 FORMAT(1H0)
200 FORMAT(1H0,1X,'COMPARISON OF LOGNORMAL PROBABILITY MODEL (',
     1F7.4,',',F7.4,') WITH OBSERVED DATA')
201 FORMAT(1H0,4X,'ARITH. MEAN =',F15.8,6X,'ARITH. STD. DEV. =',F15.8)
202 FORMAT(1H0,4X,'MEDIAN =',F15.8,8X,'MODE =',F16.8,8X,
     1'GEO. STD. DEV. =',F16.8)
203 FORMAT(1H0,4X,'MEAN OF NORMAL DIST. =',F15.8,
     17X,'STD. DEV. OF NORMAL DIST. =',F15.8)
204 FORMAT(1H ,1X,'-------------------------------------------------
     1-------------------------------------------------------------
     2------')
205 FORMAT(1H ,3X,'(CUM)',6X,'LN(X)',8X,'Y',9X,'F(Y)',8X,'P(Y)',
     17X,'PREDICTED',5X,'OBSERVED',4X,'DIFFERENCE',6X,'DIFF. SQ.',
     23X,'CHI SQUARE')
206 FORMAT(1H ,1X,'INTERVAL',55X,'N',12X,'N')
207 FORMAT(1H0,4X,'COEF. OF SKEWNESS = ',F15.8,7X,
     1'COEF. OF KURTOSIS = ',F15.8)
209 FORMAT(1H ,45X,'SUBTOTAL:',7X,3(I5,8X),I9,5X,F7.2)
210 FORMAT(2X,3(F7.4,4X),2(F9.7,3X),3X,3(I5,8X),I9,5X,F7.2)
211 FORMAT(/,6X,'SUM OF SQUARES =',F12.2,8X,'ROOT MEAN SQUARE OF DIFFE
     1RENCE =',F9.3,8X,'CHI SQUARE = ',F9.2)
212 FORMAT(1H0,5X,'RESULTS FOR REMAINING INTERVALS:',9X,F9.7,
     16X,3(I5,8X),I9,5X,F7.2)
213 FORMAT(1H0,5X,'DEGREES OF FREDOM = ',I4)
214 FORMAT(1H0,///,1X,'KOLMOGOROV SMIRNOV RELATED MEASURES:')
215 FORMAT(1H0,1X,'MAXIMUM ABSOLUTE DIFFERENCE BETWEEN CDF AND CUMULAT
     1IVE FREQUENCY OF OBSERVATIONS =',F12.7,' (AT INTERVAL NO.',I3,')')
216 FORMAT(1H0,1X,'SUM OF ABSOLUTE DIFFERENCES BETWEEN CDF AND CUMULAT
     1IVE FREQUENCY OF OBSERVATIONS =',F12.7)
220 FORMAT(1H0,1X,'COMPARISON OF X FOR OBSERVED FREQUENCIES WITH X FOR
     1 LOGNORMAL PROBABILITY MODEL (',F7.4,',',F7.4,')')
221 FORMAT(3X,'CUM. FREQ.',4X,'PROBITS',6X,'PREDICTED X',4X,'OBSERVED
     1X',5X,'DIFFERENCE',4X,'DIFF. SQUARED')
222 FORMAT(1H ,1X,'-------------------------------------------------
     1-----------------------------------')
223 FORMAT(4X,F7.3,6X,F7.4,3(4X,F11.6),5X,F13.6)
224 FORMAT(/,9X,'ROOT MEAN SQUARE OF DIFFERENCE =',F7.3)
225 FORMAT(1H0,/,1X,'LOG-LIKELIHOOD FUNCTION =',F18.7)
      RETURN
      END
```

```
      FUNCTION GAUCDF(X)
      DIMENSION A(5)
      DATA A/.31938153,-.356563782,1.781477937,
     *-1.821255978,1.330274429/
      IF(X.NE.0.0)  GO TO 1
      GAUCDF=0.5
      RETURN
    1 MEMORY=0
      PDFN=0.39894228*EXP(-0.5*X*X)
      IF(X.GT.0.0)   MEMORY=1
      T=1.0/(1.0+0.2316419*ABS(X))
```

```
      SUM=0.0
      V=1.0
      DO 10 I=1,5
      V=T*V
      SUM=SUM+A(I)*V
   10 CONTINUE
      GAUCDF=SUM*PDFN
      IF(MEMORY.EQ.1)    GAUCDF=1.0-GAUCDF
      RETURN
      END
```

```
      FUNCTION GCDFI(Q)
      DIMENSION A(3),B(3)
C INVERSE GAUSSIAN CUMULATIVE DISTRIBUTION FUNCTION
      DATA A/2.515517,0.802853,0.010328/
      DATA B/1.432788,0.189269,0.001308/
      P=Q
      IF(P.GT..5) P=1.-P
      ETA=SQRT(ALOG(1./(P*P)))
      XNUM=0.
      DENOM=1.
      ETAP=1.0
      DO 10 I=1,3
      XNUM=XNUM+A(I)*ETAP
      ETAP=ETAP*ETA
   10 DENOM=DENOM+B(I)*ETAP
      GCDFI=XNUM/DENOM-ETA
      IF(P.NE.Q) GCDFI=-GCDFI
      RETURN
      END
```

```
      SUBROUTINE AUTCOR(L,N,STORE,ISTART,JUMP)
C SUBROUTINE TO COMPUTE AUTOCORRELATION FUNCTION FOR LAGS 1-75
      COMMON /B/MARK1,MARK2,IBLANK,LINE(50)
      COMMON /G/DUMMY
      DIMENSION STORE(9000)
      WRITE(6,200)
      WRITE(6,201)
      WRITE(6,202)
      WRITE(6,203)
C CALCULATE MEAN AND VARIANCE
      SUM1=0.0
      SUM2=0.0
      IF(N.EQ.0)    GO TO 90
      T=FLOAT(N)
      DO 5 I=1,L
      IF(STORE(I).EQ.DUMMY)    GO TO 5
      Z=STORE(I)
      SUM1=SUM1+Z
      SUM2=SUM2+Z*Z
    5 CONTINUE
      AVE=SUM1/T
      VAR=SUM2/T-AVE*AVE
```

```
C BEGIN COMPUTATION
      J=75*JUMP
      DO 10 M=ISTART,J,JUMP
      LAST=L-M
C INITIALIZE VALUES
      COUNT=0.0
      SUM3=0.
C CALCULATE COVARIANCE FOR LAG M
      DO 20 I=1,LAST
      IF(STORE(I).EQ.DUMMY) GO TO 20
      IF(STORE(I+M).EQ.DUMMY) GO TO 20
      SUM3=SUM3+STORE(I)*STORE(I+M)
      COUNT=COUNT+1.0
   20 CONTINUE
      COV=SUM3/COUNT-AVE*AVE
      R=COV/VAR
C GENERATE GRAPH OF CORRELATION COEFFICIENT R
      IF(R) 25,26,26
   25 NUM1=IFIX(20.0*(R+1.0))
      NUM2=20
      GO TO 27
   26 NUM1=20
      NUM2=IFIX(20.0*R)+20
   27 CONTINUE
      DO 30 I2=NUM1,NUM2
      LINE(I2)=MARK1
   30 CONTINUE
      D=FLOAT(M)/24.0
      WRITE(6,204) M,D,COUNT,COV,R,(LINE(I1),I1=1,40)
      DO 40 I2=NUM1,NUM2
      LINE(I2)=IBLANK
   40 CONTINUE
   10 CONTINUE
      WRITE(6,203)
      WRITE(6,205) AVE,VAR
      RETURN
   90 WRITE(6,210)
  210 FORMAT(1H0,1X,'ERROR:  N = 0;  DIVISION IS IMPOSSIBLE')
  200 FORMAT(1H1,///,30X,'* * * * * AUTOCORRELATION RESULTS * * * * *',
     1//)
  201 FORMAT(1H ,6X,'LAG',4X,'LAG',4X,'PAIRS',3X,'COVARIANCE',3X,
     1'CORRELATION',11X,'AUTOCORRELATION FUNCTION')
  202 FORMAT(1H ,5X,'HOURS',3X,'DAY',25X,'COEFFICIENT',1X,'-1.0',1X,
     1'-8',2X,'-6',2X,'-4',2X,'-2',3X,'0',2X,'+2',2X,'+4',2X,'+6',2X,
     2'+8',1X,'+1.0')
  203 FORMAT(1H ,55X,'+---+---+---+---+---+---+---+---+---+---+')
  204 FORMAT(1H ,5X,I4,2X,F6.2,3X,F5.0,3X,F8.4,9X,F4.2,6X,'I',40A1,'I')
  205 FORMAT(1H0,15X,'FOR THIS AUTOCORRELATION:  MEAN =',F8.4,3X,
     1'AND   VARIANCE =',F9.4)
      RETURN
      END
```

APPENDIX B

SAMPLE OUTPUT FOR NEW YORK CITY

NEW YORK CITY, NY

YR=74 STATE=33 AREA=4680 SITE=023 AGCY=H PROJ=01 POLL=0001200 METH=11 UNT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 8364

BASIC STATISTICS-----

N = 8364

MEAN = 16.2074280

2ND MOMENT (VAR.) = 47.1796722 STANDARD DEV. = 6.86874580

COEFFICIENT OF VARIATION = 0.42380232

3RD MOMENT = 260.864258 COEF. OF SKEWNESS = 0.80497521

4TH MOMENT = 9293.21875 COEF. OF KURTOSIS = 4.17499828

LOWEST VALUE = 2.00000000

HIGHEST VALUE = 57.0000000

SECOND HIGHEST VALUE = 56.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 0

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	0	0.0 %	I I	0.0 %	0.0 %
0.50- 1.50	0	0.0 %	I I	0.0 %	0.0 %
1.50- 2.50	8	0.096%	I I	0.096%	0.096%
2.50- 3.50	14	0.167%	I I	0.263%	0.263%
3.50- 4.50	63	0.753%	IC I	1.016%	1.016%
4.50- 5.50	139	1.662%	IC* I	2.678%	2.678%
5.50- 6.50	174	2.080%	I*C I	4.758%	4.758%
6.50- 7.50	265	3.168%	I***C I	7.927%	7.926%
7.50- 8.50	359	4.292%	I**** C I	12.219%	12.218%
8.50- 9.50	375	4.484%	I**** C I	16.703%	16.701%
9.50- 10.50	398	4.758%	I***** C I	21.461%	21.458%

10.50- 11.50	391	4.675%	I*****	26.136%	26.133%
11.50- 12.50	457	5.464%	I*****	31.600%	31.596%
12.50- 13.50	547	6.540%	I*******	38.140%	38.135%
13.50- 14.50	498	5.954%	I******	44.094%	44.088%
14.50- 15.50	519	6.205%	I******	50.299%	50.293%
15.50- 16.50	499	5.966%	I******	56.265%	56.258%
16.50- 17.50	496	5.930%	I******	62.195%	62.188%
17.50- 18.50	473	5.655%	I******	67.850%	67.842%
18.50- 19.50	379	4.531%	I*****	72.382%	72.373%
19.50- 20.50	336	4.017%	I****	76.399%	76.390%
20.50- 21.50	252	3.013%	I***	79.412%	79.402%
21.50- 22.50	284	3.396%	I***	82.807%	82.797%
22.50- 23.50	276	3.300%	I***	86.107%	86.097%
23.50- 24.50	182	2.176%	I**	88.283%	88.272%
24.50- 25.50	192	2.296%	I**	90.579%	90.568%
25.50- 26.50	141	1.686%	I**	92.264%	92.253%
26.50- 27.50	108	1.291%	I*	93.556%	93.544%
27.50- 28.50	112	1.339%	I*	94.895%	94.883%
28.50- 29.50	72	0.861%	I*	95.755%	95.744%
29.50- 30.50	74	0.885%	I*	96.640%	96.629%
30.50- 31.50	52	0.622%	I*	97.262%	97.250%
31.50- 32.50	52	0.622%	I*	97.884%	97.872%
32.50- 33.50	53	0.634%	I*	98.517%	98.506%
33.50- 34.50	7	0.084%	I	98.601%	98.589%
34.50- 35.50	28	0.335%	I	98.936%	98.924%
35.50- 36.50	17	0.203%	I	99.139%	99.127%
36.50- 37.50	15	0.179%	I	99.318%	99.306%
37.50- 38.50	11	0.132%	I	99.450%	99.438%
38.50- 39.50	6	0.072%	I	99.522%	99.510%
39.50- 40.50	8	0.096%	I	99.617%	99.605%
40.50- 41.50	2	0.024%	I	99.641%	99.629%
41.50- 42.50	8	0.096%	I	99.737%	99.725%
42.50- 43.50	6	0.072%	I	99.808%	99.797%
43.50- 44.50	0	0.0 %	I	99.808%	99.797%
44.50- 45.50	3	0.036%	I	99.844%	99.832%
45.50- 46.50	1	0.012%	I	99.856%	99.844%
46.50- 47.50	2	0.024%	I	99.880%	99.868%
47.50- 48.50	2	0.024%	I	99.904%	99.892%
48.50- 49.50	2	0.024%	I	99.928%	99.916%
49.50- 50.50	3	0.036%	I	99.964%	99.952%
50.50- 51.50	0	0.0 %	I	99.964%	99.952%
51.50- 52.50	1	0.012%	I	99.976%	99.964%
52.50- 53.50	0	0.0 %	I	99.976%	99.964%
53.50- 54.50	0	0.0 %	I	99.976%	99.964%
54.50- 55.50	0	0.0 %	I	99.976%	99.964%
55.50- 56.50	1	0.012%	I	99.988%	99.976%
56.50- 57.50	1	0.012%	I	100.000%	99.988%
57.50- 58.50	0	0.0 %	I	100.000%	99.988%
58.50- 59.50	0	0.0 %	I	100.000%	99.988%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

PARAMETERS EST. BY THE METHOD OF MOMENTS-----

COMPARISON OF LOGNORMAL PROBABILITY MODEL (16.2074, 6.8683) WITH OBSERVED DATA

ARITH. MEAN = 16.2073975 ARITH. STD. DEV. = 6.86831379

MEDIAN = 14.9227457 MODE = 12.6508274 GEO. STD. DEV. = 1.50140858

MEAN OF NORMAL DIST. = 2.70288658 STD. DEV. OF NORMAL DIST. = 0.40640360

COEF. OF SKEWNESS = 1.34743309 COEF. OF KURTOSIS = 6.39294052

(CUM) INTERVAL	LN(X)	Y	F(Y)	P(Y)	PREDICTED N	OBSERVED N	DIFFERENCE	DIFF. SQ.	CHI SQUARE
0.5000	-0.6931	-8.3563	0.0000000	0.0000000	0	0	0	0	0.0
1.5000	0.4055	-5.6531	0.0000000	0.0000000	0	0	0	0	0.0
2.5000	0.9163	-4.3961	0.0000055	0.0000055	0	8	-8	64	0.0
3.5000	1.2528	-3.5682	0.0001798	0.0001743	1	14	-13	169	169.00
4.5000	1.5041	-2.9498	0.0015900	0.0014102	12	63	-51	2601	216.75
5.5000	1.7047	-2.4560	0.0070241	0.0054342	45	139	-94	8836	196.36
6.5000	1.8718	-2.0450	0.0204287	0.0134045	112	174	-62	3844	34.32
7.5000	2.0149	-1.6929	0.0452412	0.0248126	208	265	-57	3249	15.62
8.5000	2.1401	-1.3849	0.0830446	0.0378034	316	359	-43	1849	5.85
9.5000	2.2513	-1.1112	0.1332417	0.0501970	420	375	45	2025	4.82
10.5000	2.3514	-0.8649	0.1935375	0.0602958	504	398	106	11236	22.29
11.5000	2.4423	-0.6411	0.2607332	0.0671957	562	391	171	29241	52.03
12.5000	2.5257	-0.4359	0.3314488	0.0707156	591	457	134	17956	30.38
13.5000	2.6027	-0.2465	0.4026301	0.0711813	595	547	48	2304	3.87
14.5000	2.6741	-0.0707	0.4718137	0.0691836	579	498	81	6561	11.33
15.5000	2.7408	0.0934	0.5372024	0.0653886	547	519	28	784	1.43
16.5000	2.8034	0.2472	0.5976328	0.0604304	505	499	6	36	0.07
17.5000	2.8622	0.3920	0.6524739	0.0548411	459	496	-37	1369	2.98
18.5000	2.9178	0.5287	0.7015091	0.0490352	410	473	-63	3969	9.68
19.5000	2.9704	0.6583	0.7448209	0.0433118	362	379	-17	289	0.80
20.5000	3.0204	0.7813	0.7826983	0.0378774	317	336	-19	361	1.14
21.5000	3.0681	0.8985	0.8155482	0.0328500	275	252	23	529	1.92
22.5000	3.1135	1.0104	0.8438470	0.0282987	237	284	-47	2209	9.32
23.5000	3.1570	1.1174	0.8680875	0.0242406	203	276	-73	5329	26.25
24.5000	3.1987	1.2199	0.8887554	0.0206679	173	182	-9	81	0.47

25.5000	3.2387	1.3184	0.9063102	0.0175548	147	192	-45	2025	13.78
26.5000	3.2771	1.4130	0.9211754	0.0148652	124	141	-17	289	2.33
27.5000	3.3142	1.5042	0.9337309	0.0125555	105	108	-3	9	0.09
28.5000	3.3499	1.5921	0.9443140	0.0105831	89	112	-23	529	5.94
29.5000	3.3844	1.6769	0.9532202	0.0089062	74	72	2	4	0.05
30.5000	3.4177	1.7589	0.9607063	0.0074860	63	74	-11	121	1.92
31.5000	3.4500	1.8383	0.9669926	0.0062863	53	52	1	1	0.02
32.5000	3.4812	1.9152	0.9722680	0.0052755	44	52	-8	64	1.45
33.5000	3.5115	1.9898	0.9766932	0.0044251	37	53	-16	256	6.92
34.5000	3.5410	2.0622	0.9804042	0.0037110	31	7	24	576	18.58
35.5000	3.5695	2.1325	0.9835162	0.0031120	26	28	-2	4	0.15
36.5000	3.5973	2.2008	0.9861260	0.0026098	22	17	5	25	1.14
37.5000	3.6243	2.2673	0.9883153	0.0021893	18	15	3	9	0.50
38.5000	3.6507	2.3321	0.9901521	0.0018368	15	11	4	16	1.07
39.5000	3.6763	2.3952	0.9916941	0.0015420	13	6	7	49	3.77
40.5000	3.7013	2.4567	0.9929892	0.0012951	11	8	3	9	0.82
41.5000	3.7257	2.5167	0.9940775	0.0010883	9	2	7	49	5.44
42.5000	3.7495	2.5753	0.9949926	0.0009151	8	8	0	0	0.0
43.5000	3.7728	2.6325	0.9957625	0.0007700	6	6	0	0	0.0
44.5000	3.7955	2.6885	0.9964109	0.0006484	5	0	5	25	5.00
45.5000	3.8177	2.7431	0.9969573	0.0005464	5	3	2	4	0.80
				SUBTOTAL:	8338	8351	-13	108955	886.47
RESULTS FOR REMAINING INTERVALS:				0.0031086	26	13	13	169	6.50

SUM OF SQUARES = 109124.00 ROOT MEAN SQUARE OF DIFFERENCE = 48.148 CHI SQUARE = 892.97

DEGREES OF FREDOM = 44

KOLMOGOROV SMIRNOV RELATED MEASURES:

MAXIMUM ABSOLUTE DIFFERENCE BETWEEN CDF AND CUMULATIVE FREQUENCY OF OBSERVATIONS = 0.0391456 (AT INTERVAL NO. 9)

SUM OF ABSOLUTE DIFFERENCES BETWEEN CDF AND CUMULATIVE FREQUENCY OF OBSERVATIONS = 0.5496979

LOG-LIKELIHOOD FUNCTION = -27797.6641

COMPARISON OF X FOR OBSERVED FREQUENCIES WITH X FOR LOGNORMAL PROBABILITY MODEL (16.2074, 6.8683)

CUM. FREQ.	PROBITS	PREDICTED X	OBSERVED X	DIFFERENCE	DIFF. SQUARED
0.0	0.0	0.0	0.500000	0.0	**********
0.0	0.0	0.0	1.500000	0.0	**********
0.096	-3.1037	4.227151	2.500000	1.727151	2.983049
0.263	-2.7910	4.800014	3.500000	1.300014	1.690035
1.016	-2.3207	5.810892	4.500000	1.310892	1.718437
2.678	-1.9308	6.808754	5.500000	1.308754	1.712836
4.758	-1.6691	7.572763	6.500000	1.072763	1.150821
7.927	-1.4103	8.412766	7.500000	0.912766	0.833141
12.219	-1.1642	9.297539	8.500000	0.797539	0.636068
16.703	-0.9659	10.077744	9.500000	0.577744	0.333789
21.461	-0.7903	10.823261	10.500000	0.323261	0.104498
26.136	-0.6388	11.510506	11.500000	0.010506	0.000110
31.600	-0.4785	12.285516	12.500000	-0.214484	0.046003
38.140	-0.3014	13.202462	13.500000	-0.297538	0.088529
44.094	-0.1483	14.050035	14.500000	-0.449965	0.202469
50.299	0.0075	14.968099	15.500000	-0.531901	0.282919
56.265	0.1574	15.908269	16.500000	-0.591731	0.350146
62.195	0.3102	16.927521	17.500000	-0.572479	0.327732
67.850	0.4631	18.012894	18.500000	-0.487106	0.237273
72.382	0.5939	18.996063	19.500000	-0.503937	0.253952
76.399	0.7189	19.986542	20.500000	-0.513458	0.263639
79.412	0.8206	20.829773	21.500000	-0.670227	0.449204
82.807	0.9465	21.923203	22.500000	-0.576797	0.332695
86.107	1.0852	23.194275	23.500000	-0.305725	0.093468
88.283	1.1894	24.197495	24.500000	-0.302505	0.091510
90.579	1.3154	25.469589	25.500000	-0.030411	0.000925
92.264	1.4233	26.611298	26.500000	0.111298	0.012387
93.556	1.5188	27.664093	27.500000	0.164093	0.026927
94.895	1.6351	29.002670	28.500000	0.502670	0.252677
95.755	1.7234	30.062241	29.500000	0.562241	0.316114
96.640	1.8308	31.403473	30.500000	0.903473	0.816263
97.262	1.9212	32.578857	31.500000	1.078857	1.163933
97.884	2.0307	34.061890	32.500000	1.561890	2.439499
98.517	2.1751	36.120575	33.500000	2.620575	6.867413
98.601	2.1980	36.457993	34.500000	1.957993	3.833735
98.936	2.3033	38.052414	35.500000	2.552414	6.514816
99.139	2.3824	39.295166	36.500000	2.795166	7.812953
99.318	2.4672	40.672928	37.500000	3.172928	10.067471
99.450	2.5430	41.945450	38.500000	3.445450	11.871124
99.522	2.5914	42.779358	39.500000	3.279358	10.754188
99.617	2.6673	44.118225	40.500000	3.618225	13.091553
99.641	2.6889	44.507172	41.500000	3.007172	9.043080
99.737	2.7907	46.388184	42.500000	3.888184	15.117971
99.808	2.8921	48.340332	43.500000	4.840332	23.428802
99.808	2.8921	48.340332	44.500000	3.840332	14.748150
99.844	2.9567	49.624512	45.500000	4.124512	17.011597
99.856	2.9812	50.121719	46.500000	3.621719	13.116851
99.880	3.0365	51.260635	47.500000	3.760635	14.142378
99.904	3.1030	52.664352	48.500000	4.164352	17.341827
99.928	3.1869	54.491104	49.500000	4.991104	24.911118

ROOT MEAN SQUARE OF DIFFERENCE = 2.186

```
                    *  *  *   DISTRIBUTION OF DATA (INTERVAL WIDTH =  0.5000)   *  *  *

NO. OF VALUES LESS THAN  -5.0000=      0      NO. OF ZERO VALUES =      2

   INTERVAL      NUMBER    PERCENT  +---------------------HISTOGRAM-------------------+ CUM. FREQ.   PLOT POS.
 -5.00- -4.50        0     0.0  %  I                                                I    0.0  %       0.0  %
 -4.50- -4.00        0     0.0  %  I                                                I    0.0  %       0.0  %
 -4.00- -3.50        0     0.0  %  I                                                I    0.0  %       0.0  %
 -3.50- -3.00        0     0.0  %  I                                                I    0.0  %       0.0  %
 -3.00- -2.50        0     0.0  %  I                                                I    0.0  %       0.0  %
 -2.50- -2.00        0     0.0  %  I                                                I    0.0  %       0.0  %
 -2.00- -1.50        0     0.0  %  I                                                I    0.0  %       0.0  %
 -1.50- -1.00        0     0.0  %  I                                                I    0.0  %       0.0  %
 -1.00- -0.50        7    14.000%  I******C*******                                  I   14.000%      13.725%
 -0.50-  0.0         7    14.000%  I*************C                                   I   28.000%      27.451%
  0.0 -  0.50        6    12.000%  I************       C                            I   40.000%      39.216%
  0.50-  1.00        6    12.000%  I************              C                     I   52.000%      50.980%
  1.00-  1.50        5    10.000%  I**********                   C                 I   62.000%      60.784%
  1.50-  2.00        3     6.000%  I******                          C              I   68.000%      66.667%
  2.00-  2.50        0     0.0  %  I                                C              I   68.000%      66.667%
  2.50-  3.00        3     6.000%  I******                             C           I   74.000%      72.549%
  3.00-  3.50        4     8.000%  I********                              C        I   82.000%      80.392%
  3.50-  4.00        5    10.000%  I**********                                 C    I   92.000%      90.196%
  4.00-  4.50        2     4.000%  I****                                        C   I   96.000%      94.118%
  4.50-  5.00        2     4.000%  I****                                          CI  100.000%      98.039%
                                   +=================================================+

NO. OF VALUES EQUAL TO OR GREATER THAN    5.0000 =    0  ( 0.0  %)

   N =      50

   MEAN =        1.35716152

   2ND MOMENT (VAR.) =          2.99574184              STANDARD DEV. =         1.73082161

        COEFFICIENT OF VARIATION =     1.27532387

   3RD MOMENT =          2.78621960          COEF. OF SKEWNESS =     0.53735220

   4TH MOMENT =         16.7045898           COEF. OF KURTOSIS =     1.86134529

   LOWEST VALUE =       -0.67022705

   HIGHEST VALUE =       4.99110413

        SECOND HIGHEST VALUE =       4.84033203
```

PARAMETERS EST. BY THE METHOD OF LARSEN-----

FOR THE 70 PERCENTILE, X = 19.000 AND Y = 0.52440; FOR THE 99.9 PERCENTILE, X = 48.000 AND Y = 3.09023

COMPARISON OF LOGNORMAL PROBABILITY MODEL (16.7812, 6.2644) WITH OBSERVED DATA

ARITH. MEAN = 16.7812042 ARITH. STD. DEV. = 6.26444054

MEDIAN = 15.7214928 MODE = 13.7985983 GEO. STD. DEV. = 1.43504143

MEAN OF NORMAL DIST. = 2.75502872 STD. DEV. OF NORMAL DIST. = 0.36119395

COEF. OF SKEWNESS = 1.17192364 COEF. OF KURTOSIS = 5.53756428

(CUM) INTERVAL	LN(X)	Y	F(Y)	P(Y)	PREDICTED N	OBSERVED N	DIFFERENCE	DIFF. SQ.	CHI SQUARE
0.5000	-0.6931	-9.5466	0.0000000	0.0000000	0	0	0	0	0.0
1.5000	0.4055	-6.5050	0.0000000	0.0000000	0	0	0	0	0.0
2.5000	0.9163	-5.0907	0.0000002	0.0000002	0	8	-8	64	0.0
3.5000	1.2528	-4.1592	0.0000160	0.0000158	0	14	-14	196	0.0
4.5000	1.5041	-3.4634	0.0002668	0.0002508	2	63	-61	3721	1860.50
5.5000	1.7047	-2.9078	0.0018200	0.0015532	13	139	-126	15876	1221.23
6.5000	1.8718	-2.4453	0.0072366	0.0054166	45	174	-129	16641	369.80
7.5000	2.0149	-2.0491	0.0202257	0.0129891	109	265	-156	24336	223.27
8.5000	2.1401	-1.7026	0.0443231	0.0240974	202	359	-157	24649	122.02
9.5000	2.2513	-1.3946	0.0815615	0.0372384	311	375	-64	4096	13.17
10.5000	2.3514	-1.1176	0.1318787	0.0503172	421	398	23	529	1.26
11.5000	2.4423	-0.8657	0.1933298	0.0614510	514	391	123	15129	29.43
12.5000	2.5257	-0.6348	0.2627665	0.0694367	581	457	124	15376	26.46
13.5000	2.6027	-0.4218	0.3365983	0.0738318	618	547	71	5041	8.16
14.5000	2.6741	-0.2239	0.4114087	0.0748104	626	498	128	16384	26.17
15.5000	2.7408	-0.0393	0.4843334	0.0729247	610	519	91	8281	13.58
16.5000	2.8034	0.1338	0.5532229	0.0688895	576	499	77	5929	10.29
17.5000	2.8622	0.2967	0.6166579	0.0634350	531	496	35	1225	2.31
18.5000	2.9178	0.4506	0.6738479	0.0571900	478	473	5	25	0.05
19.5000	2.9704	0.5963	0.7245178	0.0506698	424	379	45	2025	4.78
20.5000	3.0204	0.7348	0.7687616	0.0442439	370	336	34	1156	3.12
21.5000	3.0681	0.8666	0.8069293	0.0381677	319	252	67	4489	14.07
22.5000	3.1135	0.9925	0.8395239	0.0325946	273	284	-11	121	0.44
23.5000	3.1570	1.1129	0.8671236	0.0275997	231	276	-45	2025	8.77
24.5000	3.1987	1.2283	0.8903272	0.0232036	194	182	12	144	0.74

25.5000	3.2387	1.3390	0.9097192	0.0193920	162	192	-30	900	5.56
26.5000	3.2771	1.4455	0.9258450	0.0161258	135	141	-6	36	0.27
27.5000	3.3142	1.5481	0.9391985	0.0133535	112	108	4	16	0.14
28.5000	3.3499	1.6470	0.9502178	0.0110193	92	112	-20	400	4.35
29.5000	3.3844	1.7424	0.9592848	0.0090670	76	72	4	16	0.21
30.5000	3.4177	1.8347	0.9667280	0.0074432	62	74	-12	144	2.32
31.5000	3.4500	1.9241	0.9728265	0.0060984	51	52	-1	1	0.02
32.5000	3.4812	2.0106	0.9778154	0.0049890	42	52	-10	100	2.38
33.5000	3.5115	2.0945	0.9818918	0.0040764	34	53	-19	361	10.62
34.5000	3.5410	2.1759	0.9852195	0.0033277	28	7	21	441	15.75
35.5000	3.5695	2.2550	0.9879344	0.0027148	23	28	-5	25	1.09
36.5000	3.5973	2.3319	0.9901481	0.0022138	19	17	2	4	0.21
37.5000	3.6243	2.4068	0.9919530	0.0018048	15	15	0	0	0.0
38.5000	3.6507	2.4796	0.9934241	0.0014712	12	11	1	1	0.08
39.5000	3.6763	2.5506	0.9946235	0.0011994	10	6	4	16	1.60
40.5000	3.7013	2.6198	0.9956015	0.0009780	8	8	0	0	0.0
41.5000	3.7257	2.6874	0.9963992	0.0007977	7	2	5	25	3.57
42.5000	3.7495	2.7533	0.9970501	0.0006509	5	8	-3	9	1.80
				SUBTOTAL:	8341	8342	-1	169953	4009.59
RESULTS FOR REMAINING INTERVALS:				0.0027499	23	22	1	1	0.04

SUM OF SQUARES = 169954.00 ROOT MEAN SQUARE OF DIFFERENCE = 62.150 CHI SQUARE = 4009.63

DEGREES OF FREDOM = 41

KOLMOGOROV SMIRNOV RELATED MEASURES:

MAXIMUM ABSOLUTE DIFFERENCE BETWEEN CDF AND CUMULATIVE FREQUENCY OF OBSERVATIONS = 0.0854638 (AT INTERVAL NO. 10)

SUM OF ABSOLUTE DIFFERENCES BETWEEN CDF AND CUMULATIVE FREQUENCY OF OBSERVATIONS = 0.6866601

LOG-LIKELIHOOD FUNCTION = -28284.4297

COMPARISON OF X FOR OBSERVED FREQUENCIES WITH X FOR LOGNORMAL PROBABILITY MODEL (16.7812, 6.2644)

CUM. FREQ.	PROBITS	PREDICTED X	OBSERVED X	DIFFERENCE	DIFF. SQUARED
0.0	0.0	0.0	0.500000	0.0	24.911118
0.0	0.0	0.0	1.500000	0.0	24.911118
0.096	-3.1037	5.124271	2.500000	2.624271	6.886800
0.263	-2.7910	5.737024	3.500000	2.237024	5.004277
1.016	-2.3207	6.799137	4.500000	2.299137	5.286031
2.678	-1.9308	7.827483	5.500000	2.327483	5.417177
4.758	-1.6691	8.603411	6.500000	2.103411	4.424336
7.927	-1.4103	9.446548	7.500000	1.946548	3.789046
12.219	-1.1642	10.324555	8.500000	1.824555	3.329002
16.703	-0.9659	11.091081	9.500000	1.591081	2.531537
21.461	-0.7903	11.817367	10.500000	1.317367	1.735455
26.136	-0.6388	12.481959	11.500000	0.981959	0.964244
31.600	-0.4785	13.226154	12.500000	0.726154	0.527300
38.140	-0.3014	14.099941	13.500000	0.599941	0.359930
44.094	-0.1483	14.901627	14.500000	0.401627	0.161304
50.299	0.0075	15.763950	15.500000	0.263950	0.069670
56.265	0.1574	16.640961	16.500000	0.140961	0.019870
62.195	0.3102	17.585236	17.500000	0.085236	0.007265
67.850	0.4631	18.583862	18.500000	0.083862	0.007033
72.382	0.5939	19.482666	19.500000	-0.017334	0.000300
76.399	0.7189	20.382950	20.500000	-0.117050	0.013701
79.412	0.8206	21.145477	21.500000	-0.354523	0.125686
82.807	0.9465	22.129181	22.500000	-0.370819	0.137507
86.107	1.0852	23.265854	23.500000	-0.234146	0.054824
88.283	1.1894	24.158127	24.500000	-0.341873	0.116877
90.579	1.3154	25.283630	25.500000	-0.216370	0.046816
92.264	1.4233	26.288452	26.500000	-0.211548	0.044752
93.556	1.5188	27.210770	27.500000	-0.289230	0.083654
94.895	1.6351	28.377838	28.500000	-0.122162	0.014924
95.755	1.7234	29.297424	29.500000	-0.202576	0.041037
96.640	1.8308	30.456284	30.500000	-0.043716	0.001911
97.262	1.9212	31.467300	31.500000	-0.032700	0.001069
97.884	2.0307	32.737228	32.500000	0.237228	0.056277
98.517	2.1751	34.489975	33.500000	0.989975	0.980050
98.601	2.1980	34.776154	34.500000	0.276154	0.076261
98.936	2.3033	36.124603	35.500000	0.624603	0.390129
99.139	2.3824	37.171295	36.500000	0.671295	0.450637
99.318	2.4672	38.327362	37.500000	0.827362	0.684528
99.450	2.5430	39.391266	38.500000	0.891266	0.794355
99.522	2.5914	40.086517	39.500000	0.586517	0.344003
99.617	2.6673	41.199600	40.500000	0.699600	0.489440
99.641	2.6889	41.522263	41.500000	0.022263	0.000496
99.737	2.7907	43.078308	42.500000	0.578308	0.334440
99.808	2.8921	44.685760	43.500000	1.185760	1.406028
99.808	2.8921	44.685760	44.500000	0.185760	0.034507
99.844	2.9567	45.739258	45.500000	0.239258	0.057244
99.856	2.9812	46.146332	46.500000	-0.353668	0.125081
99.880	3.0365	47.077103	47.500000	-0.422897	0.178842
99.904	3.1030	48.221130	48.500000	-0.278870	0.077768
99.928	3.1869	49.704880	49.500000	0.204880	0.041976

ROOT MEAN SQUARE OF DIFFERENCE = 0.977

