1996

Analysis of environmental pollutant data using generalized log-logistic distribution.

Warsono
University of Alabama at Birmingham

ANALYSIS OF ENVIRONMENTAL POLLUTANT DATA USING GENERALIZED LOG-LOGISTIC DISTRIBUTION

by

WARSONO

A DISSERTATION

Submitted in partial fulfillment of the requirement for the
degree of Doctor of Philosophy in the Department
of Biostatistics in the Graduate School, The
University of Alabama at Birmingham

BIRMINGHAM, ALABAMA

1996

UMI Number: 9704155

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

ABSTRACT OF DISSERTATION
GRADUATE SCHOOL, UNIVERSITY OF ALABAMA AT BIRMINGHAM

Degree Ph.D. Major Subject Biostatistics

Name of Candidate Warsono

Title Analysis of Environmental Pollutant Data Using Generalized Log-Logistic Distribution

Environmental pollution studies conducted to monitor ambient levels and to quantify the concentration of various pollutants entering a given environmental area are of great interest for possible adverse-health effects. Of particular importance in environmental data analysis is to select appropriate probability models. The previous studies indicated that none of the probability models, including the classical lognormal, has been identified to be superior to others in a general sense. To address this problem, the purpose of this study is twofold. Firstly, we introduce a generalized log-logistics distribution as a general model in fitting environmental pollutant data, and develop maximum-likelihood techniques for estimating parameters of the proposed distribution. The family of the generalized log-logistics distribution includes several well-known distributions in modeling data of environmental pollutant concentrations, such as lognormal, weibull, and gamma as special cases. Secondly, by applying the proposed model to seven data sets, we explore the possibilities of using this model as a general probability model for representing environmental-quality data.

The results of applications indicate that the generalized log-logistics distribution could be a good alternative to the classical lognormal distribution for fitting environmental quality data. For all of data sets, generally, the four-parameter GLL distribution fits better than the lognormal, log-logistic, and three-parameter GLL distributions, or provides at least as good a fit as the other distributions.

Abstract Approved by: Committee Chairman Ken P Smith

Program Director Edwin Bradley

Date 8/28/96 Dean of Graduate School John Rodden

DEDICATION

To my parents, Mrs. Onah and Mr. Wadri,

and to my wife, Sri Retno, and daughter, Nisa Larasati.

ACKNOWLEDGMENTS

I am deeply grateful to a number of persons and institutions for their support throughout the years of my academic career. First, I would like to express my sincere gratitude to my major professor, Dr. Karan P. Singh, for enthusiastically inspiring, mentoring, encouraging, and motivating me throughout my dissertation process and for spending countless hours in advising and refining my work. His patience, enthusiasm, and optimism have given me the confidence to complete this endeavor.

My appreciation extended to each member of my dissertation committee, Dr. Alfred A. Bartolucci, Dr. Edwin L. Bradley, Dr. Charles R. Katholi, Dr. Deodutta Roy, and Dr. Pauline Jolly for their helpful comments and suggestions, which served to improve this dissertation. I would also like to express appreciation to the faculty, staff, and classmates in the Department of Biostatistics for their help and encouragement.

The University of Lampung and Higher Education Development Support Project, both in Indonesia, deserve a word of thanks for giving me the opportunity to pursue a graduate program in the United States. I would also like to thank the United States Agency for International Development for providing me a scholarship during my doctoral study.

I thankfully acknowledge Dr. Singh and his family. Their kind hospitality, affection, and encouragement throughout my graduate study and stay in Birmingham, USA, have made me and my family feel close to home in this foreign land.

My special love and recognition must go to all my family members in Indonesia. The continuing love, support, and prayers of my parents, Mrs. Onah and Mr. Wadri, sisters, and brothers have been to me a mainstay of strength throughout the years of my education. The long-distance love and prayers of my parents-in-law, Mrs. Sri Suyati and Mr. Suratno, sisters-in-law, and brothers-in-law have also been a tremendous inspiration for me.

Last, but definitely not least, to my wife, Sri Retno, I cannot find enough words to express my genuine love and gratefulness for her unconditional love, endless encouragement, and prayerful support. The past six years since we have been married have been one of the happiest times of my life, though we have been through some difficult times together. My genuine love also goes to my little lovely daughter, Nisa Larasati, whose cuteness has made me laugh when life seems too serious and encouraged me to expedite the completion of this dissertation. There are others too numerous to mention, and I would like to thank each of you.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

CHAPTER I

INTRODUCTION

Environmental pollution studies conducted to monitor ambient levels and to quantify the concentration of various pollutants entering a given environmental area are of great interest for possible adverse-health effects. Selecting appropriate probability models for the data is an important step in environmental data analysis. These probability models may become the basis for estimating the parameters to meet the evolving information needs of environmental quality management. Unfortunately, the environmental pollution data are frequently skewed to the right; that is, they have a long tail toward high concentration. Therefore, the validity of applying the normal distribution for curve fitting of these type of data may be questioned. One way of modeling this type of distribution is to find a transformation of the data values so that the transformed values conform more closely to the normal distribution, and the logarithmic transformation is often applied in this context to pollution data. However, parameter estimates of the transformed data are rarely of interest. The estimate of the mean, for example, in the original scale of measurement is the primary purpose of environmental study.

A further complication is introduced by the fact that there are a number of observations measured as less than detection limit (DL) established by analytical laboratories. The analysts may report them as nondetect (ND) or less than detection limit (LDL) rather than numerical values. In statistical terminology, data sets containing such data are called left-censored data because the values of data below the detection limit are not available. Even if the data are normally distributed, the presence of left-censoring creates some difficulties when applying classical methods because one will be uncertain as to what to use for censored values. In practical applications, to handle the censored data, many analysts ignore the values of observations below the DL or set them equal to zero, the DL or the DL divided by two (DL/2) prior to parameters estimation. Replacing with the DL/2 implicitly assumes a uniform distribution between zero and the DL. But the deletion or the replacement gives biased estimates of the parameters, and the intensity of the bias will be worse as the degree of censoring increases (Newman et al., 1989). This is the main reason Newman et al. (1989) do not recommend the use of such techniques.

Many investigators in different environmental fields have reported that the pollutant concentrations measured in the environment have a lognormal distribution, or nearly so. Since environmental pollution data are inherently positive as well as highly skewed, the lognormal distribution is an ideal descriptor of such data, with a positively skewed, positive range, and heavy right tail. A significant number of technical reports have been published suggesting that empirical distributions of pollutant concentration data tend to be lognormally distributed. An excellent review of applications of lognormal models to aerometric data is presented by Mage (1981). Recently, Ott (1995) provided a review of applications of lognormal distribution for analysis of water-, soil-, and air-quality data.

Although the lognormal distribution has been widely employed to represent pollution concentration data, a fact that also should be pointed out is that it is possible that other distributions might work better. The "same" lognormal distribution with different parameters is sometimes appropriate. Dealing with air-quality data, Larsen (1977) added a third parameter, an increment, to the lognormal distribution. The third parameter is either a positive or negative increment that is added to every observed concentration until a curved log-probability plot is transformed into a fairly straight line. However, Mage and Ott (1978) called their model the censored three-parameter lognormal model, criticized in Larsen's 1977 article. Mage and Ott do not suggest the automatic use of a particular model, because failure to consider the validity of the model, if hypothesis tests are involved, can lead to predictions that are not supported under scientific scrutiny.

Other parametric distributions can also be used successfully with positively skewed data. Using the sum-of-squares error as the goodness-of-fit criterion, Bencala and Seinfeld (1976) showed that the Weibull model produces lower values than that of the lognormal for five of eight carbon monoxide (CO) data sets. Apt (1976) applied the Weibull distribution in relation to the distribution of atmospheric radioactivity data, with some success. He concluded that the Weibull distribution function can be used empirically to describe spatial and temporal distributions of atmospheric tritium oxide, gross-beta, and plutonium-239 concentrations.

Berger, Melice, and Demuth (1982) examined the goodness-of-fit based on the extreme values and the median in fitting a gamma distribution to daily atmospheric sulfur dioxide (SO2) concentrations in the Gent region of Belgium. They found that the gamma distribution provided a better representation of the whole ensemble than the usual lognormal. Jakeman and Taylor (1985) also observed that gamma models provide a better representation of acid-gas concentrations in an industrial airshed than does the lognormal model.

Comprehensive studies to report a comparison of the fits of several distributions to pollutant concentrations were carried out by a number of researchers, such as Bencala and Seinfeld (1976), Holland and Fitz-Simons (1982), Georgopulos and Seinfeld (1982), and Gilliom and Helsel (1986). Using procedures of goodness-of-fit tests, such as chi-square and log-likelihood tests, Taylor, Jakeman, and Simpson (1986) conducted a comparison study of the fits of lognormal, gamma, exponential, and Weibull distributions to extensive air-quality data in Melbourne, Australia. In their study, the lognormal distribution was the best for the majority of the NO, NOx, and SO2 data sets; the gamma distribution for O3, NO2, and CO; and the Weibull distribution for CO and SO2.

Obviously, none of the probability models, including the classical lognormal, has been identified to be superior to others in a general sense. One approach to overcoming this problem could be to use a very general model that includes most of the distributions. Among the general models, the generalized log-logistic (GLL) distribution has good potential for fitting environmental pollutant data. The GLL distribution is an extension of the log-logistic distribution. The log-logistic distribution is similar in shape to the lognormal distribution, but it may be more convenient to apply. This is because of its greater mathematical simplicity, especially when dealing with the censored data. As noted by Singh (1989), therefore, one advantage of the proposed model as a general model is the potential improvement in the fit to the data, while retaining mathematical simplicity.

In modeling lung cancer survival data, Singh (1989) demonstrated the application of three-parameter GLL models as alternatives to a log-logistic model. He considered that one, either the first or the second shape parameter of the shape parameters, has been assumed to be unity. Singh's work showed that the three-parameter GLL model fits the data better than log-logistic model. More recently, when applying breast cancer survival data, a similar conclusion was made by Singh, Bartolucci, and Burgard (1994).

Of particular interest is the fact that the GLL distribution may have attractive features. The family of the GLL distribution includes several well-known distributions in modeling contaminant concentrations, such as lognormal, Weibull, and gamma distributions as limiting distributions or special cases. Another interesting feature of the considered GLL model is that it might be linked to the other well-known generalized models reported upon in the literature.

It is clear that the common characteristics of environmental pollutant concentration data are skewed to the right, and the presence of left-censored data caused by the detection limit make it difficult to analyze. Though the lognormal distribution is often used for fitting pollutant concentration data, the data analyst should point out that it is possible that other distributions might work better. Moreover, the values of the parameter estimates depend considerably on the validity of the assumptions regarding the underlying distribution. Thus, selecting an appropriate probability model for the data set under study is a very important step in the analysis of environmental quality data.

One approach to determining an appropriate model is to use a very general model that includes a suitable model as a special case. Although in environmental studies the GLL distribution is a relatively "unknown" distribution, as mentioned earlier the skewness and the heavy tail of the GLL distributions seem to make it suitable for modeling environmental pollution data. Also, the family of the GLL distribution is quite rich and includes a number of submodels that are very common distributions in fitting pollutant concentration data. Therefore, the GLL distribution has desirable features and seems to be a promising distribution for environmental modeling. Thus, in this study we propose to consider the use of the GLL distribution in fitting pollutant concentration data. The proposed model may provide more flexibility to fit environmental data when the skewness, kurtosis, or other moments of the distribution fail to conform to lognormality. Thus, the GLL distribution may become a good alternative to the log-normal distribution. The overall objective is to provide analysts, especially those who work in environmental areas, more latitude in selecting various models.

The objective of this study is twofold. First, we develop maximum-likelihood techniques for estimating parameters of the GLL distributions. In particular, we consider the four-parameter GLL where we assume that $m_1 \neq m_2$, denoted by GLL(m_1,m_2). We also study the three-parameter GLL where we assume that $m_1 = m_2 = m$, denoted by GLL(m,m). Second, by applying the proposed model to various sets of data, we explore the possibility of using GLL distribution as a general probability model for representing environmental quality data. For purposes of comparison, we also consider the three-parameter GLL distribution where $m_1 = m$, and $m_2 = 1$, denoted by GLL(m,1); the three-parameter GLL distribution where $m_1 = 1$, and $m_2 = m$, denoted by GLL(1,m); the log-logistic distribution, denoted by GLL(1,1); and lognormal distribution.

The rest of the dissertation is conveniently organized as follows. Chapter II provides a brief historical review of the lognormal distribution and discusses the application of this distribution in fitting environmental pollutant data. Chapter III is devoted to the details of the generalized log-logistic distribution, especially to a discussion of the theoretical development of the maximum-likelihood estimation procedure for estimating parameters of the proposed distribution. Along with chapter IV, chapter III serves as the core of this dissertation. Chapter IV explores the possibilities of using the generalized log-logistic distributions as a general probability model in fitting data of environmental pollutant concentration and discusses a comparison with the lognormal distribution. Finally, chapter V gives directions and suggestions for further efforts and future research.

CHAPTER II

LOGNORMAL DISTRIBUTION

In this chapter, we provide a brief review of lognormal (LN) distribution. The LN distribution has been known and discussed for almost a century, but it was not until Aitchison and Brown (1957) published a monograph devoted entirely to the LN distribution. In this landmark book, they deal rigorously with the theory, beginning with a discussion of the genesis of the LN distribution, passing through estimation problems of parameters, and ending with a review of applications. Moreover, Johnson and Kotz (1970) summarized the history of theory and applications of LN distributions and the estimation procedures for the two-parameter lognormal (LN2), three-parameter lognormal (LN3), and related distributions.

Since the publication of the books by Aitchison and Brown and by Johnson and Kotz, the theory of LN distributions has steadily progressed and fields of application have greatly increased. A book, edited by Crow and Shimizu (1988), containing contributions from several experts has comprehensively discussed the more recent developments in the genesis, properties, and applications of LN distributions, estimation and test theories, and some results for related distributions. Recently, Johnson, Kotz, and Balakrishnan (1994) revised the book published by Johnson and Kotz (1970).

Shimizu and Crow (1988) stated three reasons why there is much to discuss about the LN distribution if the data analysis can be referred to the intensively studied normal distribution by taking the logarithm:

1. The parameter estimates resulting from the inverse transformation are biased.

2. The two-parameter distribution is often not a sufficient description; a third parameter, the threshold or location parameter is needed.

3. The distribution may be censored or truncated, or the data may be classified into groups, so that special methods are needed.

The LN distribution in its simplest form may be defined as the distribution of a random variable whose logarithm is normally distributed. Such a variable is essentially positive. The LN distribution is positively skewed and completely characterized by two parameters, a geometric mean and a standard geometric deviation, but may be generalized by a translation parameter, truncation and censoring, adjoining a point probability mass, extension to two or more dimensions, and transformation (Shimizu and Crow, 1988).

The random variable X is said to have a two-parameter lognormal distribution if the random variable $Y = \ln X$, where $0<X$, is normally distributed with mean μ and variance σ^2. The probability density function (PDF) of X is given by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x-\mu)^2}, \quad 0<x.$$

The rth moment of X about the origin is expressed as

$$\mu_r = E(X^r) = e^{r\mu+\frac{1}{2}r^2\sigma^2}.$$

From the properties of the moment generating function of the normal distribution, the corresponding mean and variance are respectively

$$E(X) = e^{\mu+\frac{1}{2}\sigma^2}$$

and

$$Var(X) = e^{2\mu} e^{\sigma^2} [e^{\sigma^2}-1].$$

The median is

$$Med(X) = e^{\mu},$$

and the mode is

$$Mode(X) = e^{\mu-\sigma^2}.$$

The relation among the mean, median, and mode of the lognormal distribution is

$$\mathrm{Mode}(X) < \mathrm{Med}(X) < \mathrm{E}(X).$$

The cumulative distribution function (CDF) of the lognormal distribution is

$$F(x)=\Phi\left(\frac{\ln x-\mu}{\sigma}\right)$$

where $\Phi(z)$ is the CDF of standardized normally distributed random variable Z, expressed by

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\ , \quad -\infty<z<\infty.$$

Now we review the estimation of the parameters of the LN distribution by the method of maximum-likelihood estimation, that is, by determining values of these parameters that maximize the likelihood function. In order to get the maximum-likelihood estimates, we first need to form the likelihood function determined by the product of the PDF values for the n observations. Thus, the likelihood function can be expressed by

$$
\begin{aligned}
L(\mu,\sigma|\underline{x}) &= \prod_{i=1}^{n} f(x_i) \\
&= \prod_{i=1}^{n} \frac{1}{x_i\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(\ln x_i-\mu)^2} \\
&= [\frac{1}{\sigma\sqrt{2\pi}}]^n\, [\prod_{i=1}^{n} \frac{1}{x_i}]\, e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln x_i-\mu)^2} .
\end{aligned}
$$

To find the values of the parameters that maximize the likelihood function, it is generally more convenient to work with the logarithm of the likelihood function. Maximizing the log-likelihood function is equivalent to maximizing the likelihood function because the logarithmic transformation is monotonic. This practice simplifies problems in statistical inference because many probability models contain exponential terms. Thus, we form the log-likelihood function by taking the logarithm of L defined as follows

$$
l(\mu,\sigma\underline{x}) = n \ln[\frac{1}{\sigma\sqrt{2\pi}}] - \sum_{i=1}^{n} \ln(x_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} [\ln(x_i) - \mu]^2.
$$

Taking the first-order partial derivatives of the log-likelihood function with respect to μ and σ, and then setting the resulting functions equal to zero, we will obtain the maximum-likelihood estimates of parameters of the LN distribution. Then the maximum-likelihood estimates of μ and σ, respectively, are

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

and

$$
\hat{\sigma} = [\frac{1}{n}\sum_{i=1}^{n} (\ln x_i-\mu)^2]^{1/2}.
$$

Detailed discussions of the maximum-likelihood estimation methodology for LN2 distribution of censored data sets are provided by Aitchinson and Brown (1957), Kushner (1976), Gilbert and Kinnison (1981), Holland and Fitz-Simons (1982), and Gilbert (1987). Similar discussions for three-parameter lognormally distributed data sets are available in literatures by Harter and Moore (1966), Tiku (1968), Gilbert (1987), and Cohen (1988). Recently, Johnson et al. (1994) furnished an excellent review of the maximum-likelihood methods for estimation of the lognormally distributed parameters.

Mage and Ott (1984) made an attempt to show an effectiveness of three different approaches (i.e., method of fractiles, method of moment, and method of maximum likelihood) for estimating parameters using the LN distribution as a parent distribution. Their results showed that if air-quality observations actually arise from a stationary LN distribution, then the maximum-likelihood approach gives minimum variance estimates, as also noted elsewhere (Holland and Fitz-Simons, 1982). Furthermore, Ott (1995) stated that if one is sampling from a true LN distribution, then it can be shown that the minimum-variance estimate of the parameters is obtained by the maximum-likelihood estimation. In detail, Kendall and Stuart (1979) discussed the properties of the maximum-likelihood estimators.

The LN distribution has a long history of application in the field of environmental pollution. A rich literature has been published over the past two decades, suggesting that pollutant concentration data tend to be lognormally distributed. The decision to apply the LN distribution in fitting pollutant concentration data can be attributed to the work of Larsen (1969, 1973, and 1974). Using graphical techniques, he concluded that regardless of pollutant, city, or averaging time, the air concentration distributions are approximately lognormally distributed. An excellent review of the history of the applications of probability models, especially lognormal models, to aerometric data is given by Mage (1981). Under lognormally assumption, El-Shaarawi (1989) examined several methods for making inferences about the levels of many metals and organic contaminants in ambient water samples from the Niagara River. More recently, the applications of the lognormal model to air-, soil-, and water-quality data are presented in considerable detail by Ott (1995).

Having developed physical mechanisms generating environmental quality data, Ott (1995) provided an argument as to why the LN distribution is so ubiquitous in environmental phenomena. The LN distribution has been fitted not only for air quality data, as mentioned in this book, but also for water quality and geological data . Ott's explanations involve the central limit theorem and the diffusion law.

The other examples of the LN distribution can be found in the articles of Kalpasanov and Kurchatova (1976), Kushner (1976), Owen and DeRouen (1980), and Mage and Ott (1984) for air-quality data; Schubert, Brodsky, and Tyler (1967) and Gilbert and Kinnison (1980) for radionuclide data; and Gilliom and Helsel (1986), Helsel and Gilliom (1986), Newman et al. (1989), and Stoline (1991) for water-quality data.

A number of investigators also considered other distributional forms for environmental-quality data. Using the sum-of-squares error as the goodness-of-fit criterion, Bencala and Seinfeld (1976) showed that Weibull models produce lower values than that of the lognormal model for five of eight CO data sets. But they stated that the LN model is convenient from a practical point of view. A similar study comparing the LN model with other probability models is also carried out by other reseachers, for example, by Berger et al. (1982), Simpson, Butt, and Jakeman (1984), Jakeman and Taylor (1985), and Taylor et al. (1986). Georgopoulos and Seinfeld (1982) presented a critical review of statistical distributions, such as Weibull, gamma, and many others, and stated that the LN distribution has been the most popular in representing urban air pollutant concentration data.

CHAPTER III

GENERALIZED LOG-LOGISTIC DISTRIBUTION

In the statistical literature, the modeling of data by generalized probability models has been noted to be advantageous by numerous authors, because selecting the best probability model in a particular case is not an easy task. Each generalization usually includes normal, lognormal, and gamma distributions as its either limiting or special cases; consequently, the generalized models must provide at least as good a fit as other special models. A detailed discussion of these as well as many other related distributions are provided by McDonald and Richards (1987). In practice, however, the flexibility of the generalized model to include the distributional shapes involves intensive computations.

A general guideline of model selection of some generalized models, such as the generalized beta of the first and second kind (GB1, GB2), has been outlined by McDonald and Richards (1987). Applications of the distributions of returns on stocks, failure times of ball bearings, and incomes are discussed in their work. To fit survival data, Ciampi, Hogg, and Kates (1986) proposed a generalized F (GF) family as an alternative to the proportional hazards model. Moreover, Singh (1989) suggested a generalized log-logistic (GLL) distribution, which is a natural extension of the log-logistic (LL) distribution in modeling data of lung and other cancers.

In relation to the LN distribution, it may be worthwhile to represent a brief discussion of the GLL distribution as an extension of the LL distribution provided by Singh (1989). The LL distribution is roughly similar in shape to the LN distribution, but a mathematical simplicity of the LL model seems to make it more attractive than the LN model. This is especially true when there are some censored observations. He then proposed a GLL model that reflects the skewness and the structure of the heavy tail of the model in fitting skewed hazard functions, while retaining mathematical simplicity. He also demonstrated the flexibility of the GLL model in fitting lung cancer survival data. Further illustrations of the GLL application in modeling breast cancer survival data are given by Singh et al. (1994).

Four-Parameter Generalized Log-Logistic Distribution

Let a random variable X have four-parameter GLL distribution with shape parameters m_1 and m_2, denoted by $X \sim \mathrm{GLL}(m_1,m_2)$, and then the CDF can be written in the form

$$G(x) = \frac{1}{\mathrm{B}(m_1,m_2)} \int_0^{F(x)} w^{m_1-1} (1-w)^{m_2-1} \, dw \quad ; x,\ m_1,\ m_2 > 0$$

where $\mathrm{B}(m_1,m_2)$ represents the complete beta function, which is defined as follows

$$\mathrm{B}(m_1,m_2) = \frac{\Gamma(m_1)\,\Gamma(m_2)}{\Gamma(m_1+m_2)}$$

where Γ is the gamma function, and

$$F(x) = [1 + e^{-[\beta+\alpha\ln(x)]}]^{-1}$$

is the log-logistic distribution function.

The corresponding PDF can be obtained by differentiating the CDF with respect to x as follows:

$$g(x) = \frac{d}{dx} G(x)$$
$$= \frac{d}{dx} [\frac{1}{B(m_1,m_2)} \int_0^{F(x)} w^{m_1-1} (1-w)^{m_2-1} dw]$$
$$= \frac{1}{B(m_1,m_2)} [F(x)]^{m_1-1} [1-F(x)]^{m_2-1} \frac{d}{dx}F(x).$$

But

$$\frac{d}{dx}F(x) = \frac{d}{dx} [1 + e^{-(\beta+\alpha\ln(x))}]^{-1}$$
$$= -[1+e^{-(\beta+\alpha\ln(x))}]^{-2} \cdot e^{-(\beta+\alpha\ln(x))} \cdot -\frac{\alpha}{x}$$
$$= \frac{\alpha}{x} [F(x)]^2 [\frac{1}{F(x)}-1]$$
$$= \frac{\alpha}{x} [F(x)]^2 [\frac{1-F(x)}{F(x)}]$$
$$= \frac{\alpha}{x} [F(x)] [1-F(x)].$$

After simplification, the PDF of the GLL(m_1,m_2) distribution then becomes expressible as

$$g(x) = \frac{\alpha}{xB(m_1,m_2)} [F(x)]^{m_1} [1-F(x)]^{m_2}.$$

In the PDF of GLL distribution, β and α are the location parameter and the scale parameter, respectively, and (m_1,m_2) are shape parameters. Note that if $m_1 = m_2 = 1$, GLL(m_1,m_2) distribution reduces to the LL distribution. Observe that g(x) is positively skewed if $m_1 > m_2$ and negatively skewed if $m_1 < m_2$.

In order to gain more insight into the flexibility of GLL distribution, it is worthwhile to discuss its properties in terms of a hazard function. Let T be a random variable of the survival time. Setting $X = T$, the hazard function is given by

$$h(t) = \frac{g(t)}{[1-G(t)]} .$$

The hazard functions of the GLL model include increasing (I), decreasing (D), unimodal (∩), and bathtub (U) shaped hazar rates (Singh, 1988; Singh et al., 1994). Graphs of hazard functions of GLL (m_1,m_2) model for various values of parameters can be shown in Figures 1-3.

The family of GLL distributions is quite rich and contains many distributions as limiting distributions or special cases. These distributions include the Weibull, LN, and gamma distribution, which are very popular in modeling environmental pollution data. Singh (1989) stated that the GLL model is a reparameterization of other well-known generalized models, such as the generalized F (GF) model discussed by Ciampi et al. (1986) and the generalized beta of second type (GB2) model considered by McDonald (1984) and McDonald and Richards (1987). The GLL model may also be linked to the generalized gamma (GG) model (Singh et al., 1994).

Before proceeding to generate the special case of the GLL distribution, it may be useful to give a discussion about its relationship to the other generalized models. Letting $\alpha = 1/\sigma$ and $\beta = -\mu/\sigma + \ln(m_1/m_2)$, it can easily be shown that the PDF of GLL(m_1,m_2) becomes

$$g(x) = \frac{1}{\sigma B(m_1,m_2)} \frac{e^{(-\mu m_1/\sigma)}\,(m_1/m_2)^{m_1}\,x^{(m_1/\sigma)-1}}{[1+(m_1/m_2)\,(e^{-\mu}x)^{1/\sigma}]^{m_1+m_2}}$$

which is the GF model deliberated by Hogg and Ciampi (1985) and Ciampi et al. (1986).

It can be seen that the GLL distribution and the GF distribution are similar. Following Hogg and Ciampi (1985), Ciampi et al. (1986) showed that the GG distribution is a special case of the GF distribution. Note that, in fitting air pollutant concentrations, the GG model

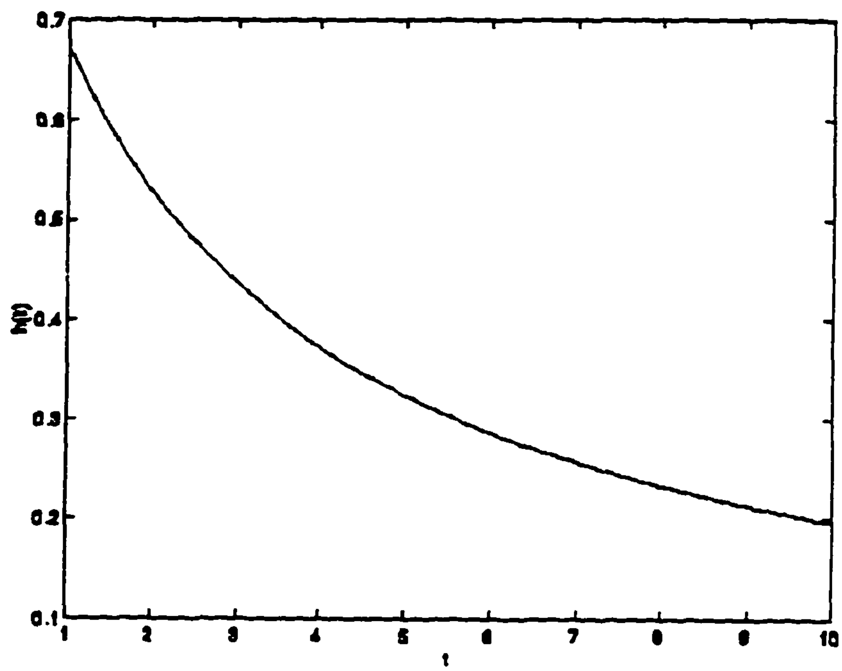


Figure 1. Graph of the hazard function for $\alpha = 1$, $\beta = -1$, $m_1 = 1$, and $m_2 = 2.5$.

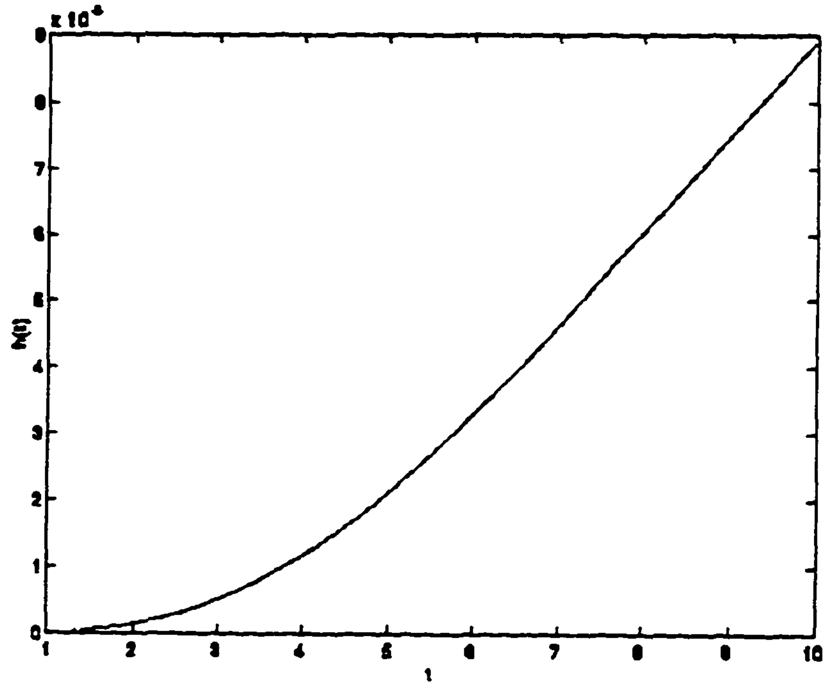


Figure 2. Graph of the hazard function for $\alpha = 1$, $\beta = -3$, $m_1 = 5$, and $m_2 = 2.5$.

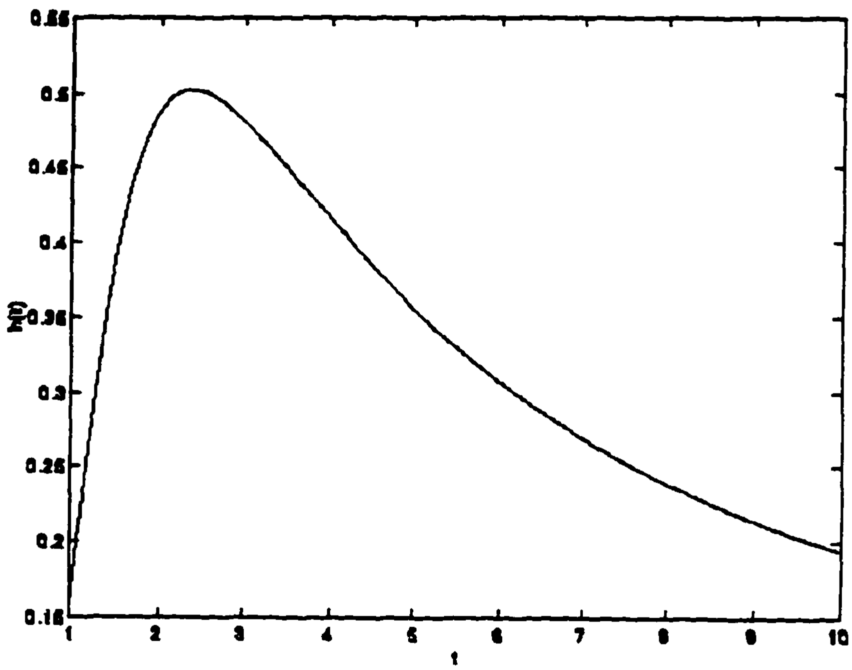


Figure 3. Graph of the hazard function for $\alpha = 2$, $\beta = -0.5$, $m_1 = 5$, and $m_2 = 1$.

has been proposed by Marani, Lavagni, and Buttazzoni (1986) and Okur (1988). Furthermore, a number of authors, such as Ciampi et al. (1986) and DiCiccio (1987), have demonstrated the relationship between the GG distribution with other parametric distributions including the LN, Weibull, and gamma distributions as limiting or special cases.

It can also be shown that if we let α=a and β=-a ln(b) in GLL(m_1,m_2), the PDF of GLL(m_1,m_2) becomes

$$g(x) = \frac{a\, x^{am_1 - 1}}{b^{am_1}\, \mathrm{B}(m_1,m_2)\, [1+(x/b)^a]^{m_1+m_2}}$$

which is the pdf of GB2 distribution considered by McDonald (1984) and McDonald and Richards (1987). Moreover, McDonald (1984) showed that for $m_2 \to \infty$ and b=γ(m_2)(1/a), the GB2 model will reduce to the GG model, that is, the GG distribution, which is a special case of the GB2 distribution as well.

Therefore, it is clear that the GG, LN, Weibull, and gamma distributions are the limiting or special cases of the GLL distribution, because the GLL distribution is similar to the GF and GB2 distributions. Consequently, the GLL model should provide at least as good a fit as these models.

Now we discuss the estimation of the parameters of the GLL distribution by the method of maximum-likelihood estimation. For k^{th} sample, let n_1 and n_2 be, respectively, the number uncensored and left censored observations. Let n^* be the number of uncensored observations in N random samples, where

$$n^* = \sum_{k=1}^{N} n_{1k}.$$

The likelihood function can be defined as

$$L(\alpha,\beta,m_1,m_2|\underline{x}) = \prod_{k=1}^{N}\{\prod_{i=1}^{n_{1k}} g_k(x_i) \,.\, \prod_{j=1}^{n_{2k}} G_k(x_j)\}.$$

If there is no left censored observation, the likelihood function is expressed by

$$L(\alpha,\beta,m_1,m_2|\underline{x}) = \prod_{k=1}^{N}\{\prod_{i=1}^{n_{1k}} g_k(x_i)\}.$$

The likelihood function of the GLL(m_1,m_2) is defined as follows

$$\begin{aligned} L(\alpha,\beta,m_1,m_2|\underline{x}) &= \prod_{k=1}^{N}\{\prod_{i=1}^{n_{1k}} \frac{\alpha}{x_i}\frac{1}{B(m_1,m_2)}[F_k(x_i)]^{m_1}\,[1-F_k(x_i)]^{m_2}\,. \\ &\quad \prod_{j=1}^{n_{2k}}\frac{1}{B(m_1,m_2)}\int_0^{F_k(x_j)} w^{m_1-1}\,(1-w)^{m_2-1}\,dw\,\} \\ &= \prod_{k=1}^{N}\{\prod_{i=1}^{n_{1k}} \frac{\alpha}{x_i}\frac{1}{B(m_1,m_2)}[(F_k(x_i)]^{m_1}\,[1-F_k(x_i)]^{m_2}. \\ &\quad \prod_{j=1}^{n_{2k}} I_{(m_1,m_2)}\,[F_k(x_j)]\,\} \end{aligned}$$

where

$$I_{(m_1,m_2)}\,(x) = \frac{1}{B(m_1,m_2)}\int_0^{x} w^{m_1-1}\,(1-w)^{m_2-1}\,dw\,; \quad 0\le x\le 1$$

is the incomplete beta function with parameters m_1 and m_2.

The log-likelihood function then can be expressed by

$$\begin{aligned} l\,(\alpha,\beta,m_1,m_2|\underline{x}) &= \sum_{k=1}^{N}\{\,n_{1k}\,[\,[\ln(\alpha)] - [\ln[B(m_1,m_2)]]\,] + \\ &\quad m_1\sum_{i=1}^{n_{1k}}\ln[F_k(x_i)] + m_2\sum_{i=1}^{n_{1k}}\ln[1-F_k(x_i)] - \sum_{i=1}^{n_{1k}}\ln(x_i) + \\ &\quad \sum_{j=1}^{n_{2k}}\ln(\,I_{(m_1,m_2)}\,[F_k(x_j)]\,)\,\}. \end{aligned}$$

In order to obtain the maximum-likelihood estimators, we shall take first order derivatives of the log-likelihood function with respect to α, β, m_1, and m_2, and we then set the resulting functions equal to zero. Therefore, the maximum likelihood estimates of parameters are given by the solutions of the following equations

$$\frac{\partial\, l\,(\alpha,\beta,m_1,m_2|\underline{x})}{\partial\alpha} = 0 \quad , \quad \frac{\partial\, l\,(\alpha,\beta,m_1,m_2|\underline{x})}{\partial\beta} = 0$$

and

$$\frac{\partial\, l\,(\alpha,\beta,m_1,m_2|\underline{x})}{\partial m_1} = 0 \quad , \quad \frac{\partial\, l\,(\alpha,\beta,m_1,m_2|\underline{x})}{\partial m_2} = 0.$$

The first-order derivatives of the log-likelihood function with respect to α, β, m_1, and m_2 are

$$\frac{\partial l}{\partial\alpha} = \sum_{k=1}^{N}\{ \frac{n_{1k}}{\alpha} + m_1 \sum_{i=1}^{n_{1k}} [\ln(x_i)]\ [1-F_k(x_i)] - m_2 \sum_{i=1}^{n_{1k}} [\ln(x_i)]\ [F_k(x_i)] + \frac{1}{B(m_1,m_2)} \sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} [\ln(x_j)]\ [F_k(x_j)]^{m_1}\ [1-F_k(x_j)]^{m_2} \}$$

$$\frac{\partial l}{\partial\beta} = \sum_{k=1}^{N}\{ m_1 n_{1k} - (m_1+m_2) \sum_{i=1}^{n_{1k}} [F_k(x_i)] + \frac{1}{B(m_1,m_2)} \sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} [F_k(x_j)]^{m_1}\ [1-F_k(x_j)]^{m_2} \}$$

$$\frac{\partial l}{\partial m_1} = \sum_{k=1}^{N}\{ -n_{1k}\ [\Psi(m_1)-\Psi(m_1+m_2)] + \sum_{i=1}^{n_{1k}} \ln[F_k(x_i)] + \sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1} \}$$

and

$$\frac{\partial l}{\partial m_2} = \sum_{k=1}^{N} \{ -n_{1k} \, [\Psi(m_2)-\Psi(m_1+m_2)] + \sum_{i=1}^{n_{1k}} \ln[1-F_k(x_i)] +$$
$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2} \}$$

where $\Psi(m)$ is the psi (digamma) function defined by

$$\Psi(m) = \frac{d}{dm}[\ln \Gamma(m)]$$
$$= \frac{\Gamma'(m)}{\Gamma(m)} .$$

Let $\underline{\theta}$ be the parameter vector, that is, $\underline{\theta}' = (\alpha,\beta,m_1,m_2)$. An asymptotic variance-covariance matrix, $V(\underline{\theta})$, can be obtained by taking inverse of Fisher information matrix, $I(\underline{\theta})$. The Fisher information matrix has (i,j) element

$$- E\left\{ \frac{\partial^2 l}{\partial\theta_i \partial\theta_j} \right\} .$$

The derivation of the first- and second-order derivatives of the log-likelihood function with respect to α, β, m_1, and m_2 are given in Appendix A.

The GLL model is attractive in form, but solving for the estimates of above parameters using maximum-likelihood techniques is not at all straightforward. Differentiating the gamma or beta function and dealing with some nonlinear equations simultaneously cause some computational difficulties. Because the log-likelihood equations are nonlinear in α, β, m_1, and m_2, the maximum-likelihood estimates can not be solved analytically; therefore, iteration techniques are needed.

Fortunately, to solve nonlinear equations, which typically involves an optimization (maximization or minimization) of a function, there are good computer programs available in some packages of computer programs, such as MATLAB version 4, particularly in the

FMINS procedure used for the purpose of minimizing a function of several variables. The algorithm utilized in this procedure is the Nelder-Mead simplex (Hanselman and Littlefield, 1995). Special functions, such as incomplete beta, natural logarithm of the beta function, incomplete gamma, and natural logarithm of the gamma function, are available in MATLAB.

Three-Parameter Generalized Log-logistic Distribution

The CDF of GLL(m,m) can be written in the form

$$G(x) = \frac{1}{\mathrm{B}(m,m)} \int_0^{F(x)} [w\,(1-w)]^{m-1}\, dw.$$

The corresponding PDF can be obtained by differentiating the CDF with respect to x:

$$\begin{aligned} g(x) &= \frac{d}{dx}\, G(x) \\ &= \frac{d}{dx}\left[\frac{1}{\mathrm{B}(m,m)} \int_0^{F(x)} [w\,(1-w)^{m-1} dw\right] \\ &= \frac{1}{\mathrm{B}(m,m)}\, [F(x)]^{m-1}\, [1-F(x)]^{m-1}\, \frac{d}{dx}F(x)\,. \end{aligned}$$

But

$$\begin{aligned} \frac{d}{dx}F(x) &= \frac{d}{dx}\,[1 + e^{-(\beta+\alpha\ln(x))}]^{-1} \\ &= -[1+e^{-(\beta+\alpha\ln(x))}]^{-2}\,.\, e^{-(\beta+\alpha\ln(x))}\,.\, -\frac{\alpha}{x} \\ &= \frac{\alpha}{x}\, [F(x)]^2 \left[\frac{1}{F(x)}-1\right] \\ &= \frac{\alpha}{x}\, [F(x)]^2 \left[\frac{1-F(x)}{F(x)}\right] \\ &= \frac{\alpha}{x}\, [F(x)]\, [1-F(x)]. \end{aligned}$$

After simplification, the PDF of the GLL(m,m) distribution is given by

$$g(x) = \frac{\alpha}{x\mathrm{B}(m,m)}\, [F(x)]^m\, [1-F(x)]^m.$$

The likelihood function is expressed as follows:

$$L(\alpha,\beta,m|\underline{x}) = \prod_{k=1}^{N}\{\prod_{i=1}^{n_{1k}} \frac{\alpha}{x_i}\frac{1}{B(m,m)} [(F_k(x_i))\ (1-F_k(x_i))]^m \cdot$$
$$\prod_{j=1}^{n_{2k}} \frac{1}{B(m,m)} \int_0^{F_k(x_j)} [w(1-w)]^{m-1}dw\ \}.$$

The above equation can be rewritten as

$$L(\alpha,\beta,m|\underline{x}) = \prod_{k=1}^{N}\left\{\prod_{i=1}^{n_{1k}} \frac{\alpha}{x_i}\frac{1}{B(m,m)} [F_k(x_i)]^m\ [1-F_k(x_i)]^m \prod_{j=1}^{n_{2k}} I_m\ [F_k(x_j)]\right\}$$

where

$$I_{(a,b)}(x) = \frac{1}{B(a,b)}\int_0^x (w)^{a-1}\ (1-w)^{b-1}dw$$

is the incomplete beta function with parameters a and b. For notation simplicity, we write $I_c(x)$ for $I_{(c,c)}(x)$ whenever a = b = c.

The log-likelihood function then can be defined as

$$l\ (\alpha,\beta,m|\underline{x}) = \sum_{k=1}^{N}\{\ n_{1k}\ [\ \ln(\alpha)-\ln[B(m,m)]]\ +$$
$$m\sum_{i=1}^{n_{1k}}[\ [\ln(F_k(x_i)]+\ln[1-F_k(x_i)]\ ]\ -\ \sum_{i=1}^{n_{1k}}\ln(x_i)\ +$$
$$\sum_{j=1}^{n_{2k}}\ln[I_m\ [F_k(x_j)]]\ \}.$$

The first-order derivatives of the log-likelihood function with respect to α, β, and m are

$$\frac{\partial l}{\partial \alpha} = \sum_{k=1}^{N}\{\ \frac{n_{1k}}{\alpha} + m\sum_{i=1}^{n_{1k}}[\ln(x_i)]\ [1-F_k(x_i)]\ +$$
$$\frac{1}{B(m,m)}\sum_{j=1}^{n_{2k}}\frac{1}{I_{(m)}[F_k(x_j)]}[\ln(x_j)]\ [F_k(x_j)]^m\ [1-F_k(x_j)]^m\ \ \}$$

$$\frac{\partial l}{\partial \beta} = \sum_{k=1}^{N}\{ mn_{1k} -2m \sum_{i=1}^{n_{1k}} [F_k(x_i)] + \frac{1}{B(m,m)} \sum_{j=1}^{n_{2k}} \frac{1}{I_m[F_k(x_j)]} [F_k(x_j)]^m [1-F_k(x_j)]^m \}$$

and

$$\begin{aligned}\frac{\partial l}{\partial m} = \sum_{k=1}^{N}\{ -\frac{n_{1k}}{B(m,m)} \frac{\partial B(m,m)}{\partial m} + \sum_{i=1}^{n_{1k}}[\ln[F_k(x_i)] + \ln[1-F_k(x_i)]] +\\ m \sum_{i=1}^{n_{1k}} [\frac{1}{F_k(x_i)} \frac{\partial F_k(x_i)}{\partial m} + \frac{-1}{1-F_k(x_i)} \frac{\partial F_k(x_i)}{\partial m}] +\\ \sum_{j=1}^{n_{2k}} \frac{1}{I_m[F_k(x_j)]} \frac{\partial I_m[F_k(x_j)]}{\partial m} \}.\end{aligned}$$

But

$$\begin{aligned}\frac{\partial B(m,m)}{\partial m} &= B(m,m) \frac{\partial}{\partial m}[\log(B(m,m)]\\ &= B(m,m) \frac{\partial}{\partial m}[\log\frac{[\Gamma(m)]^2}{\Gamma(2m)}]\\ &= B(m,m) \frac{\partial}{\partial m}[2\log(\Gamma(m) - \log(\Gamma(2m)]\\ &= B(m,m)\ [2 \frac{\Gamma'(m)}{\Gamma(m)} - 2 \frac{\Gamma'(2m)}{\Gamma(2m)}]\\ &= 2\ B(m,m)\ [\Psi(m) - \Psi(2m)]\end{aligned}$$

where $\Psi(m) = \Gamma'(m) / \Gamma(m)$ is the digamma function. Then

$$\begin{aligned}\frac{\partial l}{\partial m} &= \sum_{k=1}^{N}\{ -2\ n_{1k}\ [\Psi(m) - \Psi(2m)] + \sum_{i=1}^{n_{1k}}[\ln[F_k(x_i)] + \ln[1-F_k(x_i)]] +\\ &\quad \sum_{j=1}^{n_{2k}} \frac{1}{I_m[F_k(x_j)]} \frac{\partial I_m[F_k(x_j)]}{\partial m} \}\\ &= \sum_{k=1}^{N}\{ -2\ n_{1k}\ [\Psi(m) - \Psi(2m)] + \sum_{i=1}^{n_{1k}} \ln[\ [F_k(x_i)]\ [1-F_k(x_i)]\ ] +\\ &\quad \sum_{j=1}^{n_{2k}} \frac{1}{I_m[F_k(x_j)]} \frac{\partial I_m[F_k(x_j)]}{\partial m} \}.\end{aligned}$$

The derivation of the first- and second-order derivatives of the log-likelihood function with respect to α, β, and m are given in Appendix A.