```
                    *  *  *   DISTRIBUTION OF DATA (INTERVAL WIDTH =  0.5000)   *  *  *

NO. OF VALUES LESS THAN  -5.0000=     0     NO. OF ZERO VALUES =     2

  INTERVAL      NUMBER   PERCENT  +---------------------HISTOGRAM---------------------+ CUM. FREQ.  PLOT POS.
-5.00- -4.50        0     0.0  %  I                                                  I   0.0  %      0.0  %
-4.50- -4.00        0     0.0  %  I                                                  I   0.0  %      0.0  %
-4.00- -3.50        0     0.0  %  I                                                  I   0.0  %      0.0  %
-3.50- -3.00        0     0.0  %  I                                                  I   0.0  %      0.0  %
-3.00- -2.50        0     0.0  %  I                                                  I   0.0  %      0.0  %
-2.50- -2.00        0     0.0  %  I                                                  I   0.0  %      0.0  %
-2.00- -1.50        0     0.0  %  I                                                  I   0.0  %      0.0  %
-1.50- -1.00        0     0.0  %  I                                                  I   0.0  %      0.0  %
-1.00- -0.50        0     0.0  %  I                                                  I   0.0  %      0.0  %
-0.50-  0.0        16    32.000%  I***************C****************                   I  32.000%     31.373%
 0.0 -  0.50       13    26.000%  I*************************  C                       I  58.000%     56.863%
 0.50-  1.00       11    22.000%  I**********************                 C          I  80.000%     78.431%
 1.00-  1.50        2     4.000%  I****                                   C         I  84.000%     82.353%
 1.50-  2.00        3     6.000%  I******                                     C     I  90.000%     88.235%
 2.00-  2.50        4     8.000%  I********                                       C I  98.000%     96.078%
 2.50-  3.00        1     2.000%  I**                                            CI 100.000%     98.039%
 3.00-  3.50        0     0.0  %  I                                              CI 100.000%     98.039%
 3.50-  4.00        0     0.0  %  I                                              CI 100.000%     98.039%
 4.00-  4.50        0     0.0  %  I                                              CI 100.000%     98.039%
 4.50-  5.00        0     0.0  %  I                                              CI 100.000%     98.039%
                                  +-------------------------------------------------+

NO. OF VALUES EQUAL TO OR GREATER THAN    5.0000 =    0  ( 0.0  %)

  N =      50

  MEAN =        0.52330536

  2ND MOMENT (VAR.) =          0.694542050            STANDARD DEV. =          0.833391905

      COEFFICIENT OF VARIATION =     1.59255314

  3RD MOMENT =          0.581009209          COEF. OF SKEWNESS =     1.00377178

  4TH MOMENT =          1.40156555           COEF. OF KURTOSIS =     2.90546894

  LOWEST VALUE =       -0.42289734

  HIGHEST VALUE =       2.62427139

      SECOND HIGHEST VALUE =       2.32748318
```

PARAMETERS EST. BY APPROXIMATE M. L. E.-----

COMPARISON OF LOGNORMAL PROBABILITY MODEL (16.3493, 7.8954) WITH OBSERVED DATA

ARITH. MEAN = 16.3492737 ARITH. STD. DEV. = 7.89539242

MEDIAN = 14.7224560 MODE = 11.9383097 GEO. STD. DEV. = 1.58066273

MEAN OF NORMAL DIST. = 2.68937397 STD. DEV. OF NORMAL DIST. = 0.45784432

COEF. OF SKEWNESS = 1.56138134 COEF. OF KURTOSIS = 7.62625694

(CUM) INTERVAL	LN(X)	Y	F(Y)	P(Y)	PREDICTED N	OBSERVED N	DIFFERENCE	DIFF. SQ.	CHI SQUARE
0.5000	-0.6931	-7.3879	0.0000000	0.0000000	0	0	0	0	0.0
1.5000	0.4055	-4.9884	0.0000003	0.0000003	0	0	0	0	0.0
2.5000	0.9163	-3.8727	0.0000538	0.0000535	0	8	-8	64	0.0
3.5000	1.2528	-3.1378	0.0008513	0.0007974	7	14	-7	49	7.0C
4.5000	1.5041	-2.5889	0.0048147	0.0039634	33	63	-30	900	27.27
5.5000	1.7047	-2.1506	0.0157551	0.0109404	92	139	-47	2209	24.01
6.5000	1.8718	-1.7857	0.0370739	0.0213188	178	174	4	16	0.09
7.5000	2.0149	-1.4731	0.0703561	0.0332822	278	265	13	169	0.61
8.5000	2.1401	-1.1998	0.1151147	0.0447587	374	359	15	225	0.60
9.5000	2.2513	-0.9568	0.1693246	0.0542099	453	375	78	6084	13.43
10.5000	2.3514	-0.7382	0.2301838	0.0608591	509	398	111	12321	24.21
11.5000	2.4423	-0.5395	0.2947556	0.0645719	540	391	149	22201	41.11
12.5000	2.5257	-0.3574	0.3603871	0.0656315	549	457	92	8464	15.42
13.5000	2.6027	-0.1893	0.4249167	0.0645296	540	547	-7	49	0.09
14.5000	2.6741	-0.0333	0.4867357	0.0618190	517	498	19	361	0.70
15.5000	2.7408	0.1124	0.5447499	0.0580142	485	519	-34	1156	2.38
16.5000	2.8034	0.2490	0.5983048	0.0535550	448	499	-51	2601	5.81
17.5000	2.8622	0.3775	0.6470907	0.0487859	408	496	-88	7744	18.98
18.5000	2.9178	0.4989	0.6910583	0.0439676	368	473	-105	11025	29.96
19.5000	2.9704	0.6138	0.7303371	0.0392788	329	379	-50	2500	7.60
20.5000	3.0204	0.7231	0.7651800	0.0348428	291	336	-45	2025	6.96
21.5000	3.0681	0.8271	0.7959071	0.0307271	257	252	5	25	0.10
22.5000	3.1135	0.9264	0.8228772	0.0269701	226	284	-58	3364	14.88
23.5000	3.1570	1.0214	0.8464591	0.0235819	197	276	-79	6241	31.68
24.5000	3.1987	1.1124	0.8670135	0.0205544	172	182	-10	100	0.58

25.5000	3.2387	1.1998	0.8848837	0.0178702	149	192	-43	1849	12.41
26.5000	3.2771	1.2838	0.9003899	0.0155061	130	141	-11	121	0.93
27.5000	3.3142	1.3647	0.9138235	0.0134336	112	108	4	16	0.14
28.5000	3.3499	1.4427	0.9254468	0.0116233	97	112	-15	225	2.32
29.5000	3.3844	1.5180	0.9354950	0.0100482	84	72	12	144	1.71
30.5000	3.4177	1.5908	0.9441761	0.0086811	73	74	-1	1	0.01
31.5000	3.4500	1.6613	0.9516726	0.0074965	63	52	11	121	1.92
32.5000	3.4812	1.7296	0.9581450	0.0064724	54	52	2	4	0.07
33.5000	3.5115	1.7957	0.9637324	0.0055874	47	53	-6	36	0.77
34.5000	3.5410	1.8600	0.9685565	0.0048240	40	7	33	1089	27.22
35.5000	3.5695	1.9224	0.9727221	0.0041656	35	28	7	49	1.40
36.5000	3.5973	1.9831	0.9763203	0.0035982	30	17	13	169	5.63
37.5000	3.6243	2.0421	0.9794296	0.0031093	26	15	11	121	4.65
38.5000	3.6507	2.0996	0.9821174	0.0026878	22	11	11	121	5.50
39.5000	3.6763	2.1556	0.9844423	0.0023249	19	6	13	169	8.89
40.5000	3.7013	2.2102	0.9864544	0.0020121	17	8	9	81	4.76
41.5000	3.7257	2.2635	0.9881968	0.0017424	15	2	13	169	11.27
42.5000	3.7495	2.3155	0.9897068	0.0015099	13	8	5	25	1.92
43.5000	3.7728	2.3663	0.9910160	0.0013093	11	6	5	25	2.27
44.5000	3.7955	2.4159	0.9921522	0.0011362	10	0	10	100	10.00
45.5000	3.8177	2.4645	0.9931390	0.0009868	8	3	5	25	3.13
46.5000	3.8395	2.5119	0.9939965	0.0008575	7	1	6	36	5.14
47.5000	3.8607	2.5584	0.9947425	0.0007460	6	2	4	16	2.67
48.5000	3.8816	2.6039	0.9953917	0.0006493	5	2	3	9	1.80
49.5000	3.9020	2.6485	0.9959574	0.0005657	5	2	3	9	1.80
				SUBTOTAL:	8329	8358	-29	94623	391.83
RESULTS FOR REMAINING INTERVALS:				0.0041846	35	6	29	841	24.03

SUM OF SQUARES = 95464.00 ROOT MEAN SQUARE OF DIFFERENCE = 43.074 CHI SQUARE = 415.85

DEGREES OF FREDOM = 48

KOLMOGOROV SMIRNOV RELATED MEASURES:

MAXIMUM ABSOLUTE DIFFERENCE BETWEEN CDF AND CUMULATIVE FREQUENCY OF OBSERVATIONS = 0.0457984 (AT INTERVAL NO. 15)

SUM OF ABSOLUTE DIFFERENCES BETWEEN CDF AND CUMULATIVE FREQUENCY OF OBSERVATIONS = 0.7321321

LOG-LIKELIHOOD FUNCTION = -27710.3086

COMPARISON OF X FOR OBSERVED FREQUENCIES WITH X FOR LOGNORMAL PROBABILITY MODEL (16.3493, 7.8954)

CUM. FREQ.	PROBITS	PREDICTED X	OBSERVED X	DIFFERENCE	DIFF. SQUARED
0.0	0.0	0.0	0.500000	0.0	0.041976
0.0	0.0	0.0	1.500000	0.0	0.041976
0.096	-3.1037	3.555015	2.500000	1.055015	1.113055
0.263	-2.7910	4.102253	3.500000	0.602253	0.362709
1.016	-2.3207	5.087785	4.500000	0.587785	0.345491
2.678	-1.9308	6.082258	5.500000	0.582258	0.339025
4.758	-1.6691	6.856419	6.500000	0.356419	0.127034
7.927	-1.4103	7.719060	7.500000	0.219060	0.047987
12.219	-1.1642	8.639547	8.500000	0.139547	0.019473
16.703	-0.9659	9.460540	9.500000	-0.039460	0.001557
21.461	-0.7903	10.252595	10.500000	-0.247405	0.061209
26.136	-0.6388	10.988902	11.500000	-0.511098	0.261221
31.600	-0.4785	11.825923	12.500000	-0.674077	0.454380
38.140	-0.3014	12.824889	13.500000	-0.675111	0.455775
44.094	-0.1483	13.756133	14.500000	-0.743867	0.553338
50.299	0.0075	14.772878	15.500000	-0.727122	0.528707
56.265	0.1574	15.822313	16.500000	-0.677687	0.459259
62.195	0.3102	16.968903	17.500000	-0.531097	0.282064
67.850	0.4631	18.199554	18.500000	-0.300446	0.090267
72.382	0.5939	19.322449	19.500000	-0.177551	0.031524
76.399	0.7189	20.461151	20.500000	-0.038849	0.001509
79.412	0.8206	21.436249	21.500000	-0.063751	0.004064
82.807	0.9465	22.708099	22.500000	0.208099	0.043305
86.107	1.0852	24.196686	23.500000	0.696686	0.485371
88.283	1.1894	25.378906	24.500000	0.878906	0.772476
90.579	1.3154	26.886932	25.500000	1.386932	1.923581
92.264	1.4233	28.248535	26.500000	1.748535	3.057375
93.556	1.5188	29.510620	27.500000	2.010620	4.042593
94.895	1.6351	31.124161	28.500000	2.624161	6.886219
95.755	1.7234	32.408112	29.500000	2.908112	8.457112
96.640	1.8308	34.041565	30.500000	3.541565	12.542682
97.262	1.9212	35.480286	31.500000	3.980286	15.842673
97.884	2.0307	37.305038	32.500000	4.805038	23.088394
98.517	2.1751	39.854675	33.500000	6.354675	40.381897
98.601	2.1980	40.274353	34.500000	5.774353	33.343140
98.936	2.3033	42.264038	35.500000	6.764038	45.752197
99.139	2.3824	43.822220	36.500000	7.322220	53.614899
99.318	2.4672	45.556992	37.500000	8.056992	64.915100
99.450	2.5430	47.165894	38.500000	8.665894	75.097702
99.522	2.5914	48.223618	39.500000	8.723618	76.101501
99.617	2.6673	49.927216	40.500000	9.427216	88.872391
99.641	2.6889	50.423386	41.500000	8.923386	79.626801
99.737	2.7907	52.830551	42.500000	10.330551	106.720276
99.808	2.8921	55.341766	43.500000	11.841766	140.227417
99.808	2.8921	55.341766	44.500000	10.841766	117.543884
99.844	2.9567	57.000793	45.500000	11.500793	132.268250
99.856	2.9812	57.644592	46.500000	11.144592	124.201935
99.880	3.0365	59.122345	47.500000	11.622345	135.078888
99.904	3.1030	60.949432	48.500000	12.449432	154.988358
99.928	3.1869	63.336349	49.500000	13.836349	191.444565

ROOT MEAN SQUARE OF DIFFERENCE = 5.904

```
                    *  *  *   DISTRIBUTION OF DATA (INTERVAL WIDTH =  0.5000)   *  *  *

NO. OF VALUES LESS THAN  -5.0000=     0     NO. OF ZERO VALUES =     2

   INTERVAL      NUMBER   PERCENT  +---------------------HISTOGRAM--------------------+ CUM. FREQ.  PLOT POS.
 -5.00- -4.50         0    0.0  %  I                                                I   0.0  %      0.0  %
 -4.50- -4.00         0    0.0  %  I                                                I   0.0  %      0.0  %
 -4.00- -3.50         0    0.0  %  I                                                I   0.0  %      0.0  %
 -3.50- -3.00         0    0.0  %  I                                                I   0.0  %      0.0  %
 -3.00- -2.50         0    0.0  %  I                                                I   0.0  %      0.0  %
 -2.50- -2.00         0    0.0  %  I                                                I   0.0  %      0.0  %
 -2.00- -1.50         0    0.0  %  I                                                I   0.0  %      0.0  %
 -1.50- -1.00         0    0.0  %  I                                                I   0.0  %      0.0  %
 -1.00- -0.50         7   14.000%  I******C*******                                  I  14.000%     13.725%
 -0.50-  0.0          6   12.000%  I************C                                     I  26.000%     25.490%
  0.0 -  0.50         6   12.000%  I************      C                              I  38.000%     37.255%
  0.50-  1.00         5   10.000%  I**********              C                       I  48.000%     47.059%
  1.00-  1.50         2    4.000%  I****                       C                    I  52.000%     50.980%
  1.50-  2.00         1    2.000%  I**                          C                 I  54.000%     52.941%
  2.00-  2.50         1    2.000%  I**                           C                I  56.000%     54.902%
  2.50-  3.00         2    4.000%  I****                            C              I  60.000%     58.824%
  3.00-  3.50         0    0.0  %  I                                C              I  60.000%     58.824%
  3.50-  4.00         2    4.000%  I****                              C            I  64.000%     62.745%
  4.00-  4.50         0    0.0  %  I                                 C            I  64.000%     62.745%
  4.50-  5.00         1    2.000%  I**                                C           I  66.000%     64.706%
                                   +-------------------------------------------------+

NO. OF VALUES EQUAL TO OR GREATER THAN    5.0000 =   17  (34.000%)

   N =      50

   MEAN =        3.73007393

   2ND MOMENT (VAR.) =         21.3710937              STANDARD DEV. =         4.62288761

        COEFFICIENT OF VARIATION =     1.23935509

   3RD MOMENT =        71.8648689            COEF. OF SKEWNESS =     0.72740561

   4TH MOMENT =       893.707764             COEF. OF KURTOSIS =     1.95677948

   LOWEST VALUE =       -0.74386692

   HIGHEST VALUE =      13.8363495

        SECOND HIGHEST VALUE =      12.4494324
```

PARAMETERS EST. BY OPTIMIZATION OF M. L. E.-----

NUMBER	L	K1	MEM	SUM1	SUM3	MEAN	SIGMA	CHANGE
0	2	0	0	-27710.3086	-28045.8398	0.268937E+01	0.457844E+00	0.100000E+00
1	2	0	0	-27710.3086	-27822.6719	0.268937E+01	0.457844E+00	0.500000E-01
2	2	0	0	-27710.3086	-27748.6992	0.268937E+01	0.457844E+00	0.250000E-01
3	2	0	0	-27710.3086	-27724.4453	0.268937E+01	0.457844E+00	0.125000E-01
4	2	0	2	-27707.7852	-27710.3086	0.268937E+01	0.445344E+00	-0.125000E-01
5	2	0	0	-27707.7852	-27707.4922	0.268937E+01	0.445344E+00	0.625000E-02
6	2	0	1	-27707.4922	-27710.3086	0.268937E+01	0.451594E+00	0.625000E-02
7	2	0	0	-27707.4922	-27708.5117	0.268937E+01	0.451594E+00	0.312500E-02
8	2	0	2	-27707.2305	-27707.4922	0.268937E+01	0.448469E+00	-0.312500E-02
9	2	0	0	-27707.2305	-27707.2617	0.268937E+01	0.448469E+00	0.156250E-02
10	2	0	0	-27707.2305	-27707.2305	0.268937E+01	0.448469E+00	0.781250E-03
11	2	0	0	-27707.2305	-27707.2461	0.268937E+01	0.448469E+00	0.390625E-03
12	2	0	0	-27707.2305	-27707.2422	0.268937E+01	0.448469E+00	0.195313E-03
13	2	0	0	-27707.2305	-27707.2305	0.268937E+01	0.448469E+00	0.976563E-04
14	2	0	0	-27707.2305	-27707.2305	0.268937E+01	0.448469E+00	0.488281E-04
15	2	0	0	-27707.2305	-27707.2344	0.268937E+01	0.448469E+00	0.244141E-04
16	2	0	0	-27707.2305	-27707.2500	0.268937E+01	0.448469E+00	0.122070E-04
17	2	0	0	-27707.2305	-27707.2344	0.268937E+01	0.448469E+00	0.610352E-05
18	2	0	0	-27707.2305	-27707.2422	0.268937E+01	0.448469E+00	0.305176E-05
19	2	0	0	-27707.2305	-27707.2383	0.268937E+01	0.448469E+00	0.152588E-05
20	2	0	0	-27707.2305	-27707.2344	0.268937E+01	0.448469E+00	0.762940E-06
21	1	0	0	-27707.2305	-27905.3906	0.268937E+01	0.448469E+00	0.100000E+00
22	1	0	0	-27707.2305	-27754.4766	0.268937E+01	0.448469E+00	0.500000E-01
23	1	0	0	-27707.2305	-27717.8906	0.268937E+01	0.448469E+00	0.250000E-01
24	1	0	0	-27707.2305	-27709.3320	0.268937E+01	0.448469E+00	0.125000E-01
25	1	0	0	-27707.2305	-27707.4844	0.268937E+01	0.448469E+00	0.625000E-02
26	1	0	0	-27707.2305	-27707.1484	0.268937E+01	0.448469E+00	0.312500E-02
27	1	0	1	-27707.1484	-27707.4844	0.269250E+01	0.448469E+00	0.312500E-02
28	1	0	0	-27707.1484	-27707.2695	0.269250E+01	0.448469E+00	0.156250E-02
29	1	0	2	-27707.1367	-27707.1484	0.269094E+01	0.448469E+00	-0.156250E-02
30	1	0	0	-27707.1367	-27707.1367	0.269094E+01	0.448469E+00	0.781250E-03
31	1	0	0	-27707.1367	-27707.1328	0.269094E+01	0.448469E+00	0.390625E-03
32	1	0	1	-27707.1328	-27707.1367	0.269133E+01	0.448469E+00	0.390625E-03
33	1	0	0	-27707.1328	-27707.1367	0.269133E+01	0.448469E+00	0.195313E-03
34	1	0	0	-27707.1328	-27707.1406	0.269133E+01	0.448469E+00	0.976563E-04
35	1	0	0	-27707.1328	-27707.1445	0.269133E+01	0.448469E+00	0.488281E-04
36	1	0	0	-27707.1328	-27707.1406	0.269133E+01	0.448469E+00	0.244141E-04
37	1	0	0	-27707.1328	-27707.1367	0.269133E+01	0.448469E+00	0.122070E-04
38	1	0	0	-27707.1328	-27707.1289	0.269133E+01	0.448469E+00	0.610352E-05
39	1	0	1	-27707.1289	-27707.1367	0.269133E+01	0.448469E+00	0.610352E-05
40	1	0	0	-27707.1289	-27707.1367	0.269133E+01	0.448469E+00	0.305176E-05
41	1	0	0	-27707.1289	-27707.1289	0.269133E+01	0.448469E+00	0.152588E-05
42	1	0	0	-27707.1289	-27707.1289	0.269133E+01	0.448469E+00	0.762940E-06
43	2	0	0	-27707.1289	-27997.9492	0.269133E+01	0.448469E+00	0.100000E+00
44	2	0	0	-27707.1289	-27791.2812	0.269133E+01	0.448469E+00	0.500000E-01
45	2	0	0	-27707.1289	-27729.5195	0.269133E+01	0.448469E+00	0.250000E-01
46	2	0	0	-27707.1289	-27712.7187	0.269133E+01	0.448469E+00	0.125000E-01
47	2	0	0	-27707.1289	-27708.4336	0.269133E+01	0.448469E+00	0.625000E-02
48	2	0	0	-27707.1289	-27707.5945	0.269133E+01	0.448469E+00	0.312500E-02
49	2	0	0	-27707.1289	-27707.1680	0.269133E+01	0.448469E+00	0.156250E-02
50	2	0	0	-27707.1289	-27707.1211	0.269133E+01	0.448469E+00	0.781250E-03

51	2	0	1	-27707.1211	-27707.1680	0.269133E+01	0.449250E+00	0.781250E-03
52	2	0	0	-27707.1211	-27707.1523	0.269133E+01	0.449250E+00	0.390625E-03
53	2	0	2	-27707.1133	-27707.1172	0.269133E+01	0.448860E+00	-0.390625E-03
54	2	0	0	-27707.1133	-27707.1328	0.269133E+01	0.448860E+00	0.195313E-03
55	2	0	0	-27707.1133	-27707.1289	0.269133E+01	0.448860E+00	0.976563E-04
56	2	0	0	-27707.1133	-27707.1328	0.269133E+01	0.448860E+00	0.488281E-04
57	2	0	0	-27707.1133	-27707.1289	0.269133E+01	0.448860E+00	0.244141E-04
58	2	0	0	-27707.1133	-27707.1250	0.269133E+01	0.448860E+00	0.122070E-04
59	2	0	0	-27707.1133	-27707.1133	0.269133E+01	0.448860E+00	0.610352E-05
60	2	0	0	-27707.1133	-27707.1211	0.269133E+01	0.448860E+00	0.305176E-05
61	2	0	0	-27707.1133	-27707.1289	0.269133E+01	0.448860E+00	0.152588E-05
62	2	0	0	-27707.1133	-27707.1133	0.269133E+01	0.448860E+00	0.762940E-06
63	1	0	0	-27707.1133	-27913.0352	0.269133E+01	0.448860E+00	0.100000E+00
64	1	0	0	-27707.1133	-27758.3398	0.269133E+01	0.448860E+00	0.500000E-01
65	1	0	0	-27707.1133	-27719.8125	0.269133E+01	0.448860E+00	0.250000E-01
66	1	0	0	-27707.1133	-27710.2383	0.269133E+01	0.448860E+00	0.125000E-01
67	1	0	0	-27707.1133	-27707.8789	0.269133E+01	0.448860E+00	0.625000E-02
68	1	0	0	-27707.1133	-27707.2969	0.269133E+01	0.448860E+00	0.312500E-02
69	1	0	0	-27707.1133	-27707.1680	0.269133E+01	0.448860E+00	0.156250E-02
70	1	0	0	-27707.1133	-27707.1289	0.269133E+01	0.448860E+00	0.781250E-03
71	1	0	0	-27707.1133	-27707.1211	0.269133E+01	0.448860E+00	0.390625E-03
72	1	0	0	-27707.1133	-27707.1328	0.269133E+01	0.448860E+00	0.195313E-03
73	1	0	0	-27707.1133	-27707.1250	0.269133E+01	0.448860E+00	0.976563E-04
74	1	0	0	-27707.1133	-27707.1250	0.269133E+01	0.448860E+00	0.488281E-04
75	1	0	0	-27707.1133	-27707.1133	0.269133E+01	0.448860E+00	0.244141E-04
76	1	0	0	-27707.1133	-27707.1250	0.269133E+01	0.448860E+00	0.122070E-04
77	1	0	0	-27707.1133	-27707.1211	0.269133E+01	0.448860E+00	0.610352E-05
78	1	0	0	-27707.1133	-27707.1328	0.269133E+01	0.448860E+00	0.305176E-05
79	1	0	0	-27707.1133	-27707.1133	0.269133E+01	0.448860E+00	0.152588E-05
80	1	0	0	-27707.1133	-27707.1133	0.269133E+01	0.448860E+00	0.762940E-06
81	2	0	0	-27707.1094	-27999.6984	0.269133E+01	0.448860E+00	0.100000E+00
82	2	0	0	-27707.1094	-27792.5000	0.269133E+01	0.448860E+00	0.500000E-01
83	2	0	0	-27707.1094	-27730.2070	0.269133E+01	0.448860E+00	0.250000E-01
84	2	0	0	-27707.1094	-27713.0859	0.269133E+01	0.448860E+00	0.125000E-01
85	2	0	0	-27707.1094	-27708.6133	0.269133E+01	0.448860E+00	0.625000E-02
86	2	0	0	-27707.1094	-27707.4805	0.269133E+01	0.448860E+00	0.312500E-02
87	2	0	0	-27707.1094	-27707.2109	0.269133E+01	0.448860E+00	0.156250E-02
88	2	0	0	-27707.1094	-27707.1445	0.269133E+01	0.448860E+00	0.781250E-03
89	2	0	0	-27707.1094	-27707.1211	0.269133E+01	0.448860E+00	0.390625E-03
90	2	0	0	-27707.1094	-27707.1289	0.269133E+01	0.448860E+00	0.195313E-03
91	2	0	0	-27707.1094	-27707.1328	0.269133E+01	0.448860E+00	0.976563E-04
92	2	0	0	-27707.1094	-27707.1289	0.269133E+01	0.448860E+00	0.488281E-04
93	2	0	0	-27707.1094	-27707.1133	0.269133E+01	0.448860E+00	0.244141E-04
94	2	0	0	-27707.1094	-27707.1172	0.269133E+01	0.448860E+00	0.122070E-04
95	2	0	0	-27707.1094	-27707.1250	0.269133E+01	0.448860E+00	0.610352E-05
96	2	0	0	-27707.1094	-27707.1133	0.269133E+01	0.448860E+00	0.305176E-05
97	2	0	0	-27707.1094	-27707.1172	0.269133E+01	0.448860E+00	0.152588E-05
98	2	0	0	-27707.1094	-27707.1250	0.269133E+01	0.448860E+00	0.762940E-06
99	1	0	0	-27707.1094	-27913.0391	0.269133E+01	0.448860E+00	0.100000E+00
100	1	0	0	-27707.1094	-27758.3516	0.269133E+01	0.448860E+00	0.500000E-01

101	1	0	0	-27707.1094	-27719.8125	0.269133E+01	0.448860E+00	0.250000E-01
102	1	0	0	-27707.1094	-27710.2305	0.269133E+01	0.448860E+00	0.125000E-01
103	1	0	0	-27707.1094	-27707.8672	0.269133E+01	0.448860E+00	0.625000E-02
104	1	0	0	-27707.1094	-27707.5008	0.269133E+01	0.448860E+00	0.312500E-02
105	1	0	0	-27707.1094	-27707.1680	0.269133E+01	0.448860E+00	0.156250E-02
106	1	0	0	-27707.1094	-27707.1250	0.269133E+01	0.448860E+00	0.781250E-03
107	1	0	0	-27707.1094	-27707.1289	0.269133E+01	0.448860E+00	0.390625E-03
108	1	0	0	-27707.1094	-27707.1250	0.269133E+01	0.448860E+00	0.195313E-03
109	1	0	0	-27707.1094	-27707.1055	0.269133E+01	0.448860E+00	0.976563E-04
110	1	0	1	-27707.1055	-27707.1250	0.269143E+01	0.448860E+00	0.976563E-04
111	1	0	0	-27707.1055	-27707.1153	0.269143E+01	0.448860E+00	0.488281E-04
112	1	0	0	-27707.1055	-27707.1289	0.269143E+01	0.448860E+00	0.244141E-04
113	1	0	0	-27707.1055	-27707.1172	0.269143E+01	0.448860E+00	0.122070E-04
114	1	0	0	-27707.1055	-27707.1172	0.269143E+01	0.448860E+00	0.610352E-05
115	1	0	0	-27707.1055	-27707.1016	0.269143E+01	0.448860E+00	0.305176E-05
116	1	0	1	-27707.1016	-27707.1172	0.269143E+01	0.448860E+00	0.305176E-05
117	1	0	0	-27707.1016	-27707.1153	0.269143E+01	0.448860E+00	0.152588E-05
118	1	0	0	-27707.1016	-27707.1016	0.269143E+01	0.448860E+00	0.762940E-06
119	2	0	0	-27707.1016	-27999.9062	0.269143E+01	0.448860E+00	0.100000E+00
120	2	0	0	-27707.1016	-27792.4922	0.269143E+01	0.448860E+00	0.500000E-01
121	2	0	0	-27707.1016	-27730.2109	0.269143E+01	0.448860E+00	0.250000E-01
122	2	0	0	-27707.1016	-27713.0937	0.269143E+01	0.448860E+00	0.125000E-01
123	2	0	0	-27707.1016	-27708.6055	0.269143E+01	0.448860E+00	0.625000E-02
124	2	0	0	-27707.1016	-27707.4844	0.269143E+01	0.448860E+00	0.312500E-02
125	2	0	0	-27707.1016	-27707.2109	0.269143E+01	0.448860E+00	0.156250E-02
126	2	0	0	-27707.1016	-27707.1406	0.269143E+01	0.448860E+00	0.781250E-03
127	2	0	0	-27707.1016	-27707.1328	0.269143E+01	0.448860E+00	0.390625E-03
128	2	0	0	-27707.1016	-27707.1211	0.269143E+01	0.448860E+00	0.195313E-03
129	2	0	0	-27707.1016	-27707.1289	0.269143E+01	0.448860E+00	0.976563E-04
130	2	0	0	-27707.1016	-27707.1289	0.269143E+01	0.448860E+00	0.488281E-04
131	2	0	0	-27707.1016	-27707.1172	0.269143E+01	0.448860E+00	0.244141E-04
132	2	0	0	-27707.1016	-27707.1153	0.269143E+01	0.448860E+00	0.122070E-04
133	2	0	0	-27707.1016	-27707.1133	0.269143E+01	0.448860E+00	0.610352E-05
134	2	0	0	-27707.1016	-27707.1211	0.269143E+01	0.448860E+00	0.305176E-05
135	2	0	0	-27707.1016	-27707.1211	0.269143E+01	0.448860E+00	0.152588E-05
136	2	0	0	-27707.1016	-27707.1172	0.269143E+01	0.448860E+00	0.762940E-06
137	1	0	0	-27707.1016	-27913.4570	0.269143E+01	0.448860E+00	0.100000E+00
138	1	0	0	-27707.1016	-27758.5547	0.269143E+01	0.448860E+00	0.500000E-01
139	1	0	0	-27707.1016	-27719.9023	0.269143E+01	0.448860E+00	0.250000E-01
140	1	0	0	-27707.1016	-27710.2852	0.269143E+01	0.448860E+00	0.125000E-01
141	1	0	0	-27707.1016	-27707.8984	0.269143E+01	0.448860E+00	0.625000E-02
142	1	0	0	-27707.1016	-27707.2969	0.269143E+01	0.448860E+00	0.312500E-02
143	1	0	0	-27707.1016	-27707.1758	0.269143E+01	0.448860E+00	0.156250E-02
144	1	0	0	-27707.1016	-27707.1445	0.269143E+01	0.448860E+00	0.781250E-03
145	1	0	0	-27707.1016	-27707.1367	0.269143E+01	0.448860E+00	0.390625E-03
146	1	0	0	-27707.1016	-27707.1250	0.269143E+01	0.448860E+00	0.195313E-03
147	1	0	0	-27707.1016	-27707.1250	0.269143E+01	0.448860E+00	0.976563E-04
148	1	0	0	-27707.1016	-27707.1211	0.269143E+01	0.448860E+00	0.488281E-04
149	1	0	0	-27707.1016	-27707.1328	0.269143E+01	0.448860E+00	0.244141E-04
150	1	0	0	-27707.1016	-27707.1250	0.269143E+01	0.448860E+00	0.122070E-04

151	1	0	0	-27707.1016	-27707.1289	0.269143E+01	0.448860E+00	0.610352E-05
152	1	0	0	-27707.1016	-27707.1172	0.269143E+01	0.448860E+00	0.305176E-05
153	1	0	0	-27707.1016	-27707.1133	0.269143E+01	0.448860E+00	0.152588E-05
154	1	0	0	-27707.1016	-27707.1016	0.269143E+01	0.448860E+00	0.762940E-06
155	1	0	0	-27707.1250	-27707.1445	0.269227E+01	0.449139E+00	0.390625E-03
156	1	0	0	-27707.1250	-27707.1406	0.269227E+01	0.449139E+00	0.195313E-03
157	1	0	0	-27707.1250	-27707.1328	0.269227E+01	0.449139E+00	0.976563E-04
158	1	0	0	-27707.1250	-27707.1328	0.269227E+01	0.449139E+00	0.488281E-04
159	1	0	0	-27707.1250	-27707.1367	0.269227E+01	0.449139E+00	0.244141E-04
160	1	0	0	-27707.1250	-27707.1328	0.269227E+01	0.449139E+00	0.122070E-04
161	1	0	0	-27707.1250	-27707.1367	0.269227E+01	0.449139E+00	0.610352E-05
162	1	0	0	-27707.1250	-27707.1211	0.269227E+01	0.449139E+00	0.305176E-05
163	1	0	1	-27707.1211	-27707.1367	0.269228E+01	0.449139E+00	0.305176E-05
164	1	0	0	-27707.1211	-27707.1406	0.269228E+01	0.449139E+00	0.152588E-05
165	1	0	0	-27707.1211	-27707.1211	0.269228E+01	0.449139E+00	0.762940E-06
166	2	0	0	-27707.1211	-28001.3086	0.269228E+01	0.449139E+00	0.100000E+00
167	2	0	0	-27707.1211	-27793.4062	0.269228E+01	0.449139E+00	0.500000E-01
168	2	0	0	-27707.1211	-27730.7148	0.269228E+01	0.449139E+00	0.250000E-01
169	2	0	0	-27707.1211	-27713.3516	0.269228E+01	0.449139E+00	0.125000E-01
170	2	0	0	-27707.1211	-27708.7617	0.269228E+01	0.449139E+00	0.625000E-02
171	2	0	0	-27707.1211	-27707.5586	0.269228E+01	0.449139E+00	0.312500E-02
172	2	0	0	-27707.1211	-27707.2500	0.269228E+01	0.449139E+00	0.156250E-02
173	2	0	0	-27707.1211	-27707.1758	0.269228E+01	0.449139E+00	0.781250E-03
174	2	0	0	-27707.1211	-27707.1562	0.269228E+01	0.449139E+00	0.390625E-03
175	2	0	0	-27707.1211	-27707.1523	0.269228E+01	0.449139E+00	0.195313E-03
176	2	0	0	-27707.1211	-27707.1367	0.269228E+01	0.449139E+00	0.976563E-04
177	2	0	0	-27707.1211	-27707.1367	0.269228E+01	0.449139E+00	0.488281E-04
178	2	0	0	-27707.1211	-27707.1211	0.269228E+01	0.449139E+00	0.244141E-04
179	2	0	0	-27707.1211	-27707.1328	0.269228E+01	0.449139E+00	0.122070E-04
180	2	0	0	-27707.1211	-27707.1328	0.269228E+01	0.449139E+00	0.610352E-05
181	2	0	0	-27707.1211	-27707.1367	0.269228E+01	0.449139E+00	0.305176E-05
182	2	0	0	-27707.1211	-27707.1406	0.269228E+01	0.449139E+00	0.152588E-05
183	2	0	0	-27707.1211	-27707.1367	0.269228E+01	0.449139E+00	0.762940E-06
184	1	0	0	-27707.1211	-27916.7070	0.269228E+01	0.449139E+00	0.100000E+00
185	1	0	0	-27707.1211	-27760.2383	0.269228E+01	0.449139E+00	0.500000E-01
186	1	0	0	-27707.1211	-27720.7812	0.269228E+01	0.449139E+00	0.250000E-01
187	1	0	0	-27707.1211	-27710.7266	0.269228E+01	0.449139E+00	0.125000E-01
188	1	0	0	-27707.1211	-27708.1211	0.269228E+01	0.449139E+00	0.625000E-02
189	1	0	0	-27707.1211	-27707.4141	0.269228E+01	0.449139E+00	0.312500E-02
190	1	0	0	-27707.1211	-27707.2383	0.269228E+01	0.449139E+00	0.156250E-02
191	1	0	0	-27707.1211	-27707.1680	0.269228E+01	0.449139E+00	0.781250E-03
192	1	0	0	-27707.1211	-27707.1367	0.269228E+01	0.449139E+00	0.390625E-03
193	1	0	0	-27707.1211	-27707.1484	0.269228E+01	0.449139E+00	0.195313E-03
194	1	0	0	-27707.1211	-27707.1289	0.269228E+01	0.449139E+00	0.976563E-04
195	1	0	0	-27707.1211	-27707.1367	0.269228E+01	0.449139E+00	0.488281E-04
196	1	0	0	-27707.1211	-27707.1328	0.269228E+01	0.449139E+00	0.244141E-04
197	1	0	0	-27707.1211	-27707.1250	0.269228E+01	0.449139E+00	0.122070E-04
198	1	0	0	-27707.1211	-27707.1289	0.269228E+01	0.449139E+00	0.610352E-05
199	1	0	2	-27707.1172	-27707.1289	0.269227E+01	0.449139E+00	-0.610352E-05
200	1	0	0	-27707.1172	-27707.1289	0.269227E+01	0.449139E+00	0.305176E-05

COMPARISON OF LOGNORMAL PROBABILITY MODEL (16.3163, 7.7086) WITH OBSERVED DATA

ARITH. MEAN = 16.3163300 ARITH. STD. DEV. = 7.70859432

MEDIAN = 14.7527590 MODE = 12.0607328 GEO. STD. DEV. = 1.56652546

MEAN OF NORMAL DIST. = 2.69143009 STD. DEV. OF NORMAL DIST. = 0.44885981

COEF. OF SKEWNESS = 1.52279091 COEF. OF KURTOSIS = 7.38779926

(CUM) INTERVAL	LN(X)	Y	F(Y)	P(Y)	PREDICTED N	OBSERVED N	DIFFERENCE	DIFF. SQ.	CHI SQUARE
0.5000	-0.6931	-7.5404	0.0000000	0.0000000	0	0	0	0	0.0
1.5000	0.4055	-5.0928	0.0000002	0.0000002	0	0	0	0	0.0
2.5000	0.9163	-3.9548	0.0000383	0.0000381	0	8	-8	64	0.0
3.5000	1.2528	-3.2052	0.0006750	0.0006367	5	14	-9	81	16.20
4.5000	1.5041	-2.6453	0.0040814	0.0034064	28	63	-35	1225	43.75
5.5000	1.7047	-2.1982	0.0139676	0.0098862	83	139	-56	3136	37.78
6.5000	1.8718	-1.8260	0.0339233	0.0199558	167	174	-7	49	0.29
7.5000	2.0149	-1.5072	0.0658783	0.0319550	267	265	2	4	0.01
8.5000	2.1401	-1.2284	0.1096551	0.0437768	366	359	7	49	0.13
9.5000	2.2513	-0.9806	0.1634025	0.0537474	450	375	75	5625	12.50
10.5000	2.3514	-0.7576	0.2243455	0.0609430	510	398	112	12544	24.60
11.5000	2.4423	-0.5549	0.2894729	0.0651274	545	391	154	23716	43.52
12.5000	2.5257	-0.3692	0.3560036	0.0665308	556	457	99	9801	17.63
13.5000	2.6027	-0.1977	0.4216397	0.0656360	549	547	2	4	0.01
14.5000	2.6741	-0.0385	0.4846447	0.0630050	527	498	29	841	1.60
15.5000	2.7408	0.1101	0.5438263	0.0591816	495	519	-24	576	1.16
16.5000	2.8034	0.2494	0.5984609	0.0546346	457	499	-42	1764	3.86
17.5000	2.8622	0.3805	0.6481957	0.0497348	416	496	-80	6400	15.38
18.5000	2.9178	0.5043	0.6929590	0.0447633	374	473	-99	9801	26.21
19.5000	2.9704	0.6215	0.7328777	0.0399187	334	379	-45	2025	6.06
20.5000	3.0204	0.7330	0.7682076	0.0353299	295	336	-41	1681	5.70
21.5000	3.0681	0.8391	0.7992833	0.0310757	260	252	8	64	0.25
22.5000	3.1135	0.9403	0.8264808	0.0271975	227	284	-57	3249	14.31
23.5000	3.1570	1.0372	0.8501851	0.0237043	198	276	-78	6084	30.73
24.5000	3.1987	1.1301	0.8707765	0.0205914	172	182	-10	100	0.58

25.5000	3.2387	1.2192	0.8886149	0.0176384	149	192	-43	1849	12.41
26.5000	3.2771	1.3049	0.9040353	0.0154204	129	141	-12	144	1.12
27.5000	3.3142	1.3874	0.9173426	0.0133073	111	108	3	9	0.08
28.5000	3.3499	1.4670	0.9288108	0.0114682	96	112	-16	256	2.67
29.5000	3.3844	1.5438	0.9386842	0.0098734	83	72	11	121	1.46
30.5000	3.4177	1.6181	0.9471785	0.0084943	71	74	-3	9	0.13
31.5000	3.4500	1.6900	0.9544826	0.0073041	61	52	9	81	1.33
32.5000	3.4812	1.7596	0.9607615	0.0062789	53	52	1	1	0.02
33.5000	3.5115	1.8271	0.9661582	0.0053967	45	53	-8	64	1.42
34.5000	3.5410	1.8926	0.9707971	0.0046389	39	7	32	1024	26.26
35.5000	3.5695	1.9563	0.9747849	0.0039878	33	28	5	25	0.76
36.5000	3.5973	2.0182	0.9782140	0.0034291	29	17	12	144	4.97
37.5000	3.6243	2.0784	0.9811638	0.0029498	25	15	10	100	4.00
38.5000	3.6507	2.1370	0.9837023	0.0025385	21	11	10	100	4.76
39.5000	3.6763	2.1942	0.9858881	0.0021858	18	6	12	144	8.00
40.5000	3.7013	2.2499	0.9877711	0.0018830	16	8	8	64	4.00
41.5000	3.7257	2.3042	0.9893944	0.0016233	14	2	12	144	10.29
42.5000	3.7495	2.3572	0.9907945	0.0014002	12	8	4	16	1.33
43.5000	3.7728	2.4091	0.9920032	0.0012087	10	6	4	16	1.60
44.5000	3.7955	2.4597	0.9930473	0.0010441	9	0	9	81	9.00
45.5000	3.8177	2.5092	0.9939498	0.0009025	8	3	5	25	3.13
46.5000	3.8395	2.5576	0.9947307	0.0007809	7	1	6	36	5.14
47.5000	3.8607	2.6050	0.9954069	0.0006762	6	2	4	16	2.67
48.5000	3.8816	2.6515	0.9959927	0.0005859	5	2	3	9	1.80
				SUBTOTAL:	8331	8356	-25	93361	410.58
RESULTS FOR REMAINING INTERVALS:				0.0039455	33	8	25	625	18.94

SUM OF SQUARES = 93986.00 ROOT MEAN SQUARE OF DIFFERENCE = 43.211 CHI SQUARE = 429.52

DEGREES OF FREDOM = 47

KOLMOGOROV SMIRNOV RELATED MEASURES:

MAXIMUM ABSOLUTE DIFFERENCE BETWEEN CDF AND CUMULATIVE FREQUENCY OF OBSERVATIONS = 0.0437074 (AT INTERVAL NO. 15)

SUM OF ABSOLUTE DIFFERENCES BETWEEN CDF AND CUMULATIVE FREQUENCY OF OBSERVATIONS = 0.6775879

LOG-LIKELIHOOD FUNCTION = -27707.1016

COMPARISON OF X FOR OBSERVED FREQUENCIES WITH X FOR LOGNORMAL PROBABILITY MODEL (16.3163, 7.7086)

CUM. FREQ.	PROBITS	PREDICTED X	OBSERVED X	DIFFERENCE	DIFF. SQUARED
0.0	0.0	0.0	0.500000	0.0	0.0
0.0	0.0	0.0	1.500000	0.0	0.0
0.096	-3.1037	3.663063	2.500000	1.163063	1.352715
0.263	-2.7910	4.215078	3.500000	0.715078	0.511337
1.016	-2.3207	5.205677	4.500000	0.705677	0.497980
2.678	-1.9308	6.201427	5.500000	0.701427	0.492000
4.758	-1.6691	6.974341	6.500000	0.474341	0.225000
7.927	-1.4103	7.833577	7.500000	0.333577	0.111274
12.219	-1.1642	8.748359	8.500000	0.248359	0.061682
16.703	-0.9659	9.562633	9.500000	0.062633	0.003923
21.461	-0.7903	10.346901	10.500000	-0.153099	0.023439
26.136	-0.6388	11.074899	11.500000	-0.425101	0.180711
31.600	-0.4785	11.901320	12.500000	-0.598680	0.358418
38.140	-0.3014	12.886128	13.500000	-0.613872	0.376838
44.094	-0.1483	13.802810	14.500000	-0.697190	0.486074
50.299	0.0075	14.802296	15.500000	-0.697704	0.486791
56.265	0.1574	15.832472	16.500000	-0.667528	0.445594
62.195	0.3102	16.956512	17.500000	-0.543488	0.295379
67.850	0.4631	18.161285	18.500000	-0.338715	0.114728
72.382	0.5939	19.259186	19.500000	-0.240814	0.057991
76.399	0.7189	20.371262	20.500000	-0.128738	0.016574
79.412	0.8206	21.322586	21.500000	-0.177414	0.031476
82.807	0.9465	22.562149	22.500000	0.062149	0.003863
86.107	1.0852	24.011246	23.500000	0.511246	0.261372
88.283	1.1894	25.160828	24.500000	0.660828	0.436693
90.579	1.3154	26.625717	25.500000	1.125717	1.267239
92.264	1.4233	27.946976	26.500000	1.446976	2.093739
93.556	1.5188	29.170593	27.500000	1.670593	2.790881
94.895	1.6351	30.733398	28.500000	2.233398	4.988069
95.755	1.7234	31.975861	29.500000	2.475861	6.129886
96.640	1.8308	33.555115	30.500000	3.055115	9.333726
97.262	1.9212	34.944916	31.500000	3.444916	11.867444
97.884	2.0307	36.705978	32.500000	4.205978	17.690247
98.517	2.1751	39.163849	33.500000	5.663849	32.079178
98.601	2.1980	39.568100	34.500000	5.068100	25.685623
98.936	2.3033	41.483597	35.500000	5.983597	35.803421
99.139	2.3824	42.982483	36.500000	6.482483	42.022583
99.318	2.4672	44.649979	37.500000	7.149979	51.122192
99.450	2.5430	46.195389	38.500000	7.695389	59.218994
99.522	2.5914	47.210770	39.500000	7.710770	59.455963
99.617	2.6673	48.845322	40.500000	8.345322	69.644379
99.641	2.6889	49.321136	41.500000	7.821136	61.170166
99.737	2.7907	51.628387	42.500000	9.128387	83.327454
99.808	2.8921	54.033234	43.500000	10.533234	110.949005
99.808	2.8921	54.033234	44.500000	9.533234	90.882538
99.844	2.9567	55.620758	45.500000	10.120758	102.429733
99.856	2.9812	56.236603	46.500000	9.736603	94.801422
99.880	3.0365	57.649658	47.500000	10.149658	103.015549
99.904	3.1030	59.395706	48.500000	10.895706	118.716400
99.928	3.1869	61.675293	49.500000	12.175293	148.237747

ROOT MEAN SQUARE OF DIFFERENCE = 5.199

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 0.5000) * * *

NO. OF VALUES LESS THAN -5.0000= 0 NO. OF ZERO VALUES = 2