CHAPTER IV

APPLICATIONS OF MODELS TO DATA SETS

In order to explore the possibilities of using the GLL distribution as a general probability model for representing some statistical characteristics of environmental-quality data, in this section, the four-parameter and three-parameter GLL distributions are applied to several environmental pollutant data sets and the results are compared with those obtained using log-logistic and lognormal distributions. The comparison indicates that the GLL(m_1,m_2) distribution seems to be a promising probability model for fitting environmental-quality data.

Example 4.1

This example uses uncensored data of mercury concentration in ppm in 115 sample swordfish published by Lee and Krutchkoff (1980). The maximum-likelihood estimates of the parameters obtained by the GLL(m_1,m_2) model are $\hat{\alpha}$ = 14.0428, $\hat{\beta}$ = -5.4922, $\hat{m}_1$ = 0.1192, and $\hat{m}_2$ = 0.4342. The 95% asymptotic confidence intervals for α, β, m_1, and m_2 are [13.7578,14.3278], [-5.6749,-5.3095], [0.1167,0.1217], and [0.4094,0.4580], respectively. The GLL(m,m) fit yields $\hat{\alpha}$ = 17.5504, $\hat{\beta}$ = -1.7498, and $\hat{m}$ = 0.1303. The 95% asymptotic confidence intervals for α, β, and m are [17.1572,17.9436], [-1.9612,-1.5384], and [0.1257,0.1349], respectively. Given in Table 1 are the log-likelihood function, and Akaike Information Criterion (AIC) under GLL(m_1,m_2), GLL(m,m), GLL(m,1), GLL(1,m), and GLL(1,1) distributions. For the purpose of comparison, the data set is also fitted by the lognormal distribution.

According to Table 1, it is clear that the value of the log-likelihood of the GLL(m_1,m_2) distribution is considerably larger than those of the lognormal and GLL(1,1) distributions and slightly larger than those of the GLL(m,1), GLL(1,m), and GLL(m,m) distributions. Note, also, that these values of the GLL(m,1), GLL(m,1), and GLL(m,m) distributions are remarkably larger than that of the traditional lognormal distribution and the GLL(1,1) distribution is slightly larger than that of the lognormal distribution. Therefore, by looking of the maximum log-likelihood values, the GLL(m_1,m_2) distribution seems to be a better statistical model in fitting data of mercury concentration. Moreover, the AIC of the GLL(m_1,m_2) and GLL(m,1) distributions are considerably lower than those of log-logistic, GLL(1,m), and GLL(m,m) distributions. Consequently, from the AIC value standpoint, the GLL(m_1,m_2) and GLL(m,1) distributions may provide better description of the data of the mercury concentration.

Figures 4-6 present graphs of the CDFs of compared models superimposed on the empirical distribution function, which is defined as the proportion of sample observations less than or equal to x. The graphs suggest that there is improvement in fit using the GLL distributions. In particular, the GLL(m_1,m_2) performs considerably better than the lognormal, GLL(1,1), GLL(1,m), GLL(m,m) distributions and slightly better than GLL(m,1) distribution. Notice that the other GLL distributions also appear to do better than the classical lognormal distribution.

Table 1

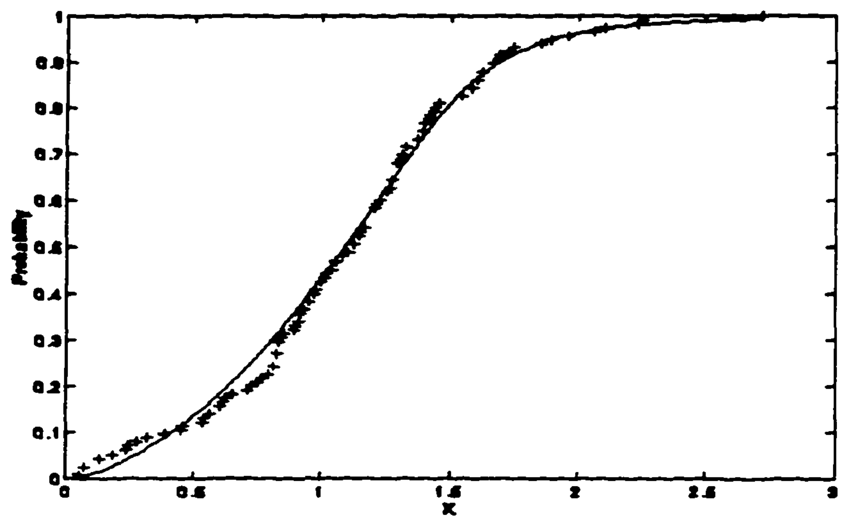
Values of the log-likelihood functions and of AIC for models fitted to mercury data

Model	Log-likelihood	AIC
GLL(1,1)	-100.4040	204.8080
GLL(m,1)	-81.1847	168.3694
GLL(1,m)	-85.5972	177.1944
GLL(m,m)	-93.9988	193.9976
GLL(m_1,m_2)	-80.9181	169.8362
Lognormal	-114.1681	-

Example 4.2

This example uses uncensored data of tritium oxide concentration of 26 air-sampling stations in the Los Alamos Scientific Laboratory, published by Apt (1976). The maximum-likelihood estimates of the parameters obtained by the GLL(m_1,m_2) model are $\hat{\alpha}$ = 0.6063, $\hat{\beta}$ = -0.3132, $\hat{m}_1$ = 28.3270, and $\hat{m}_2$ = 5.6582. The 95% asymptotic confidence intervals for α, β, m_1, and m_2 are [0.5730,0.6396], [-0.4189,-0.2075], [28.3094,28.3446], and [5.6573,5.6591], respectively. The GLL(m,m) fit yields $\hat{\alpha}$ = 0.1160, $\hat{\beta}$ = -0.3832, and $\hat{m}$ = 239.7601. The 95% asymptotic confidence intervals for α, β, and m are [0.1039,0.1281], [-0.7881,0.0235], and [150.9256,328.4866], respectively. The maximum log-likelihood values and AIC under GLL(m_1,m_2), GLL(m,m), GLL(m,1), GLL(1,m), and GLL(1,1) distributions are given in Table 2. For the purpose of comparison, we also

GLL(m_1,m_2) Distribution



GLL(m,m) Distribution

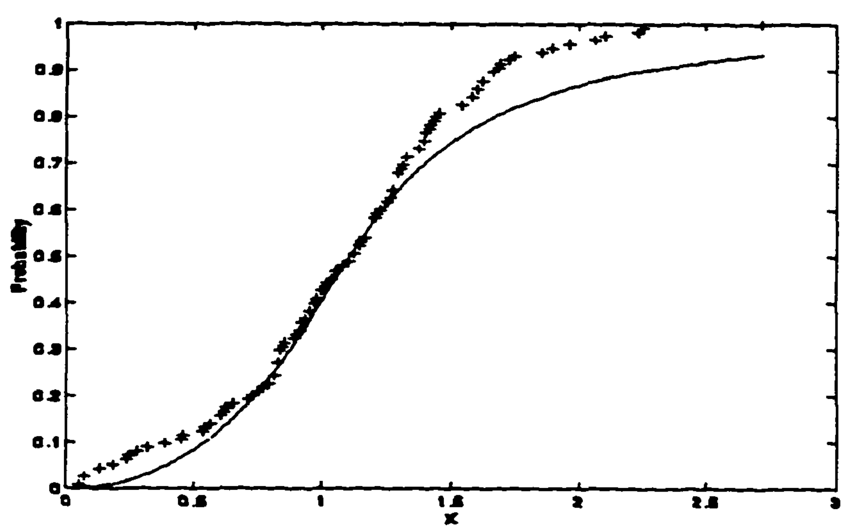
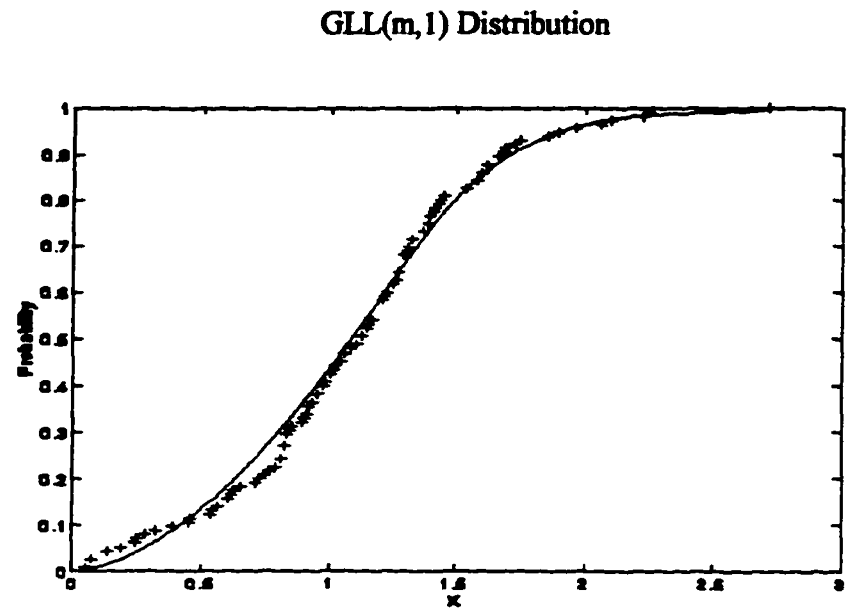


Figure 4. Graphs of the CDFs of the GLL(m_1,m_2) and GLL(m,m) distributions and the empirical distribution function (+ line) of mercury concentrations.



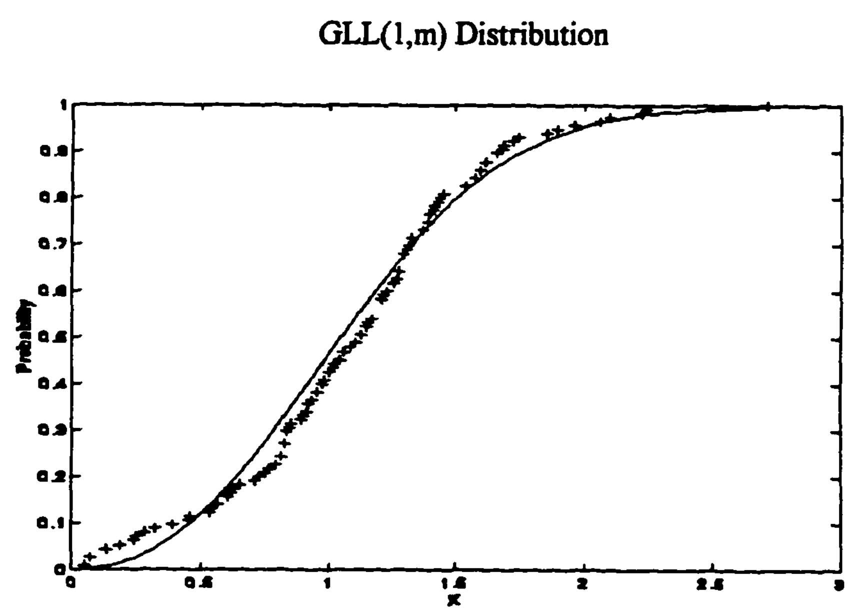
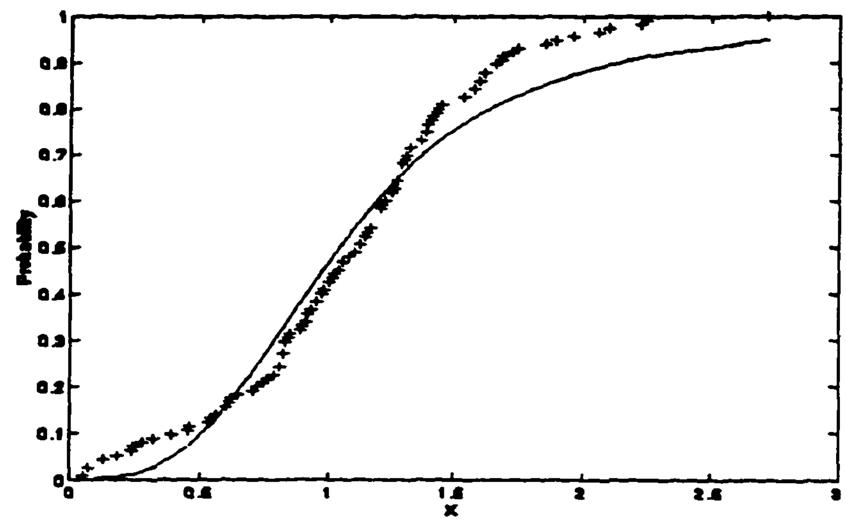


Figure 5. Graphs of the CDFs of the GLL(m,1) and GLL(1,m) distributions and the empirical distribution function (+ line) of mercury concentrations.

GLL(1,1) Distribution



Lognormal Distribution

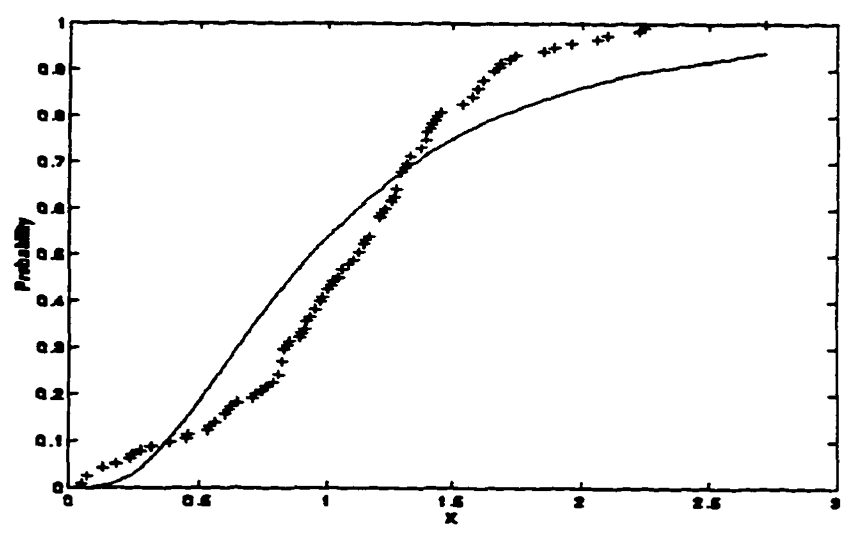


Figure 6. Graphs of the CDFs of the GLL(1,1) and lognormal distributions and the empirical distribution function (+ line) of mercury concentrations.

calculate the value of the log-likelihood function of the lognormal distribution when data are fitted by the lognormal distribution.

Table 2 shows that the value of log-likelihood of GLL(m_1,m_2) distribution is slightly larger than those of the lognormal, GLL(1,1), GLL(m,1), GLL(1,m), and GLL(m,m) distributions. Note that the value of log-likelihood of the lognormal is somewhat larger than those of the other GLL distributions. From the value of log-likelihood function standpoint, therefore, the GLL(m_1,m_2) model seems to be a promising model in fitting data of tritium oxide concentration.

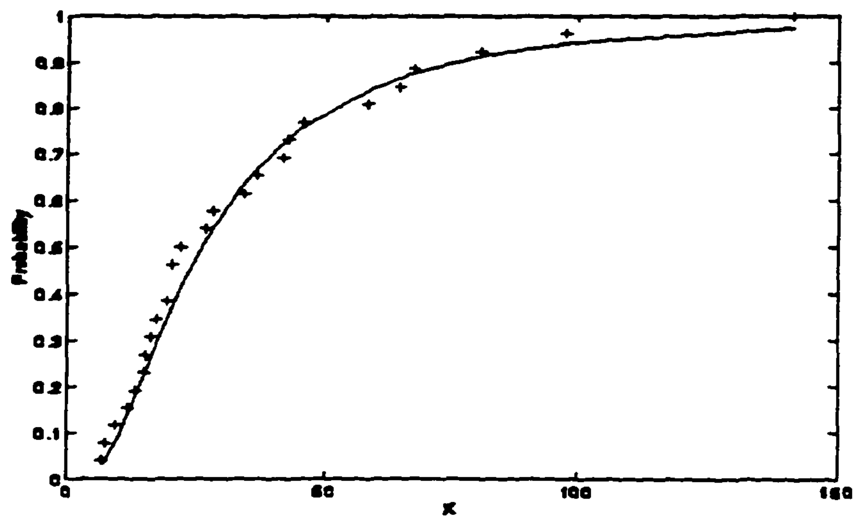
Table 2

Values of the log-likelihood functions and of AIC for models fitted to tritium oxide data

Model	Log-likelihood	AIC
GLL(1,1)	-117.0293	238.0586
GLL(m,1)	-116.5560	239.1120
GLL(1,m)	-116.9560	239.9120
GLL(m,m)	-116.2774	238.5548
GLL(m_1,m_2)	-116.1760	240.3520
Lognormal	-116.2738	-

Graphs of CDFs of models superimposed on the empirical distribution function suggest that there is a slight improvement in fit using the GLL(m_1,m_2) distribution. The GLL(1,1), GLL(m,1), GLL(m,1), and lognormal distribution show a similar performance in

GLL(m_1,m_2) Distribution



GLL(m,m) Distribution

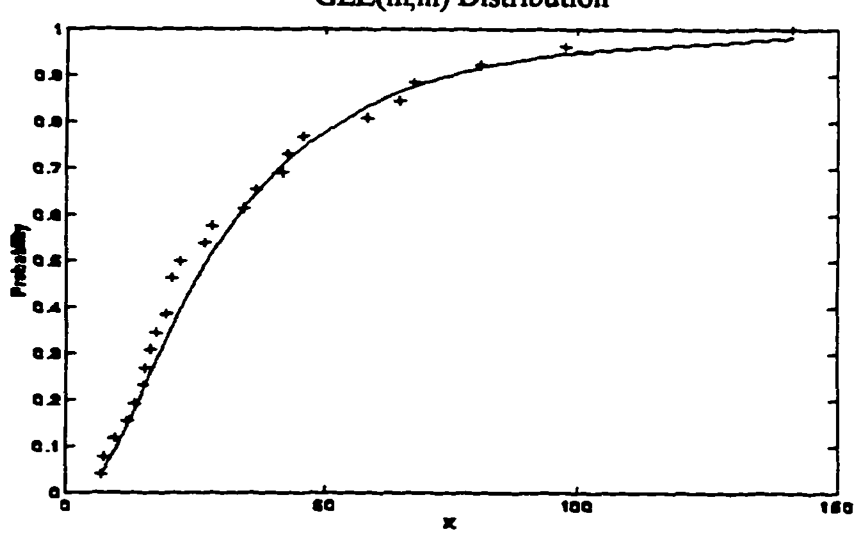
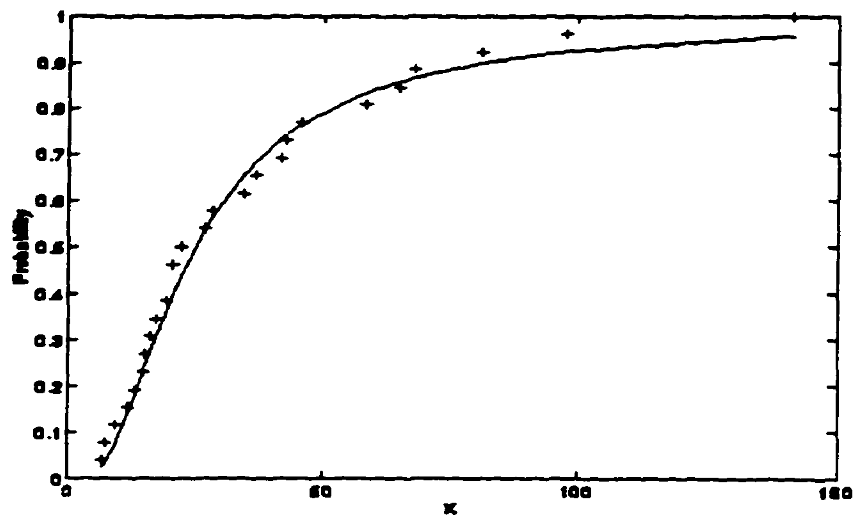


Figure 7. Graphs of the CDFs of the GLL(m_1,m_2) and GLL(m,m) distributions and the empirical distribution function (+ line) of tritium oxide concentrations.

GLL(m,1) Distribution



GLL(1,m) Distribution

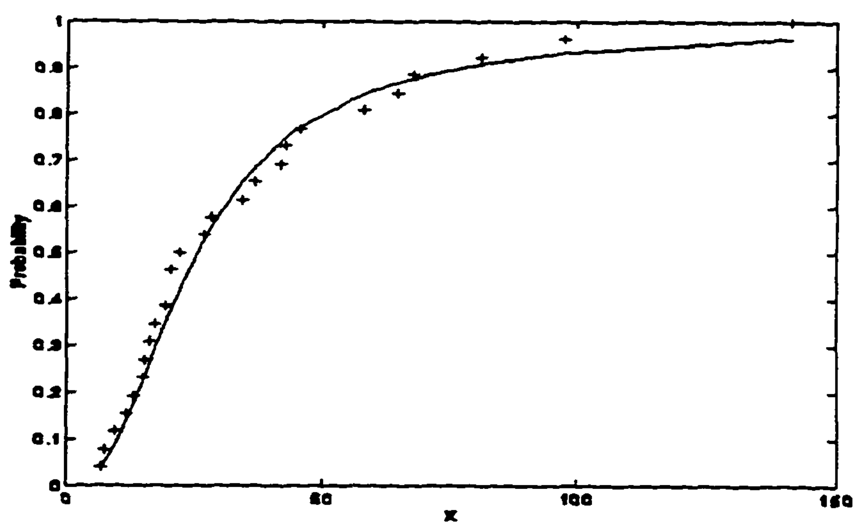
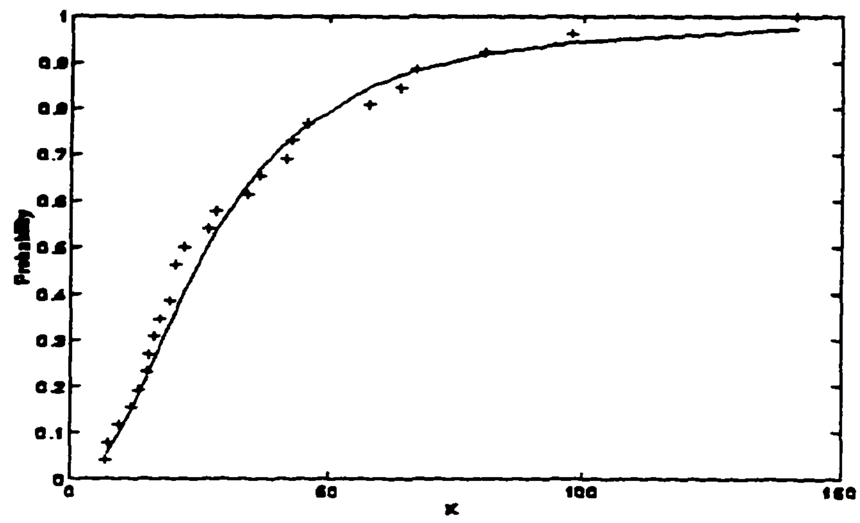


Figure 8. Graphs of the CDFs of the GLL(m,1) and GLL(1,m) distributions and the empirical distribution function (+ line) of tritium oxide concentrations.