```
  INTERVAL      NUMBER   PERCENT  +---------------------HISTOGRAM---------------------+ CUM. FREQ.  PLOT POS.
-5.00- -4.50        0     0.0  %  I                                                   I   0.0  %      0.0  %
-4.50- -4.00        0     0.0  %  I                                                   I   0.0  %      0.0  %
-4.00- -3.50        0     0.0  %  I                                                   I   0.0  %      0.0  %
-3.50- -3.00        0     0.0  %  I                                                   I   0.0  %      0.0  %
-3.00- -2.50        0     0.0  %  I                                                   I   0.0  %      0.0  %
-2.50- -2.00        0     0.0  %  I                                                   I   0.0  %      0.0  %
-2.00- -1.50        0     0.0  %  I                                                   I   0.0  %      0.0  %
-1.50- -1.00        0     0.0  %  I                                                   I   0.0  %      0.0  %
-1.00- -0.50        6    12.000%  I*****C******                                        I  12.000%    11.765%
-0.50-  0.0         6    12.000%  I***********C                                        I  24.000%    23.529%
 0.0 -  0.50        7    14.000%  I**************    C                                  I  38.000%    37.255%
 0.50-  1.00        5    10.000%  I**********            C                              I  48.000%    47.059%
 1.00-  1.50        3     6.000%  I******                   C                           I  54.000%    52.941%
 1.50-  2.00        1     2.000%  I**                        C                          I  56.000%    54.902%
 2.00-  2.50        2     4.000%  I****                        C                        I  60.000%    58.824%
 2.50-  3.00        0     0.0  %  I                            C                       I  60.000%    58.824%
 3.00-  3.50        2     4.000%  I****                          C                     I  64.000%    62.745%
 3.50-  4.00        0     0.0  %  I                              C                     I  64.000%    62.745%
 4.00-  4.50        1     2.000%  I**                             C                    I  66.000%    64.706%
 4.50-  5.00        0     0.0  %  I                              C                    I  66.000%    64.706%
                                  +-----------------------------------------------------+
```

NO. OF VALUES EQUAL TO OR GREATER THAN 5.0000 = 17 (34.000%)

N = 50

MEAN = 3.28416157

2ND MOMENT (VAR.) = 16.5768585 STANDARD DEV. = 4.07146931

COEFFICIENT OF VARIATION = 1.23972747

3RD MOMENT = 48.8780975 COEF. OF SKEWNESS = 0.72420472

4TH MOMENT = 535.812500 COEF. OF KURTOSIS = 1.94988155

LOWEST VALUE = -0.69770432

HIGHEST VALUE = 12.1752930

SECOND HIGHEST VALUE = 10.8957062

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	8344.	40.2979	0.85	I ********** I
2	0.08	8329.	30.4541	0.65	I ********** I
3	0.13	8316.	20.5601	0.44	I ********* I
4	0.17	8303.	12.5139	0.27	I ****** I
5	0.21	8292.	7.0474	0.15	I *** I
6	0.25	8282.	4.2896	0.09	I ** I
7	0.29	8273.	3.3521	0.07	I ** I
8	0.33	8264.	2.9170	0.06	I ** I
9	0.38	8255.	1.4814	0.03	I * I
10	0.42	8247.	-1.3271	-.03	I ** I
11	0.46	8239.	-4.2395	-.09	I *** I
12	0.50	8231.	-5.6270	-.12	I **** I
13	0.54	8223.	-5.1177	-.11	I **** I
14	0.58	8215.	-2.7568	-.06	I *** I
15	0.63	8208.	-0.0955	-.00	I ** I
16	0.67	8201.	1.5415	0.03	I * I
17	0.71	8194.	2.2419	0.05	I * I
18	0.75	8187.	2.8704	0.06	I ** I
19	0.79	8181.	4.7144	0.10	I ** I
20	0.83	8175.	8.6309	0.18	I **** I
21	0.88	8169.	14.5095	0.31	I ******* I
22	0.92	8163.	21.6313	0.46	I ********** I
23	0.96	8157.	28.1580	0.60	I ********** I
24	1.00	8151.	31.4460	0.67	I ********** I
25	1.04	8142.	29.1252	0.62	I ********** I
26	1.08	8135.	23.1028	0.49	I ********** I
27	1.13	8128.	15.8843	0.34	I ******* I
28	1.17	8121.	9.4268	0.20	I **** I
29	1.21	8115.	4.7346	0.10	I *** I
30	1.25	8109.	2.1035	0.04	I * I
31	1.29	8103.	1.0566	0.02	I * I
32	1.33	8098.	0.5125	0.01	I * I
33	1.38	8093.	-0.8267	-.02	I ** I
34	1.42	8088.	-3.1582	-.07	I *** I
35	1.46	8084.	-5.6675	-.12	I **** I
36	1.50	8080.	-7.0406	-.15	I **** I
37	1.54	8076.	-6.9749	-.15	I **** I
38	1.58	8072.	-5.2219	-.11	I **** I
39	1.63	8068.	-2.9182	-.06	I *** I
40	1.67	8064.	-1.1904	-.03	I ** I

41	1.71	8060.	-0.0168	-.00	I ** I
42	1.75	8056.	1.2385	0.03	I * I
43	1.79	8052.	3.2424	0.07	I ** I
44	1.83	8048.	6.7607	0.14	I *** I
45	1.88	8044.	11.7698	0.25	I ***** I
46	1.92	8040.	17.5083	0.37	I ******** I
47	1.96	8037.	22.5261	0.48	I ********** I
48	2.00	8034.	24.7637	0.52	I *********** I
49	2.04	8032.	22.7544	0.48	I ********** I
50	2.08	8030.	17.8567	0.38	I ******** I
51	2.13	8028.	12.0422	0.26	I ****** I
52	2.17	8026.	6.8835	0.15	I *** I
53	2.21	8023.	3.0854	0.07	I ** I
54	2.25	8021.	0.8657	0.02	I * I
55	2.29	8019.	-0.2698	-.01	I ** I
56	2.33	8018.	-1.1499	-.02	I ** I
57	2.38	8017.	-2.7568	-.06	I *** I
58	2.42	8016.	-4.9517	-.10	I **** I
59	2.46	8015.	-6.8293	-.14	I **** I
60	2.50	8014.	-7.6889	-.16	I ***** I
61	2.54	8013.	-7.4138	-.16	I ***** I
62	2.58	8012.	-5.8638	-.12	I **** I
63	2.63	8011.	-3.7773	-.08	I *** I
64	2.67	8010.	-2.2947	-.05	I ** I
65	2.71	8009.	-1.3433	-.03	I ** I
66	2.75	8009.	-0.1687	-.00	I ** I
67	2.79	8008.	1.9507	0.04	I * I
68	2.83	8007.	5.4138	0.11	I *** I
69	2.88	8006.	9.9563	0.21	I ***** I
70	2.92	8004.	15.2744	0.32	I ******* I
71	2.96	8001.	19.9636	0.42	I ********* I
72	3.00	7999.	22.1375	0.47	I ********** I
73	3.04	7999.	20.2937	0.43	I ********* I
74	3.08	7997.	15.9343	0.34	I ******* I
75	3.13	7995.	10.4851	0.22	I ***** I

+---+---+---+---+---+---+---+---+---+---+

FOR THIS AUTOCORRELATION: MEAN = 16.2074 AND VARIANCE = 47.2019

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -8 -6 -4 -2 0 +2 +4 +6 +8 +1.0
6	0.25	8282.	4.2896	0.09	I ** I
12	0.50	8231.	-5.6270	-.12	I **** I
18	0.75	8187.	2.8704	0.06	I ** I
24	1.00	8151.	31.4460	0.67	I ********** I
30	1.25	8109.	2.1035	0.04	I * I
36	1.50	8080.	-7.0406	-.15	I **** I
42	1.75	8056.	1.2385	0.03	I * I
48	2.00	8034.	24.7637	0.52	I ********** I
54	2.25	8021.	0.8657	0.02	I * I
60	2.50	8014.	-7.6889	-.16	I ***** I
66	2.75	8009.	-0.1687	-.00	I ** I
72	3.00	7999.	22.1375	0.47	I ********** I
78	3.25	7988.	-0.5220	-.01	I ** I
84	3.50	7988.	-9.1828	-.19	I ***** I
90	3.75	7988.	-0.9343	-.02	I ** I
96	4.00	7986.	21.1116	0.45	I ********* I
102	4.25	7984.	-1.0769	-.02	I ** I
108	4.50	7984.	-9.8192	-.21	I ****** I
114	4.75	7984.	-1.3311	-.03	I ** I
120	5.00	7983.	22.7319	0.48	I ********** I
126	5.25	7981.	-1.3169	-.03	I ** I
132	5.50	7975.	-9.9982	-.21	I ****** I
138	5.75	7964.	-1.8933	-.04	I ** I
144	6.00	7952.	26.1492	0.55	I ************ I
150	6.25	7944.	-1.1599	-.02	I ** I
156	6.50	7932.	-10.1460	-.21	I ****** I
162	6.75	7920.	-1.6121	-.03	I ** I
168	7.00	7907.	31.0938	0.66	I ********** I
174	7.25	7899.	-1.2932	-.03	I ** I
180	7.50	7899.	-10.5775	-.22	I ****** I
186	7.75	7899.	-2.2854	-.05	I ** I
192	8.00	7897.	26.1238	0.55	I ************ I
198	8.25	7891.	-1.8696	-.04	I ** I
204	8.50	7880.	-10.5211	-.22	I ****** I
210	8.75	7869.	-2.9265	-.06	I *** I
216	9.00	7857.	21.5134	0.46	I ********** I
222	9.25	7851.	-1.7456	-.04	I ** I
228	9.50	7849.	-9.8804	-.21	I ****** I
234	9.75	7843.	-1.5466	-.03	I ** I
240	10.00	7837.	20.9663	0.44	I ********* I

246	10.25	7840.	-1.6902	-.04	I ** I
252	10.50	7842.	-8.5768	-.18	I ***** I
258	10.75	7846.	-0.9131	-.02	I ** I
264	11.00	7844.	21.0776	0.45	I ********* I
270	11.25	7847.	-1.3406	-.03	I ** I
276	11.50	7847.	-9.2144	-.20	I ***** I
282	11.75	7847.	-1.7612	-.04	I ** I
288	12.00	7848.	21.7439	0.46	I ********** I
294	12.25	7847.	-1.9290	-.04	I ** I
300	12.50	7842.	-10.3130	-.22	I ****** I
306	12.75	7837.	-3.0771	-.07	I *** I
312	13.00	7830.	24.5962	0.52	I *********** I
318	13.25	7821.	-2.6370	-.06	I *** I
324	13.50	7805.	-10.9281	-.23	I ****** I
330	13.75	7797.	-3.2648	-.07	I *** I
336	14.00	7791.	28.6819	0.61	I ************** I
342	14.25	7786.	-2.0867	-.04	I ** I
348	14.50	7770.	-9.9015	-.21	I ****** I
354	14.75	7765.	-1.5012	-.03	I ** I
360	15.00	7763.	25.8796	0.55	I *********** I
366	15.25	7757.	-0.8015	-.02	I ** I
372	15.50	7753.	-8.5327	-.18	I ***** I
378	15.75	7752.	-0.7927	-.02	I ** I
384	16.00	7752.	22.7117	0.48	I ********** I
390	16.25	7753.	-0.2830	-.01	I ** I
396	16.50	7758.	-8.0335	-.17	I ***** I
402	16.75	7762.	-1.3513	-.03	I ** I
408	17.00	7766.	20.3301	0.43	I ********* I
414	17.25	7760.	-1.4336	-.03	I ** I
420	17.50	7749.	-8.9643	-.19	I ***** I
426	17.75	7746.	-2.0137	-.04	I ** I
432	18.00	7735.	19.3838	0.41	I ********* I
438	18.25	7723.	-1.3596	-.03	I ** I
444	18.50	7707.	-8.7542	-.19	I ***** I
450	18.75	7699.	-1.4587	-.03	I ** I

+---+---+---+---+---+---+---+---+---+---+

FOR THIS AUTOCORRELATION: MEAN = 16.2074 AND VARIANCE = 47.2019

APPENDIX C

BASIC STATISTICS, HISTOGRAMS, AND AUTOCORRELATIONS FOR THE GEOGRAPHICAL DATA GROUP

NEW YORK CITY, NY

YR=74 STATE=33 AREA=4680 SITE=023 AGCY=H PROJ=01 POLL=0001200 METH=11 UNT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 8364

BASIC STATISTICS-----

N = 8364

MEAN = 16.2074280

2ND MOMENT (VAR.) = 47.1796722 STANDARD DEV. = 6.86874580

COEFFICIENT OF VARIATION = 0.42380232

3RD MOMENT = 260.864258 COEF. OF SKEWNESS = 0.80497521

4TH MOMENT = 9293.21875 COEF. OF KURTOSIS = 4.17499828

LOWEST VALUE = 2.00000000

HIGHEST VALUE = 57.0000000

SECOND HIGHEST VALUE = 56.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 0

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	0	0.0 %	I I	0.0 %	0.0 %
0.50- 1.50	0	0.0 %	I I	0.0 %	0.0 %
1.50- 2.50	8	0.096%	I I	0.096%	0.096%
2.50- 3.50	14	0.167%	I I	0.263%	0.263%
3.50- 4.50	63	0.753%	IC I	1.016%	1.016%
4.50- 5.50	139	1.662%	IC* I	2.678%	2.678%
5.50- 6.50	174	2.080%	I*C I	4.758%	4.758%
6.50- 7.50	265	3.168%	I***C I	7.927%	7.926%
7.50- 8.50	359	4.292%	I**** C I	12.219%	12.218%
8.50- 9.50	375	4.484%	I**** C I	16.703%	16.701%
9.50- 10.50	398	4.758%	I***** C I	21.461%	21.458%

10.50- 11.50	391	4.675%	26.136%	26.133%
11.50- 12.50	457	5.464%	31.600%	31.596%
12.50- 13.50	547	6.540%	38.140%	38.135%
13.50- 14.50	498	5.954%	44.094%	44.088%
14.50- 15.50	519	6.205%	50.299%	50.293%
15.50- 16.50	499	5.966%	56.265%	56.258%
16.50- 17.50	496	5.930%	62.195%	62.188%
17.50- 18.50	473	5.655%	67.850%	67.842%
18.50- 19.50	379	4.531%	72.382%	72.373%
19.50- 20.50	336	4.017%	76.399%	76.390%
20.50- 21.50	252	3.013%	79.412%	79.402%
21.50- 22.50	284	3.396%	82.807%	82.797%
22.50- 23.50	276	3.300%	86.107%	86.097%
23.50- 24.50	182	2.176%	88.283%	88.272%
24.50- 25.50	192	2.296%	90.579%	90.568%
25.50- 26.50	141	1.686%	92.264%	92.253%
26.50- 27.50	108	1.291%	93.556%	93.544%
27.50- 28.50	112	1.339%	94.895%	94.883%
28.50- 29.50	72	0.861%	95.755%	95.744%
29.50- 30.50	74	0.885%	96.640%	96.629%
30.50- 31.50	52	0.622%	97.262%	97.250%
31.50- 32.50	52	0.622%	97.884%	97.872%
32.50- 33.50	53	0.634%	98.517%	98.506%
33.50- 34.50	7	0.084%	98.601%	98.589%
34.50- 35.50	28	0.335%	98.936%	98.924%
35.50- 36.50	17	0.203%	99.139%	99.127%
36.50- 37.50	15	0.179%	99.318%	99.306%
37.50- 38.50	11	0.132%	99.450%	99.438%
38.50- 39.50	6	0.072%	99.522%	99.510%
39.50- 40.50	8	0.096%	99.617%	99.605%
40.50- 41.50	2	0.024%	99.641%	99.629%
41.50- 42.50	8	0.096%	99.737%	99.725%
42.50- 43.50	6	0.072%	99.808%	99.797%
43.50- 44.50	0	0.0 %	99.808%	99.797%
44.50- 45.50	3	0.036%	99.844%	99.832%
45.50- 46.50	1	0.012%	99.856%	99.844%
46.50- 47.50	2	0.024%	99.880%	99.868%
47.50- 48.50	2	0.024%	99.904%	99.892%
48.50- 49.50	2	0.024%	99.928%	99.916%
49.50- 50.50	3	0.036%	99.964%	99.952%
50.50- 51.50	0	0.0 %	99.964%	99.952%
51.50- 52.50	1	0.012%	99.976%	99.964%
52.50- 53.50	0	0.0 %	99.976%	99.964%
53.50- 54.50	0	0.0 %	99.976%	99.964%
54.50- 55.50	0	0.0 %	99.976%	99.964%
55.50- 56.50	1	0.012%	99.988%	99.976%
56.50- 57.50	1	0.012%	100.000%	99.988%
57.50- 58.50	0	0.0 %	100.000%	99.988%
58.50- 59.50	0	0.0 %	100.000%	99.988%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	8344	40.2979	0.85	**********
2	0.08	8329	30.4541	0.65	**********
3	0.13	8316	20.5601	0.44	*********
4	0.17	8303	12.5139	0.27	******
5	0.21	8292	7.0474	0.15	***
6	0.25	8282	4.2896	0.09	**
7	0.29	8273	3.3521	0.07	**
8	0.33	8264	2.9170	0.06	**
9	0.38	8255	1.4814	0.03	*
10	0.42	8247	-1.3271	-.03	**
11	0.46	8239	-4.2395	-.09	***
12	0.50	8231	-5.6270	-.12	****
13	0.54	8223	-5.1177	-.11	****
14	0.58	8215	-2.7568	-.06	***
15	0.63	8208	-0.0955	-.00	**
16	0.67	8201	1.5415	0.03	*
17	0.71	8194	2.2419	0.05	*
18	0.75	8187	2.8704	0.06	**
19	0.79	8181	4.7144	0.10	**
20	0.83	8175	8.6369	0.18	****
21	0.88	8169	14.5095	0.31	*******
22	0.92	8163	21.6313	0.46	**********
23	0.96	8157	28.1580	0.60	**********
24	1.00	8151	31.4460	0.67	**********
25	1.04	8142	29.1252	0.62	**********
26	1.08	8135	23.1028	0.49	**********
27	1.13	8128	15.8843	0.34	*******
28	1.17	8121	9.4268	0.20	****
29	1.21	8115	4.7346	0.10	***
30	1.25	8109	2.1035	0.04	*
31	1.29	8103	1.0566	0.02	*
32	1.33	8098	0.5125	0.01	*
33	1.38	8093	-0.8267	-.02	**
34	1.42	8088	-3.1582	-.07	***
35	1.46	8084	-5.6675	-.12	****
36	1.50	8080	-7.0406	-.15	****
37	1.54	8076	-6.9749	-.15	****
38	1.58	8072	-5.2219	-.11	****
39	1.63	8068	-2.9182	-.06	***
40	1.67	8064	-1.1904	-.03	**
41	1.71	8060	-0.0168	-.00	**
42	1.75	8056	1.2385	0.03	*
43	1.79	8052	3.2424	0.07	**
44	1.83	8048	6.7607	0.14	***
45	1.88	8044	11.7698	0.25	*****
46	1.92	8040	17.5083	0.37	********
47	1.96	8037	22.5261	0.48	**********
48	2.00	8034	24.7637	0.52	**********
49	2.04	8032	22.7554	0.48	**********
50	2.08	8030	17.8567	0.38	********

51	2.13	8028.	12.0422	0.26	******
52	2.17	8026.	6.8835	0.15	***
53	2.21	8023.	3.0854	0.07	**
54	2.25	8021.	0.8657	0.02	*
55	2.29	8019.	-0.2698	-.01	**
56	2.33	8018.	-1.1499	-.02	**
57	2.38	8017.	-2.7568	-.06	***
58	2.42	8016.	-4.9517	-.10	****
59	2.46	8015.	-6.8293	-.14	****
60	2.50	8014.	-7.6889	-.16	*****
61	2.54	8013.	-7.4138	-.16	*****
62	2.58	8012.	-5.8638	-.12	****
63	2.63	8011.	-3.7773	-.08	***
64	2.67	8010.	-2.2947	-.05	**
65	2.71	8009.	-1.3433	-.03	**
66	2.75	8009.	-0.1687	-.00	**
67	2.79	8008.	1.9507	0.04	*
68	2.83	8007.	5.4138	0.11	***
69	2.88	8006.	9.9563	0.21	*****
70	2.92	8004.	15.2744	0.32	*******
71	2.96	8001.	19.9636	0.42	*********
72	3.00	7999.	22.1375	0.47	**********
73	3.04	7999.	20.2937	0.43	*********
74	3.08	7997.	15.9343	0.34	*******
75	3.13	7995.	10.4851	0.22	*****

FOR THIS AUTOCORRELATION: MEAN = 16.2074 AND VARIANCE = 47.2019

PASADANA, CA

YR=74 STATE=05 AREA=5760 SITE=004 AGCY=I PROJ=01 POLL=0000300 METH=11 UNT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 8632

BASIC STATISTICS=====

N = 8632

MEAN = 4.42805767

2ND MOMENT (VAR.) = 12.0125704 STANDARD DEV. = 3.46591568

COEFFICIENT OF VARIATION = 0.78271687

3RD MOMENT = 108.401703 COEF. OF SKEWNESS = 2.60364914

4TH MOMENT = 1780.89233 COEF. OF KURTOSIS = 12.3414392

LOWEST VALUE = 1.00000000

HIGHEST VALUE = 33.0000000

SECOND HIGHEST VALUE = 29.0000000

HISTOGRAM OF RAW DATA=====

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 0

INTERVAL	NUMBER	PERCENT	HISTOGRAM	CUM. FREQ.	PLOT POS.
-0.50- 0.50	0	0.0 %	I	0.0 %	0.0 %
0.50- 1.50	695	8.051%	I***C****	8.051%	8.051%
1.50- 2.50	1704	19.740%	I**********C******	27.792%	27.789%
2.50- 3.50	2034	23.563%	I********** C	51.355%	51.349%
3.50- 4.50	1418	16.427%	I********** C	67.783%	67.775%
4.50- 5.50	820	9.500%	I********* C	77.282%	77.273%
5.50- 6.50	566	6.557%	I******* C	83.839%	83.829%
6.50- 7.50	371	4.298%	I**** C	88.137%	88.127%
7.50- 8.50	230	2.665%	I*** C	90.802%	90.791%
8.50- 9.50	139	1.610%	I** C	92.412%	92.401%
9.50- 10.50	134	1.552%	I** C	93.964%	93.953%

10.50- 11.50	86	0.996%	I*	94.961%	94.950%
11.50- 12.50	89	1.031%	I*	95.992%	95.980%
12.50- 13.50	72	0.834%	I*	96.826%	96.814%
13.50- 14.50	59	0.684%	I*	97.509%	97.498%
14.50- 15.50	49	0.568%	I*	98.077%	98.065%
15.50- 16.50	33	0.382%	I	98.459%	98.448%
16.50- 17.50	17	0.197%	I	98.656%	98.645%
17.50- 18.50	18	0.209%	I	98.865%	98.853%
18.50- 19.50	21	0.243%	I	99.108%	99.096%
19.50- 20.50	15	0.174%	I	99.282%	99.270%
20.50- 21.50	16	0.185%	I	99.467%	99.455%
21.50- 22.50	7	0.081%	I	99.548%	99.536%
22.50- 23.50	9	0.104%	I	99.652%	99.641%
23.50- 24.50	10	0.116%	I	99.768%	99.757%
24.50- 25.50	5	0.058%	I	99.826%	99.814%
25.50- 26.50	7	0.081%	I	99.907%	99.896%
26.50- 27.50	3	0.035%	I	99.942%	99.930%
27.50- 28.50	3	0.035%	I	99.977%	99.965%
28.50- 29.50	1	0.012%	I	99.988%	99.977%
29.50- 30.50	0	0.0 %	I	99.988%	99.977%
30.50- 31.50	0	0.0 %	I	99.988%	99.977%
31.50- 32.50	0	0.0 %	I	99.988%	99.977%
32.50- 33.50	1	0.012%	I	100.000%	99.988%
33.50- 34.50	0	0.0 %	I	100.000%	99.988%
34.50- 35.50	0	0.0 %	I	100.000%	99.988%
35.50- 36.50	0	0.0 %	I	100.000%	99.988%
36.50- 37.50	0	0.0 %	I	100.000%	99.988%
37.50- 38.50	0	0.0 %	I	100.000%	99.988%
38.50- 39.50	0	0.0 %	I	100.000%	99.988%
39.50- 40.50	0	0.0 %	I	100.000%	99.988%
40.50- 41.50	0	0.0 %	I	100.000%	99.988%
41.50- 42.50	0	0.0 %	I	100.000%	99.988%
42.50- 43.50	0	0.0 %	I	100.000%	99.988%
43.50- 44.50	0	0.0 %	I	100.000%	99.988%
44.50- 45.50	0	0.0 %	I	100.000%	99.988%
45.50- 46.50	0	0.0 %	I	100.000%	99.988%
46.50- 47.50	0	0.0 %	I	100.000%	99.988%
47.50- 48.50	0	0.0 %	I	100.000%	99.988%
48.50- 49.50	0	0.0 %	I	100.000%	99.988%
49.50- 50.50	0	0.0 %	I	100.000%	99.988%
50.50- 51.50	0	0.0 %	I	100.000%	99.988%
51.50- 52.50	0	0.0 %	I	100.000%	99.988%
52.50- 53.50	0	0.0 %	I	100.000%	99.988%
53.50- 54.50	0	0.0 %	I	100.000%	99.988%
54.50- 55.50	0	0.0 %	I	100.000%	99.988%
55.50- 56.50	0	0.0 %	I	100.000%	99.988%
56.50- 57.50	0	0.0 %	I	100.000%	99.988%
57.50- 58.50	0	0.0 %	I	100.000%	99.988%
58.50- 59.50	0	0.0 %	I	100.000%	99.988%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	8541	9.9783	0.83	**********
2	0.08	8528	6.8666	0.57	**********
3	0.13	8520	4.6281	0.39	********
4	0.17	8516	3.1719	0.26	******
5	0.21	8510	2.2143	0.18	****
6	0.25	8509	1.6846	0.14	***
7	0.29	8504	1.6239	0.14	***
8	0.33	8500	2.1250	0.18	****
9	0.38	8497	3.1051	0.26	******
10	0.42	8495	4.1300	0.34	*******
11	0.46	8493	4.7624	0.40	********
12	0.50	8492	4.9412	0.41	*********
13	0.54	8491	4.6492	0.39	********
14	0.58	8490	3.8696	0.32	*******
15	0.63	8489	2.7628	0.23	*****
16	0.67	8490	1.6996	0.14	***
17	0.71	8491	1.1312	0.09	**
18	0.75	8491	1.1045	0.09	**
19	0.79	8488	1.5033	0.13	***
20	0.83	8487	2.2129	0.18	****
21	0.88	8487	3.2565	0.27	******
22	0.92	8484	4.8447	0.40	*********
23	0.96	8485	6.9350	0.58	**********
24	1.00	8498	8.1181	0.68	**********
25	1.04	8485	6.9929	0.58	**********
26	1.08	8481	4.9155	0.41	*********
27	1.13	8480	3.2482	0.27	******
28	1.17	8477	2.1002	0.17	****
29	1.21	8475	1.2983	0.11	***
30	1.25	8476	0.8175	0.07	**
31	1.29	8476	0.7499	0.06	**
32	1.33	8473	1.1298	0.09	**
33	1.38	8471	1.9386	0.16	****
34	1.42	8470	2.8863	0.24	*****
35	1.46	8469	3.4854	0.29	******
36	1.50	8468	3.6462	0.30	*******
37	1.54	8467	3.3573	0.28	******
38	1.58	8466	2.6615	0.22	*****
39	1.63	8466	1.7246	0.14	***
40	1.67	8466	0.9318	0.08	**
41	1.71	8465	0.5523	0.05	*
42	1.75	8465	0.5546	0.05	*
43	1.79	8465	0.9280	0.08	**
44	1.83	8464	1.5748	0.13	***
45	1.88	8461	2.5253	0.21	*****
46	1.92	8460	3.7805	0.31	*******
47	1.96	8466	5.3316	0.44	*********
48	2.00	8465	6.2416	0.52	**********
49	2.04	8460	5.3941	0.45	*********
50	2.08	8456	3.7430	0.31	*******

51	2.13	8455.	2.4310	0.20	*****
52	2.17	8453.	1.5206	0.13	***
53	2.21	8453.	0.9493	0.08	**
54	2.25	8453.	0.6181	0.05	**
55	2.29	8452.	0.5587	0.05	*
56	2.33	8449.	0.8392	0.07	**
57	2.38	8448.	1.5397	0.13	***
58	2.42	8447.	2.3814	0.20	****
59	2.46	8446.	2.9867	0.25	*****
60	2.50	8445.	3.2763	0.27	******
61	2.54	8444.	3.0755	0.26	******
62	2.58	8443.	2.4577	0.20	*****
63	2.63	8442.	1.6155	0.13	***
64	2.67	8441.	0.8867	0.07	**
65	2.71	8440.	0.5550	0.05	*
66	2.75	8441.	0.5596	0.05	*
67	2.79	8441.	0.8605	0.07	**
68	2.83	8440.	1.4024	0.12	***
69	2.88	8439.	2.2466	0.19	****
70	2.92	8436.	3.4894	0.29	******
71	2.96	8441.	5.0428	0.42	*********
72	3.00	8439.	5.9226	0.49	**********
73	3.04	8434.	5.0753	0.42	*********
74	3.08	8431.	3.4354	0.29	******
75	3.13	8435.	2.1866	0.18	****

FOR THIS AUTOCORRELATION: MEAN = 4.4281 AND VARIANCE = 12.0168

SPRINGFIELD, MA

YR=74 STATE=22 AREA=2160 SITE=005 AGCY=F PROJ=01 POLL=0000300 METH=11 UNT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 8339

BASIC STATISTICS-----

N = 8339

MEAN = 3.70587349

2ND MOMENT (VAR.) = 6.74194527 STANDARD DEV. = 2.59652519

COEFFICIENT OF VARIATION = 0.70065129

3RD MOMENT = 37.3594055 COEF. OF SKEWNESS = 2.13413811

4TH MOMENT = 88.601807 COEF. OF KURTOSIS = 12.9494486

LOWEST VALUE = 0.0

HIGHEST VALUE = 30.0000000

SECOND HIGHEST VALUE = 28.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 287

INTERVAL	NUMBER	PERCENT	HISTOGRAM	CUM. FREQ.	PLOT POS.
-0.50- 0.50	292	3.502%	I*C** I	3.502%	3.501%
0.50- 1.50	976	11.704%	I*******C**** I	15.206%	15.204%
1.50- 2.50	1716	20.578%	I**********C*** I	35.784%	35.779%
2.50- 3.50	1615	19.367%	I********** C I	55.150%	55.144%
3.50- 4.50	1367	16.393%	I********** C I	71.543%	71.535%
4.50- 5.50	829	9.941%	I********** C I	81.485%	81.475%
5.50- 6.50	666	7.987%	I******** C I	89.471%	89.460%
6.50- 7.50	290	3.478%	I*** C I	92.949%	92.938%
7.50- 8.50	226	2.710%	I*** C I	95.659%	95.647%
8.50- 9.50	104	1.247%	I* C I	96.906%	96.894%
9.50- 10.50	82	0.983%	I* C I	97.889%	97.878%

10.50- 11.50	45	0.540%	I*	98.429%	98.417%
11.50- 12.50	47	0.564%	I*	98.993%	98.981%
12.50- 13.50	21	0.252%	I	99.244%	99.233%
13.50- 14.50	10	0.120%	I	99.364%	99.352%
14.50- 15.50	15	0.180%	I	99.544%	99.532%
15.50- 16.50	10	0.120%	I	99.664%	99.652%
16.50- 17.50	4	0.048%	I	99.712%	99.700%
17.50- 18.50	7	0.084%	I	99.796%	99.784%
18.50- 19.50	1	0.012%	I	99.808%	99.796%
19.50- 20.50	3	0.036%	I	99.844%	99.832%
20.50- 21.50	3	0.036%	I	99.880%	99.868%
21.50- 22.50	1	0.012%	I	99.892%	99.880%
22.50- 23.50	1	0.012%	I	99.904%	99.892%
23.50- 24.50	2	0.024%	I	99.928%	99.916%
24.50- 25.50	0	0.0 %	I	99.928%	99.916%
25.50- 26.50	4	0.048%	I	99.976%	99.964%
26.50- 27.50	0	0.0 %	I	99.976%	99.964%
27.50- 28.50	1	0.012%	I	99.988%	99.976%
28.50- 29.50	0	0.0 %	I	99.988%	99.976%
29.50- 30.50	1	0.012%	I	100.000%	99.988%
30.50- 31.50	0	0.0 %	I	100.000%	99.988%
31.50- 32.50	0	0.0 %	I	100.000%	99.988%
32.50- 33.50	0	0.0 %	I	100.000%	99.988%
33.50- 34.50	0	0.0 %	I	100.000%	99.988%
34.50- 35.50	0	0.0 %	I	100.000%	99.988%
35.50- 36.50	0	0.0 %	I	100.000%	99.988%
36.50- 37.50	0	0.0 %	I	100.000%	99.988%
37.50- 38.50	0	0.0 %	I	100.000%	99.988%
38.50- 39.50	0	0.0 %	I	100.000%	99.988%
39.50- 40.50	0	0.0 %	I	100.000%	99.988%
40.50- 41.50	0	0.0 %	I	100.000%	99.988%
41.50- 42.50	0	0.0 %	I	100.000%	99.988%
42.50- 43.50	0	0.0 %	I	100.000%	99.988%
43.50- 44.50	0	0.0 %	I	100.000%	99.988%
44.50- 45.50	0	0.0 %	I	100.000%	99.988%
45.50- 46.50	0	0.0 %	I	100.000%	99.988%
46.50- 47.50	0	0.0 %	I	100.000%	99.988%
47.50- 48.50	0	0.0 %	I	100.000%	99.988%
48.50- 49.50	0	0.0 %	I	100.000%	99.988%
49.50- 50.50	0	0.0 %	I	100.000%	99.988%
50.50- 51.50	0	0.0 %	I	100.000%	99.988%
51.50- 52.50	0	0.0 %	I	100.000%	99.988%
52.50- 53.50	0	0.0 %	I	100.000%	99.988%
53.50- 54.50	0	0.0 %	I	100.000%	99.988%
54.50- 55.50	0	0.0 %	I	100.000%	99.988%
55.50- 56.50	0	0.0 %	I	100.000%	99.988%
56.50- 57.50	0	0.0 %	I	100.000%	99.988%
57.50- 58.50	0	0.0 %	I	100.000%	99.988%
58.50- 59.50	0	0.0 %	I	100.000%	99.988%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	8271.	5.1640	0.77	**********
2	0.08	8228.	3.8366	0.57	**********
3	0.13	8197.	2.8605	0.42	*********
4	0.17	8172.	2.2334	0.33	*******
5	0.21	8150.	1.7615	0.26	******
6	0.25	8134.	1.4701	0.22	*****
7	0.29	8121.	1.3503	0.20	*****
8	0.33	8110.	1.3809	0.20	*****
9	0.38	8102.	1.1586	0.17	****
10	0.42	8094.	0.7604	0.11	***
11	0.46	8083.	0.4159	0.06	**
12	0.50	8075.	0.3137	0.05	*
13	0.54	8067.	0.2576	0.04	*
14	0.58	8060.	0.4219	0.06	**
15	0.63	8054.	0.5956	0.09	**
16	0.67	8045.	0.6403	0.09	**
17	0.71	8040.	0.4902	0.07	**
18	0.75	8035.	0.4300	0.06	**
19	0.79	8032.	0.4959	0.07	**
20	0.83	8030.	0.7155	0.11	***
21	0.88	8029.	0.9018	0.13	***
22	0.92	8022.	1.2718	0.19	****
23	0.96	8021.	1.7648	0.26	******
24	1.00	8021.	2.1496	0.32	*******
25	1.04	8018.	1.6639	0.25	*****
26	1.08	8016.	1.1047	0.16	****
27	1.13	8009.	0.7254	0.11	***
28	1.17	8005.	0.4930	0.07	**
29	1.21	8003.	0.2997	0.04	*
30	1.25	7999.	0.2189	0.03	*
31	1.29	7994.	0.2681	0.04	*
32	1.33	7990.	0.3608	0.05	**
33	1.38	7986.	0.2632	0.04	*
34	1.42	7984.	0.0568	0.01	*
35	1.46	7980.	-0.0978	-.01	**
36	1.50	7976.	-0.1391	-.02	**
37	1.54	7972.	-0.3029	-.04	**
38	1.58	7969.	-0.2221	-.03	**
39	1.63	7964.	-0.1893	-.03	**
40	1.67	7962.	-0.1905	-.03	**
41	1.71	7963.	-0.3038	-.05	**
42	1.75	7962.	-0.3555	-.05	***
43	1.79	7958.	-0.2983	-.04	**
44	1.83	7956.	-0.1378	-.02	**
45	1.88	7950.	0.0195	0.00	*
46	1.92	7944.	0.3060	0.05	*
47	1.96	7940.	0.6658	0.10	**
48	2.00	7935.	1.0016	0.15	***
49	2.04	7928.	0.6292	0.09	**
50	2.08	7921.	0.2401	0.04	*

51	2.13	7917.	-0.0521	-.01	**
52	2.17	7912.	-0.1924	-.03	**
53	2.21	7907.	-0.3111	-.05	**
54	2.25	7905.	-0.2908	-.04	**
55	2.29	7904.	-0.1214	-.02	**
56	2.33	7902.	0.0558	0.01	*
57	2.38	7900.	0.0119	0.00	*
58	2.42	7898.	-0.0738	-.01	**
59	2.46	7899.	-0.1488	-.02	**
60	2.50	7897.	-0.1408	-.02	**
61	2.54	7894.	-0.2049	-.03	**
62	2.58	7891.	-0.1500	-.02	**
63	2.63	7887.	-0.1348	-.02	**
64	2.67	7885.	-0.1914	-.03	**
65	2.71	7882.	-0.3758	-.06	***
66	2.75	7879.	-0.4639	-.07	***
67	2.79	7877.	-0.4366	-.06	***
68	2.83	7878.	-0.3692	-.05	***
69	2.88	7876.	-0.3076	-.05	**
70	2.92	7875.	-0.0721	-.01	**
71	2.96	7873.	0.3112	0.05	*
72	3.00	7872.	0.5791	0.09	**
73	3.04	7871.	0.2818	0.04	*
74	3.08	7869.	-0.0305	-.00	**
75	3.13	7867.	-0.2440	-.04	**

FOR THIS AUTOCORRELATION: MEAN = 3.7059 AND VARIANCE = 6.7432

NEWHALL, CA

YR=74 STATE=05 AREA=5120 SITE=001 AGCY=I PROJ=01 POLL=0000300 METH=11 UNT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 8623

BASIC STATISTICS-----

N = 8623

MEAN = 3.33004665

2ND MOMENT (VAR.) = 2.88537216 STANDARD DEV. = 1.69863796

COEFFICIENT OF VARIATION = 0.51009429

3RD MOMENT = 6.15525055 COEF. OF SKEWNESS = 1.25586605

4TH MOMENT = 50.8687592 COEF. OF KURTOSIS = 6.11008835

LOWEST VALUE = 1.00000000

HIGHEST VALUE = 14.0000000

SECOND HIGHEST VALUE = 14.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 0

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	0	0.0 %	I	0.0 %	0.0 %
0.50- 1.50	938	10.878%	I****C******	10.878%	10.877%
1.50- 2.50	2053	23.808%	I****************C*******	34.686%	34.682%
2.50- 3.50	2169	25.154%	I************************** C	59.840%	59.833%
3.50- 4.50	1732	20.086%	I********************* C	79.926%	79.916%
4.50- 5.50	958	11.110%	I*********** C	91.036%	91.025%
5.50- 6.50	421	4.882%	I***** C	95.918%	95.907%
6.50- 7.50	144	1.670%	I** C	97.588%	97.576%
7.50- 8.50	88	1.021%	I* C	98.608%	98.597%
8.50- 9.50	54	0.626%	I* CI	99.235%	99.223%
9.50- 10.50	29	0.336%	I CI	99.571%	99.559%

10.50- 11.50	21	0.244%	I	CI 99.814%	99.803%
11.50- 12.50	10	0.116%	I	CI 99.930%	99.919%
12.50- 13.50	3	0.035%	I	CI 99.965%	99.954%
13.50- 14.50	3	0.035%	I	CI 100.000%	99.988%
14.50- 15.50	0	0.0 %	I	CI 100.000%	99.988%
15.50- 16.50	0	0.0 %	I	CI 100.000%	99.988%
16.50- 17.50	0	0.0 %	I	CI 100.000%	99.988%
17.50- 18.50	0	0.0 %	I	CI 100.000%	99.988%
18.50- 19.50	0	0.0 %	I	CI 100.000%	99.988%
19.50- 20.50	0	0.0 %	I	CI 100.000%	99.988%
20.50- 21.50	0	0.0 %	I	CI 100.000%	99.988%
21.50- 22.50	0	0.0 %	I	CI 100.000%	99.988%
22.50- 23.50	0	0.0 %	I	CI 100.000%	99.988%
23.50- 24.50	0	0.0 %	I	CI 100.000%	99.988%
24.50- 25.50	0	0.0 %	I	CI 100.000%	99.988%
25.50- 26.50	0	0.0 %	I	CI 100.000%	99.988%
26.50- 27.50	0	0.0 %	I	CI 100.000%	99.988%
27.50- 28.50	0	0.0 %	I	CI 100.000%	99.988%
28.50- 29.50	0	0.0 %	I	CI 100.000%	99.988%
29.50- 30.50	0	0.0 %	I	CI 100.000%	99.988%
30.50- 31.50	0	0.0 %	I	CI 100.000%	99.988%
31.50- 32.50	0	0.0 %	I	CI 100.000%	99.988%
32.50- 33.50	0	0.0 %	I	CI 100.000%	99.988%
33.50- 34.50	0	0.0 %	I	CI 100.000%	99.988%
34.50- 35.50	0	0.0 %	I	CI 100.000%	99.988%
35.50- 36.50	0	0.0 %	I	CI 100.000%	99.988%
36.50- 37.50	0	0.0 %	I	CI 100.000%	99.988%
37.50- 38.50	0	0.0 %	I	CI 100.000%	99.988%
38.50- 39.50	0	0.0 %	I	CI 100.000%	99.988%
39.50- 40.50	0	0.0 %	I	CI 100.000%	99.988%
40.50- 41.50	0	0.0 %	I	CI 100.000%	99.988%
41.50- 42.50	0	0.0 %	I	CI 100.000%	99.988%
42.50- 43.50	0	0.0 %	I	CI 100.000%	99.988%
43.50- 44.50	0	0.0 %	I	CI 100.000%	99.988%
44.50- 45.50	0	0.0 %	I	CI 100.000%	99.988%
45.50- 46.50	0	0.0 %	I	CI 100.000%	99.988%
46.50- 47.50	0	0.0 %	I	CI 100.000%	99.988%
47.50- 48.50	0	0.0 %	I	CI 100.000%	99.988%
48.50- 49.50	0	0.0 %	I	CI 100.000%	99.988%
49.50- 50.50	0	0.0 %	I	CI 100.000%	99.988%
50.50- 51.50	0	0.0 %	I	CI 100.000%	99.988%
51.50- 52.50	0	0.0 %	I	CI 100.000%	99.988%
52.50- 53.50	0	0.0 %	I	CI 100.000%	99.988%
53.50- 54.50	0	0.0 %	I	CI 100.000%	99.988%
54.50- 55.50	0	0.0 %	I	CI 100.000%	99.988%
55.50- 56.50	0	0.0 %	I	CI 100.000%	99.988%
56.50- 57.50	0	0.0 %	I	CI 100.000%	99.988%
57.50- 58.50	0	0.0 %	I	CI 100.000%	99.988%
58.50- 59.50	0	0.0 %	I	CI 100.000%	99.988%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	8554.	2.5251	0.87	**********
2	0.08	8548.	1.9858	0.69	**********
3	0.13	8543.	1.5198	0.53	**********
4	0.17	8540.	1.2034	0.42	*********
5	0.21	8537.	1.0225	0.35	********
6	0.25	8536.	0.9412	0.33	*******
7	0.29	8531.	0.9428	0.33	*******
8	0.33	8528.	1.0112	0.35	********
9	0.38	8525.	1.1140	0.39	********
10	0.42	8523.	1.2097	0.42	*********
11	0.46	8521.	1.2809	0.44	*********
12	0.50	8519.	1.3000	0.45	**********
13	0.54	8517.	1.2522	0.43	*********
14	0.58	8515.	1.1376	0.39	********
15	0.63	8513.	0.9978	0.35	*******
16	0.67	8511.	0.8423	0.29	******
17	0.71	8510.	0.7204	0.25	*****
18	0.75	8507.	0.6457	0.22	*****
19	0.79	8505.	0.6327	0.22	*****
20	0.83	8503.	0.6909	0.24	*****
21	0.88	8501.	0.8444	0.29	******
22	0.92	8500.	1.0931	0.38	********
23	0.96	8500.	1.3557	0.47	**********
24	1.00	8497.	1.4792	0.51	**********
25	1.04	8493.	1.3421	0.46	**********
26	1.08	8492.	1.0533	0.36	********
27	1.13	8491.	0.7882	0.27	******
28	1.17	8488.	0.6028	0.21	*****
29	1.21	8485.	0.5107	0.18	****
30	1.25	8484.	0.4833	0.17	****
31	1.29	8482.	0.4992	0.17	****
32	1.33	8481.	0.5481	0.19	****
33	1.38	8479.	0.6141	0.21	*****
34	1.42	8478.	0.6847	0.24	*****
35	1.46	8476.	0.7276	0.25	******
36	1.50	8474.	0.7400	0.26	******
37	1.54	8472.	0.7067	0.24	*****
38	1.58	8470.	0.6479	0.22	*****
39	1.63	8468.	0.5570	0.19	****
40	1.67	8465.	0.4671	0.16	****
41	1.71	8464.	0.3890	0.13	***
42	1.75	8462.	0.3504	0.12	***
43	1.79	8460.	0.3454	0.12	***
44	1.83	8457.	0.3831	0.13	***
45	1.88	8455.	0.4880	0.17	****
46	1.92	8453.	0.6647	0.23	*****
47	1.96	8451.	0.8603	0.30	******
48	2.00	8450.	0.9510	0.33	*******
49	2.04	8449.	0.8487	0.29	******
50	2.08	8446.	0.6304	0.22	*****

51	2.13	8444.	0.4332	0.15	***
52	2.17	8441.	0.3158	0.11	***
53	2.21	8439.	0.2690	0.09	**
54	2.25	8438.	0.2693	0.09	**
55	2.29	8435.	0.2937	0.10	***
56	2.33	8433.	0.3600	0.12	***
57	2.38	8431.	0.4192	0.15	***
58	2.42	8430.	0.4780	0.17	****
59	2.46	8429.	0.5194	0.18	****
60	2.50	8428.	0.5362	0.19	****
61	2.54	8427.	0.5302	0.18	****
62	2.58	8426.	0.4844	0.17	****
63	2.63	8425.	0.4126	0.14	***
64	2.67	8424.	0.3383	0.12	***
65	2.71	8423.	0.2693	0.09	**
66	2.75	8425.	0.2330	0.08	**
67	2.79	8421.	0.2368	0.08	**
68	2.83	8420.	0.2767	0.10	**
69	2.88	8420.	0.3781	0.13	***
70	2.92	8418.	0.5469	0.19	****
71	2.96	8420.	0.7300	0.25	******
72	3.00	8419.	0.8045	0.28	******
73	3.04	8417.	0.6852	0.24	*****
74	3.08	8414.	0.4773	0.17	****
75	3.13	8415.	0.3020	0.10	***

FOR THIS AUTOCORRELATION; MEAN = 3.3300 AND VARIANCE = 2.8882

PHOENIX, AZ

YR=74 STATE=03 AREA=0600 SITE=002 AGCY=G PROJ=01 POLL=0000300 METH=11 UNIT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 8246

BASIC STATISTICS-----

N = 8246

MEAN = 2.93699646

2ND MOMENT (VAR.) = 16.8246613 STANDARD DEV. = 4.10178757

COEFFICIENT OF VARIATION = 1.39659214

3RD MOMENT = 197.331665 COEF. OF SKEWNESS = 2.85941696

4TH MOMENT = 4761.04687 COEF. OF KURTOSIS = 16.8193665

LOWEST VALUE = 0.0

HIGHEST VALUE = 61.0000000

SECOND HIGHEST VALUE = 50.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 2212

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	2218	26.898%	I************C************

```
10.50- 11.50     65    0.788%  I*                                          C   I  94.761%     94.750%
11.50- 12.50     85    1.031%  I*                                           C  I  95.792%     95.780%
12.50- 13.50     74    0.897%  I*                                           C  I  96.689%     96.677%
13.50- 14.50     37    0.449%  I                                             C I  97.138%     97.126%
14.50- 15.50     46    0.558%  I*                                            C I  97.696%     97.684%
15.50- 16.50     40    0.485%  I                                             C I  98.181%     98.169%
16.50- 17.50     34    0.412%  I                                             C I  98.593%     98.581%
17.50- 18.50     22    0.267%  I                                             C I  98.860%     98.848%
18.50- 19.50     15    0.182%  I                                              CI  99.042%     99.030%
19.50- 20.50     27    0.327%  I                                              CI  99.369%     99.357%
20.50- 21.50     10    0.121%  I                                              CI  99.491%     99.478%
21.50- 22.50      9    0.109%  I                                              CI  99.600%     99.588%
22.50- 23.50      7    0.085%  I                                              CI  99.685%     99.672%
23.50- 24.50      4    0.049%  I                                              CI  99.733%     99.721%
24.50- 25.50      8    0.097%  I                                              CI  99.830%     99.818%
25.50- 26.50      2    0.024%  I                                              CI  99.854%     99.842%
26.50- 27.50      2    0.024%  I                                              CI  99.879%     99.866%
27.50- 28.50      2    0.024%  I                                              CI  99.903%     99.891%
28.50- 29.50      1    0.012%  I                                              CI  99.915%     99.903%
29.50- 30.50      2    0.024%  I                                              CI  99.939%     99.927%
30.50- 31.50      0    0.0  %  I                                              CI  99.939%     99.927%
31.50- 32.50      1    0.012%  I                                              CI  99.951%     99.939%
32.50- 33.50      1    0.012%  I                                              CI  99.963%     99.951%
33.50- 34.50      1    0.012%  I                                              CI  99.975%     99.963%
34.50- 35.50      0    0.0  %  I                                              CI  99.975%     99.963%
35.50- 36.50      0    0.0  %  I                                              CI  99.975%     99.963%
36.50- 37.50      0    0.0  %  I                                              CI  99.975%     99.963%
37.50- 38.50      0    0.0  %  I                                              CI  99.975%     99.963%
38.50- 39.50      0    0.0  %  I                                              CI  99.975%     99.963%
39.50- 40.50      0    0.0  %  I                                              CI  99.975%     99.963%
40.50- 41.50      0    0.0  %  I                                              CI  99.975%     99.963%
41.50- 42.50      0    0.0  %  I                                              CI  99.975%     99.963%
42.50- 43.50      0    0.0  %  I                                              CI  99.975%     99.963%
43.50- 44.50      0    0.0  %  I                                              CI  99.975%     99.963%
44.50- 45.50      0    0.0  %  I                                              CI  99.975%     99.963%
45.50- 46.50      0    0.0  %  I                                              CI  99.975%     99.963%
46.50- 47.50      0    0.0  %  I                                              CI  99.975%     99.963%
47.50- 48.50      0    0.0  %  I                                              CI  99.975%     99.963%
48.50- 49.50      0    0.0  %  I                                              CI  99.975%     99.963%
49.50- 50.50      1    0.012%  I                                              CI  99.988%     99.975%
50.50- 51.50      0    0.0  %  I                                              CI  99.988%     99.975%
51.50- 52.50      0    0.0  %  I                                              CI  99.988%     99.975%
52.50- 53.50      0    0.0  %  I                                              CI  99.988%     99.975%
53.50- 54.50      0    0.0  %  I                                              CI  99.988%     99.975%
54.50- 55.50      0    0.0  %  I                                              CI  99.988%     99.975%
55.50- 56.50      0    0.0  %  I                                              CI  99.988%     99.975%
56.50- 57.50      0    0.0  %  I                                              CI  99.988%     99.975%
57.50- 58.50      0    0.0  %  I                                              CI  99.988%     99.975%
58.50- 59.50      0    0.0  %  I                                              CI  99.988%     99.975%
                                +-----------------------------------------------+
```

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 1 (0.012%)

LISTING OF VALUES ABOVE RANGE: 1 61.000000

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	8215.	14.5772	0.87	**********
2	0.08	8188.	11.5765	0.69	**********
3	0.13	8161.	8.8622	0.53	**********
4	0.17	8138.	6.4732	0.38	**********
5	0.21	8117.	4.5426	0.27	**********
6	0.25	8098.	3.0751	0.18	**********
7	0.29	8082.	2.3095	0.14	**********
8	0.33	8065.	2.0709	0.12	**********
9	0.38	8050.	2.1607	0.13	**********
10	0.42	8038.	2.3772	0.14	**********
11	0.46	8026.	2.6184	0.16	**********
12	0.50	8014.	2.8259	0.17	**********
13	0.54	8003.	2.8603	0.17	**********
14	0.58	7995.	2.6990	0.16	**********
15	0.63	7989.	2.4057	0.14	**********
16	0.67	7983.	2.1558	0.13	**********
17	0.71	7978.	2.2451	0.13	**********
18	0.75	7973.	2.7957	0.17	**********
19	0.79	7969.	3.8067	0.23	**********
20	0.83	7963.	5.2387	0.31	**********
21	0.88	7958.	6.8194	0.41	**********
22	0.92	7951.	8.3674	0.50	**********
23	0.96	7944.	9.5040	0.56	**********
24	1.00	7937.	9.8357	0.58	**********
25	1.04	7930.	9.1481	0.54	**********
26	1.08	7924.	7.7858	0.46	**********
27	1.13	7916.	6.2134	0.37	**********
28	1.17	7908.	4.6555	0.28	**********
29	1.21	7900.	3.2328	0.19	**********
30	1.25	7892.	2.1783	0.13	**********
31	1.29	7884.	1.5498	0.09	**********
32	1.33	7875.	1.3288	0.08	**********
33	1.38	7866.	1.3767	0.08	**********
34	1.42	7856.	1.6600	0.10	**********
35	1.46	7846.	1.9162	0.11	**********
36	1.50	7838.	2.1183	0.13	**********
37	1.54	7831.	2.0972	0.12	**********
38	1.58	7823.	1.8646	0.11	**********
39	1.63	7815.	1.6352	0.10	**********
40	1.67	7808.	1.4475	0.09	**********
41	1.71	7801.	1.5929	0.09	**********
42	1.75	7794.	2.1641	0.13	**********
43	1.79	7787.	3.1304	0.19	**********
44	1.83	7781.	4.2425	0.25	**********
45	1.88	7775.	5.4908	0.33	**********
46	1.92	7771.	6.6908	0.40	**********
47	1.96	7768.	7.6216	0.45	**********
48	2.00	7765.	7.9678	0.47	**********
49	2.04	7762.	7.4092	0.44	**********
50	2.08	7757.	6.4493	0.38	**********

51	2.13	7752.	5.3183	0.32	I ******* I
52	2.17	7748.	4.0768	0.24	I ***** I
53	2.21	7745.	2.9546	0.18	I **** I
54	2.25	7742.	2.0426	0.12	I *** I
55	2.29	7738.	1.4955	0.09	I ** I
56	2.33	7734.	1.3146	0.08	I ** I
57	2.38	7729.	1.4347	0.09	I ** I
58	2.42	7724.	1.6610	0.10	I ** I
59	2.46	7720.	1.9337	0.11	I *** I
60	2.50	7715.	2.1525	0.13	I *** I
61	2.54	7710.	2.0910	0.12	I *** I
62	2.58	7705.	1.8058	0.11	I *** I
63	2.63	7701.	1.5211	0.09	I ** I
64	2.67	7697.	1.4078	0.08	I ** I
65	2.71	7694.	1.5687	0.09	I ** I
66	2.75	7693.	1.9399	0.12	I *** I
67	2.79	7692.	2.6223	0.16	I ****
68	2.83	7691.	3.5886	0.21	I ***** I
69	2.88	7689.	4.7541	0.28	I ****** I
70	2.92	7687.	5.9056	0.35	I ******** I
71	2.96	7686.	6.7670	0.40	I ********* I
72	3.00	7685.	7.0988	0.42	I ********* I
73	3.04	7685.	6.6925	0.40	I ******** I
74	3.08	7685.	5.8878	0.35	I ******* I
75	3.13	7684.	4.7184	0.28	I ****** I

+---+---+---+---+---+---+---+---+---+---+

FOR THIS AUTOCORRELATION: MEAN = 2.9370 AND VARIANCE = 16.8255

NAPA, CA

YR=74 STATE=05 AREA=5000 SITE=003 AGCY=I PROJ=01 POLL=0000500 METH=11 UNT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 8558

BASIC STATISTICS-----

N = 8558

MEAN = 2.73112869

2ND MOMENT (VAR.) = 1.94335270 STANDARD DEV. = 1.39404202

COEFFICIENT OF VARIATION = 0.51042706

3RD MOMENT = 5.69019604 COEF. OF SKEWNESS = 2.10038853

4TH MOMENT = 42.5159607 COEF. OF KURTOSIS = 11.2576780

LOWEST VALUE = 0.0

HIGHEST VALUE = 15.0000000

SECOND HIGHEST VALUE = 15.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 18

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	18	0.210%	I I	0.210%	0.210%
0.50- 1.50	965	11.276%	I*****C***** I	11.486%	11.485%
1.50- 2.50	3450	40.313%	I**************************C************** I	51.799%	51.793%
2.50- 3.50	2511	29.341%	I*******************

10.50- 11.50	11	0.129%	I	CI 99.883%	99.871%
11.50- 12.50	5	0.058%	I	CI 99.941%	99.930%
12.50- 13.50	1	0.012%	I	CI 99.953%	99.941%
13.50- 14.50	2	0.023%	I	CI 99.977%	99.965%
14.50- 15.50	2	0.023%	I	CI 100.000%	99.988%
15.50- 16.50	0	0.0 %	I	CI 100.000%	99.988%
16.50- 17.50	0	0.0 %	I	CI 100.000%	99.988%
17.50- 18.50	0	0.0 %	I	CI 100.000%	99.988%
18.50- 19.50	0	0.0 %	I	CI 100.000%	99.988%
19.50- 20.50	0	0.0 %	I	CI 100.000%	99.988%
20.50- 21.50	0	0.0 %	I	CI 100.000%	99.988%
21.50- 22.50	0	0.0 %	I	CI 100.000%	99.988%
22.50- 23.50	0	0.0 %	I	CI 100.000%	99.988%
23.50- 24.50	0	0.0 %	I	CI 100.000%	99.988%
24.50- 25.50	0	0.0 %	I	CI 100.000%	99.988%
25.50- 26.50	0	0.0 %	I	CI 100.000%	99.988%
26.50- 27.50	0	0.0 %	I	CI 100.000%	99.988%
27.50- 28.50	0	0.0 %	I	CI 100.000%	99.988%
28.50- 29.50	0	0.0 %	I	CI 100.000%	99.988%
29.50- 30.50	0	0.0 %	I	CI 100.000%	99.988%
30.50- 31.50	0	0.0 %	I	CI 100.000%	99.988%
31.50- 32.50	0	0.0 %	I	CI 100.000%	99.988%
32.50- 33.50	0	0.0 %	I	CI 100.000%	99.988%
33.50- 34.50	0	0.0 %	I	CI 100.000%	99.988%
34.50- 35.50	0	0.0 %	I	CI 100.000%	99.988%
35.50- 36.50	0	0.0 %	I	CI 100.000%	99.988%
36.50- 37.50	0	0.0 %	I	CI 100.000%	99.988%
37.50- 38.50	0	0.0 %	I	CI 100.000%	99.988%
38.50- 39.50	0	0.0 %	I	CI 100.000%	99.988%
39.50- 40.50	0	0.0 %	I	CI 100.000%	99.988%
40.50- 41.50	0	0.0 %	I	CI 100.000%	99.988%
41.50- 42.50	0	0.0 %	I	CI 100.000%	99.988%
42.50- 43.50	0	0.0 %	I	CI 100.000%	99.988%
43.50- 44.50	0	0.0 %	I	CI 100.000%	99.988%
44.50- 45.50	0	0.0 %	I	CI 100.000%	99.988%
45.50- 46.50	0	0.0 %	I	CI 100.000%	99.988%
46.50- 47.50	0	0.0 %	I	CI 100.000%	99.988%
47.50- 48.50	0	0.0 %	I	CI 100.000%	99.988%
48.50- 49.50	0	0.0 %	I	CI 100.000%	99.988%
49.50- 50.50	0	0.0 %	I	CI 100.000%	99.988%
50.50- 51.50	0	0.0 %	I	CI 100.000%	99.988%
51.50- 52.50	0	0.0 %	I	CI 100.000%	99.988%
52.50- 53.50	0	0.0 %	I	CI 100.000%	99.988%
53.50- 54.50	0	0.0 %	I	CI 100.000%	99.988%
54.50- 55.50	0	0.0 %	I	CI 100.000%	99.988%
55.50- 56.50	0	0.0 %	I	CI 100.000%	99.988%
56.50- 57.50	0	0.0 %	I	CI 100.000%	99.988%
57.50- 58.50	0	0.0 %	I	CI 100.000%	99.988%
58.50- 59.50	0	0.0 %	I	CI 100.000%	99.988%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	8423.	1.6020	0.82	**********
2	0.08	8413.	1.2470	0.64	**********
3	0.13	8409.	0.9808	0.50	**********
4	0.17	8406.	0.7805	0.40	**********
5	0.21	8404.	0.6035	0.31	*******
6	0.25	8400.	0.4595	0.24	*****
7	0.29	8397.	0.4222	0.22	*****
8	0.33	8394.	0.4755	0.24	*****
9	0.38	8391.	0.5685	0.29	******
10	0.42	8388.	0.6541	0.34	*******
11	0.46	8385.	0.7154	0.37	********
12	0.50	8382.	0.7380	0.38	********
13	0.54	8379.	0.7048	0.36	********
14	0.58	8376.	0.6262	0.32	*******
15	0.63	8373.	0.5141	0.26	******
16	0.67	8370.	0.4192	0.22	*****
17	0.71	8367.	0.3289	0.17	****
18	0.75	8365.	0.3290	0.17	****
19	0.79	8362.	0.4114	0.21	*****
20	0.83	8359.	0.5304	0.27	******
21	0.88	8356.	0.6641	0.34	*******
22	0.92	8353.	0.8133	0.42	*********
23	0.96	8351.	0.9787	0.50	**********
24	1.00	8349.	1.0552	0.54	**********
25	1.04	8347.	0.9803	0.50	**********
26	1.08	8342.	0.8007	0.41	*********
27	1.13	8339.	0.6527	0.34	*******
28	1.17	8337.	0.5147	0.26	******
29	1.21	8333.	0.3907	0.20	*****
30	1.25	8330.	0.2936	0.15	****
31	1.29	8327.	0.2728	0.14	***
32	1.33	8326.	0.3456	0.18	****
33	1.38	8325.	0.4238	0.22	*****
34	1.42	8325.	0.5023	0.26	******
35	1.46	8324.	0.5760	0.30	******
36	1.50	8323.	0.6037	0.31	*******
37	1.54	8322.	0.5902	0.30	*******
38	1.58	8321.	0.5162	0.27	******
39	1.63	8321.	0.4319	0.22	*****
40	1.67	8320.	0.3566	0.18	****
41	1.71	8319.	0.2939	0.15	****
42	1.75	8318.	0.3035	0.16	****
43	1.79	8319.	0.3636	0.19	****
44	1.83	8317.	0.4654	0.24	*****
45	1.88	8315.	0.5696	0.29	******
46	1.92	8317.	0.7024	0.36	********
47	1.96	8319.	0.8497	0.44	*********
48	2.00	8367.	0.9064	0.47	**********
49	2.04	8317.	0.8454	0.43	*********
50	2.08	8312.	0.7096	0.36	********

51	2.13	8308.	0.5975	0.31	I ******* I
52	2.17	8308.	0.4903	0.25	I ****** I
53	2.21	8308.	0.3756	0.19	I **** I
54	2.25	8304.	0.2995	0.15	I **** I
55	2.29	8303.	0.3074	0.16	I **** I
56	2.33	8302.	0.3700	0.19	I **** I
57	2.38	8301.	0.4406	0.23	I ***** I
58	2.42	8300.	0.5137	0.26	I ****** I
59	2.46	8299.	0.5758	0.30	I ****** I
60	2.50	8298.	0.5903	0.30	I ******* I
61	2.54	8297.	0.5611	0.29	I ****** I
62	2.58	8296.	0.5059	0.26	I ****** I
63	2.63	8295.	0.4385	0.23	I ***** I
64	2.67	8294.	0.3718	0.19	I **** I
65	2.71	8293.	0.3111	0.16	I **** I
66	2.75	8294.	0.3186	0.16	I **** I
67	2.79	8293.	0.3888	0.20	I **** I
68	2.83	8291.	0.4758	0.24	I ***** I
69	2.88	8290.	0.5864	0.30	I ******* I
70	2.92	8290.	0.6967	0.36	I ******** I
71	2.96	8294.	0.8313	0.43	I ********* I
72	3.00	8316.	0.8823	0.45	I ********** I
73	3.04	8293.	0.8183	0.42	I ********* I
74	3.08	8289.	0.6866	0.35	I ******** I
75	3.13	8286.	0.5677	0.29	I ****** I

+---+---+---+---+---+---+---+---+---+

FOR THIS AUTOCORRELATION: MEAN = 2.7311 AND VARIANCE = 1.9451

PHILADELPHIA, PA

YR=74 STATE=39 AREA=7140 SITE=002 AGCY=A PROJ=10 POLL=9999999 METH=11 UNT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 6836

BASIC STATISTICS-----

N = 6836

MEAN = 2.35532475

2ND MOMENT (VAR.) = 3.59263706 STANDARD DEV. = 1.89542484

COEFFICIENT OF VARIATION = 0.80474031

3RD MOMENT = 25.4414215 COEF. OF SKEWNESS = 3.73612976

4TH MOMENT = 424.209229 COEF. OF KURTOSIS = 32.8664856

LOWEST VALUE = 0.0

HIGHEST VALUE = 29.0000000

SECOND HIGHEST VALUE = 29.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 460

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	460	6.729%	I**C**** I	6.729%	6.728%
0.50- 1.50	1736	25.395%	I***************C********* I	32.124%	32.119%
1.50- 2.50	2195	32.109%	I*********************************C I	64.233%	64.224%
2.50- 3.50	1351	19.763%	I********************* C I	83.996%	83.984%
3.50- 4.50	546	7.987%	I******** C I	91.984%	91.970%
4.50- 5.50	254	3.716%	I**** C I	95.699%	95.685%
5.50- 6.50	104	1.521%	I** C I	97.221%	97.206%
6.50- 7.50	76	1.112%	I* C I	98.332%	98.318%
7.50- 8.50	32	0.468%	I C I	98.800%	98.786%
8.50- 9.50	18	0.263%	I CI	99.064%	99.049%
9.50- 10.50	16	0.234%	I CI	99.298%	99.283%

10.50- 11.50	9	0.132%	I	99.429%	99.415%
11.50- 12.50	7	0.102%	I	99.532%	99.517%
12.50- 13.50	4	0.059%	I	99.590%	99.576%
13.50- 14.50	4	0.059%	I	99.649%	99.634%
14.50- 15.50	7	0.102%	I	99.751%	99.737%
15.50- 16.50	4	0.059%	I	99.810%	99.795%
16.50- 17.50	2	0.029%	I	99.839%	99.824%
17.50- 18.50	4	0.059%	I	99.897%	99.883%
18.50- 19.50	0	0.0 %	I	99.897%	99.883%
19.50- 20.50	0	0.0 %	I	99.897%	99.883%
20.50- 21.50	4	0.059%	I	99.956%	99.941%
21.50- 22.50	0	0.0 %	I	99.956%	99.941%
22.50- 23.50	0	0.0 %	I	99.956%	99.941%
23.50- 24.50	0	0.0 %	I	99.956%	99.941%
24.50- 25.50	0	0.0 %	I	99.956%	99.941%
25.50- 26.50	1	0.015%	I	99.971%	99.956%
26.50- 27.50	0	0.0 %	I	99.971%	99.956%
27.50- 28.50	0	0.0 %	I	99.971%	99.956%
28.50- 29.50	2	0.029%	I	100.000%	99.985%
29.50- 30.50	0	0.0 %	I	100.000%	99.985%
30.50- 31.50	0	0.0 %	I	100.000%	99.985%
31.50- 32.50	0	0.0 %	I	100.000%	99.985%
32.50- 33.50	0	0.0 %	I	100.000%	99.985%
33.50- 34.50	0	0.0 %	I	100.000%	99.985%
34.50- 35.50	0	0.0 %	I	100.000%	99.985%
35.50- 36.50	0	0.0 %	I	100.000%	99.985%
36.50- 37.50	0	0.0 %	I	100.000%	99.985%
37.50- 38.50	0	0.0 %	I	100.000%	99.985%
38.50- 39.50	0	0.0 %	I	100.000%	99.985%
39.50- 40.50	0	0.0 %	I	100.000%	99.985%
40.50- 41.50	0	0.0 %	I	100.000%	99.985%
41.50- 42.50	0	0.0 %	I	100.000%	99.985%
42.50- 43.50	0	0.0 %	I	100.000%	99.985%
43.50- 44.50	0	0.0 %	I	100.000%	99.985%
44.50- 45.50	0	0.0 %	I	100.000%	99.985%
45.50- 46.50	0	0.0 %	I	100.000%	99.985%
46.50- 47.50	0	0.0 %	I	100.000%	99.985%
47.50- 48.50	0	0.0 %	I	100.000%	99.985%
48.50- 49.50	0	0.0 %	I	100.000%	99.985%
49.50- 50.50	0	0.0 %	I	100.000%	99.985%
50.50- 51.50	0	0.0 %	I	100.000%	99.985%
51.50- 52.50	0	0.0 %	I	100.000%	99.985%
52.50- 53.50	0	0.0 %	I	100.000%	99.985%
53.50- 54.50	0	0.0 %	I	100.000%	99.985%
54.50- 55.50	0	0.0 %	I	100.000%	99.985%
55.50- 56.50	0	0.0 %	I	100.000%	99.985%
56.50- 57.50	0	0.0 %	I	100.000%	99.985%
57.50- 58.50	0	0.0 %	I	100.000%	99.985%
58.50- 59.50	0	0.0 %	I	100.000%	99.985%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	6781.	2.9830	0.83	**********
2	0.08	6758.	2.3431	0.65	**********
3	0.13	6749.	1.8784	0.52	**********
4	0.17	6737.	1.5713	0.44	*********
5	0.21	6726.	1.3713	0.38	********
6	0.25	6714.	1.2731	0.35	********
7	0.29	6704.	1.2485	0.35	*******
8	0.33	6694.	1.2503	0.35	*******
9	0.38	6684.	1.2514	0.35	*******
10	0.42	6675.	1.1431	0.32	*******
11	0.46	6666.	0.9958	0.28	******
12	0.50	6658.	0.8297	0.23	*****
13	0.54	6650.	0.7168	0.20	****
14	0.58	6641.	0.6733	0.19	****
15	0.63	6633.	0.6043	0.17	****
16	0.67	6627.	0.4998	0.14	***
17	0.71	6618.	0.4116	0.11	***
18	0.75	6611.	0.3534	0.10	**
19	0.79	6604.	0.3057	0.09	**
20	0.83	6600.	0.3373	0.09	**
21	0.88	6594.	0.4168	0.12	***
22	0.92	6589.	0.5273	0.15	***
23	0.96	6582.	0.6664	0.19	****
24	1.00	6580.	0.7743	0.22	*****
25	1.04	6570.	0.6773	0.19	****
26	1.08	6563.	0.5195	0.14	***
27	1.13	6557.	0.3684	0.10	***
28	1.17	6552.	0.2514	0.07	**
29	1.21	6549.	0.1930	0.05	**
30	1.25	6542.	0.2008	0.06	**
31	1.29	6536.	0.2271	0.06	**
32	1.33	6532.	0.2948	0.08	**
33	1.38	6527.	0.3568	0.10	**
34	1.42	6523.	0.3396	0.09	**
35	1.46	6519.	0.2989	0.08	**
36	1.50	6515.	0.2690	0.07	**
37	1.54	6513.	0.2512	0.07	**
38	1.58	6510.	0.2412	0.07	**
39	1.63	6506.	0.2127	0.06	**
40	1.67	6501.	0.1446	0.04	*
41	1.71	6496.	0.0828	0.02	*
42	1.75	6495.	0.0474	0.01	*
43	1.79	6491.	0.0240	0.01	*
44	1.83	6486.	0.0355	0.01	*
45	1.88	6485.	0.1152	0.03	*
46	1.92	6481.	0.2847	0.08	**
47	1.96	6481.	0.4506	0.13	***
48	2.00	6481.	0.5199	0.14	***
49	2.04	6478.	0.4172	0.12	***
50	2.08	6474.	0.2683	0.07	**

51	2.13	6473.	0.1188	0.03	I * I
52	2.17	6472.	0.0153	0.00	I * I
53	2.21	6474.	-0.0279	-.01	I ** I
54	2.25	6470.	-0.0170	-.00	I ** I
55	2.29	6468.	0.0097	0.00	I * I
56	2.33	6465.	0.0549	0.02	I * I
57	2.38	6465.	0.0888	0.02	I * I
58	2.42	6464.	0.0898	0.02	I * I
59	2.46	6462.	0.0787	0.02	I * I
60	2.50	6462.	0.0685	0.02	I * I
61	2.54	6463.	0.0409	0.01	I * I
62	2.58	6463.	0.0282	0.01	I * I
63	2.63	6461.	0.0022	0.00	I * I
64	2.67	6459.	-0.0636	-.02	I ** I
65	2.71	6457.	-0.1249	-.03	I ** I
66	2.75	6455.	-0.1784	-.05	I ** I
67	2.79	6455.	-0.2184	-.06	I *** I
68	2.83	6454.	-0.1996	-.06	I *** I
69	2.88	6452.	-0.1529	-.04	I ** I
70	2.92	6452.	-0.0722	-.02	I ** I
71	2.96	6449.	0.0135	0.00	I * I
72	3.00	6443.	0.0469	0.01	I * I
73	3.04	6440.	-0.0413	-.01	I ** I
74	3.08	6439.	-0.1484	-.04	I ** I
75	3.13	6433.	-0.2309	-.06	I *** I

FOR THIS AUTOCORRELATION: MEAN = 2.3553 AND VARIANCE = 3.5933

DENVER, CO

YR=74 STATE=06 AREA=0580 SITE=001 AGCY=F PROJ=01 POLL=0000300 METH=11 UNT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 7952

BASIC STATISTICS-----

N = 7952

MEAN = 1.91322899

2ND MOMENT (VAR.) = 5.57618904 STANDARD DEV. = 2.36139584

COEFFICIENT OF VARIATION = 1.23424625

3RD MOMENT = 41.9626617 COEF. OF SKEWNESS = 3.18681908

4TH MOMENT = 625.271973 COEF. OF KURTOSIS = 20.1091614

LOWEST VALUE = 0.0

HIGHEST VALUE = 30.0000000

SECOND HIGHEST VALUE = 25.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 2133