GLL(1,1) Distribution



Lognormal Distribution

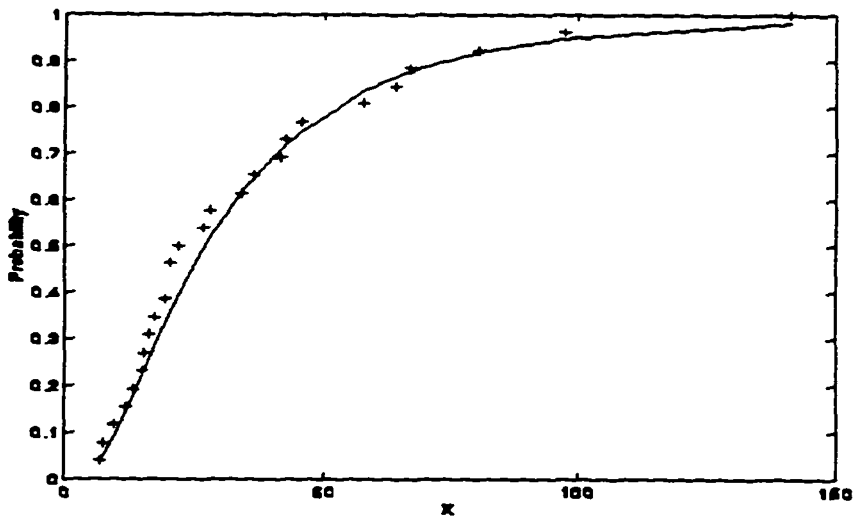


Figure 9. Graphs of the CDFs of the GLL(1,1) and lognormal distributions and the empirical distribution function (+ line) of tritium oxide concentrations.

fitting data. See Figures 7-9. Hence, graphically, the GLL(m_1,m_2) distribution seems to be a better distribution in fitting data of tritium oxide concentration.

Example 4.3

This example uses data of copper concentrations with 34 uncensored and 14 censored observations in the San Joaquin Valley, California, published by Millard and Deverel (1988). The maximum-likelihood estimates of the parameters obtained by the GLL(m_1,m_2) model are $\hat{\alpha}$ = 0.6560, $\hat{\beta}$ = 1.8748, $\hat{m}_1$ = 41.9214, and $\hat{m}_2$ = 3.6089. The 95% asymptotic confidence intervals for α, β, m_1, and m_2 are [0.6551,0.6569], [1.8739,1.8758], [41.5963, 42.2465], and [3.5251,3.6927], respectively. The GLL(m,m) fit yields $\hat{\alpha}$ = 0.1103, $\hat{\beta}$ = -0.1184, and $\hat{m}$ = 199.5571. The 95% asymptotic confidence intervals for of α, β, and m are [0.11029,0.11031], [-0.707,0.4702], and [198.4403,200.6739], respectively. Table 3 contains the log-likelihood function and AIC under GLL(m_1,m_2), GLL(m,m), GLL(m,1), GLL(1,m), and GLL(1,1) or log-logistic distributions. Also, for the purpose of comparison, we fitted the lognormal distribution to the the same data set.

From Table 3, it can be seen that the value of log-likelihood of GLL(m_1,m_2) distribution is larger than those of the lognormal, GLL(1,1), GLL(m,1), GLL(1,m), and GLL(m,m) distributions. Also note that the these values of the other GLL distributions are larger than that of the lognormal distribution. Hence, on the basis of the values of log-likelihood functions, the GLL(m_1,m_2) model seems to be a promising model in fitting data of copper concentration. However, the AIC of the GLL(1,1) distribution is slightly lower than the compared distribution. Thus, in this case, because of mathematical simplicity, the log-logistic distribution is preferable over the three-parameter and four-parameter GLL distributions.

Table 3

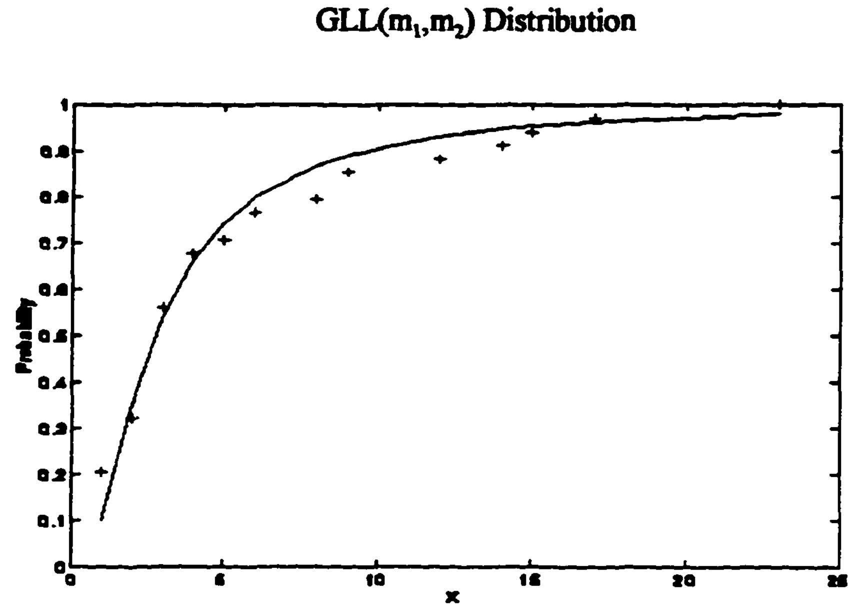
Values of the log-likelihood functions and of AIC for models fitted to copper data

Model	Log-likelihood	AIC
GLL(1,1)	- 96.9766	238.0586
GLL(m,1)	- 96.4600	239.1120
GLL(1,m)	- 96.9296	239.9120
GLL(m,m)	- 96.3422	238.5548
GLL(m_1,m_2)	- 96.0383	240.3520
Lognormal	-100.0472	-

Figures 10-12 show graphs of CDFs of models superimposed on the empirical distribution function. The graphs suggest that the performances of the GLL(m_1,m_2), GLL(m,m), GLL(m,1), GLL(1,m), GLL(1,1), and lognormal are similar. In other words, the GLL(m_1,m_2) distribution fits the data as well as the GLL(m,m), GLL(m,1), GLL(1,m), GLL(1,1), and lognormal distributions in fitting data of copper concentrations.

Example 4.4

This example is selected from Roberts (1979). For uncensored data of sulfur dioxide (SO2) concentrations with 19 observations, the maximum-likelihood estimates of the parameters obtained by using the GLL(m_1,m_2) model are $\hat{\alpha}$ = 1.0919, $\hat{\beta}$ = -2.6069, $\hat{m}_1$ = 11.5423, and $\hat{m}_2$ = 61.4388. The 95% asymptotic confidence intervals for α, β, m_1, and m_2 are [1.0108,1.173], [-2.6761,-2.5377], [11.5391,11.5455], and [61.4295,61.4481], respectively. The GLL(m,m) fit yields $\hat{\alpha}$ = 0.2793, $\hat{\beta}$ = -0.2301, and $\hat{m}$ = 286.9014.



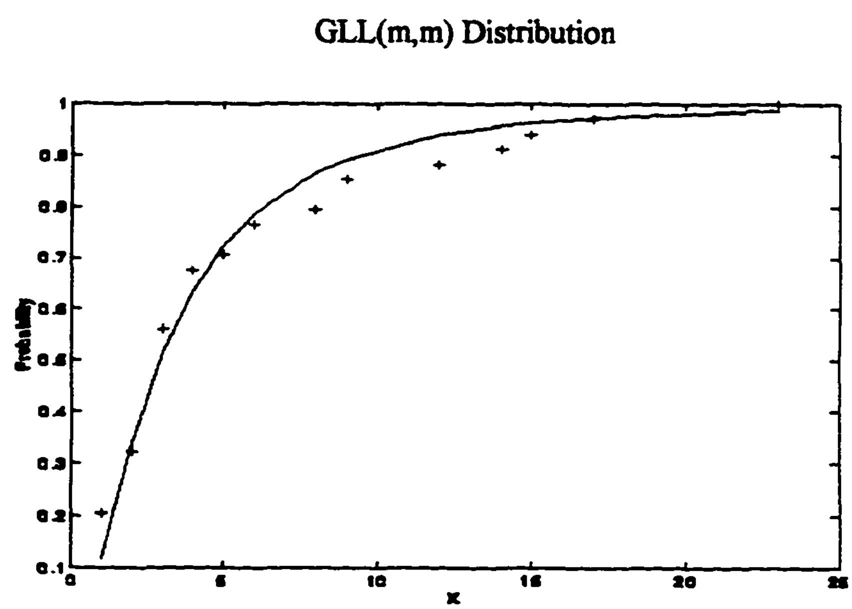
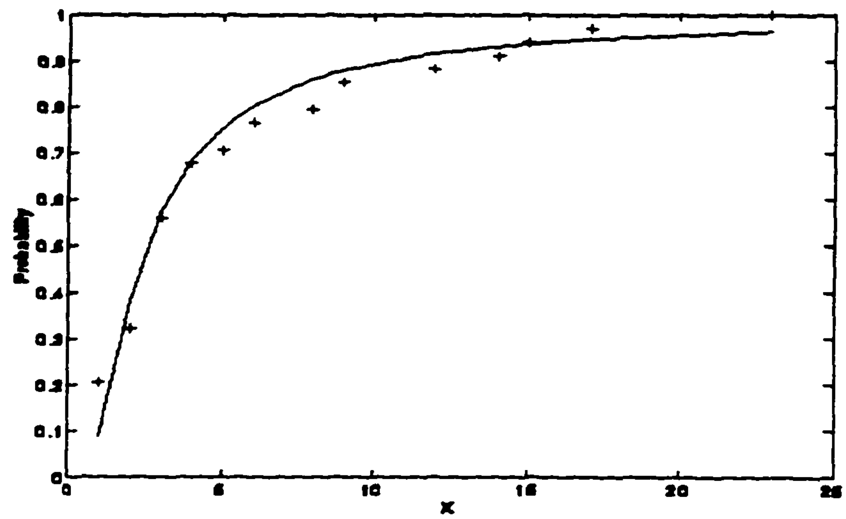


Figure 10. Graphs of the CDFs of the GLL(m_1,m_2) and GLL(m,m) distributions and the empirical distribution function (+ line) of copper concentrations.

GLL(m,1) Distribution



GLL(1,m) Distribution

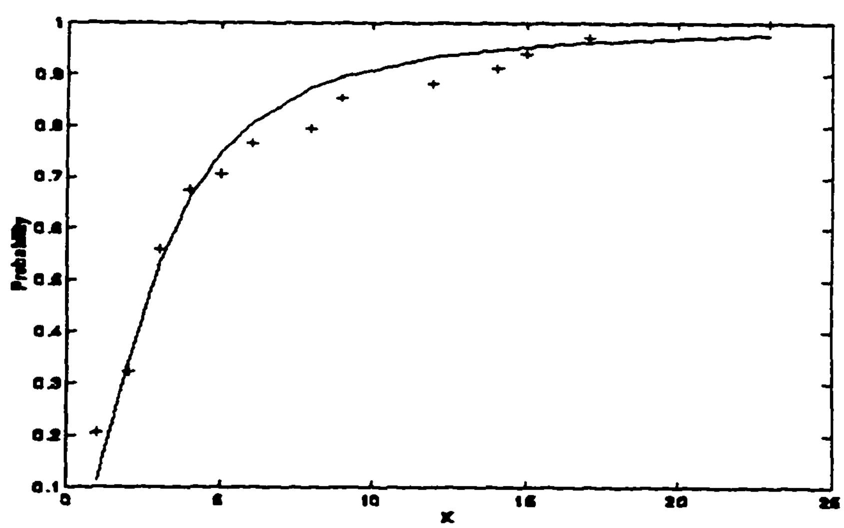
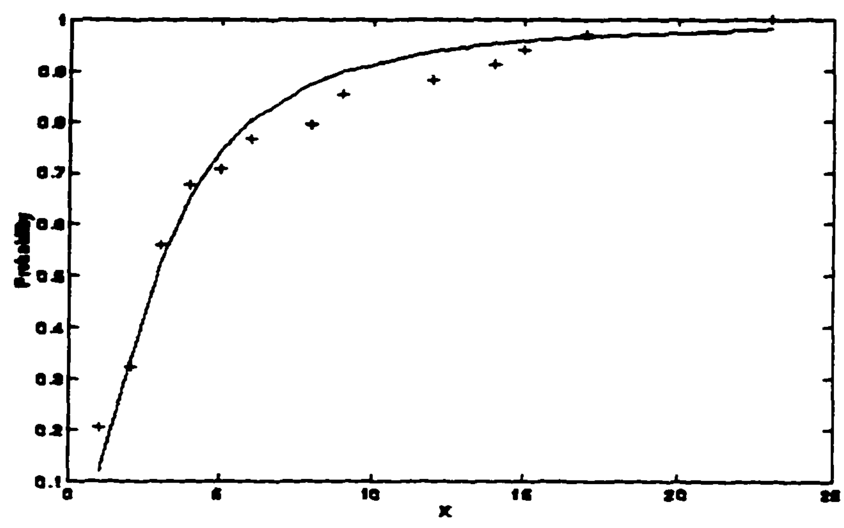


Figure 11. Graphs of the CDFs of the GLL(m,1) and GLL(1,m) distributions and the empirical distribution function (+ line) of copper concentrations.

GLL(1,1) Distribution



Lognormal Distribution

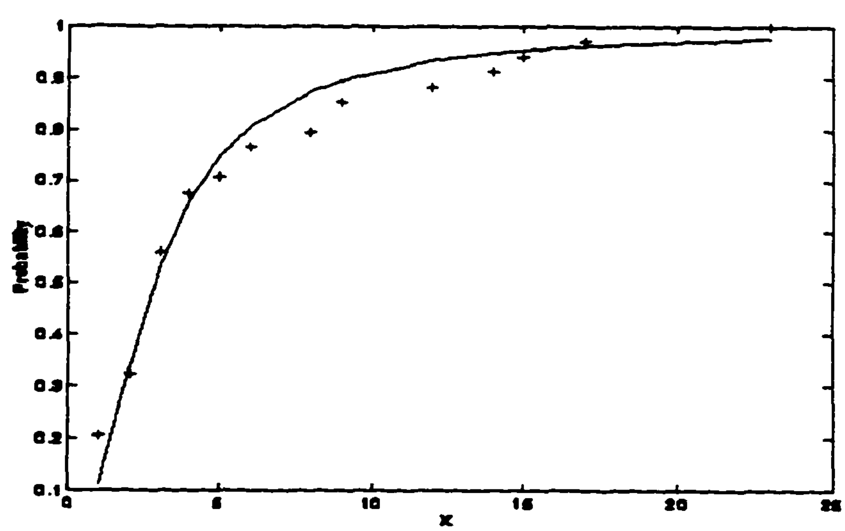


Figure 12. Graphs of the CDFs of the GLL(1,1) and lognormal distributions and the empirical distribution function (+ line) of copper concentrations.

The 95% asymptotic confidence intervals for the MLE of α, β, and m are [0.2376,0.321], [-0.7262,0.266], and [136.324,437.4788], respectively. The maximum log-likelihood values and AIC under GLL(m_1,m_2), GLL(m,m), GLL(m,1), GLL(1,m), and GLL(1,1) distributions are exhibited in Table 4.4. For the purpose of comparison, we also fitted the lognormal distribution to the same data set.

Table 4 shows that the value of log-likelihood of GLL(m_1,m_2) distribution is larger than those of the lognormal, GLL(1,1), GLL(m,1), GLL(1,m), and GLL(m,m) distributions. Thus, on the basis of the values of log-likelihood functions, the GLL(m_1,m_2) model shows better fit than the other models in fitting data of SO2 concentration. But, as can be seen from Table 4, the AIC of the GLL(1,1) distribution is somewhat lower than those of the other GLL distributions. Therefore, in terms of the AIC criterion, there is a slight improvement when the data are fitted by the GLL(1,1) distribution.

Figures 13-15 display graphs of the CDFs of models superimposed on the empirical distribution function. The graphs suggest that the performances of the GLL(m_1,m_2), GLL(m,m), GLL(m,1), GLL(1,m), GLL(1,1), and lognormal in fitting data of SO2 concentration are similar. In other words, the GLL(m_1,m_2) distribution fits the data as well as the lognormal and the other compared distribution.

Example 4.5

This example is selected from U.S. Environmental Protection Agency (EPA) (1979). For data of chloride concentrations, the maximum-likelihood estimates of the parameters obtained by the GLL(m_1,m_2) model are $\hat{\alpha}$ = 2.8195, $\hat{\beta}$ = -5.5151, $\hat{m}_1$ = 147.002, and $\hat{m}_2$ = 0.7440. The 95% asymptotic confidence intervals for of α, β, m_1, and m_2 are [2.3728,3.2662], [-7.2449,-3.7853], [143.5248,150.4792], and [0.739,0.7489], respectively.

Table 4

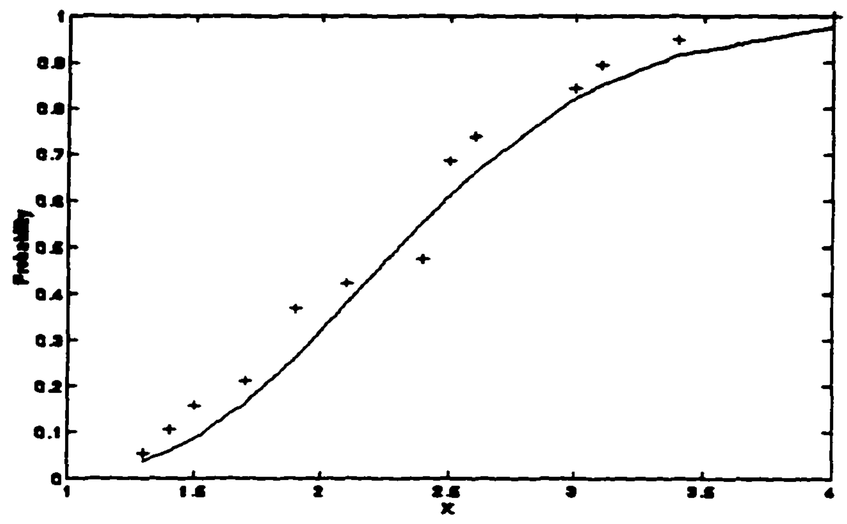
Values of the log-likelihood functions and of AIC for models fitted to SO2 data

Model	Log-likelihood	AIC
GLL(1,1)	-20.1639	44.3278
GLL(m,1)	-20.0832	46.1664
GLL(1,m)	-19.8655	45.7310
GLL(m,m)	-19.6455	45.2910
GLL(m_1,m_2)	-19.5852	47.1704
Lognormal	-19.6431	-

The GLL(m,m) fit yields $\hat{\alpha} = 20.2786$, $\hat{\beta} = -82.0547$, and $\hat{m} = 0.1222$. The 95% asymptotic confidence intervals for α, β, and m are [20.0454,20.5118], [-83.0929,-81.0165], and [0.1199,0.1245], respectively. The maximum log-likelihood values and AIC under GLL(m_1,m_2), GLL(m,m), GLL(m,1), GLL(1,m), and GLL(1,1) distributions are given in Table 5. For the purpose of comparison, we also calculate the value of the log-likelihood function of the lognormal distribution when data are fitted by using the lognormal distribution.

It can be seen that the value of log-likelihood of GLL(m_1,m_2) distribution is larger than those obtained by using the lognormal, GLL(1,1), GLL(m,1), GLL(1,m), and GLL(m,m) distributions. In terms of the values of log-likelihood functions, therefore, the GLL(m_1,m_2) model seems to be a better distribution than the compared distributions in fitting data of chloride concentration. From Table 5, however, it is clear that the AIC of the GLL(1,1) is

GLL(m_1,m_2) Distribution



GLL(m,m) Distribution

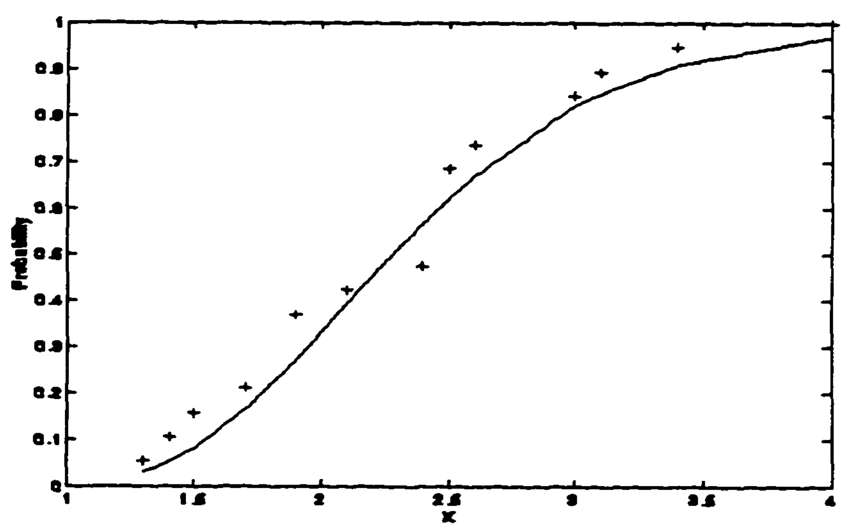
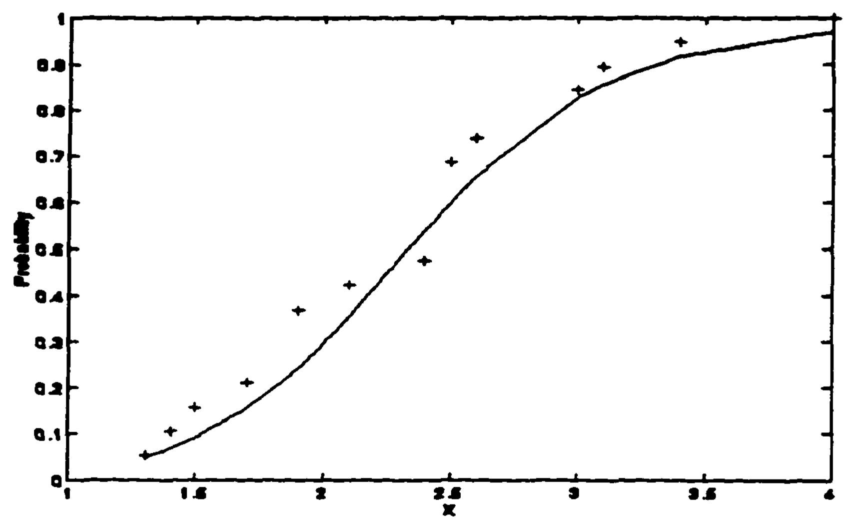


Figure 13. Graphs of the CDFs of the GLL(m_1,m_2) and GLL(m,m) distributions and the empirical distribution function (+ line) of SO2 concentrations.