```
   INTERVAL      NUMBER   PERCENT  +---------------------HISTOGRAM---------------------+ CUM. FREQ.  PLOT POS.
 -0.50-  0.50     2133   26.823%  I***********C**************                         I  26.823%     26.820%
  0.50-  1.50     2168   27.264%  I*************************C                         I  54.087%     54.080%
  1.50-  2.50     1646   20.699%  I*********************              C              I  74.786%     74.777%
  2.50-  3.50      868   10.915%  I***********                                  C     I  85.702%     85.691%
  3.50-  4.50      410    5.156%  I*****                                          C   I  90.858%     90.846%
  4.50-  5.50      237    2.980%  I***                                              C I  93.838%     93.826%
  5.50-  6.50      146    1.836%  I**                                               C I  95.674%     95.662%
  6.50-  7.50       90    1.132%  I*                                                C I  96.806%     96.794%
  7.50-  8.50       63    0.792%  I*                                                 C I 97.598%     97.586%
  8.50-  9.50       49    0.616%  I*                                                 C I 98.214%     98.202%
  9.50- 10.50       39    0.490%  I                                                  C I 98.705%     98.692%
```

10.50- 11.50	21	0.264%	I	C I	98.969%	98.956%
11.50- 12.50	21	0.264%	I	CI	99.233%	99.220%
12.50- 13.50	15	0.189%	I	CI	99.421%	99.409%
13.50- 14.50	11	0.138%	I	CI	99.560%	99.547%
14.50- 15.50	6	0.075%	I	CI	99.635%	99.623%
15.50- 16.50	8	0.101%	I	CI	99.736%	99.723%
16.50- 17.50	4	0.050%	I	CI	99.786%	99.774%
17.50- 18.50	2	0.025%	I	CI	99.811%	99.799%
18.50- 19.50	3	0.038%	I	CI	99.849%	99.836%
19.50- 20.50	3	0.038%	I	CI	99.887%	99.874%
20.50- 21.50	2	0.025%	I	CI	99.912%	99.899%
21.50- 22.50	1	0.013%	I	CI	99.924%	99.912%
22.50- 23.50	2	0.025%	I	CI	99.950%	99.937%
23.50- 24.50	2	0.025%	I	CI	99.975%	99.962%
24.50- 25.50	1	0.013%	I	CI	99.987%	99.975%
25.50- 26.50	0	0.0 %	I	CI	99.987%	99.975%
26.50- 27.50	0	0.0 %	I	CI	99.987%	99.975%
27.50- 28.50	0	0.0 %	I	CI	99.987%	99.975%
28.50- 29.50	0	0.0 %	I	CI	99.987%	99.975%
29.50- 30.50	1	0.013%	I	CI	100.000%	99.987%
30.50- 31.50	0	0.0 %	I	CI	100.000%	99.987%
31.50- 32.50	0	0.0 %	I	CI	100.000%	99.987%
32.50- 33.50	0	0.0 %	I	CI	100.000%	99.987%
33.50- 34.50	0	0.0 %	I	CI	100.000%	99.987%
34.50- 35.50	0	0.0 %	I	CI	100.000%	99.987%
35.50- 36.50	0	0.0 %	I	CI	100.000%	99.987%
36.50- 37.50	0	0.0 %	I	CI	100.000%	99.987%
37.50- 38.50	0	0.0 %	I	CI	100.000%	99.987%
38.50- 39.50	0	0.0 %	I	CI	100.000%	99.987%
39.50- 40.50	0	0.0 %	I	CI	100.000%	99.987%
40.50- 41.50	0	0.0 %	I	CI	100.000%	99.987%
41.50- 42.50	0	0.0 %	I	CI	100.000%	99.987%
42.50- 43.50	0	0.0 %	I	CI	100.000%	99.987%
43.50- 44.50	0	0.0 %	I	CI	100.000%	99.987%
44.50- 45.50	0	0.0 %	I	CI	100.000%	99.987%
45.50- 46.50	0	0.0 %	I	CI	100.000%	99.987%
46.50- 47.50	0	0.0 %	I	CI	100.000%	99.987%
47.50- 48.50	0	0.0 %	I	CI	100.000%	99.987%
48.50- 49.50	0	0.0 %	I	CI	100.000%	99.987%
49.50- 50.50	0	0.0 %	I	CI	100.000%	99.987%
50.50- 51.50	0	0.0 %	I	CI	100.000%	99.987%
51.50- 52.50	0	0.0 %	I	CI	100.000%	99.987%
52.50- 53.50	0	0.0 %	I	CI	100.000%	99.987%
53.50- 54.50	0	0.0 %	I	CI	100.000%	99.987%
54.50- 55.50	0	0.0 %	I	CI	100.000%	99.987%
55.50- 56.50	0	0.0 %	I	CI	100.000%	99.987%
56.50- 57.50	0	0.0 %	I	CI	100.000%	99.987%
57.50- 58.50	0	0.0 %	I	CI	100.000%	99.987%
58.50- 59.50	0	0.0 %	I	CI	100.000%	99.987%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	7709.	4.5073	0.81	**********
2	0.08	7660.	3.1360	0.56	**********
3	0.13	7636.	2.0867	0.37	********
4	0.17	7612.	1.3782	0.25	*****
5	0.21	7590.	0.9382	0.17	****
6	0.25	7569.	0.7354	0.13	***
7	0.29	7549.	0.7207	0.13	***
8	0.33	7531.	0.9262	0.17	****
9	0.38	7515.	1.1938	0.21	*****
10	0.42	7507.	1.3198	0.24	*****
11	0.46	7498.	1.3608	0.24	*****
12	0.50	7492.	1.3327	0.24	*****
13	0.54	7480.	1.2448	0.22	*****
14	0.58	7467.	1.0295	0.18	****
15	0.63	7455.	0.7743	0.14	***
16	0.67	7445.	0.4857	0.09	**
17	0.71	7438.	0.2641	0.05	*
18	0.75	7431.	0.1907	0.03	*
19	0.79	7420.	0.2661	0.05	*
20	0.83	7412.	0.4693	0.08	**
21	0.88	7408.	0.8352	0.15	***
22	0.92	7409.	1.3628	0.24	*****
23	0.96	7405.	1.9062	0.34	*******
24	1.00	7488.	2.1387	0.38	********
25	1.04	7409.	1.9031	0.34	*******
26	1.08	7389.	1.3543	0.24	*****
27	1.13	7373.	0.8258	0.15	***
28	1.17	7363.	0.3859	0.07	**
29	1.21	7356.	0.0717	0.01	*
30	1.25	7355.	-0.0916	-.02	**
31	1.29	7348.	-0.1023	-.02	**
32	1.33	7339.	0.0662	0.01	*
33	1.38	7334.	0.3322	0.06	**
34	1.42	7330.	0.5349	0.10	**
35	1.46	7327.	0.6935	0.12	***
36	1.50	7324.	0.7937	0.14	***
37	1.54	7322.	0.7502	0.13	***
38	1.58	7315.	0.5575	0.10	**
39	1.63	7308.	0.3168	0.06	**
40	1.67	7305.	0.0722	0.01	*
41	1.71	7303.	-0.1004	-.02	**
42	1.75	7303.	-0.0904	-.02	**
43	1.79	7300.	0.0085	0.00	*
44	1.83	7296.	0.1821	0.03	*
45	1.88	7300.	0.4604	0.08	**
46	1.92	7302.	0.8371	0.15	****
47	1.96	7312.	1.2269	0.22	*****
48	2.00	7386.	1.3929	0.25	*****
49	2.04	7313.	1.2620	0.23	*****
50	2.08	7306.	0.8502	0.15	****

51	2.13	7296.	0.4018	0.07	I ** I
52	2.17	7291.	0.0795	0.01	I * I
53	2.21	7285.	-0.1107	-.02	I ** I
54	2.25	7283.	-0.2219	-.04	I ** I
55	2.29	7275.	-0.1796	-.03	I ** I
56	2.33	7270.	-0.0124	-.00	I ** I
57	2.38	7265.	0.1678	0.03	I * I
58	2.42	7263.	0.3002	0.05	I ** I
59	2.46	7261.	0.3861	0.07	I ** I
60	2.50	7262.	0.4333	0.08	I ** I
61	2.54	7260.	0.4402	0.08	I ** I
62	2.58	7258.	0.3633	0.07	I ** I
63	2.63	7253.	0.1413	0.03	I * I
64	2.67	7255.	-0.0948	-.02	I ** I
65	2.71	7255.	-0.2752	-.05	I ** I
66	2.75	7256.	-0.2509	-.04	I ** I
67	2.79	7253.	-0.1310	-.02	I ** I
68	2.83	7252.	0.0282	0.01	I * I
69	2.88	7253.	0.2852	0.05	I ** I
70	2.92	7253.	0.6397	0.11	I *** I
71	2.96	7254.	0.9649	0.17	I **** I
72	3.00	7337.	1.1182	0.20	I ***** I
73	3.04	7263.	0.9341	0.17	I **** I
74	3.08	7257.	0.6012	0.11	I *** I
75	3.13	7248.	0.2880	0.05	I ** I

+---+---+---+---+---+---+---+---+---+---+

FOR THIS AUTOCORRELATION: MEAN = 1.9132 AND VARIANCE = 5.5772

ALEXANDRIA, VA

YR=74 STATE=48 AREA=0080 SITE=009 AGCY=H PROJ=01 POLL=0000200 METH=11 UNT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 8606

BASIC STATISTICS-----

N = 8606

MEAN = 1.63871670

2ND MOMENT (VAR.) = 2.93676567 STANDARD DEV. = 1.71369934

COEFFICIENT OF VARIATION = 1.04575634

3RD MOMENT = 16.7698364 COEF. OF SKEWNESS = 3.33215332

4TH MOMENT = 172.063431 COEF. OF KURTOSIS = 19.9503174

LOWEST VALUE = 0.0

HIGHEST VALUE = 21.0000000

SECOND HIGHEST VALUE = 20.5000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000 0 NO. OF ZERO VALUES = 488

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	488	5.670%	I**C*** I	5.670%	5.670%
0.50- 1.50	4511	52.417%	I****************************C********************I	58.087%	58.081%
1.50- 2.50	1996	23.193%	I************************* C I	81.280%	81.271%
2.50- 3.50	690	8.018%	I******** C I	89.298%	89.288%
3.50- 4.50	341	3.962%	I**** C I	93.260%	93.250%
4.50- 5.50	216	2.510%	I*** C I	95.770%	95.759%
5.50- 6.50	108	1.255%	I* C I	97.025%	97.014%
6.50- 7.50	74	0.860%	I* C I	97.885%	97.874%
7.50- 8.50	71	0.825%	I* C I	98.710%	98.699%
8.50- 9.50	37	0.430%	I CI	99.140%	99.129%
9.50- 10.50	33	0.383%	I CI	99.524%	99.512%

10.50- 11.50	13	0.151%	I	CI 99.675%	99.663%
11.50- 12.50	7	0.081%	I	CI 99.756%	99.744%
12.50- 13.50	7	0.081%	I	CI 99.837%	99.826%
13.50- 14.50	5	0.058%	I	CI 99.895%	99.884%
14.50- 15.50	0	0.0 %	I	CI 99.895%	99.884%
15.50- 16.50	2	0.023%	I	CI 99.919%	99.907%
16.50- 17.50	2	0.023%	I	CI 99.942%	99.930%
17.50- 18.50	3	0.035%	I	CI 99.977%	99.965%
18.50- 19.50	0	0.0 %	I	CI 99.977%	99.965%
19.50- 20.50	0	0.0 %	I	CI 99.977%	99.965%
20.50- 21.50	2	0.023%	I	CI 100.000%	99.988%
21.50- 22.50	0	0.0 %	I	CI 100.000%	99.988%
22.50- 23.50	0	0.0 %	I	CI 100.000%	99.988%
23.50- 24.50	0	0.0 %	I	CI 100.000%	99.988%
24.50- 25.50	0	0.0 %	I	CI 100.000%	99.988%
25.50- 26.50	0	0.0 %	I	CI 100.000%	99.988%
26.50- 27.50	0	0.0 %	I	CI 100.000%	99.988%
27.50- 28.50	0	0.0 %	I	CI 100.000%	99.988%
28.50- 29.50	0	0.0 %	I	CI 100.000%	99.988%
29.50- 30.50	0	0.0 %	I	CI 100.000%	99.988%
30.50- 31.50	0	0.0 %	I	CI 100.000%	99.988%
31.50- 32.50	0	0.0 %	I	CI 100.000%	99.988%
32.50- 33.50	0	0.0 %	I	CI 100.000%	99.988%
33.50- 34.50	0	0.0 %	I	CI 100.000%	99.988%
34.50- 35.50	0	0.0 %	I	CI 100.000%	99.988%
35.50- 36.50	0	0.0 %	I	CI 100.000%	99.988%
36.50- 37.50	0	0.0 %	I	CI 100.000%	99.988%
37.50- 38.50	0	0.0 %	I	CI 100.000%	99.988%
38.50- 39.50	0	0.0 %	I	CI 100.000%	99.988%
39.50- 40.50	0	0.0 %	I	CI 100.000%	99.988%
40.50- 41.50	0	0.0 %	I	CI 100.000%	99.988%
41.50- 42.50	0	0.0 %	I	CI 100.000%	99.988%
42.50- 43.50	0	0.0 %	I	CI 100.000%	99.988%
43.50- 44.50	0	0.0 %	I	CI 100.000%	99.988%
44.50- 45.50	0	0.0 %	I	CI 100.000%	99.988%
45.50- 46.50	0	0.0 %	I	CI 100.000%	99.988%
46.50- 47.50	0	0.0 %	I	CI 100.000%	99.988%
47.50- 48.50	0	0.0 %	I	CI 100.000%	99.988%
48.50- 49.50	0	0.0 %	I	CI 100.000%	99.988%
49.50- 50.50	0	0.0 %	I	CI 100.000%	99.988%
50.50- 51.50	0	0.0 %	I	CI 100.000%	99.988%
51.50- 52.50	0	0.0 %	I	CI 100.000%	99.988%
52.50- 53.50	0	0.0 %	I	CI 100.000%	99.988%
53.50- 54.50	0	0.0 %	I	CI 100.000%	99.988%
54.50- 55.50	0	0.0 %	I	CI 100.000%	99.988%
55.50- 56.50	0	0.0 %	I	CI 100.000%	99.988%
56.50- 57.50	0	0.0 %	I	CI 100.000%	99.988%
57.50- 58.50	0	0.0 %	I	CI 100.000%	99.988%
58.50- 59.50	0	0.0 %	I	CI 100.000%	99.988%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -8 -6 -4 -2 0 +2 +4 +6 +8 +1.0
1	0.04	8581.	2.3496	0.80	I ********** I
2	0.08	8564.	1.7579	0.60	I ********** I
3	0.13	8551.	1.3100	0.45	I ********* I
4	0.17	8538.	1.0389	0.35	I ******** I
5	0.21	8528.	0.8679	0.30	I ****** I
6	0.25	8519.	0.7348	0.25	I ****** I
7	0.29	8512.	0.6604	0.22	I ***** I
8	0.33	8506.	0.6391	0.22	I ***** I
9	0.38	8500.	0.6180	0.21	I ***** I
10	0.42	8494.	0.5595	0.19	I **** I
11	0.46	8488.	0.4934	0.17	I **** I
12	0.50	8482.	0.4314	0.15	I *** I
13	0.54	8478.	0.3820	0.13	I *** I
14	0.58	8474.	0.3525	0.12	I *** I
15	0.63	8470.	0.2930	0.10	I ** I
16	0.67	8468.	0.2136	0.07	I ** I
17	0.71	8466.	0.1478	0.05	I ** I
18	0.75	8466.	0.1204	0.04	I * I
19	0.79	8465.	0.1243	0.04	I * I
20	0.83	8463.	0.1633	0.06	I ** I
21	0.88	8461.	0.2593	0.09	I ** I
22	0.92	8459.	0.4171	0.14	I *** I
23	0.96	8456.	0.6252	0.21	I ***** I
24	1.00	8453.	0.7115	0.24	I ***** I
25	1.04	8451.	0.6258	0.21	I ***** I
26	1.08	8448.	0.4366	0.15	I *** I
27	1.13	8445.	0.2922	0.10	I ** I
28	1.17	8442.	0.2121	0.07	I ** I
29	1.21	8440.	0.1776	0.06	I ** I
30	1.25	8439.	0.1840	0.06	I ** I
31	1.29	8438.	0.2234	0.08	I ** I
32	1.33	8438.	0.2830	0.10	I ** I
33	1.38	8438.	0.3141	0.11	I *** I
34	1.42	8437.	0.3312	0.11	I *** I
35	1.46	8436.	0.3426	0.12	I *** I
36	1.50	8435.	0.3334	0.11	I *** I
37	1.54	8434.	0.3139	0.11	I *** I
38	1.58	8433.	0.2779	0.09	I ** I
39	1.63	8432.	0.2295	0.08	I ** I
40	1.67	8431.	0.1822	0.06	I ** I
41	1.71	8430.	0.1392	0.05	I * I
42	1.75	8429.	0.1062	0.04	I * I
43	1.79	8428.	0.1024	0.03	I * I
44	1.83	8428.	0.1215	0.04	I * I
45	1.88	8428.	0.1830	0.06	I ** I
46	1.92	8429.	0.3073	0.10	I *** I
47	1.96	8432.	0.4596	0.16	I **** I
48	2.00	8431.	0.5535	0.19	I **** I
49	2.04	8429.	0.4956	0.17	I **** I
50	2.08	8426.	0.3719	0.13	I *** I

51	2.13	8424.	0.2831	0.10	I ** I
52	2.17	8421.	0.2469	0.08	I ** I
53	2.21	8419.	0.2492	0.08	I ** I
54	2.25	8418.	0.2585	0.09	I ** I
55	2.29	8417.	0.2681	0.09	I ** I
56	2.33	8416.	0.2818	0.10	I ** I
57	2.38	8415.	0.3083	0.10	I *** I
58	2.42	8414.	0.3314	0.11	I *** I
59	2.46	8412.	0.2933	0.10	I ** I
60	2.50	8411.	0.2580	0.09	I ** I
61	2.54	8410.	0.2300	0.08	I ** I
62	2.58	8409.	0.2118	0.07	I ** I
63	2.63	8408.	0.1807	0.06	I ** I
64	2.67	8408.	0.1216	0.04	I * I
65	2.71	8409.	0.0641	0.02	I * I
66	2.75	8410.	0.0358	0.01	I * I
67	2.79	8411.	0.0258	0.01	I * I
68	2.83	8411.	0.0170	0.01	I * I
69	2.88	8411.	0.0406	0.01	I * I
70	2.92	8409.	0.1077	0.04	I * I
71	2.96	8406.	0.2075	0.07	I ** I
72	3.00	8404.	0.2579	0.09	I ** I
73	3.04	8402.	0.2034	0.07	I ** I
74	3.08	8400.	0.1123	0.04	I * I
75	3.13	8397.	0.0455	0.02	I * I

FOR THIS AUTOCORRELATION: MEAN = 1.6387 AND VARIANCE = 2.9387

NORFORK, VA

YR=74 STATE=48 AREA=2140 SITE=010 AGCY=F PROJ=01 POLL=0000150 METH=11 UNT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 8241

BASIC STATISTICS-----

N = 8241

MEAN = 1.37737465

2ND MOMENT (VAR.) = 1.12630749 STANDARD DEV. = 1.06127644

COEFFICIENT OF VARIATION = 0.77050668

3RD MOMENT = 5.94860172 COEF. OF SKEWNESS = 4.97655964

4TH MOMENT = 65.2872925 COEF. OF KURTOSIS = 51.4653320

LOWEST VALUE = 0.0

HIGHEST VALUE = 19.0000000

SECOND HIGHEST VALUE = 17.5000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 600

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	600	7.281%	I***C*** I	7.281%	7.280%
0.50- 1.50	3403	41.294%	I************************C****************** I	48.574%	48.568%
1.50- 2.50	3600	43.684%	I** C I	92.258%	92.247%
2.50- 3.50	368	4.465%	I**** C I	96.724%	96.712%
3.50- 4.50	127	1.541%	I** C I	98.265%	98.253%
4.50- 5.50	50	0.607%	I* C I	98.871%	98.859%
5.50- 6.50	36	0.437%	I CI	99.308%	99.296%
6.50- 7.50	12	0.146%	I CI	99.454%	99.442%
7.50- 8.50	14	0.170%	I CI	99.624%	99.612%
8.50- 9.50	10	0.121%	I CI	99.745%	99.733%
9.50- 10.50	5	0.061%	I CI	99.806%	99.794%

10.50- 11.50	4	0.049%	I	CI 99.854%	99.842%
11.50- 12.50	3	0.036%	I	CI 99.891%	99.879%
12.50- 13.50	1	0.012%	I	CI 99.903%	99.891%
13.50- 14.50	5	0.061%	I	CI 99.963%	99.951%
14.50- 15.50	0	0.0 %	I	CI 99.963%	99.951%
15.50- 16.50	0	0.0 %	I	CI 99.963%	99.951%
16.50- 17.50	1	0.012%	I	CI 99.976%	99.963%
17.50- 18.50	1	0.012%	I	CI 99.988%	99.976%
18.50- 19.50	1	0.012%	I	CI 100.000%	99.988%
19.50- 20.50	0	0.0 %	I	CI 100.000%	99.988%
20.50- 21.50	0	0.0 %	I	CI 100.000%	99.988%
21.50- 22.50	0	0.0 %	I	CI 100.000%	99.988%
22.50- 23.50	0	0.0 %	I	CI 100.000%	99.988%
23.50- 24.50	0	0.0 %	I	CI 100.000%	99.988%
24.50- 25.50	0	0.0 %	I	CI 100.000%	99.988%
25.50- 26.50	0	0.0 %	I	CI 100.000%	99.988%
26.50- 27.50	0	0.0 %	I	CI 100.000%	99.988%
27.50- 28.50	0	0.0 %	I	CI 100.000%	99.988%
28.50- 29.50	0	0.0 %	I	CI 100.000%	99.988%
29.50- 30.50	0	0.0 %	I	CI 100.000%	99.988%
30.50- 31.50	0	0.0 %	I	CI 100.000%	99.988%
31.50- 32.50	0	0.0 %	I	CI 100.000%	99.988%
32.50- 33.50	0	0.0 %	I	CI 100.000%	99.988%
33.50- 34.50	0	0.0 %	I	CI 100.000%	99.988%
34.50- 35.50	0	0.0 %	I	CI 100.000%	99.988%
35.50- 36.50	0	0.0 %	I	CI 100.000%	99.988%
36.50- 37.50	0	0.0 %	I	CI 100.000%	99.988%
37.50- 38.50	0	0.0 %	I	CI 100.000%	99.988%
38.50- 39.50	0	0.0 %	I	CI 100.000%	99.988%
39.50- 40.50	0	0.0 %	I	CI 100.000%	99.988%
40.50- 41.50	0	0.0 %	I	CI 100.000%	99.988%
41.50- 42.50	0	0.0 %	I	CI 100.000%	99.988%
42.50- 43.50	0	0.0 %	I	CI 100.000%	99.988%
43.50- 44.50	0	0.0 %	I	CI 100.000%	99.988%
44.50- 45.50	0	0.0 %	I	CI 100.000%	99.988%
45.50- 46.50	0	0.0 %	I	CI 100.000%	99.988%
46.50- 47.50	0	0.0 %	I	CI 100.000%	99.988%
47.50- 48.50	0	0.0 %	I	CI 100.000%	99.988%
48.50- 49.50	0	0.0 %	I	CI 100.000%	99.988%
49.50- 50.50	0	0.0 %	I	CI 100.000%	99.988%
50.50- 51.50	0	0.0 %	I	CI 100.000%	99.988%
51.50- 52.50	0	0.0 %	I	CI 100.000%	99.988%
52.50- 53.50	0	0.0 %	I	CI 100.000%	99.988%
53.50- 54.50	0	0.0 %	I	CI 100.000%	99.988%
54.50- 55.50	0	0.0 %	I	CI 100.000%	99.988%
55.50- 56.50	0	0.0 %	I	CI 100.000%	99.988%
56.50- 57.50	0	0.0 %	I	CI 100.000%	99.988%
57.50- 58.50	0	0.0 %	I	CI 100.000%	99.988%
58.50- 59.50	0	0.0 %	I	CI 100.000%	99.988%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	8225.	0.9191	0.82	**********
2	0.08	8211.	0.7155	0.63	**********
3	0.13	8199.	0.5510	0.49	**********
4	0.17	8188.	0.4437	0.39	********
5	0.21	8177.	0.3700	0.33	*******
6	0.25	8166.	0.3246	0.29	******
7	0.29	8157.	0.2933	0.26	******
8	0.33	8149.	0.2657	0.24	*****
9	0.38	8140.	0.2290	0.20	*****
10	0.42	8131.	0.1937	0.17	****
11	0.46	8121.	0.1566	0.14	***
12	0.50	8110.	0.1297	0.12	***
13	0.54	8103.	0.1120	0.10	**
14	0.58	8096.	0.1121	0.10	**
15	0.63	8089.	0.1105	0.10	**
16	0.67	8082.	0.0989	0.09	**
17	0.71	8075.	0.0846	0.08	**
18	0.75	8069.	0.0627	0.06	**
19	0.79	8063.	0.0525	0.05	*
20	0.83	8057.	0.0482	0.04	*
21	0.88	8051.	0.0638	0.06	**
22	0.92	8045.	0.0960	0.09	**
23	0.96	8039.	0.1284	0.11	***
24	1.00	8032.	0.1587	0.14	***
25	1.04	8025.	0.1462	0.13	***
26	1.08	8018.	0.1144	0.10	***
27	1.13	8011.	0.0845	0.07	**
28	1.17	8004.	0.0636	0.06	**
29	1.21	7998.	0.0657	0.06	**
30	1.25	7992.	0.0794	0.07	**
31	1.29	7987.	0.0923	0.08	**
32	1.33	7982.	0.1018	0.09	**
33	1.38	7977.	0.0933	0.08	**
34	1.42	7972.	0.0732	0.06	**
35	1.46	7967.	0.0473	0.04	*
36	1.50	7961.	0.0284	0.03	*
37	1.54	7956.	0.0199	0.02	*
38	1.58	7951.	0.0228	0.02	*
39	1.63	7946.	0.0182	0.02	*
40	1.67	7941.	0.0031	0.00	*
41	1.71	7936.	-0.0205	-.02	**
42	1.75	7932.	-0.0402	-.04	**
43	1.79	7928.	-0.0508	-.05	**
44	1.83	7925.	-0.0522	-.05	**
45	1.88	7922.	-0.0390	-.03	**
46	1.92	7920.	-0.0129	-.01	**
47	1.96	7918.	0.0162	0.01	*
48	2.00	7916.	0.0329	0.03	*
49	2.04	7914.	0.0183	0.02	*
50	2.08	7912.	0.0003	0.00	*

51	2.13	7910.	-0.0206	-.02	I ** I
52	2.17	7908.	-0.0307	-.03	I ** I
53	2.21	7906.	-0.0230	-.02	I ** I
54	2.25	7904.	-0.0103	-.01	I ** I
55	2.29	7902.	0.0098	0.01	I * I
56	2.33	7900.	0.0240	0.02	I * I
57	2.38	7898.	0.0364	0.03	I * I
58	2.42	7897.	0.0424	0.04	I * I
59	2.46	7894.	0.0442	0.04	I * I
60	2.50	7891.	0.0480	0.04	I * I
61	2.54	7888.	0.0596	0.05	I ** I
62	2.58	7885.	0.0715	0.06	I ** I
63	2.63	7882.	0.0701	0.06	I ** I
64	2.67	7879.	0.0554	0.05	I * I
65	2.71	7876.	0.0293	0.03	I * I
66	2.75	7873.	0.0073	0.01	I * I
67	2.79	7870.	-0.0021	-.00	I ** I
68	2.83	7867.	0.0008	0.00	I * I
69	2.88	7864.	0.0150	0.01	I * I
70	2.92	7860.	0.0424	0.04	I * I
71	2.96	7857.	0.0622	0.06	I ** I
72	3.00	7855.	0.0746	0.07	I ** I
73	3.04	7853.	0.0496	0.04	I * I
74	3.08	7851.	0.0231	0.02	I * I
75	3.13	7849.	-0.0098	-.01	I ** I

FOR THIS AUTOCORRELATION; MEAN = 1.3774 AND VARIANCE = 1.1271

BARSTOW, CA

YR=74 STATE=05 AREA=0580 SITE=001 AGCY=I PROJ=01 POLL=9999999 METH=11 UNT=7

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 6992

BASIC STATISTICS-----

N = 6992

MEAN = 0.44150454

2ND MOMENT (VAR.) = 0.738471210 STANDARD DEV. = 0.859343469

COEFFICIENT OF VARIATION = 1.94639778

3RD MOMENT = 1.861683A5 COEF. OF SKEWNESS = 2.93363094

4TH MOMENT = 8.02122879 COEF. OF KURTOSIS = 14.7086821

LOWEST VALUE = 0.0

HIGHEST VALUE = 8.00000000

SECOND HIGHEST VALUE = 8.00000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000 = 0 NO. OF ZERO VALUES = 4900

INTERVAL	NUMBER	PERCENT	HISTOGRAM	CUM. FREQ.	PLOT POS.
-0.50- 0.50	4900	70.080%	I***

10.50- 11.50	0	0.0	%	I	CI 100.000%	99.986%
11.50- 12.50	0	0.0	%	I	CI 100.000%	99.986%
12.50- 13.50	0	0.0	%	I	CI 100.000%	99.986%
13.50- 14.50	0	0.0	%	I	CI 100.000%	99.986%
14.50- 15.50	0	0.0	%	I	CI 100.000%	99.986%
15.50- 16.50	0	0.0	%	I	CI 100.000%	99.986%
16.50- 17.50	0	0.0	%	I	CI 100.000%	99.986%
17.50- 18.50	0	0.0	%	I	CI 100.000%	99.986%
18.50- 19.50	0	0.0	%	I	CI 100.000%	99.986%
19.50- 20.50	0	0.0	%	I	CI 100.000%	99.986%
20.50- 21.50	0	0.0	%	I	CI 100.000%	99.986%
21.50- 22.50	0	0.0	%	I	CI 100.000%	99.986%
22.50- 23.50	0	0.0	%	I	CI 100.000%	99.986%
23.50- 24.50	0	0.0	%	I	CI 100.000%	99.986%
24.50- 25.50	0	0.0	%	I	CI 100.000%	99.986%
25.50- 26.50	0	0.0	%	I	CI 100.000%	99.986%
26.50- 27.50	0	0.0	%	I	CI 100.000%	99.986%
27.50- 28.50	0	0.0	%	I	CI 100.000%	99.986%
28.50- 29.50	0	0.0	%	I	CI 100.000%	99.986%
29.50- 30.50	0	0.0	%	I	CI 100.000%	99.986%
30.50- 31.50	0	0.0	%	I	CI 100.000%	99.986%
31.50- 32.50	0	0.0	%	I	CI 100.000%	99.986%
32.50- 33.50	0	0.0	%	I	CI 100.000%	99.986%
33.50- 34.50	0	0.0	%	I	CI 100.000%	99.986%
34.50- 35.50	0	0.0	%	I	CI 100.000%	99.986%
35.50- 36.50	0	0.0	%	I	CI 100.000%	99.986%
36.50- 37.50	0	0.0	%	I	CI 100.000%	99.986%
37.50- 38.50	0	0.0	%	I	CI 100.000%	99.986%
38.50- 39.50	0	0.0	%	I	CI 100.000%	99.986%
39.50- 40.50	0	0.0	%	I	CI 100.000%	99.986%
40.50- 41.50	0	0.0	%	I	CI 100.000%	99.986%
41.50- 42.50	0	0.0	%	I	CI 100.000%	99.986%
42.50- 43.50	0	0.0	%	I	CI 100.000%	99.986%
43.50- 44.50	0	0.0	%	I	CI 100.000%	99.986%
44.50- 45.50	0	0.0	%	I	CI 100.000%	99.986%
45.50- 46.50	0	0.0	%	I	CI 100.000%	99.986%
46.50- 47.50	0	0.0	%	I	CI 100.000%	99.986%
47.50- 48.50	0	0.0	%	I	CI 100.000%	99.986%
48.50- 49.50	0	0.0	%	I	CI 100.000%	99.986%
49.50- 50.50	0	0.0	%	I	CI 100.000%	99.986%
50.50- 51.50	0	0.0	%	I	CI 100.000%	99.986%
51.50- 52.50	0	0.0	%	I	CI 100.000%	99.986%
52.50- 53.50	0	0.0	%	I	CI 100.000%	99.986%
53.50- 54.50	0	0.0	%	I	CI 100.000%	99.986%
54.50- 55.50	0	0.0	%	I	CI 100.000%	99.986%
55.50- 56.50	0	0.0	%	I	CI 100.000%	99.986%
56.50- 57.50	0	0.0	%	I	CI 100.000%	99.986%
57.50- 58.50	0	0.0	%	I	CI 100.000%	99.986%
58.50- 59.50	0	0.0	%	I	CI 100.000%	99.986%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	6940.	0.5871	0.79	**********
2	0.08	6921.	0.4781	0.65	**********
3	0.13	6908.	0.3927	0.53	**********
4	0.17	6897.	0.3236	0.44	*********
5	0.21	6887.	0.2710	0.37	********
6	0.25	6879.	0.2380	0.32	*******
7	0.29	6871.	0.2215	0.30	******
8	0.33	6863.	0.2311	0.31	*******
9	0.38	6856.	0.2456	0.33	*******
10	0.42	6849.	0.2584	0.35	*******
11	0.46	6842.	0.2637	0.36	********
12	0.50	6835.	0.2648	0.36	********
13	0.54	6828.	0.2595	0.35	********
14	0.58	6821.	0.2431	0.33	*******
15	0.63	6814.	0.2208	0.30	******
16	0.67	6807.	0.2047	0.28	******
17	0.71	6800.	0.1927	0.26	******
18	0.75	6793.	0.1946	0.26	******
19	0.79	6786.	0.2131	0.29	******
20	0.83	6780.	0.2440	0.33	*******
21	0.88	6776.	0.2844	0.38	********
22	0.92	6769.	0.3283	0.44	*********
23	0.96	6764.	0.3703	0.50	**********
24	1.00	6759.	0.3884	0.53	**********
25	1.04	6755.	0.3590	0.49	**********
26	1.08	6748.	0.3069	0.42	*********
27	1.13	6742.	0.2689	0.36	********
28	1.17	6736.	0.2298	0.30	*******
29	1.21	6731.	0.1968	0.27	******
30	1.25	6726.	0.1707	0.23	*****
31	1.29	6721.	0.1592	0.22	*****
32	1.33	6716.	0.1620	0.22	*****
33	1.38	6711.	0.1701	0.23	*****
34	1.42	6706.	0.1873	0.25	******
35	1.46	6701.	0.1983	0.27	******
36	1.50	6697.	0.1958	0.27	******
37	1.54	6694.	0.1890	0.26	******
38	1.58	6690.	0.1750	0.24	*****
39	1.63	6686.	0.1576	0.21	*****
40	1.67	6682.	0.1387	0.19	****
41	1.71	6678.	0.1326	0.18	****
42	1.75	6674.	0.1389	0.19	****
43	1.79	6670.	0.1586	0.21	*****
44	1.83	6666.	0.1882	0.25	******
45	1.88	6662.	0.2149	0.29	******
46	1.92	6659.	0.2502	0.34	*******
47	1.96	6657.	0.2946	0.40	********
48	2.00	6656.	0.3134	0.42	*********
49	2.04	6648.	0.2896	0.39	********
50	2.08	6647.	0.2457	0.33	*******

51	2.13	6642.	0.2116	0.29	I ****** I
52	2.17	6642.	0.1852	0.25	I ***** I
53	2.21	6640.	0.1575	0.21	I ***** I
54	2.25	6638.	0.1421	0.19	I **** I
55	2.29	6637.	0.1358	0.18	I **** I
56	2.33	6636.	0.1461	0.20	I **** I
57	2.38	6635.	0.1500	0.22	I ***** I
58	2.42	6634.	0.1714	0.23	I ***** I
59	2.46	6633.	0.1703	0.24	I ***** I
60	2.50	6632.	0.1786	0.24	I ***** I
61	2.54	6630.	0.1710	0.23	I ***** I
62	2.58	6628.	0.1586	0.21	I ***** I
63	2.63	6627.	0.1413	0.19	I **** I
64	2.67	6625.	0.1247	0.17	I **** I
65	2.71	6623.	0.1140	0.15	I **** I
66	2.75	6622.	0.1233	0.17	I **** I
67	2.79	6621.	0.1456	0.19	I **** I
68	2.83	6619.	0.1640	0.22	I ***** I
69	2.88	6619.	0.1986	0.27	I ****** I
70	2.92	6615.	0.2348	0.32	I ******* I
71	2.96	6615.	0.2702	0.37	I ******** I
72	3.00	6612.	0.2886	0.39	I ******** I
73	3.04	6613.	0.2775	0.38	I ******** I
74	3.08	6612.	0.2437	0.33	I ******* I
75	3.13	6613.	0.2081	0.28	I ****** I

+---+---+---+---+---+---+---+---+---+---+

FOR THIS AUTOCORRELATION: MEAN = 0.4415 AND VARIANCE = 0.7389

APPENDIX D

BASIC STATISTICS, HISTOGRAMS, AND AUTOCORRELATIONS FOR THE LONGITUDINAL DATA GROUP

1966 CO WASHINGTON DC CAMP

NO. OF VALUES READ = 8100 NO OF VALID VALUES STORED = 7822

BASIC STATISTICS-----

N = 7822

MEAN = 3.31079006

2ND MOMENT (VAR.) = 6.11432362 STANDARD DEV. = 2.47271538

COEFFICIENT OF VARIATION = 0.74686563

3RD MOMENT = 49.3938751 COEF. OF SKEWNESS = 3.26701355

4TH MOMENT = 886.793213 COEF. OF KURTOSIS = 23.7205811

LOWEST VALUE = 0.0

HIGHEST VALUE = 38.0000000

SECOND HIGHEST VALUE = 33.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 161

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	161	2.058%	IC* I	2.058%	2.058%
0.50- 1.50	1055	13.488%	I*******C***** I	15.546%	15.544%
1.50- 2.50	2160	27.614%	I**********C****** I	43.160%	43.155%
2.50- 3.50	1784	22.807%	I********** C I	65.968%	65.959%
3.50- 4.50	1135	14.510%	I********** C I	80.478%	80.468%
4.50- 5.50	617	7.888%	I******** C I	88.366%	88.355%
5.50- 6.50	361	4.615%	I***** C I	92.981%	92.969%
6.50- 7.50	184	2.352%	I** C I	95.334%	95.321%
7.50- 8.50	111	1.419%	I* C I	96.753%	96.740%
8.50- 9.50	61	0.780%	I* C I	97.533%	97.520%
9.50- 10.50	47	0.601%	I* C I	98.133%	98.121%

10.50- 11.50	30	0.384%	I	C I 98.517%	98.504%
11.50- 12.50	22	0.281%	I	C I 98.798%	98.786%
12.50- 13.50	23	0.294%	I	CI 99.092%	99.080%
13.50- 14.50	8	0.102%	I	CI 99.194%	99.182%
14.50- 15.50	12	0.153%	I	CI 99.348%	99.335%
15.50- 16.50	14	0.179%	I	CI 99.527%	99.514%
16.50- 17.50	8	0.102%	I	CI 99.629%	99.616%
17.50- 18.50	4	0.051%	I	CI 99.680%	99.667%
18.50- 19.50	5	0.064%	I	CI 99.744%	99.731%
19.50- 20.50	3	0.038%	I	CI 99.783%	99.770%
20.50- 21.50	4	0.051%	I	CI 99.834%	99.821%
21.50- 22.50	5	0.064%	I	CI 99.898%	99.885%
22.50- 23.50	2	0.026%	I	CI 99.923%	99.910%
23.50- 24.50	2	0.026%	I	CI 99.949%	99.936%
24.50- 25.50	1	0.013%	I	CI 99.961%	99.949%
25.50- 26.50	0	0.0 %	I	CI 99.961%	99.949%
26.50- 27.50	0	0.0 %	I	CI 99.961%	99.949%
27.50- 28.50	1	0.013%	I	CI 99.974%	99.961%
28.50- 29.50	0	0.0 %	I	CI 99.974%	99.961%
29.50- 30.50	0	0.0 %	I	CI 99.974%	99.961%
30.50- 31.50	0	0.0 %	I	CI 99.974%	99.961%
31.50- 32.50	0	0.0 %	I	CI 99.974%	99.961%
32.50- 33.50	1	0.013%	I	CI 99.987%	99.974%
33.50- 34.50	0	0.0 %	I	CI 99.987%	99.974%
34.50- 35.50	0	0.0 %	I	CI 99.987%	99.974%
35.50- 36.50	0	0.0 %	I	CI 99.987%	99.974%
36.50- 37.50	0	0.0 %	I	CI 99.987%	99.974%
37.50- 38.50	1	0.013%	I	CI 100.000%	99.987%
38.50- 39.50	0	0.0 %	I	CI 100.000%	99.987%
39.50- 40.50	0	0.0 %	I	CI 100.000%	99.987%
40.50- 41.50	0	0.0 %	I	CI 100.000%	99.987%
41.50- 42.50	0	0.0 %	I	CI 100.000%	99.987%
42.50- 43.50	0	0.0 %	I	CI 100.000%	99.987%
43.50- 44.50	0	0.0 %	I	CI 100.000%	99.987%
44.50- 45.50	0	0.0 %	I	CI 100.000%	99.987%
45.50- 46.50	0	0.0 %	I	CI 100.000%	99.987%
46.50- 47.50	0	0.0 %	I	CI 100.000%	99.987%
47.50- 48.50	0	0.0 %	I	CI 100.000%	99.987%
48.50- 49.50	0	0.0 %	I	CI 100.000%	99.987%
49.50- 50.50	0	0.0 %	I	CI 100.000%	99.987%
50.50- 51.50	0	0.0 %	I	CI 100.000%	99.987%
51.50- 52.50	0	0.0 %	I	CI 100.000%	99.987%
52.50- 53.50	0	0.0 %	I	CI 100.000%	99.987%
53.50- 54.50	0	0.0 %	I	CI 100.000%	99.987%
54.50- 55.50	0	0.0 %	I	CI 100.000%	99.987%
55.50- 56.50	0	0.0 %	I	CI 100.000%	99.987%
56.50- 57.50	0	0.0 %	I	CI 100.000%	99.987%
57.50- 58.50	0	0.0 %	I	CI 100.000%	99.987%
58.50- 59.50	0	0.0 %	I	CI 100.000%	99.987%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	7722.	5.4654	0.89	**********
2	0.08	7678.	4.5553	0.74	**********
3	0.13	7657.	3.7800	0.62	**********
4	0.17	7642.	3.2270	0.53	**********
5	0.21	7629.	2.8460	0.47	**********
6	0.25	7615.	2.6195	0.43	*********
7	0.29	7603.	2.5120	0.41	*********
8	0.33	7593.	2.4986	0.41	*********
9	0.38	7584.	2.5053	0.41	*********
10	0.42	7575.	2.4450	0.40	********
11	0.46	7565.	2.3233	0.38	********
12	0.50	7556.	2.2224	0.36	********
13	0.54	7548.	2.1807	0.36	********
14	0.58	7542.	2.1052	0.34	*******
15	0.63	7540.	1.9518	0.32	*******
16	0.67	7538.	1.7251	0.28	******
17	0.71	7535.	1.5172	0.25	*****
18	0.75	7532.	1.3978	0.23	*****
19	0.79	7535.	1.4190	0.23	*****
20	0.83	7536.	1.5486	0.25	******
21	0.88	7537.	1.7435	0.29	******
22	0.92	7540.	2.0072	0.33	*******
23	0.96	7546.	2.3118	0.38	********
24	1.00	7545.	2.4319	0.40	********
25	1.04	7545.	2.2623	0.37	********
26	1.08	7538.	1.9787	0.32	*******
27	1.13	7538.	1.7334	0.28	******
28	1.17	7538.	1.5324	0.25	******
29	1.21	7537.	1.3973	0.23	*****
30	1.25	7534.	1.3329	0.22	*****
31	1.29	7528.	1.3146	0.21	*****
32	1.33	7526.	1.3674	0.22	*****
33	1.38	7522.	1.4064	0.23	*****
34	1.42	7520.	1.4009	0.23	*****
35	1.46	7516.	1.3668	0.22	*****
36	1.50	7512.	1.3536	0.22	*****
37	1.54	7509.	1.3479	0.22	*****
38	1.58	7507.	1.2956	0.21	*****
39	1.63	7506.	1.1836	0.19	****
40	1.67	7505.	1.0159	0.17	****
41	1.71	7503.	0.8340	0.14	***
42	1.75	7505.	0.7112	0.12	***
43	1.79	7506.	0.7085	0.12	***
44	1.83	7504.	0.8145	0.13	***
45	1.88	7505.	0.9791	0.16	****
46	1.92	7505.	1.1790	0.19	****
47	1.96	7505.	1.3842	0.23	*****
48	2.00	7507.	1.4656	0.24	*****
49	2.04	7507.	1.3469	0.22	*****
50	2.08	7505.	1.1126	0.18	****

51	2.13	7505.	0.8998	0.15	I *** I
52	2.17	7504.	0.7661	0.13	I *** I
53	2.21	7502.	0.7038	0.12	I *** I
54	2.25	7500.	0.7068	0.12	I *** I
55	2.29	7499.	0.7857	0.13	I *** I
56	2.33	7496.	0.9251	0.15	I **** I
57	2.38	7493.	1.0740	0.18	I **** I
58	2.42	7492.	1.1906	0.19	I **** I
59	2.46	7490.	1.2280	0.20	I ***** I
60	2.50	7489.	1.2228	0.20	I **** I
61	2.54	7488.	1.1975	0.20	I **** I
62	2.58	7488.	1.1535	0.19	I **** I
63	2.63	7489.	1.0648	0.17	I **** I
64	2.67	7489.	0.9407	0.15	I **** I
65	2.71	7489.	0.7927	0.13	I *** I
66	2.75	7490.	0.7069	0.12	I *** I
67	2.79	7489.	0.6884	0.11	I *** I
68	2.83	7489.	0.7090	0.12	I *** I
69	2.88	7485.	0.7880	0.13	I *** I
70	2.92	7484.	0.9308	0.15	I **** I
71	2.96	7486.	1.1084	0.18	I **** I
72	3.00	7483.	1.1985	0.20	I **** I
73	3.04	7486.	1.1048	0.18	I **** I
74	3.08	7483.	0.9102	0.15	I *** I
75	3.13	7481.	0.7159	0.12	I *** I

FOR THIS AUTOCORRELATION: MEAN = 3.3108 AND VARIANCE = 6.1160

WASHINGTON, DC - 1966

1968 CO WASHINGTON DC CAMP

NO. OF VALUES READ = 8784 NO OF VALID VALUES STORED = 7266

BASIC STATISTICS-----

N = 7266.

MEAN = 3.41398239

2ND MOMENT (VAR.) = 6.88111877 STANDARD DEV. = 2.62318897

COEFFICIENT OF VARIATION = 0.76836628

3RD MOMENT = 43.9770660 COEF. OF SKEWNESS = 2.43634224

4TH MOMENT = 651.973389 COEF. OF KURTOSIS = 13.7692976

LOWEST VALUE = 0.0

HIGHEST VALUE = 27.0000000

SECOND HIGHEST VALUE = 26.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 192

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	192	2.642%	IC** I	2.642%	2.642%
0.50- 1.50	1210	16.653%	I*********C******* I	19.295%	19.293%
1.50- 2.50	1910	26.287%	I**********C*** I	45.582%	45.576%
2.50- 3.50	1253	17.245%	I********** C I	62.827%	62.818%
3.50- 4.50	881	12.125%	I********** C I	74.952%	74.941%
4.50- 5.50	692	9.524%	I********** C I	84.476%	84.464%
5.50- 6.50	451	6.207%	I****** C I	90.683%	90.670%
6.50- 7.50	251	3.454%	I*** C I	94.137%	94.124%
7.50- 8.50	149	2.051%	I** C I	96.188%	96.174%
8.50- 9.50	73	1.005%	I* C I	97.192%	97.179%
9.50- 10.50	57	0.784%	I* C I	97.977%	97.963%

10.50- 11.50	34	0.468%	C I	98.445%	98.431%
11.50- 12.50	21	0.289%	C I	98.734%	98.720%
12.50- 13.50	18	0.248%	C I	98.981%	98.968%
13.50- 14.50	17	0.234%	CI	99.215%	99.202%
14.50- 15.50	9	0.124%	CI	99.339%	99.326%
15.50- 16.50	8	0.110%	CI	99.449%	99.436%
16.50- 17.50	11	0.151%	CI	99.601%	99.587%
17.50- 18.50	7	0.096%	CI	99.697%	99.683%
18.50- 19.50	3	0.041%	CI	99.738%	99.725%
19.50- 20.50	5	0.069%	CI	99.807%	99.793%
20.50- 21.50	6	0.083%	CI	99.890%	99.876%
21.50- 22.50	2	0.028%	CI	99.917%	99.903%
22.50- 23.50	2	0.028%	CI	99.945%	99.931%
23.50- 24.50	0	0.0 %	CI	99.945%	99.931%
24.50- 25.50	1	0.014%	CI	99.958%	99.945%
25.50- 26.50	2	0.028%	CI	99.986%	99.972%
26.50- 27.50	1	0.014%	CI	100.000%	99.986%
27.50- 28.50	0	0.0 %	CI	100.000%	99.986%
28.50- 29.50	0	0.0 %	CI	100.000%	99.986%
29.50- 30.50	0	0.0 %	CI	100.000%	99.986%
30.50- 31.50	0	0.0 %	CI	100.000%	99.986%
31.50- 32.50	0	0.0 %	CI	100.000%	99.986%
32.50- 33.50	0	0.0 %	CI	100.000%	99.986%
33.50- 34.50	0	0.0 %	CI	100.000%	99.986%
34.50- 35.50	0	0.0 %	CI	100.000%	99.986%
35.50- 36.50	0	0.0 %	CI	100.000%	99.986%
36.50- 37.50	0	0.0 %	CI	100.000%	99.986%
37.50- 38.50	0	0.0 %	CI	100.000%	99.986%
38.50- 39.50	0	0.0 %	CI	100.000%	99.986%
39.50- 40.50	0	0.0 %	CI	100.000%	99.986%
40.50- 41.50	0	0.0 %	CI	100.000%	99.986%
41.50- 42.50	0	0.0 %	CI	100.000%	99.986%
42.50- 43.50	0	0.0 %	CI	100.000%	99.986%
43.50- 44.50	0	0.0 %	CI	100.000%	99.986%
44.50- 45.50	0	0.0 %	CI	100.000%	99.986%
45.50- 46.50	0	0.0 %	CI	100.000%	99.986%
46.50- 47.50	0	0.0 %	CI	100.000%	99.986%
47.50- 48.50	0	0.0 %	CI	100.000%	99.986%
48.50- 49.50	0	0.0 %	CI	100.000%	99.986%
49.50- 50.50	0	0.0 %	CI	100.000%	99.986%
50.50- 51.50	0	0.0 %	CI	100.000%	99.986%
51.50- 52.50	0	0.0 %	CI	100.000%	99.986%
52.50- 53.50	0	0.0 %	CI	100.000%	99.986%
53.50- 54.50	0	0.0 %	CI	100.000%	99.986%
54.50- 55.50	0	0.0 %	CI	100.000%	99.986%
55.50- 56.50	0	0.0 %	CI	100.000%	99.986%
56.50- 57.50	0	0.0 %	CI	100.000%	99.986%
57.50- 58.50	0	0.0 %	CI	100.000%	99.986%
58.50- 59.50	0	0.0 %	CI	100.000%	99.986%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	7144.	6.1608	0.90	**********
2	0.08	7079.	5.2803	0.77	**********
3	0.13	7058.	4.5395	0.66	**********
4	0.17	7048.	4.0346	0.59	**********
5	0.21	7037.	3.6974	0.54	**********
6	0.25	7025.	3.5313	0.51	**********
7	0.29	7010.	3.5146	0.51	**********
8	0.33	6996.	3.5577	0.52	**********
9	0.38	6985.	3.5615	0.52	**********
10	0.42	6975.	3.4752	0.50	**********
11	0.46	6964.	3.3737	0.49	**********
12	0.50	6956.	3.2831	0.48	**********
13	0.54	6948.	3.2001	0.46	**********
14	0.58	6940.	3.1466	0.46	**********
15	0.63	6933.	3.0626	0.44	**********
16	0.67	6926.	2.9067	0.42	**********
17	0.71	6919.	2.6868	0.39	**********
18	0.75	6914.	2.4983	0.36	**********
19	0.79	6910.	2.4343	0.35	**********
20	0.83	6901.	2.4691	0.36	**********
21	0.88	6893.	2.6503	0.39	**********
22	0.92	6885.	2.9325	0.43	**********
23	0.96	6881.	3.3014	0.48	**********
24	1.00	6876.	3.5133	0.51	**********
25	1.04	6868.	3.3728	0.49	**********
26	1.08	6863.	3.0191	0.44	**********
27	1.13	6858.	2.7012	0.39	**********
28	1.17	6851.	2.5244	0.37	**********
29	1.21	6843.	2.4513	0.36	**********
30	1.25	6838.	2.4997	0.36	**********
31	1.29	6830.	2.6266	0.38	**********
32	1.33	6823.	2.8005	0.41	**********
33	1.38	6816.	2.9006	0.42	**********
34	1.42	6810.	2.9284	0.43	**********
35	1.46	6804.	2.8838	0.42	**********
36	1.50	6798.	2.8587	0.42	**********
37	1.54	6795.	2.8391	0.41	**********
38	1.58	6787.	2.8183	0.41	**********
39	1.63	6782.	2.7306	0.40	**********
40	1.67	6779.	2.5592	0.37	**********
41	1.71	6773.	2.3619	0.34	**********
42	1.75	6767.	2.1899	0.32	**********
43	1.79	6764.	2.1238	0.31	**********
44	1.83	6759.	2.1765	0.32	**********
45	1.88	6760.	2.3248	0.34	**********
46	1.92	6757.	2.5886	0.38	**********
47	1.96	6751.	2.8897	0.42	**********
48	2.00	6741.	3.0431	0.44	**********
49	2.04	6734.	2.9138	0.42	**********
50	2.08	6730.	2.6272	0.38	**********

51	2.13	6729.	2.4061	0.35	I ******* I
52	2.17	6726.	2.2894	0.33	I ******* I
53	2.21	6721.	2.2404	0.33	I ******* I
54	2.25	6713.	2.2610	0.33	I ******* I
55	2.29	6707.	2.3711	0.34	I ******* I
56	2.33	6705.	2.4697	0.36	I ******** I
57	2.38	6702.	2.4977	0.36	I ******** I
58	2.42	6697.	2.4675	0.36	I ******** I
59	2.46	6695.	2.4208	0.35	I ******** I
60	2.50	6691.	2.3568	0.34	I ******* I
61	2.54	6686.	2.3301	0.34	I ******* I
62	2.58	6681.	2.3254	0.34	I ******* I
63	2.63	6679.	2.3013	0.33	I ******* I
64	2.67	6677.	2.2257	0.32	I ******* I
65	2.71	6674.	2.0982	0.30	I ******* I
66	2.75	6671.	1.9586	0.28	I ****** I
67	2.79	6668.	1.9271	0.28	I ****** I
68	2.83	6669.	1.9648	0.29	I ****** I
69	2.88	6664.	2.0716	0.30	I ******* I
70	2.92	6663.	2.2287	0.32	I ******* I
71	2.96	6665.	2.4239	0.35	I ******** I
72	3.00	6663.	2.5738	0.37	I ******** I
73	3.04	6649.	2.4681	0.36	I ******** I
74	3.08	6640.	2.2762	0.33	I ******* I
75	3.13	6640.	2.1084	0.31	I ******* I

+---+---+---

1969 CO WASHINGTON DC CAMP

NO. OF VALUES READ = 6372 NO OF VALID VALUES STORED = 6056

BASIC STATISTICS-----

N = 6056

MEAN = 3.54161072

2ND MOMENT (VAR.) = 5.72583580 STANDARD DEV. = 2.39287186

COEFFICIENT OF VARIATION = 0.67564505

3RD MOMENT = 31.5226440 COEF. OF SKEWNESS = 2.30072403

4TH MOMENT = 424.701172 COEF. OF KURTOSIS = 12.9540567

LOWEST VALUE = 0.0

HIGHEST VALUE = 25.0000000

SECOND HIGHEST VALUE = 23.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 180

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	180	2.972%	IC** I	2.972%	2.972%
0.50- 1.50	612	10.106%	I******C*** I	13.078%	13.076%
1.50- 2.50	1332	21.995%	I**********C**** I	35.073%	35.067%
2.50- 3.50	1458	24.075%	I********** C I	59.148%	59.138%
3.50- 4.50	1042	17.206%	I********** C I	76.354%	76.341%
4.50- 5.50	631	10.419%	I********** C I	86.773%	86.759%
5.50- 6.50	299	4.937%	I***** C I	91.711%	91.695%
6.50- 7.50	172	2.840%	I*** C I	94.551%	94.535%
7.50- 8.50	108	1.783%	I** C I	96.334%	96.318%
8.50- 9.50	62	1.024%	I* C I	97.358%	97.342%
9.50- 10.50	41	0.677%	I* C I	98.035%	98.019%

10.50- 11.50	38	0.627%	I*	C I 98.662%	98.646%
11.50- 12.50	17	0.281%	I	C I 98.943%	98.927%
12.50- 13.50	15	0.248%	I	CI 99.191%	99.174%
13.50- 14.50	6	0.099%	I	CI 99.290%	99.273%
14.50- 15.50	11	0.182%	I	CI 99.471%	99.455%
15.50- 16.50	13	0.215%	I	CI 99.686%	99.670%
16.50- 17.50	5	0.083%	I	CI 99.769%	99.752%
17.50- 18.50	2	0.033%	I	CI 99.802%	99.785%
18.50- 19.50	3	0.050%	I	CI 99.851%	99.835%
19.50- 20.50	5	0.083%	I	CI 99.934%	99.917%
20.50- 21.50	0	0.0 %	I	CI 99.934%	99.917%
21.50- 22.50	1	0.017%	I	CI 99.950%	99.934%
22.50- 23.50	2	0.033%	I	CI 99.983%	99.967%
23.50- 24.50	0	0.0 %	I	CI 99.983%	99.967%
24.50- 25.50	1	0.017%	I	CI 100.000%	99.983%
25.50- 26.50	0	0.0 %	I	CI 100.000%	99.983%
26.50- 27.50	0	0.0 %	I	CI 100.000%	99.983%
27.50- 28.50	0	0.0 %	I	CI 100.000%	99.983%
28.50- 29.50	0	0.0 %	I	CI 100.000%	99.983%
29.50- 30.50	0	0.0 %	I	CI 100.000%	99.983%
30.50- 31.50	0	0.0 %	I	CI 100.000%	99.983%
31.50- 32.50	0	0.0 %	I	CI 100.000%	99.983%
32.50- 33.50	0	0.0 %	I	CI 100.000%	99.983%
33.50- 34.50	0	0.0 %	I	CI 100.000%	99.983%
34.50- 35.50	0	0.0 %	I	CI 100.000%	99.983%
35.50- 36.50	0	0.0 %	I	CI 100.000%	99.983%
36.50- 37.50	0	0.0 %	I	CI 100.000%	99.983%
37.50- 38.50	0	0.0 %	I	CI 100.000%	99.983%
38.50- 39.50	0	0.0 %	I	CI 100.000%	99.983%
39.50- 40.50	0	0.0 %	I	CI 100.000%	99.983%
40.50- 41.50	0	0.0 %	I	CI 100.000%	99.983%
41.50- 42.50	0	0.0 %	I	CI 100.000%	99.983%
42.50- 43.50	0	0.0 %	I	CI 100.000%	99.983%
43.50- 44.50	0	0.0 %	I	CI 100.000%	99.983%
44.50- 45.50	0	0.0 %	I	CI 100.000%	99.983%
45.50- 46.50	0	0.0 %	I	CI 100.000%	99.983%
46.50- 47.50	0	0.0 %	I	CI 100.000%	99.983%
47.50- 48.50	0	0.0 %	I	CI 100.000%	99.983%
48.50- 49.50	0	0.0 %	I	CI 100.000%	99.983%
49.50- 50.50	0	0.0 %	I	CI 100.000%	99.983%
50.50- 51.50	0	0.0 %	I	CI 100.000%	99.983%
51.50- 52.50	0	0.0 %	I	CI 100.000%	99.983%
52.50- 53.50	0	0.0 %	I	CI 100.000%	99.983%
53.50- 54.50	0	0.0 %	I	CI 100.000%	99.983%
54.50- 55.50	0	0.0 %	I	CI 100.000%	99.983%
55.50- 56.50	0	0.0 %	I	CI 100.000%	99.983%
56.50- 57.50	0	0.0 %	I	CI 100.000%	99.983%
57.50- 58.50	0	0.0 %	I	CI 100.000%	99.983%
58.50- 59.50	0	0.0 %	I	CI 100.000%	99.983%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	5927.	4.9464	0.86	**********
2	0.08	5867.	3.9402	0.69	**********
3	0.13	5845.	3.1827	0.56	**********
4	0.17	5825.	2.7488	0.48	**********
5	0.21	5807.	2.4825	0.43	**********
6	0.25	5793.	2.3358	0.41	**********
7	0.29	5783.	2.3773	0.42	**********
8	0.33	5774.	2.4655	0.43	**********
9	0.38	5768.	2.4755	0.43	**********
10	0.42	5760.	2.2613	0.39	**********
11	0.46	5757.	1.9368	0.34	*******
12	0.50	5754.	1.7182	0.30	*******
13	0.54	5759.	1.6860	0.29	******
14	0.58	5755.	1.7609	0.31	*******
15	0.63	5761.	1.7377	0.30	*******
16	0.67	5761.	1.6011	0.28	******
17	0.71	5765.	1.3687	0.24	*****
18	0.75	5764.	1.1987	0.21	*****
19	0.79	5763.	1.1709	0.20	*****
20	0.83	5765.	1.2379	0.22	*****
21	0.88	5760.	1.4431	0.25	******
22	0.92	5761.	1.7970	0.31	*******
23	0.96	5764.	2.2649	0.40	********
24	1.00	5768.	2.5225	0.44	*********
25	1.04	5766.	2.2848	0.40	********
26	1.08	5764.	1.8326	0.32	*******
27	1.13	5761.	1.4478	0.25	******
28	1.17	5762.	1.2050	0.21	*****
29	1.21	5753.	1.1146	0.19	****
30	1.25	5750.	1.1474	0.20	*****
31	1.29	5747.	1.2889	0.23	*****
32	1.33	5740.	1.4760	0.26	******
33	1.38	5740.	1.5821	0.28	******
34	1.42	5737.	1.4959	0.26	******
35	1.46	5732.	1.3638	0.24	*****
36	1.50	5728.	1.2719	0.22	*****
37	1.54	5734.	1.2489	0.22	*****
38	1.58	5731.	1.3013	0.23	*****
39	1.63	5727.	1.2897	0.23	*****
40	1.67	5730.	1.1268	0.20	****
41	1.71	5736.	0.9214	0.16	****
42	1.75	5730.	0.8066	0.14	***
43	1.79	5729.	0.8096	0.14	***
44	1.83	5732.	0.8335	0.15	***
45	1.88	5732.	0.9533	0.17	****
46	1.92	5734.	1.2287	0.21	*****
47	1.96	5733.	1.6210	0.28	******
48	2.00	5738.	1.8216	0.32	*******
49	2.04	5737.	1.6409	0.29	******
50	2.08	5730.	1.2989	0.23	*****

51	2.13	5728.	1.0134	0.18	I **** I
52	2.17	5724.	0.8541	0.15	I *** I
53	2.21	5715.	0.8014	0.14	I *** I
54	2.25	5712.	0.8374	0.15	I *** I
55	2.29	5704.	0.9691	0.17	I **** I
56	2.33	5707.	1.1335	0.20	I **** I
57	2.38	5710.	1.2402	0.22	I ***** I
58	2.42	5710.	1.1920	0.21	I ***** I
59	2.46	5706.	1.1032	0.19	I **** I
60	2.50	5706.	1.0509	0.18	I **** I
61	2.54	5706.	1.0905	0.19	I **** I
62	2.58	5707.	1.1840	0.21	I ***** I
63	2.63	5710.	1.1896	0.21	I ***** I
64	2.67	5709.	1.0413	0.18	I **** I
65	2.71	5711.	0.8240	0.14	I *** I
66	2.75	5715.	0.6820	0.12	I *** I
67	2.79	5714.	0.6430	0.11	I *** I
68	2.83	5714.	0.6864	0.12	I *** I
69	2.88	5724.	0.7786	0.14	I *** I
70	2.92	5719.	1.0066	0.18	I **** I
71	2.96	5719.	1.3820	0.24	I ***** I
72	3.00	5717.	1.5431	0.27	I ****** I
73	3.04	5716.	1.3415	0.23	I ***** I
74	3.08	5715.	0.9839	0.17	I **** I
75	3.13	5710.	0.7255	0.13	I *** I

+---+---+---+---+---+---+---+---+---+---+

FOR THIS AUTOCORRELATION: MEAN = 3.5416 AND VARIANCE = 5.7261

WASHINGTON, DC - 1969

1970 CO WASHINGTON DC CAMP

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 7748

BASIC STATISTICS-----

N = 7748

MEAN = 3.49070644

2ND MOMENT (VAR.) = 6.68706512 STANDARD DEV. = 2.58593559

COEFFICIENT OF VARIATION = 0.74080575

3RD MOMENT = 65.8857117 COEF. OF SKEWNESS = 3.81011200

4TH MOMENT = 1679.41138 COEF. OF KURTOSIS = 37.5565796

LOWEST VALUE = 0.0

HIGHEST VALUE = 45.0000000

SECOND HIGHEST VALUE = 42.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 270

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	270	3.485%	I*C* I	3.485%	3.484%
0.50- 1.50	816	10.532%	I******C**** I	14.017%	14.015%
1.50- 2.50	1869	24.122%	I******************C***** I	38.139%	38.134%
2.50- 3.50	1778	22.948%	I************************ C I	61.087%	61.079%
3.50- 4.50	1225	15.811%	I**************** C I	76.897%	76.887%
4.50- 5.50	698	9.009%	I********* C I	85.906%	85.895%
5.50- 6.50	447	5.769%	I****** C I	91.675%	91.663%
6.50- 7.50	238	3.072%	I*** C I	94.747%	94.735%
7.50- 8.50	136	1.755%	I** C I	96.502%	96.490%
8.50- 9.50	79	1.020%	I* C I	97.522%	97.509%
9.50- 10.50	59	0.761%	I* C I	98.283%	98.271%

10.50- 11.50	32	0.413%	I	C I 98.696%	98.684%
11.50- 12.50	33	0.426%	I	CI 99.122%	99.109%
12.50- 13.50	7	0.090%	I	CI 99.213%	99.200%
13.50- 14.50	11	0.142%	I	CI 99.355%	99.342%
14.50- 15.50	8	0.103%	I	CI 99.458%	99.445%
15.50- 16.50	10	0.129%	I	CI 99.587%	99.574%
16.50- 17.50	4	0.052%	I	CI 99.638%	99.626%
17.50- 18.50	6	0.077%	I	CI 99.716%	99.703%
18.50- 19.50	4	0.052%	I	CI 99.768%	99.755%
19.50- 20.50	2	0.026%	I	CI 99.793%	99.780%
20.50- 21.50	2	0.026%	I	CI 99.819%	99.806%
21.50- 22.50	1	0.013%	I	CI 99.832%	99.819%
22.50- 23.50	4	0.052%	I	CI 99.884%	99.871%
23.50- 24.50	0	0.0 %	I	CI 99.884%	99.871%
24.50- 25.50	0	0.0 %	I	CI 99.884%	99.871%
25.50- 26.50	1	0.013%	I	CI 99.897%	99.884%
26.50- 27.50	0	0.0 %	I	CI 99.897%	99.884%
27.50- 28.50	0	0.0 %	I	CI 99.897%	99.884%
28.50- 29.50	1	0.013%	I	CI 99.909%	99.897%
29.50- 30.50	0	0.0 %	I	CI 99.909%	99.897%
30.50- 31.50	1	0.013%	I	CI 99.922%	99.909%
31.50- 32.50	1	0.013%	I	CI 99.935%	99.922%
32.50- 33.50	1	0.013%	I	CI 99.948%	99.935%
33.50- 34.50	1	0.013%	I	CI 99.961%	99.948%
34.50- 35.50	0	0.0 %	I	CI 99.961%	99.948%
35.50- 36.50	0	0.0 %	I	CI 99.961%	99.948%
36.50- 37.50	0	0.0 %	I	CI 99.961%	99.948%
37.50- 38.50	1	0.013%	I	CI 99.974%	99.961%
38.50- 39.50	0	0.0 %	I	CI 99.974%	99.961%
39.50- 40.50	0	0.0 %	I	CI 99.974%	99.961%
40.50- 41.50	0	0.0 %	I	CI 99.974%	99.961%
41.50- 42.50	1	0.013%	I	CI 99.987%	99.974%
42.50- 43.50	0	0.0 %	I	CI 99.987%	99.974%
43.50- 44.50	0	0.0 %	I	CI 99.987%	99.974%
44.50- 45.50	1	0.013%	I	CI 100.000%	99.987%
45.50- 46.50	0	0.0 %	I	CI 100.000%	99.987%
46.50- 47.50	0	0.0 %	I	CI 100.000%	99.987%
47.50- 48.50	0	0.0 %	I	CI 100.000%	99.987%
48.50- 49.50	0	0.0 %	I	CI 100.000%	99.987%
49.50- 50.50	0	0.0 %	I	CI 100.000%	99.987%
50.50- 51.50	0	0.0 %	I	CI 100.000%	99.987%
51.50- 52.50	0	0.0 %	I	CI 100.000%	99.987%
52.50- 53.50	0	0.0 %	I	CI 100.000%	99.987%
53.50- 54.50	0	0.0 %	I	CI 100.000%	99.987%
54.50- 55.50	0	0.0 %	I	CI 100.000%	99.987%
55.50- 56.50	0	0.0 %	I	CI 100.000%	99.987%
56.50- 57.50	0	0.0 %	I	CI 100.000%	99.987%
57.50- 58.50	0	0.0 %	I	CI 100.000%	99.987%
58.50- 59.50	0	0.0 %	I	CI 100.000%	99.987%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

WASHINGTON, DC - 1970

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	7520.	5.5025	0.82	**********
2	0.08	7477.	4.3315	0.65	**********
3	0.13	7461.	3.3027	0.49	**********
4	0.17	7449.	2.6645	0.40	********
5	0.21	7440.	2.1604	0.32	*******
6	0.25	7434.	1.8748	0.28	******
7	0.29	7425.	1.7420	0.26	******
8	0.33	7417.	1.9373	0.29	******
9	0.38	7411.	2.1530	0.32	*******
10	0.42	7406.	2.1762	0.33	*******
11	0.46	7402.	2.0442	0.31	*******
12	0.50	7397.	1.9092	0.29	******
13	0.54	7396.	1.8719	0.28	******
14	0.58	7394.	1.8466	0.28	******
15	0.63	7391.	1.6391	0.25	*****
16	0.67	7388.	1.2589	0.19	****
17	0.71	7387.	0.8894	0.13	***
18	0.75	7386.	0.7527	0.11	***
19	0.79	7382.	0.8114	0.12	***
20	0.83	7382.	0.9540	0.14	***
21	0.88	7377.	1.1459	0.17	****
22	0.92	7376.	1.5891	0.24	*****
23	0.96	7388.	2.0593	0.31	*******
24	1.00	7522.	2.1873	0.33	*******
25	1.04	7378.	1.8890	0.28	******
26	1.08	7365.	1.3256	0.20	****
27	1.13	7358.	0.8464	0.13	***
28	1.17	7352.	0.5835	0.09	**
29	1.21	7341.	0.3778	0.06	**
30	1.25	7336.	0.3414	0.05	**
31	1.29	7331.	0.3997	0.06	**
32	1.33	7323.	0.7121	0.11	***
33	1.38	7318.	1.0948	0.16	****
34	1.42	7314.	1.2321	0.18	****
35	1.46	7311.	1.1694	0.17	****
36	1.50	7307.	1.1013	0.16	****
37	1.54	7306.	1.0561	0.16	****
38	1.58	7303.	1.0095	0.15	****
39	1.63	7301.	0.8189	0.12	***
40	1.67	7298.	0.5164	0.08	**
41	1.71	7299.	0.2686	0.04	*
42	1.75	7298.	0.1605	0.02	*
43	1.79	7294.	0.2113	0.03	*
44	1.83	7293.	0.3408	0.05	**
45	1.88	7288.	0.5115	0.08	**
46	1.92	7282.	0.8549	0.13	***
47	1.96	7290.	1.2266	0.18	****
48	2.00	7414.	1.3455	0.20	*****
49	2.04	7281.	1.1560	0.17	****
50	2.08	7272.	0.7360	0.11	***

51	2.13	7269.	0.3666	0.05	I ** I
52	2.17	7267.	0.1952	0.03	I * I
53	2.21	7259.	0.0724	0.01	I * I
54	2.25	7256.	0.0884	0.01	I * I
55	2.29	7248.	0.1348	0.02	I * I
56	2.33	7241.	0.4309	0.06	I ** I
57	2.38	7238.	0.7466	0.11	I *** I
58	2.42	7236.	0.8716	0.13	I *** I
59	2.46	7234.	0.8157	0.12	I *** I
60	2.50	7231.	0.7536	0.11	I *** I
61	2.54	7229.	0.7295	0.11	I *** I
62	2.58	7226.	0.7408	0.11	I *** I
63	2.63	7224.	0.6198	0.09	I ** I
64	2.67	7222.	0.3108	0.05	I * I
65	2.71	7223.	-0.0142	-.00	I ** I
66	2.75	7224.	-0.1517	-.02	I ** I
67	2.79	7219.	-0.1594	-.02	I ** I
68	2.83	7220.	-0.1276	-.02	I ** I
69	2.88	7217.	-0.0152	-.00	I ** I
70	2.92	7215.	0.2633	0.04	I * I
71	2.96	7224.	0.7023	0.11	I *** I
72	3.00	7347.	0.8280	0.12	I *** I
73	3.04	7224.	0.6572	0.10	I ** I
74	3.08	7206.	0.2377	0.04	I * I
75	3.13	7202.	-0.1123	-.02	I ** I

FOR THIS AUTOCORRELATION: MEAN = 3.4907 AND VARIANCE = 6.6882

WASHINGTON, DC - 1970

1971 CO WASHINGTON DC CAMP

NO. OF VALUES READ = 8760 NO OF VALID VALUES STORED = 7254

BASIC STATISTICS-----

N = 7254

MEAN = 3.48876476

2ND MOMENT (VAR.) = 6.60787868 STANDARD DEV. = 2.57057953

COEFFICIENT OF VARIATION = 0.73681647

3RD MOMENT = 61.4603119 COEF. OF SKEWNESS = 3.61827755

4TH MOMENT = 1359.71167 COEF. OF KURTOSIS = 31.1402893

LOWEST VALUE = 0.0

HIGHEST VALUE = 46.0000000

SECOND HIGHEST VALUE = 35.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 42

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	42	0.579%	I* I	0.579%	0.579%
0.50- 1.50	925	12.752%	I******C****** I	13.331%	13.329%
1.50- 2.50	1943	26.785%	I**********C******* I	40.116%	40.110%
2.50- 3.50	1621	22.346%	I********** C I	62.462%	62.453%
3.50- 4.50	1154	15.908%	I********** C I	78.371%	78.360%
4.50- 5.50	611	8.423%	I******** C I	86.793%	86.781%
5.50- 6.50	324	4.467%	I**** C I	91.260%	91.247%
6.50- 7.50	235	3.240%	I*** C I	94.500%	94.486%
7.50- 8.50	118	1.627%	I** C I	96.126%	96.113%
8.50- 9.50	81	1.117%	I* C I	97.243%	97.229%
9.50- 10.50	43	0.593%	I* C I	97.836%	97.822%

10.50- 11.50	47	0.648%	I*	98.484%	98.470%
11.50- 12.50	24	0.331%	I	98.814%	98.801%
12.50- 13.50	19	0.262%	I	99.076%	99.063%
13.50- 14.50	17	0.234%	I	99.311%	99.297%
14.50- 15.50	7	0.096%	I	99.407%	99.393%
15.50- 16.50	7	0.096%	I	99.504%	99.490%
16.50- 17.50	5	0.069%	I	99.573%	99.559%
17.50- 18.50	9	0.124%	I	99.697%	99.683%
18.50- 19.50	4	0.055%	I	99.752%	99.738%
19.50- 20.50	4	0.055%	I	99.807%	99.793%
20.50- 21.50	2	0.028%	I	99.834%	99.821%
21.50- 22.50	1	0.014%	I	99.848%	99.834%
22.50- 23.50	1	0.014%	I	99.862%	99.848%
23.50- 24.50	1	0.014%	I	99.876%	99.862%
24.50- 25.50	1	0.014%	I	99.890%	99.876%
25.50- 26.50	1	0.014%	I	99.903%	99.890%
26.50- 27.50	1	0.014%	I	99.917%	99.903%
27.50- 28.50	2	0.028%	I	99.945%	99.931%
28.50- 29.50	0	0.0 %	I	99.945%	99.931%
29.50- 30.50	1	0.014%	I	99.958%	99.945%
30.50- 31.50	0	0.0 %	I	99.958%	99.945%
31.50- 32.50	1	0.014%	I	99.972%	99.958%
32.50- 33.50	0	0.0 %	I	99.972%	99.958%
33.50- 34.50	0	0.0 %	I	99.972%	99.958%
34.50- 35.50	1	0.014%	I	99.986%	99.972%
35.50- 36.50	0	0.0 %	I	99.986%	99.972%
36.50- 37.50	0	0.0 %	I	99.986%	99.972%
37.50- 38.50	0	0.0 %	I	99.986%	99.972%
38.50- 39.50	0	0.0 %	I	99.986%	99.972%
39.50- 40.50	0	0.0 %	I	99.986%	99.972%
40.50- 41.50	0	0.0 %	I	99.986%	99.972%
41.50- 42.50	0	0.0 %	I	99.986%	99.972%
42.50- 43.50	0	0.0 %	I	99.986%	99.972%
43.50- 44.50	0	0.0 %	I	99.986%	99.972%
44.50- 45.50	0	0.0 %	I	99.986%	99.972%
45.50- 46.50	1	0.014%	I	100.000%	99.986%
46.50- 47.50	0	0.0 %	I	100.000%	99.986%
47.50- 48.50	0	0.0 %	I	100.000%	99.986%
48.50- 49.50	0	0.0 %	I	100.000%	99.986%
49.50- 50.50	0	0.0 %	I	100.000%	99.986%
50.50- 51.50	0	0.0 %	I	100.000%	99.986%
51.50- 52.50	0	0.0 %	I	100.000%	99.986%
52.50- 53.50	0	0.0 %	I	100.000%	99.986%
53.50- 54.50	0	0.0 %	I	100.000%	99.986%
54.50- 55.50	0	0.0 %	I	100.000%	99.986%
55.50- 56.50	0	0.0 %	I	100.000%	99.986%
56.50- 57.50	0	0.0 %	I	100.000%	99.986%
57.50- 58.50	0	0.0 %	I	100.000%	99.986%
58.50- 59.50	0	0.0 %	I	100.000%	99.986%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	6907.	5.6082	0.85	**********
2	0.08	6893.	4.0731	0.62	**********
3	0.13	6888.	3.0748	0.47	**********
4	0.17	6882.	2.3911	0.36	********
5	0.21	6873.	1.8520	0.28	******
6	0.25	6863.	1.5214	0.23	*****
7	0.29	6858.	1.4864	0.22	*****
8	0.33	6850.	1.8137	0.27	******
9	0.38	6845.	2.0716	0.31	*******
10	0.42	6844.	2.0028	0.30	*******
11	0.46	6839.	1.7576	0.27	******
12	0.50	6834.	1.5722	0.24	*****
13	0.54	6830.	1.5006	0.23	*****
14	0.58	6826.	1.4988	0.23	*****
15	0.63	6823.	1.3611	0.21	*****
16	0.67	6816.	0.9305	0.14	***
17	0.71	6814.	0.4908	0.07	**
18	0.75	6810.	0.3497	0.05	**
19	0.79	6806.	0.3709	0.06	**
20	0.83	6805.	0.4930	0.07	**
21	0.88	6803.	0.6708	0.10	***
22	0.92	6796.	0.9828	0.15	***
23	0.96	6795.	1.5857	0.24	*****
24	1.00	7084.	1.7942	0.27	******
25	1.04	6786.	1.5389	0.23	*****
26	1.08	6779.	0.9147	0.14	***
27	1.13	6778.	0.5439	0.08	**
28	1.17	6773.	0.3456	0.05	**
29	1.21	6766.	0.1744	0.03	*
30	1.25	6760.	0.0372	0.01	*
31	1.29	6756.	0.0624	0.01	*
32	1.33	6749.	0.3680	0.06	**
33	1.38	6744.	0.6990	0.11	***
34	1.42	6742.	0.7473	0.11	***
35	1.46	6737.	0.6523	0.10	**
36	1.50	6732.	0.6318	0.10	**
37	1.54	6730.	0.6645	0.10	***
38	1.58	6726.	0.7303	0.11	***
39	1.63	6722.	0.6044	0.09	**
40	1.67	6717.	0.2205	0.03	*
41	1.71	6717.	-0.1860	-.03	**
42	1.75	6713.	-0.2838	-.04	**
43	1.79	6711.	-0.2006	-.03	**
44	1.83	6713.	-0.0430	-.01	**
45	1.88	6709.	0.1271	0.02	*
46	1.92	6702.	0.4612	0.07	**
47	1.96	6703.	0.9965	0.15	****
48	2.00	6987.	1.2283	0.19	****
49	2.04	6698.	1.0536	0.16	****
50	2.08	6693.	0.5501	0.08	**

51	2.13	6692.	0.2743	0.04	I * I
52	2.17	6688.	0.0849	0.01	I * I
53	2.21	6684.	-0.0747	-.01	I ** I
54	2.25	6679.	-0.1332	-.02	I ** I
55	2.29	6677.	-0.0292	-.00	I ** I
56	2.33	6670.	0.2396	0.04	I * I
57	2.38	6667.	0.5043	0.08	I ** I
58	2.42	6666.	0.5663	0.09	I ** I
59	2.46	6664.	0.5391	0.08	I ** I
60	2.50	6660.	0.5552	0.08	I ** I
61	2.54	6656.	0.6244	0.09	I ** I
62	2.58	6654.	0.7057	0.11	I *** I
63	2.63	6650.	0.5455	0.08	I ** I
64	2.67	6647.	0.1249	0.02	I * I
65	2.71	6648.	-0.2954	-.04	I ** I
66	2.75	6644.	-0.4223	-.06	I *** I
67	2.79	6643.	-0.4234	-.06	I *** I
68	2.83	6643.	-0.3066	-.05	I ** I
69	2.88	6641.	-0.1752	-.03	I ** I
70	2.92	6634.	0.0970	0.01	I * I
71	2.96	6632.	0.6127	0.09	I ** I
72	3.00	6911.	0.8237	0.12	I *** I
73	3.04	6625.	0.6548	0.10	I ** I
74	3.08	6619.	0.1991	0.03	I * I
75	3.13	6617.	-0.0264	-.00	I ** I

+---+

FOR THIS AUTOCORRELATION: MEAN = 3.4888 AND VARIANCE = 6.6087

WASHINGTON, DC - 1971

1972 CO WASHINGTON DC CAMP

NO. OF VALUES READ = 5256 NO OF VALID VALUES STORED = 4947

BASIC STATISTICS-----

N = 4947

MEAN = 4.34990883

2ND MOMENT (VAR.) = 14.2463093 STANDARD DEV. = 3.77442837

COEFFICIENT OF VARIATION = 0.86770284

3RD MOMENT = 164.217575 COEF. OF SKEWNESS = 3.05397797

4TH MOMENT = 4212.10547 COEF. OF KURTOSIS = 20.7536469

LOWEST VALUE = 0.0

HIGHEST VALUE = 55.0000000

SECOND HIGHEST VALUE = 40.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 422

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	422	8.530%	I***C***** I	8.530%	8.529%
0.50- 1.50	329	6.650%	I*******C I	15.181%	15.178%
1.50- 2.50	784	15.848%	I***************C I	31.029%	31.023%
2.50- 3.50	670	13.544%	I************** C I	44.572%	44.563%
3.50- 4.50	818	16.535%	I***************** C I	61.108%	61.095%
4.50- 5.50	828	16.737%	I***************** C I	77.845%	77.829%
5.50- 6.50	368	7.439%	I******* C I	85.284%	85.267%
6.50- 7.50	173	3.497%	I*** C I	88.781%	88.763%
7.50- 8.50	135	2.729%	I*** C I	91.510%	91.491%
8.50- 9.50	77	1.556%	I** C I	93.066%	93.048%
9.50- 10.50	66	1.334%	I* C I	94.401%	94.381%

```
10.50- 11.50      59    1.193%  I*                                          C  I  95.593%     95.574%
11.50- 12.50      45    0.910%  I*                                          C  I  96.503%     96.483%
12.50- 13.50      41    0.829%  I*                                           C I  97.332%     97.312%
13.50- 14.50      21    0.424%  I                                            C I  97.756%     97.736%
14.50- 15.50      22    0.445%  I                                            C I  98.201%     98.181%
15.50- 16.50      12    0.243%  I                                            C I  98.443%     98.423%
16.50- 17.50       9    0.182%  I                                            C I  98.625%     98.605%
17.50- 18.50       7    0.141%  I                                            C I  98.767%     98.747%
18.50- 19.50       4    0.081%  I                                            C I  98.848%     98.828%
19.50- 20.50       2    0.040%  I                                            C I  98.888%     98.868%
20.50- 21.50       7    0.141%  I                                             CI  99.030%     99.010%
21.50- 22.50       2    0.040%  I                                             CI  99.070%     99.050%
22.50- 23.50       4    0.081%  I                                             CI  99.151%     99.131%
23.50- 24.50       2    0.040%  I                                             CI  99.191%     99.171%
24.50- 25.50      14    0.283%  I                                             CI  99.474%     99.454%
25.50- 26.50      21    0.424%  I                                             CI  99.899%     99.879%
26.50- 27.50       1    0.020%  I                                             CI  99.919%     99.899%
27.50- 28.50       0    0.0  %  I                                             CI  99.919%     99.899%
28.50- 29.50       0    0.0  %  I                                             CI  99.919%     99.899%
29.50- 30.50       0    0.0  %  I                                             CI  99.919%     99.899%
30.50- 31.50       1    0.020%  I                                             CI  99.939%     99.919%
31.50- 32.50       0    0.0  %  I                                             CI  99.939%     99.919%
32.50- 33.50       0    0.0  %  I                                             CI  99.939%     99.919%
33.50- 34.50       0    0.0  %  I                                             CI  99.939%     99.919%
34.50- 35.50       1    0.020%  I                                             CI  99.959%     99.939%
35.50- 36.50       0    0.0  %  I                                             CI  99.959%     99.939%
36.50- 37.50       0    0.0  %  I                                             CI  99.959%     99.939%
37.50- 38.50       0    0.0  %  I                                             CI  99.959%     99.939%
38.50- 39.50       0    0.0  %  I                                             CI  99.959%     99.939%
39.50- 40.50       1    0.020%  I                                             CI  99.980%     99.959%
40.50- 41.50       0    0.0  %  I                                             CI  99.980%     99.959%
41.50- 42.50       0    0.0  %  I                                             CI  99.980%     99.959%
42.50- 43.50       0    0.0  %  I                                             CI  99.980%     99.959%
43.50- 44.50       0    0.0  %  I                                             CI  99.980%     99.959%
44.50- 45.50       0    0.0  %  I                                             CI  99.980%     99.959%
45.50- 46.50       0    0.0  %  I                                             CI  99.980%     99.959%
46.50- 47.50       0    0.0  %  I                                             CI  99.980%     99.959%
47.50- 48.50       0    0.0  %  I                                             CI  99.980%     99.959%
48.50- 49.50       0    0.0  %  I                                             CI  99.980%     99.959%
49.50- 50.50       0    0.0  %  I                                             CI  99.980%     99.959%
50.50- 51.50       0    0.0  %  I                                             CI  99.980%     99.959%
51.50- 52.50       0    0.0  %  I                                             CI  99.980%     99.959%
52.50- 53.50       0    0.0  %  I                                             CI  99.980%     99.959%
53.50- 54.50       0    0.0  %  I                                             CI  99.980%     99.959%
54.50- 55.50       1    0.020%  I                                             CI 100.000%     99.980%
55.50- 56.50       0    0.0  %  I                                             CI 100.000%     99.980%
56.50- 57.50       0    0.0  %  I                                             CI 100.000%     99.980%
57.50- 58.50       0    0.0  %  I                                             CI 100.000%     99.980%
58.50- 59.50       0    0.0  %  I                                             CI 100.000%     99.980%
                                +-----------------------------------------------+