GLL(m, 1) Distribution



GLL(1,m) Distribution

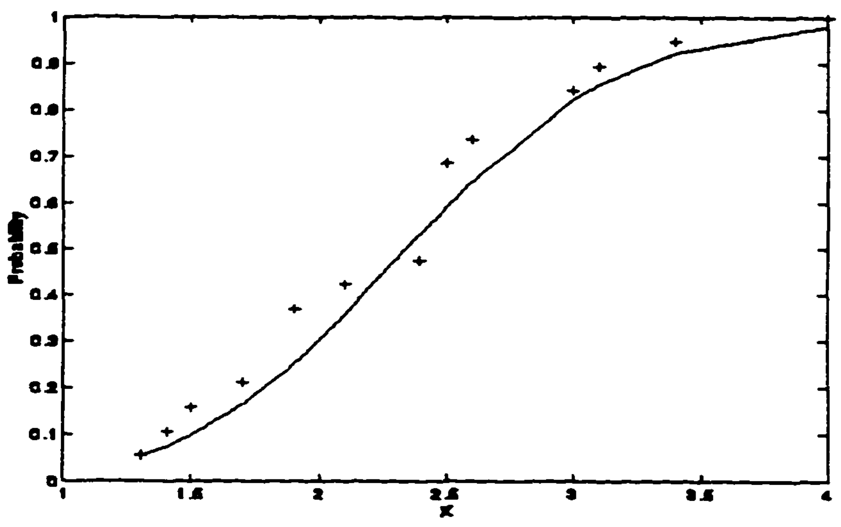
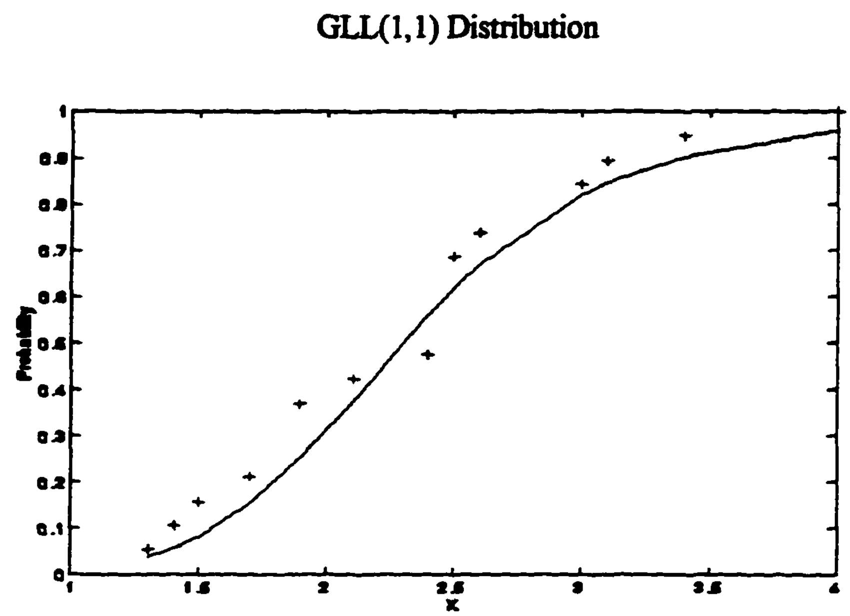


Figure 14. Graphs of the CDFs of the GLL(m,1) and GLL(1,m) distributions and the empirical distribution function (+ line) of SO2 concentrations.



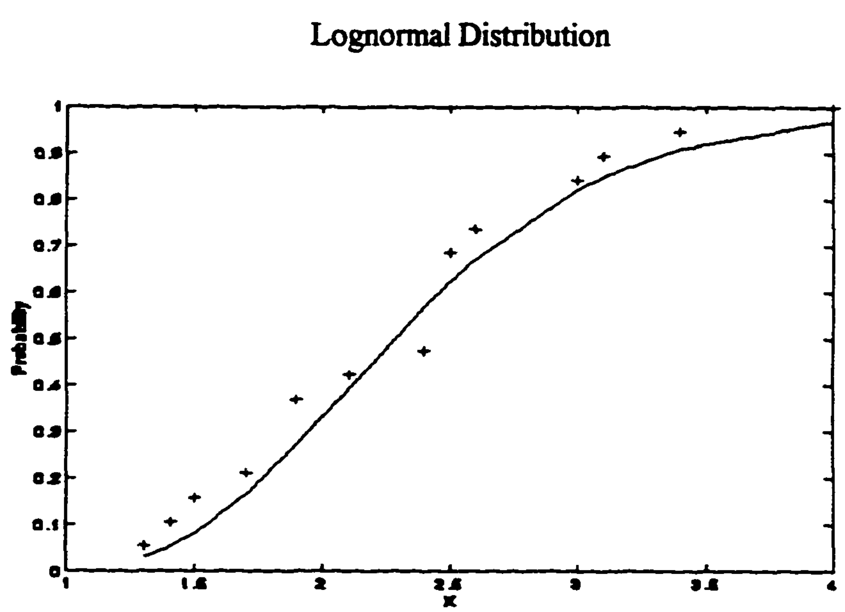


Figure 15. Graphs of the CDFs of the GLL(1,1) and lognormal distributions and the empirical distribution function (+ line) of SO2 concentrations.

relatively lower than the other GLL distributions. Thus, on the basis of the AIC criterion, the GLL(1,1) distribution is preferred to the other GLL distribution in fitting chloride data.

Table 5

Values of the log-likelihood functions and of AIC for models fitted to chloride data

Model	Log-Likelihood	AIC
GLL(1,1)	-49.2634	102.5268
GLL(m,1)	-48.5229	103.0458
GLL(1,m)	-48.7224	103.4448
GLL(m,m)	-49.1838	104.3676
GLL(m_1,m_2)	-48.5134	105.0268
Lognormal	-49.2495	-

For data of iron concentrations, the maximum-likelihood estimates of the parameters obtained by using the GLL(m_1,m_2) model are $\hat{\alpha} = 1.1036$, $\hat{\beta} = 0.1798$, $\hat{m}_1 = 32.5857$, and $\hat{m}_2 = 2.9799$. The 95% asymptotic confidence intervals for α, β, m_1, and m_2 are [0.9428,1.2644], [-0.1495,0.5091], [30.6092,34.5622], and [2.9738,2.9860], respectively. The GLL(m,m) fit yields $\hat{\alpha} = 0.2178$, $\hat{\beta} = -0.4679$, and $\hat{m} = 122.0616$. The 95% asymptotic confidence intervals for α, β, and m are [0.2175,0.2181], [-0.4721,-0.4637] and [121.4759,122.6473], respectively. The maximum log-likelihood values and AIC under GLL(m_1,m_2), GLL(m,m), GLL(m,1), GLL(1,m), and GLL(1,1) distributions are given in Table 6. For the purpose of comparison, we also calculate the value of the log-likelihood function of the lognormal distribution when data are fitted by the lognormal distribution.

It can be seen from Table 6 that because the value of log-likelihood of the GLL(m_1,m_2) distribution is larger than those of the lognormal, GLL(1,1), GLL(m,1), GLL(1,m), and GLL(m,m) distributions, the GLL(m_1,m_2) model seems to be a promising model in fitting data of iron concentration. However, in terms of the AIC criterion, the GLL(1,1) distribution is preferable over the other GLL distributions.

Table 6

Values of the log-likelihood functions and of AIC for models fitted to iron data

Model	Log-likelihood	AIC
GLL(1,1)	-30.5016	65.0032
GLL(m,1)	-30.2953	66.5906
GLL(1,m)	-30.3906	66.7812
GLL(m,m)	-30.3663	66.7326
GLL(m_1,m_2)	-30.2166	68.4332
Lognormal	-30.3648	-

For data of aluminum concentrations, the maximum-likelihood estimates of the parameters of the GLL(m_1,m_2) model are $\hat{\alpha} = 0.4105$, $\hat{\beta} = 1.5871$, $\hat{m}_1 = 242.760$, and $\hat{m}_2 = 28.7348$. The 95% asymptotic confidence intervals for α, β, m_1, and m_2 are [0.3534,0.4676], [1.5102,1.6640], [241.29,244.230], and [27.772,29.6976], respectively. The GLL(m,m) fit yields $\hat{\alpha} = 0.2864$, $\hat{\beta} = -0.3924$, and $\hat{m} = 104.2768$. The 95% asymptotic confidence intervals for α, β, and m are [0.1548,0.4180], [-2.3358,1.551], and [-79.4834,288.037], respectively. The maximum log-likelihood values and AIC under

$GLL(m_1,m_2)$, GLL(m,m), GLL(m,1), GLL(1,m), and GLL(1,1) distributions are listed in Table 7. For the purpose of comparison, we also calculate the value of the log-likelihood function of the lognormal distribution when data are fitted by the lognormal distribution.

As shown in Table 7, the value of log-likelihood of $GLL(m_1,m_2)$ distribution is larger than those of the lognormal, GLL(1,1), GLL(m,1), GLL(1,m), and GLL(m,m) distributions. From the value of log-likelihood function standpoint, therefore, the $GLL(m_1,m_2)$ model seems to be a better distribution in fitting data of aluminum concentration. But Table 7 also suggests that the AIC of the GLL(1,1) is slightly lower than the other GLL distributions. Therefore, on the basis of the AIC criterion, the GLL(1,1) distribution fits slightly better than the other GLL distribution in fitting aluminum data.

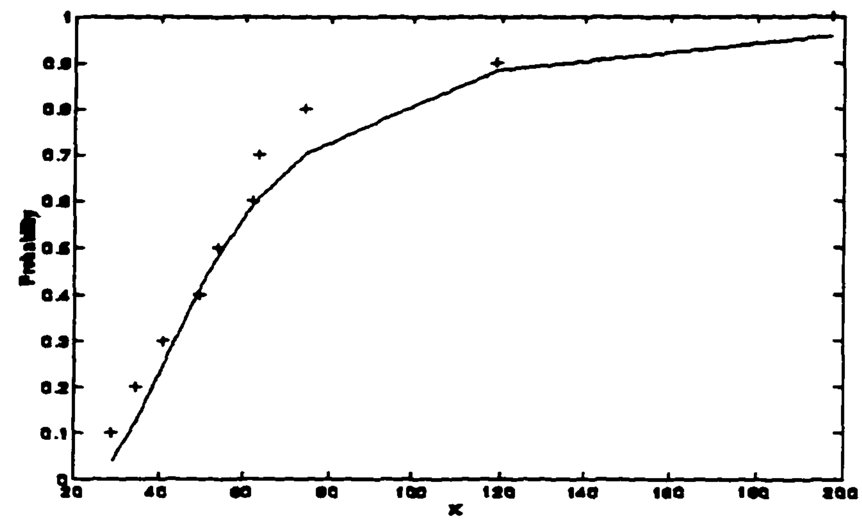
Table 7

Values of the log-likelihood functions and of AIC for models fitted to aluminum data

Model	Log-likelihood	AIC
GLL(1,1)	-20.7721	45.5442
GLL(m,1)	-20.7615	47.5230
GLL(1,m)	-20.7715	47.5430
GLL(m,m)	-20.6417	47.2834
$GLL(m_1,m_2)$	-20.6239	49.2478
Lognormal	-20.6399	-

Graphs of the CDFs of models superimposed on the empirical distribution functions for chloride data are given in Figures 16-18. As demonstrated by the graphs, all of the compared distributions show a similar performance. The similar results also are obtained for the data of iron and aluminum concentration. See Figures 19-24. Hence, the GLL(m_1,m_2) fits the data as well as the classical lognormal and the other GLL distribution in fitting data of chloride, iron, and aluminum concentrations.

GLL(m_1,m_2) Distribution



GLL(m,m) Distribution

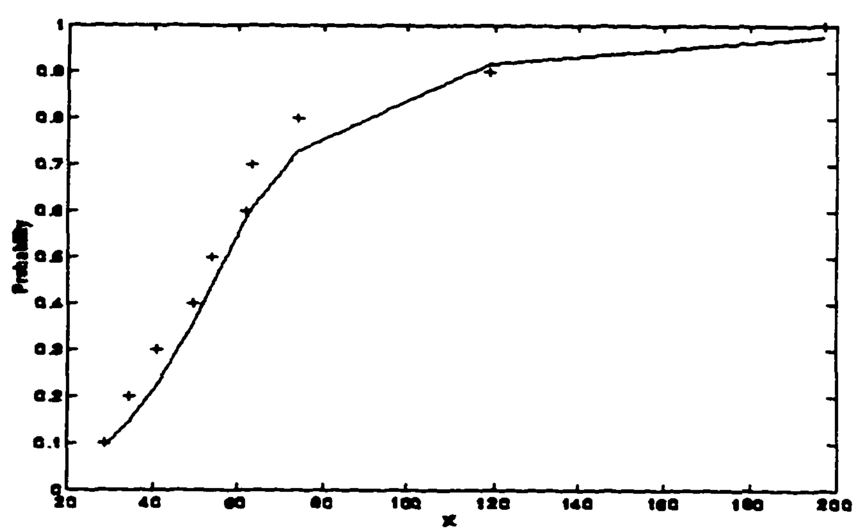
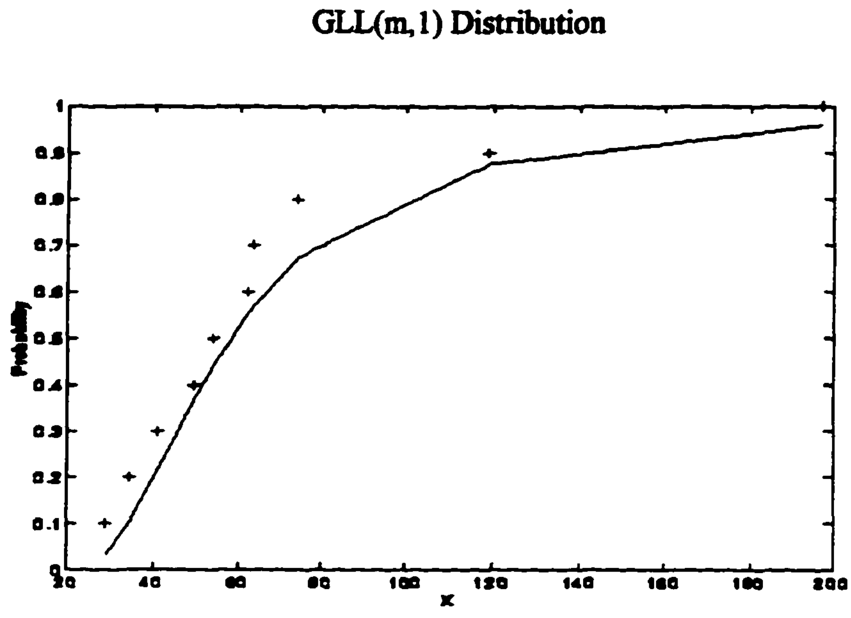


Figure 16. Graphs of the CDFs of the GLL(m_1,m_2) and GLL(m,m) distributions and the empirical distribution function (+ line) of chloride concentrations.



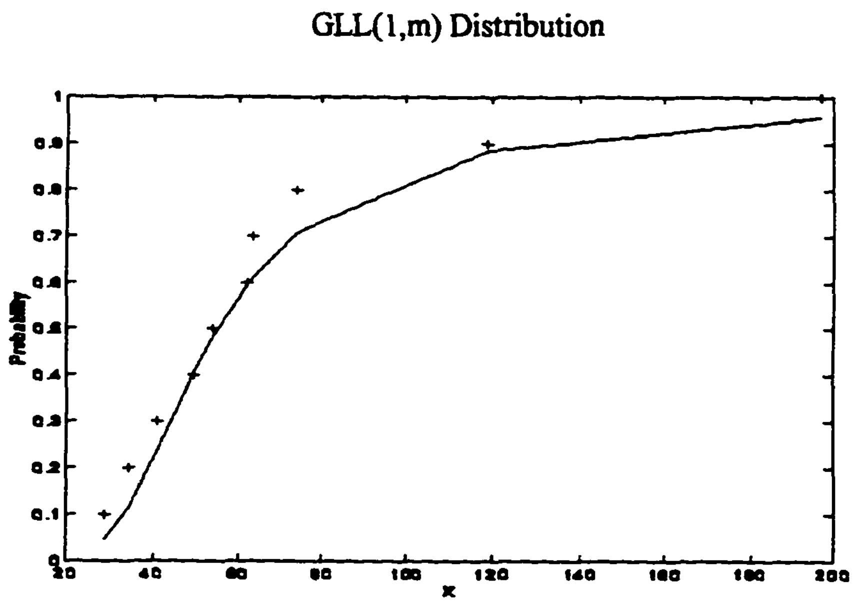
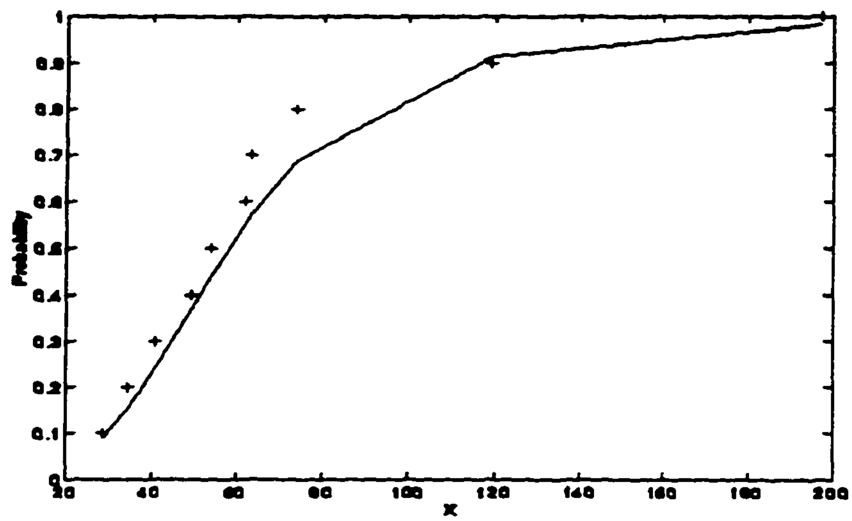


Figure 17. Graphs of the CDFs of the GLL(m,1) and GLL(1,m) distributions and the empirical distribution function (+ line) of chloride concentrations.

GLL(1,1) Distribution



Lognormal Distribution

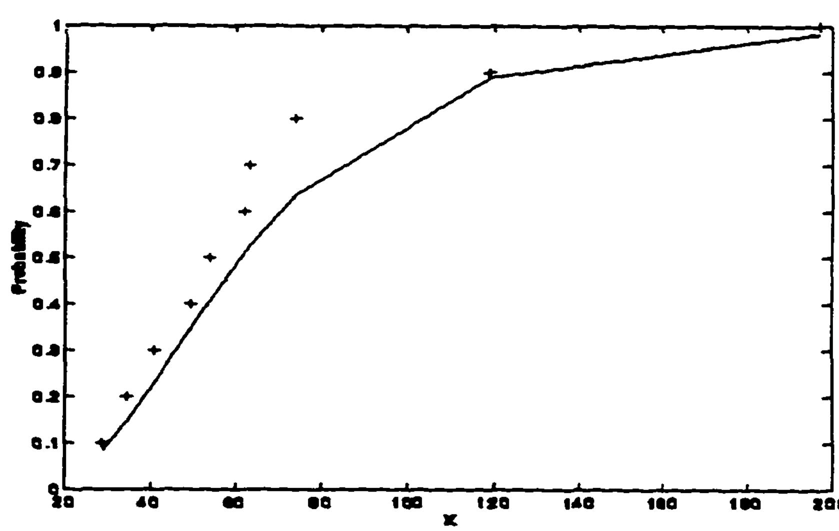
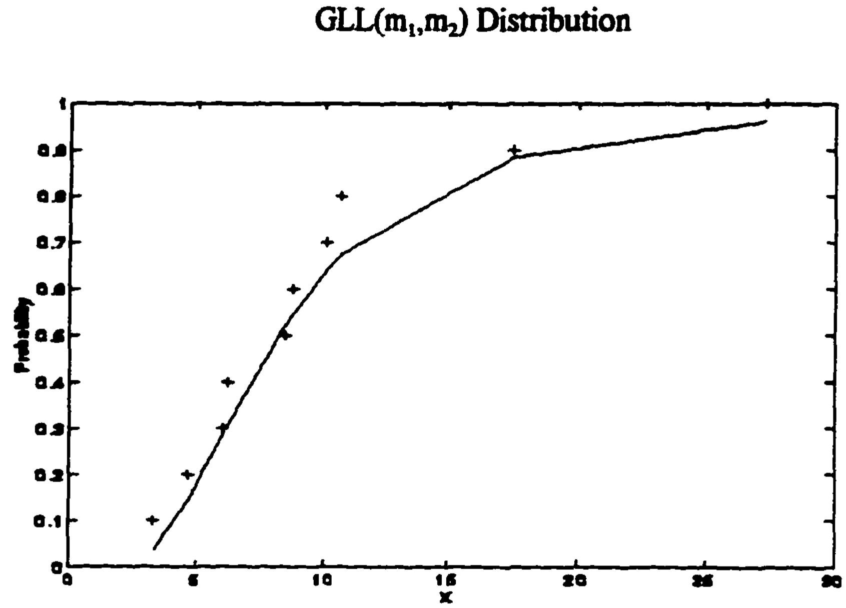


Figure 18. Graphs of the CDFs of the GLL(1,1) and lognormal distributions and the empirical distribution function (+ line) of chloride concentrations.



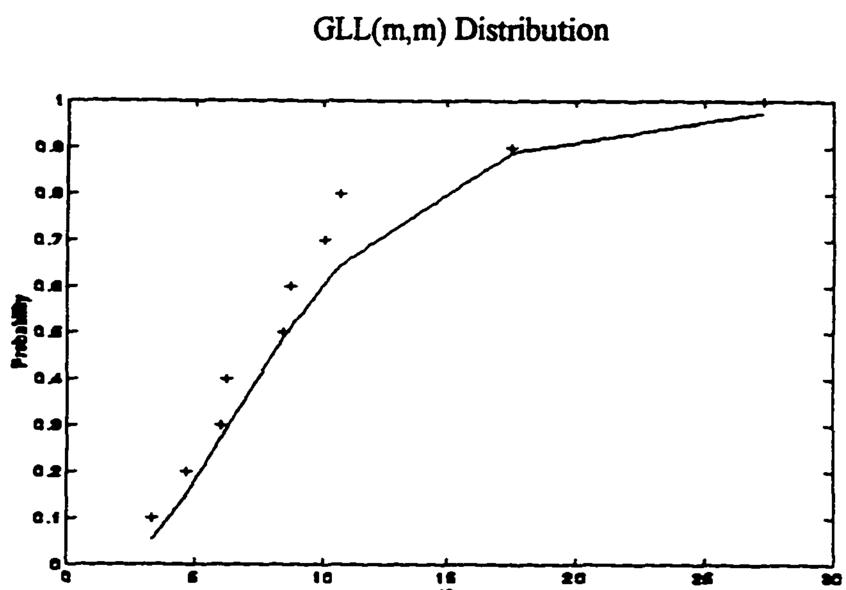
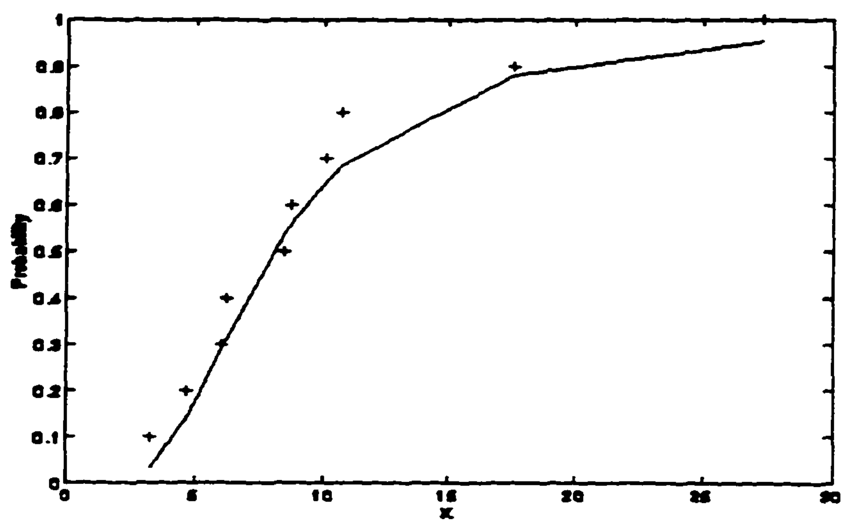


Figure 19. Graphs of the CDFs of the GLL(m_1,m_2) and GLL(m,m) distributions and the empirical distribution function (+ line) of iron concentrations.