NO. OF VALUES EQUAL TO OR GREATER THAN   59.5000 =    0  ( 0.0  %)
```

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	4722.	13.0468	0.92	**********
2	0.08	4703.	11.2633	0.79	**********
3	0.13	4699.	9.9070	0.70	**********
4	0.17	4687.	9.3023	0.65	**********
5	0.21	4682.	8.9104	0.63	**********
6	0.25	4675.	8.4561	0.59	**********
7	0.29	4671.	8.3815	0.59	**********
8	0.33	4665.	8.5480	0.60	**********
9	0.38	4659.	8.5275	0.60	**********
10	0.42	4657.	8.0609	0.57	**********
11	0.46	4655.	7.5591	0.53	**********
12	0.50	4654.	7.1625	0.50	**********
13	0.54	4654.	7.1049	0.50	**********
14	0.58	4655.	7.1498	0.50	**********
15	0.63	4653.	7.1279	0.50	**********
16	0.67	4654.	6.7287	0.47	**********
17	0.71	4653.	6.1385	0.43	**********
18	0.75	4649.	5.7982	0.41	**********
19	0.79	4649.	5.8701	0.41	**********
20	0.83	4649.	5.9077	0.41	**********
21	0.88	4650.	5.9970	0.42	**********
22	0.92	4642.	6.3635	0.45	**********
23	0.96	4647.	6.9281	0.49	**********
24	1.00	4821.	6.9043	0.48	**********
25	1.04	4641.	6.5532	0.46	**********
26	1.08	4630.	5.6517	0.40	**********
27	1.13	4635.	5.0764	0.36	**********
28	1.17	4629.	4.7532	0.33	**********
29	1.21	4631.	4.5752	0.32	**********
30	1.25	4629.	4.3712	0.31	**********
31	1.29	4628.	4.5193	0.32	**********
32	1.33	4625.	4.8069	0.34	**********
33	1.38	4623.	4.8594	0.34	**********
34	1.42	4623.	4.6457	0.33	**********
35	1.46	4622.	4.2611	0.30	**********
36	1.50	4624.	4.0158	0.28	**********
37	1.54	4619.	3.8681	0.27	**********
38	1.58	4615.	3.9214	0.28	**********
39	1.63	4614.	3.8904	0.27	**********
40	1.67	4613.	3.5895	0.25	**********
41	1.71	4610.	3.2495	0.23	**********
42	1.75	4607.	3.0883	0.22	**********
43	1.79	4608.	3.2100	0.23	**********
44	1.83	4607.	3.2331	0.23	**********
45	1.88	4609.	3.3285	0.23	**********
46	1.92	4604.	3.7180	0.26	**********
47	1.96	4611.	4.2509	0.30	**********
48	2.00	4786.	4.2991	0.30	**********
49	2.04	4610.	3.9776	0.28	**********
50	2.08	4602.	3.2491	0.23	**********

51	2.13	4604.	2.8643	0.20	I ***** I
52	2.17	4600.	2.7198	0.19	I **** I
53	2.21	4601.	2.6232	0.18	I **** I
54	2.25	4598.	2.4595	0.17	I **** I
55	2.29	4597.	2.6052	0.18	I **** I
56	2.33	4594.	2.8532	0.20	I ***** I
57	2.38	4592.	2.8956	0.20	I ***** I
58	2.42	4592.	2.6935	0.19	I **** I
59	2.46	4591.	2.4381	0.17	I **** I
60	2.50	4595.	2.3157	0.16	I **** I
61	2.54	4588.	2.3682	0.17	I **** I
62	2.58	4586.	2.5327	0.18	I **** I
63	2.63	4586.	2.5920	0.18	I **** I
64	2.67	4586.	2.2379	0.16	I **** I
65	2.71	4583.	1.8169	0.13	I *** I
66	2.75	4581.	1.6941	0.12	I *** I
67	2.79	4583.	1.8230	0.13	I *** I
68	2.83	4584.	1.9382	0.14	I *** I
69	2.88	4587.	2.1223	0.15	I *** I
70	2.92	4581.	2.5540	0.18	I **** I
71	2.96	4588.	3.1138	0.22	I ***** I
72	3.00	4755.	3.1948	0.22	I ***** I
73	3.04	4587.	2.9170	0.20	I ***** I
74	3.08	4578.	2.2716	0.16	I **** I
75	3.13	4582.	1.9535	0.14	I *** I

+---+---+---+---+---+---+---+---+---+---+

FOR THIS AUTOCORRELATION: MEAN = 4.3499 AND VARIANCE = 14.2485

WASHINGTON, DC - 1972

1973 CO WASHINGTON DC CAMP

NO. OF VALUES READ = 4716 NO OF VALID VALUES STORED = 4269

BASIC STATISTICS-----

N = 4269

MEAN = 3.67978382

2ND MOMENT (VAR.) = 9.82058430 STANDARD DEV. = 3.13378143

COEFFICIENT OF VARIATION = 0.85162103

3RD MOMENT = 108.013763 COEF. OF SKEWNESS = 3.50972843

4TH MOMENT = 2952.16772 COEF. OF KURTOSIS = 30.6102142

LOWEST VALUE = 0.0

HIGHEST VALUE = 50.0000000

SECOND HIGHEST VALUE = 45.0000000

HISTOGRAM OF RAW DATA-----

* * * DISTRIBUTION OF DATA (INTERVAL WIDTH = 1.0000) * * *

NO. OF VALUES LESS THAN -0.5000= 0 NO. OF ZERO VALUES = 150

INTERVAL	NUMBER	PERCENT	+----------HISTOGRAM----------+	CUM. FREQ.	PLOT POS.
-0.50- 0.50	150	3.514%	I*C** I	3.514%	3.513%
0.50- 1.50	661	15.484%	I********C****** I	18.997%	18.993%
1.50- 2.50	934	21.879%	I**********C** I	40.876%	40.866%
2.50- 3.50	770	18.037%	I********** C I	58.913%	58.899%
3.50- 4.50	645	15.109%	I********** C I	74.022%	74.005%
4.50- 5.50	397	9.300%	I********* C I	83.322%	83.302%
5.50- 6.50	218	5.107%	I***** C I	88.428%	88.407%
6.50- 7.50	148	3.467%	I*** C I	91.895%	91.873%
7.50- 8.50	84	1.968%	I** C I	93.863%	93.841%
8.50- 9.50	69	1.616%	I** C I	95.479%	95.457%
9.50- 10.50	38	0.890%	I* C I	96.369%	96.347%

10.50- 11.50	34	0.796%	I*	97.166%	97.143%
11.50- 12.50	27	0.632%	I*	97.798%	97.775%
12.50- 13.50	28	0.656%	I*	98.454%	98.431%
13.50- 14.50	17	0.398%	I	98.852%	98.829%
14.50- 15.50	3	0.070%	I	98.922%	98.899%
15.50- 16.50	9	0.211%	I	99.133%	99.110%
16.50- 17.50	9	0.211%	I	99.344%	99.321%
17.50- 18.50	5	0.117%	I	99.461%	99.438%
18.50- 19.50	5	0.117%	I	99.578%	99.555%
19.50- 20.50	7	0.164%	I	99.742%	99.719%
20.50- 21.50	2	0.047%	I	99.789%	99.766%
21.50- 22.50	1	0.023%	I	99.812%	99.789%
22.50- 23.50	1	0.023%	I	99.836%	99.812%
23.50- 24.50	1	0.023%	I	99.859%	99.836%
24.50- 25.50	1	0.023%	I	99.883%	99.859%
25.50- 26.50	2	0.047%	I	99.930%	99.906%
26.50- 27.50	0	0.0 %	I	99.930%	99.906%
27.50- 28.50	0	0.0 %	I	99.930%	99.906%
28.50- 29.50	0	0.0 %	I	99.930%	99.906%
29.50- 30.50	0	0.0 %	I	99.930%	99.906%
30.50- 31.50	0	0.0 %	I	99.930%	99.906%
31.50- 32.50	0	0.0 %	I	99.930%	99.906%
32.50- 33.50	0	0.0 %	I	99.930%	99.906%
33.50- 34.50	0	0.0 %	I	99.930%	99.906%
34.50- 35.50	0	0.0 %	I	99.930%	99.906%
35.50- 36.50	0	0.0 %	I	99.930%	99.906%
36.50- 37.50	0	0.0 %	I	99.930%	99.906%
37.50- 38.50	1	0.023%	I	99.953%	99.930%
38.50- 39.50	0	0.0 %	I	99.953%	99.930%
39.50- 40.50	0	0.0 %	I	99.953%	99.930%
40.50- 41.50	0	0.0 %	I	99.953%	99.930%
41.50- 42.50	0	0.0 %	I	99.953%	99.930%
42.50- 43.50	0	0.0 %	I	99.953%	99.930%
43.50- 44.50	0	0.0 %	I	99.953%	99.930%
44.50- 45.50	1	0.023%	I	99.976%	99.953%
45.50- 46.50	0	0.0 %	I	99.976%	99.953%
46.50- 47.50	0	0.0 %	I	99.976%	99.953%
47.50- 48.50	0	0.0 %	I	99.976%	99.953%
48.50- 49.50	0	0.0 %	I	99.976%	99.953%
49.50- 50.50	1	0.023%	I	100.000%	99.976%
50.50- 51.50	0	0.0 %	I	100.000%	99.976%
51.50- 52.50	0	0.0 %	I	100.000%	99.976%
52.50- 53.50	0	0.0 %	I	100.000%	99.976%
53.50- 54.50	0	0.0 %	I	100.000%	99.976%
54.50- 55.50	0	0.0 %	I	100.000%	99.976%
55.50- 56.50	0	0.0 %	I	100.000%	99.976%
56.50- 57.50	0	0.0 %	I	100.000%	99.976%
57.50- 58.50	0	0.0 %	I	100.000%	99.976%
58.50- 59.50	0	0.0 %	I	100.000%	99.976%

NO. OF VALUES EQUAL TO OR GREATER THAN 59.5000 = 0 (0.0 %)

WASHINGTON, DC - 1973

* * * * * AUTOCORRELATION RESULTS * * * * *

LAG HOURS	LAG DAY	PAIRS	COVARIANCE	CORRELATION COEFFICIENT	AUTOCORRELATION FUNCTION -1.0 -.8 -.6 -.4 -.2 0 +.2 +.4 +.6 +.8 +1.0
1	0.04	4013.	9.0552	0.92	**********
2	0.08	3969.	7.5383	0.77	**********
3	0.13	3957.	6.3346	0.65	**********
4	0.17	3934.	5.5812	0.57	**********
5	0.21	3932.	4.9620	0.51	**********
6	0.25	3923.	4.7508	0.48	**********
7	0.29	3900.	4.8453	0.49	**********
8	0.33	3895.	5.0723	0.52	**********
9	0.38	3881.	5.1392	0.52	**********
10	0.42	3866.	5.0831	0.52	**********
11	0.46	3853.	4.8573	0.49	**********
12	0.50	3845.	4.6792	0.48	**********
13	0.54	3843.	4.6192	0.47	**********
14	0.58	3849.	4.6556	0.47	**********
15	0.63	3857.	4.5377	0.46	**********
16	0.67	3853.	4.0733	0.41	**********
17	0.71	3850.	3.3883	0.35	**********
18	0.75	3866.	2.9708	0.30	**********
19	0.79	3863.	2.8490	0.29	**********
20	0.83	3856.	3.0056	0.31	**********
21	0.88	3868.	3.1792	0.32	**********
22	0.92	3862.	3.6459	0.37	**********
23	0.96	3875.	4.2858	0.44	**********
24	1.00	4043.	4.3066	0.44	**********
25	1.04	3873.	4.0277	0.41	**********
26	1.08	3869.	3.3367	0.34	**********
27	1.13	3869.	2.8402	0.29	**********
28	1.17	3857.	2.5785	0.26	**********
29	1.21	3869.	2.3925	0.24	**********
30	1.25	3869.	2.3263	0.24	**********
31	1.29	3854.	2.5368	0.26	**********
32	1.33	3853.	2.8698	0.29	**********
33	1.38	3848.	3.2087	0.33	**********
34	1.42	3841.	3.2413	0.33	**********
35	1.46	3835.	3.0756	0.31	**********
36	1.50	3835.	3.0021	0.31	**********
37	1.54	3831.	3.0452	0.31	**********
38	1.58	3828.	3.1253	0.32	**********
39	1.63	3829.	3.0727	0.31	**********
40	1.67	3825.	2.7057	0.28	**********
41	1.71	3820.	2.2930	0.23	**********
42	1.75	3836.	2.0702	0.21	**********
43	1.79	3834.	2.1045	0.21	**********
44	1.83	3828.	2.3153	0.24	**********
45	1.88	3837.	2.4402	0.25	**********
46	1.92	3825.	2.7917	0.28	**********
47	1.96	3833.	3.2215	0.33	**********
48	2.00	3995.	3.2369	0.33	**********
49	2.04	3840.	3.1428	0.32	**********
50	2.08	3837.	2.6161	0.27	**********

51	2.13	3842.	2.1432	0.22	I ***** I
52	2.17	3834.	1.9433	0.20	I **** I
53	2.21	3842.	1.7835	0.18	I **** I
54	2.25	3840.	1.7188	0.18	I **** I
55	2.29	3827.	1.8956	0.19	I **** I
56	2.33	3824.	2.2338	0.23	I ***** I
57	2.38	3824.	2.5473	0.26	I ****** I
58	2.42	3826.	2.6348	0.27	I ****** I
59	2.46	3824.	2.6268	0.27	I ****** I
60	2.50	3828.	2.6282	0.27	I ****** I
61	2.54	3816.	2.6211	0.27	I ****** I
62	2.58	3814.	2.6220	0.27	I ****** I
63	2.63	3817.	2.5768	0.26	I ****** I
64	2.67	3814.	2.2628	0.23	I ***** I
65	2.71	3811.	1.8252	0.19	I **** I
66	2.75	3827.	1.5569	0.16	I **** I
67	2.79	3824.	1.6734	0.17	I **** I
68	2.83	3818.	1.9529	0.20	I **** I
69	2.88	3828.	2.1462	0.22	I ***** I
70	2.92	3818.	2.4471	0.25	I ***** I
71	2.96	3818.	2.8345	0.29	I ****** I
72	3.00	3968.	2.7699	0.28	I ****** I
73	3.04	3818.	2.6006	0.26	I ****** I
74	3.08	3813.	2.0828	0.21	I ***** I
75	3.13	3819.	1.5563	0.16	I **** I

+---+---+---+---+---+---+---+---+---+---+

FOR THIS AUTOCORRELATION: MEAN = 3.6798 AND VARIANCE = 9.8195

WASHINGTON, DC - 1973

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE TESTING THE VALIDITY OF THE LOGNORMAL PROBABILITY MODEL: COMPUTER ANALYSIS OF CARBON MONOXIDE DATA FROM U.S. CITIES		5. REPORT DATE June 1979 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Wayne R. Ott, David T. Mage, and Victor W. Randecker		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS N/A		10. PROGRAM ELEMENT NO. 1AD809 11. CONTRACT/GRANT NO. N/A
12. SPONSORING AGENCY NAME AND ADDRESS Office of Monitoring and Technical Support Office of Research and Development U.S. Environmental Protection Agency Washington, D.C. 20460		13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE EPA/600/19
15. SUPPLEMENTARY NOTES		

16. ABSTRACT

A stratified sample consisting of 11 data sets from an original list of 166 CO air quality data sets in SAROAD was selected as a "national crossection" of observed U.S. CO concentrations, along with a longitudinal group of data sets from a single air monitoring station. The adequacy of the 2-parameter lognormal probability model (LN2) was tested using these data. A special-purpose computer program was developed for calculating the parameters of the LN2 model using four different techniques: (1) direct "method of moments," (2) Larsen's "method of fractiles," (3) "maximum likelihood estimation" (MLE) for grouped data using an MLE approximation technique, and (4) MLE for grouped data using computer optimization.

The goodness-of-fit of the LN2 model to the data was evaluated using frequency-based approaches (e.g., chi-square, magnitude of log-likelihood function, Kolmogorov-Smirnov measures) and variate-based approaches (e.g., the difference between the concentration predicted by the model for each interval and the concentration actually observed). The findings show that the method of calculating parameters for the LN2 model exerts a profound influence on the goodness-of-fit of the model to the data, particularly at the higher concentrations of greatest interest for protecting public health. The method of fractiles performs best for the variate-based tests but poorest for the frequency-based tests, suggesting that it is responding more to the randomness at the extremes than to the underlying distribution of the process.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Mathematical models, -Computer simulation, Systems analysis, Statistical analysis, Pollution - Air pollution, Sanitary engineering, Environmental engineering, Civil engineering	Environmental modeling, Data analysis, Computer Techniques, Probability Models, Environmental monitoring, Environ-metrics, Decision anysis	04B 05C 06F 12A 12B 13B

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED 20. SECURITY CLASS *(This page)* UNCLASSIFIED	21. NO. OF PAGES 22. PRICE

EPA Form 2220-1 (9-73)

★ U.S. GOVERNMENT PRINTING OFFICE: 1979-281-147/132