GLL(m,1) Distribution



GLL(1,m) Distribution

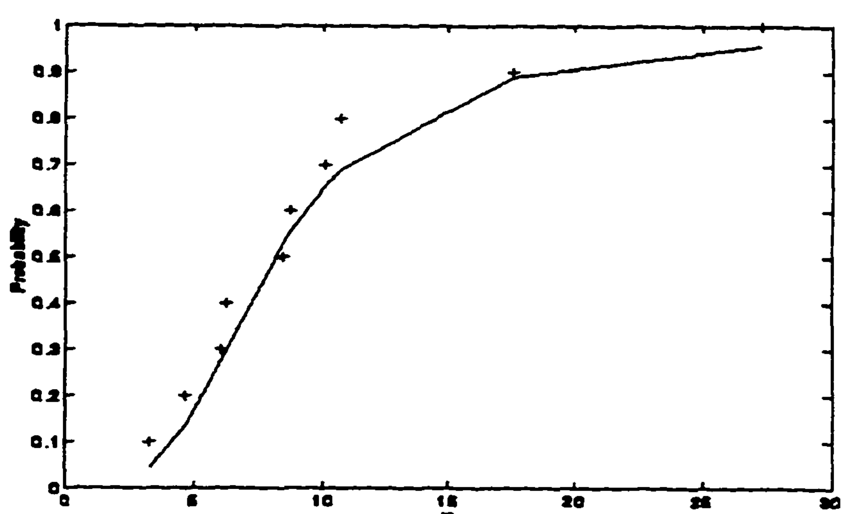
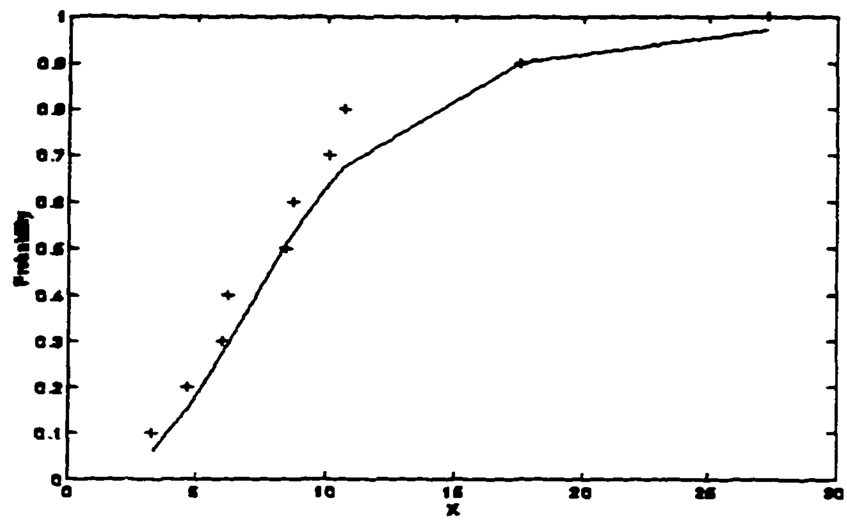


Figure 20. Graphs of the CDFs of the GLL(m,1) and GLL(1,m) distributions and the empirical distribution function (+ line) of iron concentrations.

GLL(1,1) Distribution



Lognormal Distribution

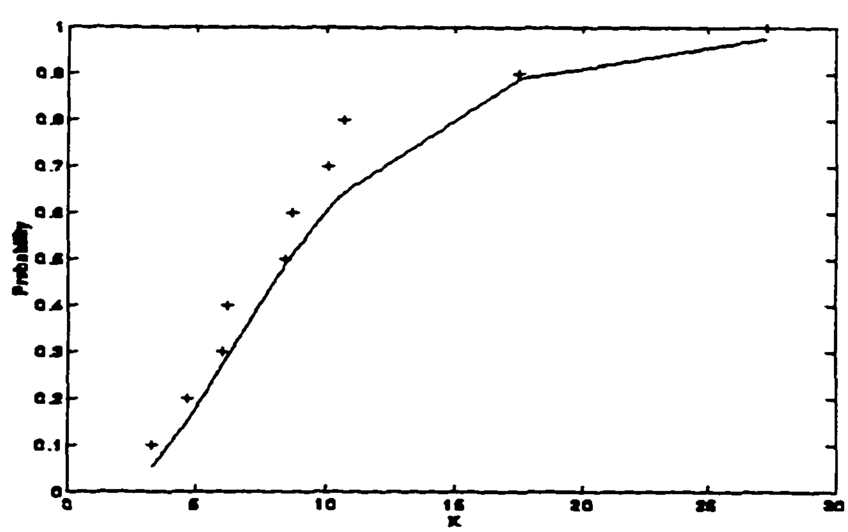
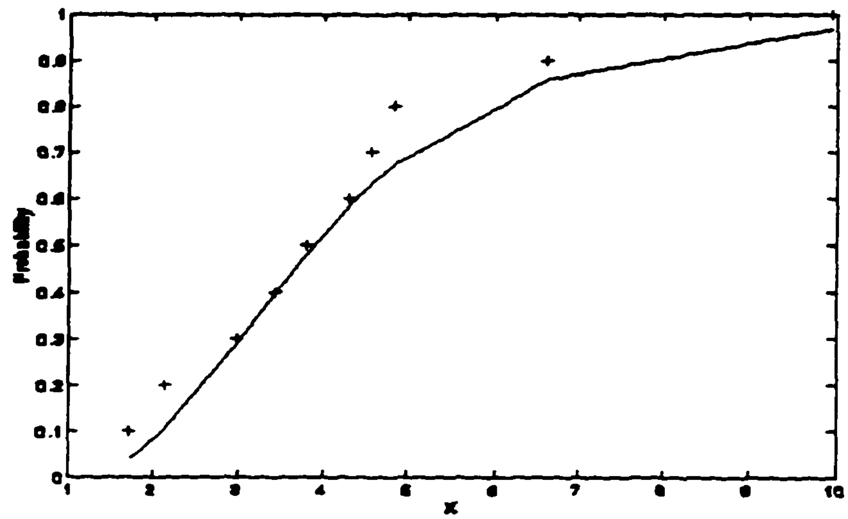


Figure 21. Graphs of the CDFs of the GLL(1,1) and lognormal distributions and the empirical distribution function (+ line) of iron concentrations.

GLL(m_1,m_2) Distribution



GLL(m,m) Distribution

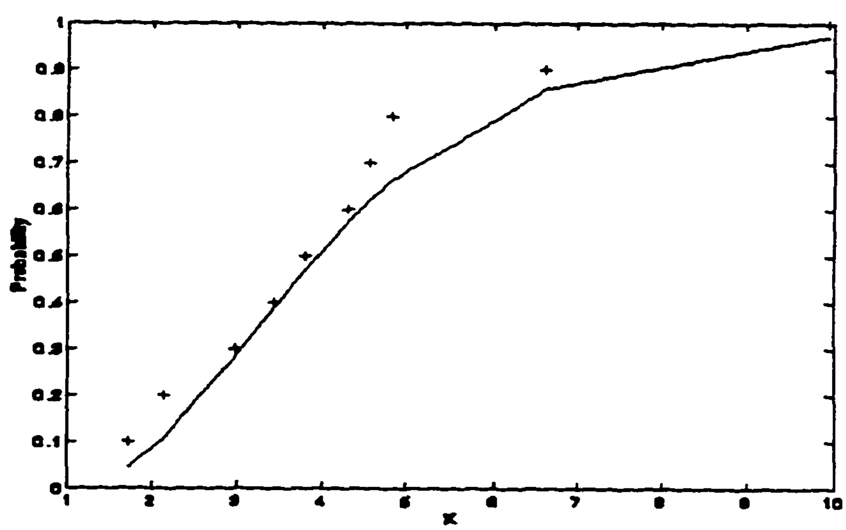
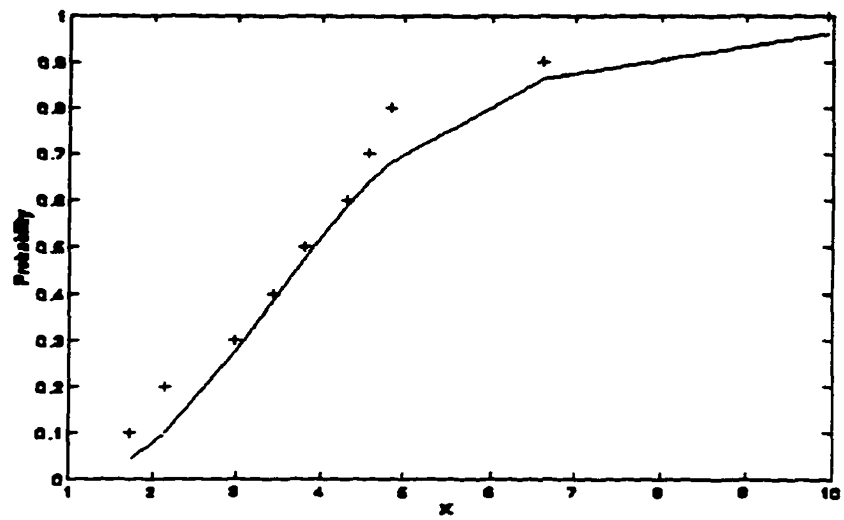


Figure 22. Graphs of the CDFs of the GLL(m_1,m_2) and GLL(m,m) distributions and the empirical distribution function (+ line) of aluminum concentrations.

GLL(m,1) Distribution



GLL(1,m) Distribution

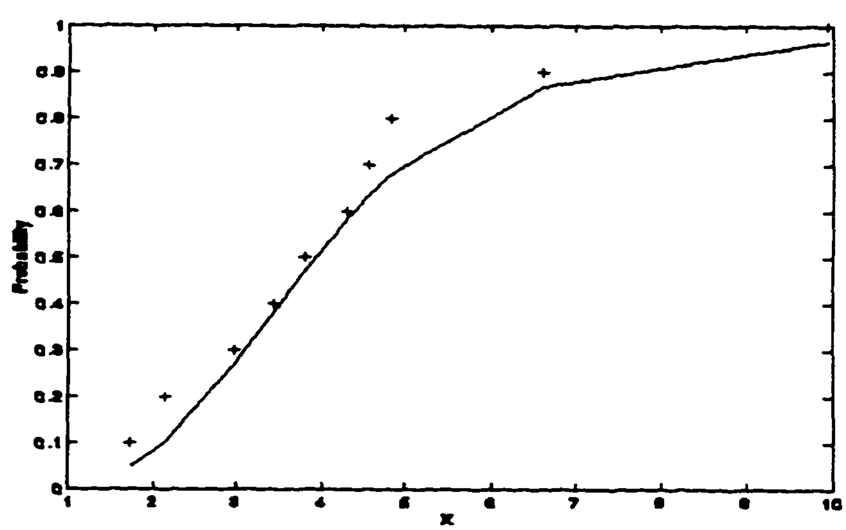
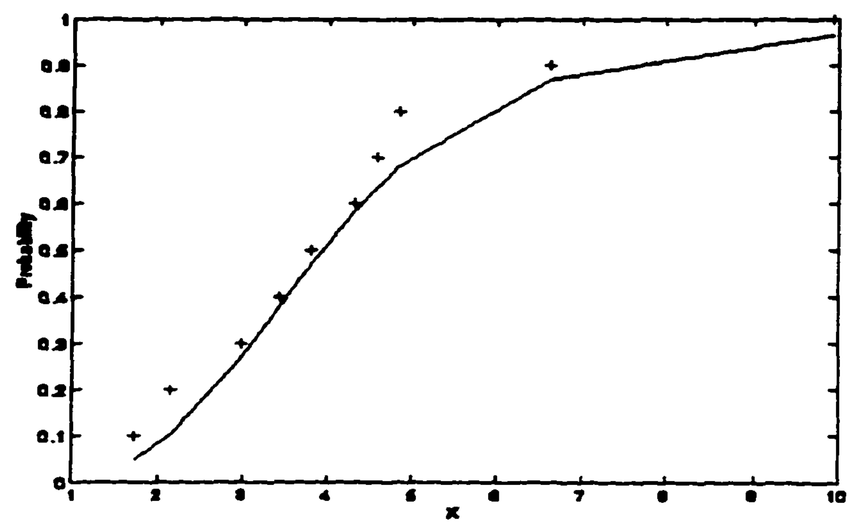


Figure 23. Graphs of the CDFs of the GLL(m,1) and GLL(1,m) distributions and the empirical distribution function (+ line) of aluminum concentrations.

GLL(1,1) Distribution



Lognormal Distribution

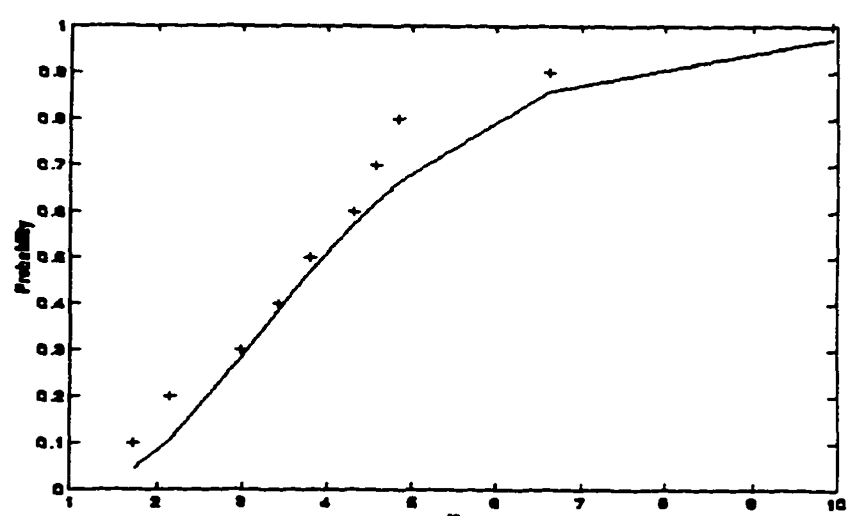


Figure 24. Graphs of the CDFs of the GLL(1,1) and lognormal distributions and the empirical distribution function (+ line) of aluminum concentrations.

CHAPTER V

FUTURE RESEARCH PROBLEMS

Of primary interest is the fact that, in fitting environmental pollutant data, none of the probability models, including the classical lognormal, has been identified to be superior to others in a general sense. It is reasonable, therefore, to use a rich family of distributions that includes several well-known distributions as special cases for fitting environmental data. Based on its desirable features, as discussed in previous chapters, we introduced a GLL distribution as a general distribution of data of environmental pollutant concentrations, and applied it to seven data sets.

As demonstrated in chapter four, in fitting environmental data, the GLL family of distributions is a good alternative to the lognormal distribution. For all of data sets, the distributions of GLL family are found generally better than the lognormal distribution. In particular, on the basis of the values of maximum log-likelihood functions, the GLL(m_1,m_2) seems to be a better probability model for all of data sets. Moreover, graphs of the CDFs of the models superimposed on the empirical distribution suggest that the GLL(m_1,m_2) is generally better than the lognormal and other families of the GLL distributions.

Hence, the use of the distribution of GLL family in fitting environmental data needs to be investigated further. Now we give a brief synopsis of problems for future research.

1. It is interesting that for data of mercury concentration with 115 sample size, the maximum log-likelihood value of GLL(m_1,m_2) distribution is considerably larger than that of the other distributions, especially the lognormal distribution. But for the other data sets with a sample size much smaller, for example, for data of tritium concentration with 26 observations, the improvement in the likelihood function is slight. What are the effects of sample size to the performance of the GLL(m_1,m_2) distribution in fitting data of environmental pollutant concentration?

2. Following the first question, what are the effects of the intensity of censoring? To answer questions 1 and 2, simulation studies may be utilized.

3. The family of probability models may change significantly for different types of pollutant, averaging time of interest, different locations, and other factors. Hence, the performance of GLL(m_1,m_2) when incorporating these factors needs to be examined further. How well does the GLL(m_1,m_2) fit these data?

4. The log-logistic is a special case of the GLL(m_1,m_2), when $m_1 = m_2 = 1$. What is the effect of estimating the shape parameters m_1 and m_2 on the estimation of the scale parameter α?

5. In this study we employed the method of maximum-likelihood estimation for estimating parameters of distributions. It is very interesting, therefore, to compare this method with others. What is the best method for estimating parameters of the GLL distributions for describing data of pollutant concentrations?

APPENDIX A

FIRST- AND SECOND-ORDER DERIVATIVES OF THE LOG-LIKELIHOOD FUNCTIONS

The log-likelihood function of the GLL(m_1,m_2) is defined as follows

$$l\,(\alpha,\beta,m_1,m_2\,|\,\underline{x}) = \sum_{k=1}^{N}\{\, n_{1k}\,[\,[\ln(\alpha)]-[\ln[B(m_1,m_2)]]\,] + m_1\sum_{i=1}^{n_{1k}}\ln[F_k(x_i)] + m_2\sum_{i=1}^{n_{1k}}\ln[1-F_k(x_i)] - \sum_{i=1}^{n_{1k}}\ln(x_i) + \sum_{j=1}^{n_{2k}}\ln(\,I_{(m_1,m_2)}\,[F_k(x_j)]\,)\,\}\,.$$

The first-order derivatives of the log-likelihood function with respect to α, β, m_1, and m_2 are listed below:

$$\frac{\partial l}{\partial \alpha} = \sum_{k=1}^{N}\{\,\frac{n_{1k}}{\alpha} + m_1\sum_{i=1}^{n_{1k}}\frac{1}{F_k(x_i)}\,[\ln(x_i)]\,[F_k(x_i)]\,[1-F_k(x_i)] + m_2\sum_{i=1}^{n_{1k}}\frac{-1}{1-F_k(x_i)}\,[\ln(x_i)]\,[F_k(x_i)]\,[1-F_k(x_i)] + \sum_{j=1}^{n_{2k}}\frac{1}{I_{(m_1,m_2)}[F_k(x_j)]}\,\frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial\alpha}\,\}$$

$$= \sum_{k=1}^{N}\{\,\frac{n_{1k}}{\alpha} + m_1\sum_{i=1}^{n_{1k}}[\ln(x_i)]\,[1-F_k(x_i)] - m_2\sum_{i=1}^{n_{1k}}[\ln(x_i)]\,[F_k(x_i)] + \sum_{j=1}^{n_{2k}}\frac{1}{I_{(m_1,m_2)}[F_k(x_j)]}\,\frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial\alpha}\,\}$$

$$= \sum_{k=1}^{N}\{\,\frac{n_{1k}}{\alpha} + m_1\sum_{i=1}^{n_{1k}}[\ln(x_i)]\,[1-F_k(x_i)] - m_2\sum_{i=1}^{n_{1k}}[\ln(x_i)]\,[F_k(x_i)] + \frac{1}{B(m_1,m_2)}\sum_{j=1}^{n_{2k}}\frac{1}{I_{(m_1,m_2)}[F_k(x_j)]}\,[\ln(x_j)]\,[F_k(x_j)]^{m_1}\,[1-F_k(x_j)]^{m_2}\,\}$$

$$\frac{\partial l}{\partial \beta} = \sum_{k=1}^{N} \{ m_1 \sum_{i=1}^{n_{1k}} \frac{1}{F_k(x_i)} [F_k(x_i)] [1-F_k(x_i)] +$$

$$m_2 \sum_{i=1}^{n_{1k}} \frac{1}{1-F_k(x_i)} [- F_k(x_i)] [1-F_k(x_i)] +$$

$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial I_m[F_k(x_j)]}{\partial F_k(x_j)} \frac{\partial F_k(x_j)}{\partial \beta} \}$$

$$= \sum_{k=1}^{N} \{ m_1 \sum_{i=1}^{n_{1k}} [1-F_k(x_i)] - m_2 \sum_{i=1}^{n_{1k}} [F_k(x_i)] +$$

$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial F_k(x_j)} \frac{\partial F_k(x_j)}{\partial \beta} \}$$

$$= \sum_{k=1}^{N} \{ m_1 \sum_{i=1}^{n_{1k}} [1-F_k(x_i)] - m_2 \sum_{i=1}^{n_{1k}} [F_k(x_i)] +$$

$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} [\frac{1}{\mathrm{B}(m_1,m_2)} [F_k(x_j)]^{m_1-1} [1-F_k(x_j)]^{m_2-1} \cdot$$

$$[F_k(x_j)] [1-F_k(x_j)]] \}$$

$$= \sum_{k=1}^{N} \{ m_1 \sum_{i=1}^{n_{1k}} [1-F_k(x_i)] - m_2 \sum_{i=1}^{n_{1k}} [F_k(x_i)] +$$

$$\frac{1}{\mathrm{B}(m_1,m_2)} \sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} [F_k(x_j)]^{m_1} [1-F_k(x_j)]^{m_2} \}$$

$$= \sum_{k=1}^{N} \{ m_1 n_{1k} - (m_1+m_2) \sum_{i=1}^{n_{1k}} [F_k(x_i)] +$$

$$\frac{1}{\mathrm{B}(m_1,m_2)} \sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} [F_k(x_j)]^{m_1} [1-F_k(x_j)]^{m_2} \}$$

$$\frac{\partial l}{\partial m_1} = \sum_{k=1}^{N}\{ -\frac{n_{1k}}{B(m_1,m_2)} \frac{\partial B(m_1,m_2)}{\partial m_1} + \sum_{i=1}^{n_{1k}} \ln[F_k(x_i)] +$$

$$m_1 \sum_{i=1}^{n_{1k}} \frac{1}{F_k(x_i)} \frac{\partial F_k(x_i)}{\partial m_1} + m_2 \sum_{i=1}^{n_{1k}} \frac{-1}{1-F_k(x_i)} \frac{\partial F_k(x_i)}{\partial m_1} +$$

$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1} \} =$$

$$\sum_{k=1}^{N}\{ -\frac{n_{1k}}{B(m_1,m_2)} B(m_1,m_2) [\Psi(m_1)-\Psi(m_1+m_2)] + \sum_{i=1}^{n_{1k}} \ln[F_k(x_i)] +$$

$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1} \}$$

$$= \sum_{k=1}^{N}\{ -n_{1k} [\Psi(m_1)-\Psi(m_1+m_2)] + \sum_{i=1}^{n_{1k}} \ln[F_k(x_i)] +$$

$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1} \}$$

$$\frac{\partial l}{\partial m_2} = \sum_{k=1}^{N}\{ -\frac{n_{1k}}{B(m_1,m_2)} \frac{\partial B(m_1,m_2)}{\partial m_2} + \sum_{i=1}^{n_{1k}} \ln[1-F_k(x_i)] +$$

$$m_1 \sum_{i=1}^{n_{1k}} \frac{1}{F_k(x_i)} \frac{\partial F_k(x_i)}{\partial m_2} + m_2 \sum_{i=1}^{n_{1k}} \frac{-1}{1-F_k(x_i)} \frac{\partial F_k(x_i)}{\partial m_2} +$$

$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2} \}$$

$$= \sum_{k=1}^{N}\{ -\frac{n_{1k}}{B(m_1,m_2)} B(m_1,m_2) [\Psi(m_2)-\Psi(m_1+m_2)] + \sum_{i=1}^{n_{1k}} \ln[1-F_k(x_i)] +$$

$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2} \}$$

$$= \sum_{k=1}^{N}\{ -n_{1k} [\Psi(m_2)-\Psi(m_1+m_2)] + \sum_{i=1}^{n_{1k}} \ln[1-F_k(x_i)] +$$

$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2} \}$$

The second-order derivatives of the log-likelihood function with respect to α, β, m_1, and m_2 are listed below:

$$\frac{\partial^2 l}{\partial\alpha\partial\alpha} = \sum_{k=1}^{N} \{ -\frac{n_{1k}}{\alpha^2} - (m_1+m_2) \sum_{i=1}^{n_{1k}} [\ln(x_i)]^2 \; [F_k(x_i)] \; [1-F_k(x_i)] -$$
$$\frac{1}{B(m_1,m_2)} \sum_{j=1}^{n_{2k}} [\ln(x_j)]^2 \, [\, \frac{m_1}{I_{(m_1,m_2)}[F_k(x_j)]} [F_k(x_j)]^{m_1} \; [1-F_k(x_j)]^{m_2+1} -$$
$$\frac{m_2}{I_{(m_1,m_2)}[F_k(x_j)]} [F_k(x_j)]^{m_1+1} \; [1-F_k(x_j)]^{m_2} +$$
$$\frac{1}{B(m_1,m_2)} \frac{1}{[I_{(m_1,m_2)}[F_k(x_j)]]^2} [F_k(x_j)]^{2m_1} \; [1-F_k(x_j)]^{2m_2} \;] \}$$

$$\frac{\partial^2 l}{\partial\beta\partial\beta} = \sum_{k=1}^{N} \{ -(m_1+m_2) \sum_{i=1}^{n_{1k}} [F_k(x_i)] \; [1-F_k(x_i)] +$$
$$\frac{1}{B(m_1,m_2)} \sum_{j=1}^{n_{2k}} [\, \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} m_1 [F_k(x_i)]^{m_1-1} \; [F_k(x_i)] \; .$$
$$[1-F_k(x_j)] \; [1-F_k(x_j)]^{m_2} +$$
$$\frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} [F_k(x_j)]^{m_1} \; m_2 [1-F_k(x_j)]^{m_2-1} \; [-F_k(x_i)] \; [1-F_k(x_j)] +$$
$$\frac{-1}{[I_{(m_1,m_2)}[F_k(x_j)]]^2} \; \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial F_k(x_j)} \; \frac{\partial F_k(x_j)}{\partial\beta} [F_k(x_i)]^{m_1} \; [1-F_k(x_j)]^{m_2} \;] \}$$
$$= \sum_{k=1}^{N} \{ -(m_1+m_2) \sum_{i=1}^{n_{1k}} [F_k(x_i)] \; [1-F_k(x_i)] +$$
$$\frac{1}{B(m_1,m_2)} \sum_{j=1}^{n_{2k}} [\, \frac{m_1}{I_{(m_1,m_2)}[F_k(x_j)]} [F_k(x_i)]^{m_1} \; [1-F_k(x_j)]^{m_2+1} -$$
$$\frac{m_2}{I_{(m_1,m_2)}[F_k(x_j)]} [F_k(x_j)]^{m_1+1} \; [1-F_k(x_j)]^{m_2} -$$
$$\frac{1}{B(m_1,m_2)} \frac{1}{[I_{(m_1,m_2)}[F_k(x_j)]]^2} [F_k(x_i)]^{2m_1} \; [1-F_k(x_j)]^{2m_2} \;] \}$$

$$\frac{\partial^2 l}{\partial m_1 \partial m_1} = \sum_{k=1}^{N} \{ -n_{1k} [\Psi'(m_1) - \Psi'(m_1+m_2)] +$$

$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial^2 I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1 \partial m_1} -$$

$$\frac{1}{(I_{(m_1,m_2)}[F_k(x_j)])^2} (\frac{\partial}{\partial m_1} I_{(m_1,m_2)}[F_k(x_j)])^2 \}$$

$$\frac{\partial^2 l}{\partial \beta \partial \alpha} = \sum_{k=1}^{N} \{ -(m_1+m_2) \sum_{i=1}^{n_{1k}} [\ln(x_i)] [F_k(x_i)] [1-F_k(x_i)] +$$

$$\frac{1}{B(m_1,m_2)} \sum_{j=1}^{n_{2k}} [\frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} m_1 [F_k(x_j)]^{m_1-1} \cdot$$

$$[\ln(x_j)] [F_k(x_j)] [1-F_k(x_j)] [1-F_k(x_j)]^{m_2} +$$

$$\frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} [F_k(x_j)]^{m_1} m_2 [1-F_k(x_j)]^{m_2-1} [\ln(x_j)] [-F_k(x_j)] [1-F_k(x_j)] +$$

$$\frac{-1}{[I_{(m_1,m_2)}[F_k(x_j)]]^2} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial F_k(x_j)} \frac{\partial F_k(x_j)}{\partial \alpha} [F_k(x_j)]^{m_1} [1-F_k(x_j)]^{m_2}] \}$$

$$= \sum_{k=1}^{N} \{ -(m_1+m_2) \sum_{i=1}^{n_{1k}} [\ln(x_i)] [F_k(x_i)] [1-F_k(x_i)] +$$

$$\frac{1}{B(m_1,m_2)} \sum_{j=1}^{n_{2k}} [\frac{m_1 \ln(x_j)}{I_{(m_1,m_2)}[F_k(x_j)]} [F_k(x_j)]^{m_1} [1-F_k(x_j)]^{m_2+1} -$$

$$\frac{m_2 \ln(x_j)}{I_{(m_1,m_2)}[F_k(x_j)]} [F_k(x_j)]^{m_1+1} [1-F_k(x_j)]^{m_2} -$$

$$\frac{1}{B(m_1,m_2)} \frac{\ln(x_j)}{[I_{(m_1,m_2)}[F_k(x_j)]]^2} [F_k(x_j)]^{2m_1} [1-F_k(x_j)]^{2m_2}] \}$$

$$
\begin{aligned}
\frac{\partial^2 l}{\partial m_1 \partial \alpha} &= \sum_{k=1}^{N} \{ \sum_{i=1}^{n_{1k}} [\frac{1}{F_k(x_i)} [\ln(x_i)] \, [F_k(x_i)] \, [1-F_k(x_i)] \,] \, + \\
&\sum_{j=1}^{n_{2k}} [\frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial^2 I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1 \partial \alpha} - \\
&\frac{1}{(I_{(m_1,m_2)}[F_k(x_j)] \,)^2} \frac{\partial I_{(m_1,m_2)}[F(x_j)]}{\partial \alpha} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1} \,] \} \\
&= \sum_{k=1}^{N} \{ \sum_{i=1}^{n_{1k}} [\ln(x_i)] \, [1-F_k(x_i)] \, + \\
&\sum_{j=1}^{n_{2k}} [\frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial^2 I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1 \partial \alpha} - \\
&\frac{1}{(I_{(m_1,m_2)}[F_k(x_j)] \,)^2} \frac{\partial I_{(m_1,m_2)}[F(x_j)]}{\partial \alpha} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1} \,] \}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 l}{\partial m_1 \partial \beta} &= \sum_{k=1}^{N} \{ \sum_{i=1}^{n_{1k}} [\frac{1}{F_k(x_i)} [F_k(x_i)] \, [1-F_k(x_i)] \, + \\
&\sum_{j=1}^{n_{2k}} [\frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial^2 I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1 \partial \beta} - \\
&\frac{1}{(I_{(m_1,m_2)}[F_k(x_j)] \,)^2} \frac{\partial I_{(m_1,m_2)}[F(x_j)]}{\partial \beta} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1} \,] \} \\
&= \sum_{k=1}^{N} \{ \sum_{i=1}^{n_{1k}} [1-F_k(x_i)] \, + \\
&\sum_{j=1}^{n_{2k}} [\frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial^2 I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1 \partial \beta} - \\
&\frac{1}{(I_{(m_1,m_2)}[F_k(x_j)] \,)^2} \frac{\partial I_{(m_1,m_2)}[F(x_j)]}{\partial \beta} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_1} \,] \}
\end{aligned}
$$

$$\frac{\partial^2 l}{\partial m_2 \partial m_2} = \sum_{k=1}^{N} \{ -n_{1k} \, [\Psi'(m_2) - \Psi'(m_1+m_2)] \; +$$

$$\sum_{j=1}^{n_{2k}} \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial^2 I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2 \partial m_2} \; -$$

$$\frac{1}{(\, I_{(m_1,m_2)}[F_k(x_j)] \,)^2} \, (\, \frac{\partial}{\partial m_2} I_{(m_1,m_2)}[F_k(x_j)] \,)^2 \, \}$$

$$\frac{\partial^2 l}{\partial m_2 \partial \alpha} = \sum_{k=1}^{N} \{ \sum_{i=1}^{n_{1k}} [\, \frac{-1}{1-F_k(x_i)} \, [\ln(x_i)] \, [F_k(x_i)] \, [1-F_k(x_i)] \; +$$

$$\sum_{j=1}^{n_{2k}} [\, \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial^2 I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2 \partial \alpha} \; -$$

$$\frac{1}{(\, I_{(m_1,m_2)}[F_k(x_j)] \,)^2} \frac{\partial I_{(m_1,m_2)}[F(x_j)]}{\partial \alpha} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2} \;] \, \}$$

$$= \sum_{k=1}^{N} \{ \sum_{i=1}^{n_{1k}} -[\ln(x_i)] \, [F_k(x_i)] \; +$$

$$\sum_{j=1}^{n_{2k}} [\, \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial^2 I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2 \partial \alpha} \; -$$

$$\frac{1}{(\, I_{(m_1,m_2)}[F_k(x_j)] \,)^2} \frac{\partial I_{(m_1,m_2)}[F(x_j)]}{\partial \alpha} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2} \;] \, \}$$

$$\frac{\partial^2 l}{\partial m_2 \partial \beta} = \sum_{k=1}^{N} \{ \sum_{i=1}^{n_{1k}} [\, \frac{-1}{1-F_k(x_i)} \, [F_k(x_i)] \, [1-F_k(x_i)] \; +$$

$$\sum_{j=1}^{n_{2k}} [\, \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial^2 I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2 \partial \beta} \; -$$

$$\frac{1}{(\, I_{(m_1,m_2)}[F_k(x_j)] \,)^2} \frac{\partial I_{(m_1,m_2)}[F(x_j)]}{\partial \beta} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2} \;] \, \}$$

$$= \sum_{k=1}^{N} \{ \sum_{i=1}^{n_{1k}} -[F_k(x_i)] \; + \; \sum_{j=1}^{n_{2k}} [\, \frac{1}{I_{(m_1,m_2)}[F_k(x_j)]} \frac{\partial^2 I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2 \partial \beta} \; -$$

$$\frac{1}{(\, I_{(m_1,m_2)}[F_k(x_j)] \,)^2} \frac{\partial I_{(m_1,m_2)}[F(x_j)]}{\partial \beta} \frac{\partial I_{(m_1,m_2)}[F_k(x_j)]}{\partial m_2} \;] \, \}$$

The log-likelihood function of the GLL(m,m) is given by

$$l\,(\alpha,\beta,m|\underline{x}) = \sum_{k=1}^{N}\{\, n_{1k}\, [\, [\ln(\alpha)-\ln(B(m,m)-] \,+$$
$$m\sum_{i=1}^{n_{1k}}[\, [\ln(F_k(x_i)]+\ln[1-F_k(x_i)]\,]\, - \sum_{i=1}^{n_{1k}}\ln(x_i)\, +$$
$$\sum_{j=1}^{n_{2k}}\ln[I_m\,[F_k(x_j)]]\,\}\,.$$

The first-order derivative of the log-likelihood function with respect to α, β, and m is shown by

$$\frac{\partial l}{\partial\alpha}\Big|_{m_1=m_2=m} = \sum_{k=1}^{N}\{\, \frac{n_{1k}}{\alpha} + m\sum_{i=1}^{n_{1k}}[\ln(x_i)]\,[1-2F_k(x_i)]\, +$$
$$\frac{1}{B(m,m)}\sum_{j=1}^{n_{2k}}\frac{1}{I_{(m)}[F_k(x_j)]}\,[\ln(x_j)]\,[F_k(x_j)]^m\,[1-F_k(x_j)]^m\,\}$$

$$\frac{\partial l}{\partial\beta}\Big|_{m_1=m_2=m} = \sum_{k=1}^{N}\{\, mn_{1k}\, -2m\sum_{i=1}^{n_{1k}}[F_k(x_i)]\, +$$
$$\frac{1}{B(m,m)}\sum_{j=1}^{n_{2k}}\frac{1}{I_m[F_k(x_j)]}\,[F_k(x_j)]^m\,[1-F_k(x_j)]^m\,\}$$

$$\frac{\partial l}{\partial m} = \sum_{k=1}^{N}\{ -\frac{n_{1k}}{B(m,m)} \frac{\partial B(m,m)}{\partial m} + \sum_{i=1}^{n_{1k}}[\ln[F_k(x_i)] + \ln[1-F_k(x_i)]] +$$
$$m \sum_{i=1}^{n_{1k}}[\frac{1}{F_k(x_i)} \frac{\partial F_k(x_i)}{\partial m} + \frac{-1}{1-F_k(x_i)} \frac{\partial F_k(x_i)}{\partial m}] +$$
$$\sum_{j=1}^{n_{2k}} \frac{1}{I_m[F_k(x_j)]} \frac{\partial I_m[F_k(x_j)]}{\partial m} \} ,$$

but

$$\frac{\partial B(m,m)}{\partial m} = B(m,m) \frac{\partial}{\partial m}[\log(B(m,m)]$$
$$= B(m,m) \frac{\partial}{\partial m}[\log\frac{[\Gamma(m)]^2}{\Gamma(2m)}]$$
$$= B(m,m) \frac{\partial}{\partial m}[2\log(\Gamma(m) - \log(\Gamma(2m)]$$
$$= B(m,m) [2 \frac{\Gamma'(m)}{\Gamma(m)} - 2 \frac{\Gamma'(2m)}{\Gamma(2m)}]$$
$$= 2 B(m,m) [\Psi(m) - \Psi(2m)]$$

where $\Psi(m) = \Gamma'(m) / \Gamma(m)$ is the digamma function. Then

$$\frac{\partial l}{\partial m} = \sum_{k=1}^{N}\{ -2 n_{1k} [\Psi(m) - \Psi(2m)] + \sum_{i=1}^{n_{1k}}[\ln[F_k(x_i)] + \ln[1-F_k(x_i)]] +$$
$$\sum_{j=1}^{n_{2k}} \frac{1}{I_m[F_k(x_j)]} \frac{\partial I_m[F_k(x_j)]}{\partial m} \}$$
$$= \sum_{k=1}^{N}\{ -2 n_{1k} [\Psi(m) - \Psi(2m)] + \sum_{i=1}^{n_{1k}} \ln[[F_k(x_i)] [1-F_k(x_i)]] +$$
$$\sum_{j=1}^{n_{2k}} \frac{1}{I_m[F_k(x_j)]} \frac{\partial I_m[F_k(x_j)]}{\partial m} \} .$$

The second-order derivatives of the log-likelihood function with respect to α, β, and m is shown by

$$
\begin{aligned}
\frac{\partial^2 l}{\partial\alpha\partial\alpha}\Big|_{m_1=m_2=m} = \sum_{k=1}^{N} \{ & -\frac{n_{1k}}{\alpha^2} - 2m \sum_{i=1}^{n_{1k}} [\ln(x_i)]^2 \;\; [F_k(x_i)] \; [1-F_k(x_i)] \; + \\
& \frac{1}{B(m,m)} \sum_{j=1}^{n_{2k}} \frac{1}{I_m[F_k(x_j)]} \; [\; [\ln(x_j)]^2 \; [F_k(x_j)]^2 \; . \\
& [1-F_k(x_j)] \;]^m \; [1-2F_k(x_i)] \; + \\
& m \; [\; [\ln(x_j)]^2 \; [F_k(x_j)]^m \; [1-F_k(x_i)]^m \; [1-2F_k(x_i)] \;] \; - \\
& \frac{1}{[B(m,m)]^2} \sum_{j=1}^{n_{2k}} \frac{1}{[I_m[F_k(x_j)]]^2} \; [\ln(x_j)]^2 \; [F_k(x_j)]^m \; [1-F_k(x_i)]^m \; \}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 l}{\partial\alpha\partial\beta}\Big|_{m_1=m_2=m} = \sum_{k=1}^{N} \{ & -2m \sum_{i=1}^{n_{1k}} [F_k(x_i)] \; [1 - 2F_k(x_i)] \; [\ln(x_i)] \; + \\
& \frac{1}{B(m,m)} \sum_{j=1}^{n_{2k}} \frac{\ln(x_j)}{I_m[F_k(x_i)]} \; [\;\; m[F_k(x_i)]^m \; [1-F_k(x_j)]^{m+1} \; - \\
& [F_k(x_i)]^{m+1} \; [1-F_k(x_j)]^m \; + \\
& \frac{1}{B(m,m)} \; \frac{1}{I_m[F_k(x_i)]} \; [F_k(x_i)]^{2m-1} \; [1-F_k(x_i)]^{2m-1} \;] \; \}
\end{aligned}
$$

$$\frac{\partial^2 l}{\partial\alpha\partial m} = \sum_{k=1}^{N}\{\sum_{i=1}^{n_{1k}}[1 - 2F_k(x_i)]\ [\ln(x_i)] +$$

$$\frac{1}{[\mathrm{B}(m,m)]^2}\ \frac{\partial B(m,m)}{\partial m}\sum_{j=1}^{n_{2k}}\frac{\ln(x_j)}{I_m[F_k(x_j)]}\ [F_k(x_j)]^m\ [1-F_k(x_j)]^m +$$

$$\frac{1}{\mathrm{B}(m,m)}\sum_{j=1}^{n_{2k}}\frac{\ln(x_j)}{I_m[F_k(x_j)]}\ [F_k(x_j)]^m\ \ln[F_k(x_j)]\ [1-F_k(x_j)] +$$

$$\frac{1}{\mathrm{B}(m,m)}\sum_{j=1}^{n_{2k}}\frac{\ln(x_j)}{I_m[F_k(x_j)]}\ [F_k(x_j)]\ [1-F_k(x_j)]^m\ \ln[1-F_k(x_j)]\ \}$$

$$= \sum_{k=1}^{N}\{\sum_{i=1}^{n_{1k}}[1 - 2F_k(x_i)]\ [\ln(x_i)] +$$

$$\frac{1}{[\mathrm{B}(m,m)]}\ [\Psi(m)-\Psi(2m)]\sum_{j=1}^{n_{2k}}\frac{\ln(x_j)}{I_m[F_k(x_j)]}\ [F_k(x_j)]^m\ [1-F_k(x_j)]^m +$$

$$\frac{1}{\mathrm{B}(m,m)}\sum_{j=1}^{n_{2k}}\frac{\ln(x_j)}{I_m[F_k(x_j)]}\ [F_k(x_j)]^m\ \ln[F_k(x_j)]\ [1-F_k(x_j)] +$$

$$\frac{1}{\mathrm{B}(m,m)}\sum_{j=1}^{n_{2k}}\frac{\ln(x_j)}{I_m[F_k(x_j)]}\ [F_k(x_j)]\ [1-F_k(x_j)]^m\ \ln[1-F_k(x_j)]\ \}$$

$$\frac{\partial^2 l}{\partial\beta\partial\beta}\Big|_{m_1=m_2=m} = \sum_{k=1}^{N}\{\ -2m\sum_{i=1}^{n_{1k}}[F_k(x_i)]\ [1-F_k(x_i)] +$$

$$\frac{1}{\mathrm{B}(m,m)}\ \sum_{j=1}^{n_{2k}}[\ \frac{m}{I_m[F_k(x_j)]}\ [F_k(x_j)]^m\ [1-F_k(x_j)]^{m+1} -$$

$$\frac{m}{I_m[F_k(x_j)]}\ [F_k(x_j)]^{m+1}\ [1-F_k(x_j)]^m -$$

$$\frac{1}{\mathrm{B}(m,m)}\ \frac{1}{[I_m[F_k(x_j)]]^2}\ [F_k(x_j)]^{2m}\ [1-F_k(x_j)]^{2m}\ ]\ \}$$

$$\frac{\partial^2 l}{\partial\beta\partial m} = \sum_{k=1}^{N}\{ n_{1k} - 2\sum_{i=1}^{n_{1k}} [F_k(x_i)] +$$

$$\frac{1}{[\mathrm{B}(m,m)]^2}\,\frac{\partial B(m,m)}{\partial m}\sum_{j=1}^{n_{2k}}\frac{1}{I_m[F_k(x_j)]}\,[F_k(x_j)]^m\,[1-F_k(x_j)]^m +$$

$$\frac{1}{\mathrm{B}(m,m)}\sum_{j=1}^{n_{2k}}\frac{1}{I_m[F_k(x_j)]}\,[F_k(x_j)]^m\,\ln[F_k(x_j)]\,[1-F_k(x_j)] +$$

$$\frac{1}{\mathrm{B}(m,m)}\sum_{j=1}^{n_{2k}}\frac{1}{I_m[F_k(x_j)]}\,[F_k(x_j)]\,[1-F_k(x_j)]^m\,\ln[1-F_k(x_j)]\,\}$$

$$= \sum_{k=1}^{N}\{ \sum_{i=1}^{n_{1k}} [F_k(x_i)] +$$

$$\frac{1}{[\mathrm{B}(m,m)]}\,[\Psi(m)-\Psi(2m)]\sum_{j=1}^{n_{2k}}\frac{1}{I_m[F_k(x_j)]}\,[F_k(x_j)]^m\,[1-F_k(x_j)]^m +$$

$$\frac{1}{\mathrm{B}(m,m)}\sum_{j=1}^{n_{2k}}\frac{1}{I_m[F_k(x_j)]}\,[F_k(x_j)]^m\,\ln[F_k(x_j)]\,[1-F_k(x_j)] +$$

$$\frac{1}{\mathrm{B}(m,m)}\sum_{j=1}^{n_{2k}}\frac{\ln(x_j)}{I_m[F_k(x_j)]}\,[F_k(x_j)]\,[1-F_k(x_j)]^m\,\ln[1-F_k(x_j)]\,\}$$

$$\frac{\partial^2 l}{\partial m\partial m} = \sum_{k=1}^{N}\{ -2\,n_{1k}\,[\Psi'(m) - 2\Psi'(2m)] +$$

$$\sum_{j=1}^{n_{2k}}[\,\frac{1}{I_m[F_k(x_j)]}\,\frac{\partial^2 I_m[F_k(x_j)]}{\partial m\partial m} -$$

$$\frac{1}{(I_m[F_k(x_j)])^2}\,(\,\frac{\partial}{\partial m}I_m[F_k(x_j)]\,)^2\,]\,\}\,.$$

APPENDIX B

DATA SETS

Below are data mercury concentration in ppm in 115 sample swordfish (Lee and Krutchkoff, 1980):

0.05
0.07
0.07
0.13
0.13
0.19
0.24
0.25
0.28
0.32
0.39
0.45
0.46
0.53
0.54
0.56
0.60
0.60
0.61
0.62
0.65
0.71
0.72
0.75
0.76
0.79
0.81
0.81
0.82
0.82
0.82
0.83
0.83
0.83
0.84
0.85
0.89
0.90
0.91
0.92

0.92
0.93
0.95
0.95
0.97
0.97
0.98
1.00
1.00
1.01
1.02
1.04
1.05
1.05
1.08
1.10
1.12
1.12
1.14
1.14
1.15
1.16
1.20
1.20
1.20
1.20
1.20
1.21
1.22
1.25
1.25
1.26
1.27
1.27
1.29
1.29
1.29
1.29
1.30
1.31
1.32
1.32
1.37
1.37
1.39

1.39
1.40
1.40
1.41
1.42
1.43
1.44
1.45
1.54
1.54
1.58
1.58
1.60
1.60
1.62
1.62
1.66
1.66
1.68
1.69
1.72
1.74
1.85
1.89
1.96
2.06
2.10
2.23
2.25
2.72

Below are data of annual average atmospheric tritium oxide concentrations (Apt, 1976):

6.8
7.2
9.5
11.8
12.9
14.7
15.2
16.2

17.1
19.3
20.0
20.0
21.8
26.7
28.0
34.0
36.3
41.5
42.6
45.4
58.2
64.4
67.5
81.2
97.7
141.3

Below are data groundwater concentrations of copper (cu) in micrograms per liter in the San Joaquin Valley, California (Millard and Deverel, 1988):

2
2
12
2
1
<10 (censored)
<10 (censored)
4
<10 (censored)
<1 (censored)
1
<2 (censored)
<2 (censored)
1
2
<10 (censored)
3
<1 (censored)
1
1

3
<5 (censored)
17
23
9
9
3
3
<15 (censored)
<5 (censored)
4
<5 (censored)
<5 (censored)
<5 (censored)
4
8
1
15
3
3
1
6
3
6
3
4
5
14
4

Below are data of annual average of sulfur dioxide (SO2) concentrations in pphm in Long Beach, California (Roberts, 1979):

4.0
3.0
3.4
2.1
1.9
1.9
1.5
1.3

1.4
2.6
3.0
2.5
3.1
2.5
2.4
2.5
2.5
1.9
1.7

Below are data of annual averages of chloride concentrations (U.S. EPA, 1979):

34.56
119.0
41.04
53.81
197.1
49.65
73.86
28.88
62.00
63.35

Below are data of annual averages of iron concentrations (U.S. EPA, 1979):

8.45
6.17
8.73
4.64
10.66
27.29
10.07
17.58
6.00
3.30

Below are data of annual averages of aluminum concentrations (U.S. EPA, 1979):

4.57
3.79
4.31

2.15
4.82
3.42
6.62
9.95
2.98
1.73

APPENDIX C

MATLAB CODES

```
% this program computes the estimates of the parameters of GLL(m1,m2) for censored
% data using method of maximum likelihood.
% the MATLAB procedure used is fmins('function', initial, options) involving
% the Nelder-Mead simplex as an algorithm.
% alpha=a, beta=b, m1=m1, and m2=m2 are the parameters.
% beta(m1,m2) : beta function with parameters m1 and m2.
% betainc(x,m1,m2) : an incomplete beta function with parameters m1 and m2.
% the user supplies data and initial values which are initial=[a0,b0,m10,m20].
% "log" stands for a natural logarithm.
% the program is saved in an M-file.
% for convenience,  a script M-file  and data can be written in a text editor.


function y=like1cu(initial,data)

        a=initial(1,1);                   % an initial value of alpha
        b=initial(2,1);                   % an initial value of beta
        m1=initial(3,1);                  % an initial value of m1
        m2=initial(4,1);                  % an initial value of m2
        number1=size(data);
        n1=number1(1,1);                  % a number of uncensored data
        laras=ones(n1,1);                 % n1-by-1 matrix of ones

logist1=1./(laras+exp(-b-a*log(data)));         % the log-logistic function for uncensored
data
temp1=log(a)-log(beta(m1,m2));
temp2=log(logist1);
temp3=log(laras-logist1);
temp4=log(data);

        censor=[10;10;10;1;2;2;10;1;5;15;5;5;5;5];  % "detection limit" of censored data
        number2=size(censor);
        n2=number2(1,1);                             % a number of censored data
        ati=ones(n2,1);                              % n2-by-1 matrix of ones

logist2=1./(ati+exp(-b-a*log(censor)));         % the log-logistic function for censored data
temp5=log(betainc(logist2,m1,m2));

y=-(n1*temp1+m1*sum(temp2)+m2*sum(temp3)
    -sum(temp4)+sum(temp5))                     % the objective function : function for
                                                % maximizing
                                                % with minus sign
```

For executing M-file in the MATLAB editor:

```
>> load 'cu.dat';              % load 'filename of data' with an extension "dat"
>> data=cu;
>> initial=[a0,b0,m10,m20];    % give an initial value for each parameter
>> fmins('like1cu',initial,[],[]) % the fmins procedure compute the estimates of parameters
                               % by maximizing the log-likelihood function
```

% this program computes the estimates of the parameters of GLL(m,m) for censored data
% using method of maximum likelihood.
% the MATLAB procedure used is fmins('function', initial, options) involving
% the Nelder-Mead simplex as an algorithm.
% alpha=a, beta=b, and m1=m2=m are the parameters.
% beta(m,m) : beta function with parameters m1=m2=m.
% betainc(x,m,m) : an incomplete beta function with parameters m1= m2=m.
% the user supplies data and initial values which are initial=[a0,b0,m0].
% "log" stands for a natural logarithm.
% the program is saved in an M-file.
% for convenience, a script M-file and data can be written in a text editor.

```
function y=like2cu(initial,data)
        a=initial(1,1);                 % an initial value of alpha
        b=initial(2,1);                 % an initial value of beta
        m=initial(3,1);                 % an initial value of m
        number1=size(data);
        n1=number1(1,1);                % a number of uncensored data
        laras=ones(n1,1);               % n1-by-1 matrix of ones

logist1=1./(laras+exp(-b-a*log(data)));       % the log-logistic function for uncensored
data
temp1=log(a)-log(beta(m,m));
temp2=log(logist1);
temp3=log(laras-logist1);
temp4=log(data);

        censor=[10;10;10;1;2;2;10;1;5;15;5;5;5;5];   % "detection limit" of censored data
        number2=size(censor);
        n2=number2(1,1);                              % a number of censored data
        ati=ones(n2,1);                               % n2-by-1 matrix of ones

logist2=1./(ati+exp(-b-a*log(censor)));       % the log-logistic function for censored data
temp5=log(betainc(logist2,m,m));
y=-(n1*temp1+m*sum(temp2)+m*sum(temp3)
    -sum(temp4)+sum(temp5))                   % the objective function : function for
                                              % maximizing
                                              % with minus sign
```

For executing M-file in the MATLAB editor:

```
>> load 'cu.dat';                    % load 'filename of data' with an extension "dat"
>> data=cu;
>> initial=[a0,b0,m0];               % give an initial value for each parameter
>> fmins('like2cu',initial,[],[])    % the fmins procedure compute the estimates of parameters
                                     % by maximizing the log-likelihood function
```

```
% this program computes the estimates of the parameters of GLL(m,1) for censored data
% using method of maximum likelihood.
% the MATLAB procedure used is fmins('function', initial, options) involving
% the Nelder-Mead simplex as an algorithm.
% alpha=a, beta=b, and m1=m are the parameters.
% beta(m,1) : beta function with parameters m1=m and set m2=1.
% betainc(x,m,1) : an incomplete beta function with parameters m1=m and set m2=1.
% the user supplies data and initial values which are initial=[a0,b0,m0].
% "log" stands for a natural logarithm.
% the program is saved in an M-file.
% for convenience, a script M-file and data can be written in a text editor.


function y=like3cu(initial,data)
        a=initial(1,1);              % an initial value of alpha
        b=initial(2,1);              % an initial value of beta
        m=initial(3,1);              % an initial value of alpha
        number1=size(data);
        n1=number1(1,1);             % a number of uncensored data
        laras=ones(n1,1);            % n1-by-1 matrix of ones

logist1=1./(laras+exp(-b-a*log(data)));      % the log-logistic function for uncensored
data
temp1=log(a)-log(beta(m,1));
temp2=log(logist1);
temp3=log(laras-logist1);
temp4=log(data);

        censor=[10;10;10;1;2;2;10;1;5;15;5;5;5;5];  % "detection limit" of censored data
        number2=size(censor);
        n2=number2(1,1);             % a number of censored data
        ati=ones(n2,1);              % n2-by-1 matrix of ones

logist2=1./(ati+exp(-b-a*log(censor)));      % the log-logistic function for censored data
temp5=log(betainc(logist2,m,1));
y=-(n1*temp1+m*sum(temp2)+sum(temp3)
    -sum(temp4)+sum(temp5))                  % the objective function : function for
                                             % maximizing
                                             % with minus sign
```

For executing M-file in the MATLAB editor:

```
>> load 'cu.dat';               % load 'filename of data' with an extension "dat"
>> data=cu;
>> initial=[a0,b0,m0];          % give initial values for each parameter
>> fmins('like3cu',initial,[],[]) % the fmins procedure compute the estimates of parameters
                                % by maximizing the log-likelihood function
```

```
% this program computes the estimates of the parameters of GLL(1,m) for censored data
% using method of maximum likelihood.
% the MATLAB procedure used is fmins('function', initial, options) involving
% the Nelder-Mead simplex  as an algorithm.
% alpha=a, beta=b, and m2=m are the parameters.
% beta(1,m) : beta function with parameters m2=m and set m1=1.
% betainc(x,1,m) : an incomplete beta function with parameters m2=m and set m1=1.
% the user supplies data and initial values which are initial=[a0,b0,m0].
% "log" stands for a natural logarithm.
% the program is saved in an M-file.
% for convenience,  a script M-file  and data can be written in a text editor.


function y=like4cu(initial,data)
        a=initial(1,1);                 % an initial value of alpha
        b=initial(2,1);                 % an initial value of beta
        m=initial(3,1);                 % an initial value of m
        number1=size(data);
        n1=number1(1,1);                % a number of uncensored data
        laras=ones(n1,1);               % n1-by-1 matrix of ones

logist1=1./(laras+exp(-b-a*log(data)));      % the log-logistic function for uncensored
data
temp1=log(a)-log(beta(1,m));
temp2=log(logist1);
temp3=log(laras-logist1);
temp4=log(data);

        censor=[10;10;10;1;2;2;10;1;5;15;5;5;5;5];  % "detection limit" of censored data
        number2=size(censor);
        n2=number2(1,1);                            % a number of censored data
        ati=ones(n2,1);                             % n2-by-1 matrix of ones

logist2=1./(ati+exp(-b-a*log(censor)));      % the log-logistic function for censored data
temp5=log(betainc(logist2,1,m));
y=-(n1*temp1+sum(temp2)+m*sum(temp3)
    -sum(temp4)+sum(temp5))                  % the objective function : function for
                                             % maximizing
                                             % with minus sign
```

For executing M-file in the MATLAB editor:

```
>> load 'cu.dat';                % load 'filename of data' with an extension "dat"
>> data=cu;
>> initial=[a0,b0,m0];           % give initial values for each parameter
>> fmins('like4cu',initial,[],[]) % the fmins procedure compute the estimates of parameters
                                 % by maximizing the log-likelihood function
```

```
% this program computes the estimates of the parameters of GLL(1,1) for censored data
% using method of maximum likelihood.
% the MATLAB procedure used is fmins('function', initial, options) involving
% the Nelder-Mead simplex as an algorithm.
% alpha=a and beta=b are the parameters.
% beta(1,1) : beta function and  set m1=1 and  m2=1.
% betainc(x,1,1) : an incomplete beta function and set m1=1 and m2=1.
% the user supplies data and initial values which are initial=[a0,b0].
% "log" stands for a natural logarithm.
% the program is saved in an M-file.
% for convenience,  a script M-file  and data can be written in a text editor.


function y=like5cu(initial,data)
        a=initial(1,1);              % an initial value of alpha
        b=initial(2,1);              % an initial value of beta
        number1=size(data);
        n1=number1(1,1);             % a number of uncensored dat
        laras=ones(n1,1);            % n1-by-1 matrix of ones

logist1=1./(laras+exp(-b-a*log(data)));      % the log-logistic function for uncensored
data
temp1=log(a)-log(beta(1,1));
temp2=log(logist1);
temp3=log(laras-logist1);
temp4=log(data);

        censor=[10;10;10;1;2;2;10;1;5;15;5;5;5;5;5];   % "detection limit" of censored data

        number2=size(censor);
        n2=number2(1,1);                    % a number of censored dat
        ati=ones(n2,1);                     % n2-by-1 matrix of ones

logist2=1./(ati+exp(-b-a*log(censor)));     % the log-logistic function for censored data
temp5=log(betainc(logist2,1,1));
y=-(n1*temp1+sum(temp2)+sum(temp3)
     -sum(temp4)+sum(temp5))                % the objective function : function for
maximizing

                                            % with minus sign
```

For executing M-file in the MATLAB editor:

```
>> load 'cu.dat';                % load 'filename of data' with an extension "dat"
>> data=cu;
>> initial=[a0,b0];              % give initial values for each parameter
>> fmins('like5cu',initial,[],[]) % the fmins procedure compute the estimates of parameters
                                 % by maximizing the log-likelihood function
```

```
% this program computes the estimates of the parameters of GLL(m1,m2) for uncensored
% data using method of maximum likelihood.
% the MATLAB procedure used is fmins('function', initial, options) involving
% the Nelder-Mead simplex as an algorithm.
% alpha=a, beta=b, m1=m1, and m2=m2 are the parameters.
% beta(m1,m2) : beta function with parameters m1 and m2.
% the user supplies data and initial values which are initial=[a0,b0,m10,m20].
% "log" stands for a natural logarithm.
% the program is saved in an M-file.
% for convenience, a script M-file and data can be written in a text editor.


function y=like1u(initial,data);
        a=initial(1,1);              % an initial value of alpha
        b=initial(2,1);              % an initial value of beta
        m1=initial(3,1);             % an initial value of m1
        m2=initial(4,1);             % an initial value of m2
        number=size(data);
        n=number(1,1);               % a number of data
        laras=ones(n,1);             % n-by-1 matrix of ones

logist=1./(laras+exp(-b-a*log(data)));      % the log-logistic function
temp1=log(a)-log(beta(m1,m2));
temp2=log(logist);
temp3=log(laras-logist);
temp4=log(data);
y=-(n*temp1+m1*sum(temp2)
    +m2*sum(temp3)-sum(temp4))   % the objective function : function for maximizing
                                 % with minus sign
```

For executing M-file in the MATLAB editor:

```
>> load 'mercu.dat';             % load 'filename of data' with an extension "dat"
>> data=mercu;
>> initial=[a0,b0,m10,m20];     % give an initial value for each parameter
>> fmins('like1u',initial,[],[]) % the fmins procedure compute the estimates of parameters
                                 % by maximizing the log-likelihood function
```

```
% this program computes the estimates of the parameters of GLL(m,m) for uncensored
% data using method of maximum likelihood.
% the MATLAB procedure used is fmins('function', initial, options) involving
% the Nelder-Mead simplex as an algorithm.
% alpha=a, beta=b, and m1=m2=m are the parameters.
% beta(m,m) : beta function with parameters m1=m2=m.
% the user supplies data and initial values which are initial=[a0,b0,m0].
% "log" stands for a natural logarithm.
% the program is saved in an M-file.
% for convenience,  a script M-file  and data can be written in a text editor.


function y=like2u(initial,data);
        a=initial(1,1);                  % an initial value of alpha
        b=initial(2,1);                  % an initial value of beta
        m=initial(3,1);                  % an initial value of m
        number=size(data);
        n=number(1,1);                   % a number of data
        laras=ones(n,1);                 % n-by-1 matrix of ones

logist=1./(laras+exp(-b-a*log(data)));        % the log-logistic function
temp1=log(a)-log(beta(m,m));
temp2=log(logist);
temp3=log(laras-logist);
temp4=log(data);
y=-(n*(log(a)-log(beta(m,m)))+m*sum(temp2)
    +m*sum(temp3)-sum(temp4))    % the objective function : function for maximizing
                                 % with minus sign
```

For executing M-file in the MATLAB editor:

```
>> load 'mercu.dat';          % load 'filename of data' with an extension "dat"
>> data=mercu;
>> initial=[a0,b0,m0];        % give an initial value for each parameter
>> fmins('like2u',initial,[],[])  % the fmins procedure compute the estimates of parameters
                              % by maximizing the log-likelihood function
```

```
% this program computes the estimates of the parameters of GLL(m,1) for uncensored
% data using method of maximum likelihood.
% the MATLAB procedure used is fmins('function', initial, options) involving
% the Nelder-Mead simplex as an algorithm.
% alpha=a, beta=b, and m1=m are the parameters.
% beta(m,1) : beta function with parameters m1=m and set m2=1.
% the user supplies data and initial values which are initial=[a0,b0,m0].
% "log" stands for a natural logarithm.
% the program is saved in an M-file.
% for convenience,  a script M-file  and data can be written in a text editor.


function y=like3u(initial,data);
        a=initial(1,1);              % an initial value of alpha
        b=initial(2,1);              % an initial value of beta
        m=initial(3,1);              % an initial value of m
        number=size(data);
        n=number(1,1);               % a number of data
        laras=ones(n,1);             % n-by-1 matrix of ones

logist=1./(laras+exp(-b-a*log(data)));          % the log-logistic function
temp1=log(a)-log(beta(m,1));
temp2=log(logist);
temp3=log(laras-logist);
temp4=log(data);
y=-(n*temp1+m*sum(temp2)
    +sum(temp3)-sum(temp4))          % the objective function : function for maximizing
                                     % with minus sign
```

For executing M-file in the MATLAB editor:

```
>> load 'mercu.dat';            % load 'filename of data' with an extension "dat"
>> data=mercu;
>> initial=[a0,b0,m0];          % give an initial value for each parameter
>> fmins('like3u',initial,[],[])  % the fmins procedure compute the estimates of parameters
                                % by maximizing the log-likelihood function
```

```
% this program computes the estimates of the parameters of GLL(1,m) for uncensored
% data using method of maximum likelihood.
% the MATLAB procedure used is fmins('function', initial, options) involving
% the Nelder-Mead simplex as an algorithm.
% alpha=a, beta=b, and m2=m are the parameters.
% beta(1,m) : beta function with parameters m2=m and set m1=1.
% the user supplies data and initial values which are initial=[a0,b0,m0].
% "log" stands for a natural logarithm.
% the program is saved in an M-file.
% for convenience,  a script M-file  and data can be written in a text editor.


function y=like4u(initial,data);
        a=initial(1,1);               % an initial value of alpha
        b=initial(2,1);               % an initial value of beta
        m=initial(3,1);               % an initial value of m
        number=size(data);
        n=number(1,1);                % a number of data
        laras=ones(n,1);              % n-by-1 matrix of ones

logist=1./(laras+exp(-b-a*log(data)));        % the log-logistic function
temp1=log(a)-log(beta(1,m));
temp2=log(logist);
temp3=log(laras-logist);
temp4=log(data);
y=-(n*temp1+sum(temp2)
    +m*sum(temp3)-sum(temp4))       % the objective function : function for maximizing
                                    % with minus sign
```

For executing M-file in the MATLAB editor:

```
>> load 'mercu.dat';             % load 'filename of data' with an extension "dat"
>> data=mercu;
>> initial=[a0,b0,m0];           % give an initial value for each parameter
>> fmins('like4u',initial,[],[]) % the fmins procedure compute the estimates of parameters
                                 % by maximizing the log-likelihood function
```

```
% this program computes the estimates of the parameters of GLL(1,1) for uncensored
% data using method of maximum likelihood.
% the MATLAB procedure used is fmins('function', initial, options) involving
% the Nelder-Mead simplex as an algorithm.
% alpha=a and beta=b are the parameters.
% beta(1,1) : beta function and set m1=1 and m2=1.
% the user supplies data and initial values which are initial=[a0,b0].
% "log" stands for a natural logarithm.
% the program is saved in M-file.
% for convenience,  a script M-file  and data can be written in a text editor.


function y=like5u(initial,data);
        a=initial(1,1);              % an initial value of alpha
        b=initial(2,1);              % an initial value of beta
        number=size(data);
        n=number(1,1);               % a number of data
        laras=ones(n,1);             % n-by-1 matrix of ones

logist=1./(laras+exp(-b-a*log(data)));        % the log-logistic function
temp1=log(a)-log(beta(1,1));
temp2=log(logist);
temp3=log(laras-logist);
temp4=log(data);
y=-(n*temp1+sum(temp2)
    +sum(temp3)-sum(temp4))          % the objective function : function for maximizing
                                     % with minus sign
```

For executing M-file in the MATLAB editor:

```
>> load 'mercu.dat';           % load 'filename of data' with an extension "dat"
>> data=mercu;
>> initial=[a0,b0,m0];         % give an initial value for each parameter
>> fmins('like5u',initial,[],[])  % the fmins procedure compute the estimates of parameters
                               % by maximizing the log-likelihood function
```

REFERENCES

Aitchison, J. and Brown, J.A.C. (1957). The lognormal distribution. London: Cambridge University Press.

Apt, K.E. (1976). Applicability of the Weibull distribution function to atmospheric radioactivity data. Atmospheric Environment, 10, 777-781.

Bencala, K.E. and Seinfeld, J.H. (1976). On frequency distributions of air pollutant concentrations. Atmospheric Environment, 10, 941-950.

Berger, A., Melice, J.L., Demuth, C. (1982). Statistical distributions of daily and high atmospheric SO2-concentrations. Atmospheric Environment, 16(12), 2863-2877.

Ciampi, A., Hogg, S.A., and Kates, L. (1986). Regression analysis of censored survival data with the generalized F family--an alternative to the proportional hazards model. Statistics in Medicine, 50, 85-96.

Cohen, A.C. (1988). Censored, truncated, and grouped estimation. In E.L. Crow and K. Shimizu (Eds.), Lognormal distributions: theory and applications (pp. 139-172). New York: Marcel Dekker.

Crow, E.L. and Shimizu, K. (1988). Lognormal distributions: theory and applications. New York: Marcel Dekker.

DiCiccio, T.J. (1987). Approximate inference for the generalized gamma distribution. Technometrics, 29(1), 33-40.

El-Shaarawi. (1989). Inferences about the mean from censored water quality data. Water Resources Research, 25(4), 685-690.

Georgopoulos, P.G. and Seinfeld, J.H. (1982). Statistical distributions of air pollutant Concentrations. Environmental Science and Technology, 16(7), 401A-415A.

Gilbert, R.O. (1987). Statistical methods for environmental pollution monitoring. New York: Van Nostrand Reinhold.

Gilbert, R.O. and Kinnison, R.R. (1981). Statistical methods for estimating the mean and variance from radionuclide data sets containing negative, unreported or less-than values. Health Physics, 40, 377-390.

Gilliom, R.J. and Helsel, D.R. (1986). Estimation of distributional parameters for censored trace level water quality data 1. Estimation techniques. Water Resources Research, 22(2), 135-146.

Hanselman, D. and Littlefield, B. (1995). The student edition of matlab version 4: user's guide. New Jersey: The Math Works.

Harter, H.L. and Moore, A.H. (1966). Local-maximum-likelihood estimation of the parameters of three-parameter lognormal populations from complete and censored samples. Journal of the American Statistical Association, 61, 842-851.

Helsel, D.R. and Gilliom, R.J. (1986). Estimation of distributional parameters for censored trace level water quality data 2. Verification and applications. Water Resources Research, 22(2), 147-155.

Hogg, S.A. and Ciampi, A. (1985). GFREG: a Computer program for maximum likelihood regression using the generalized F distribution. Computer Methods and Programs in Biomedicine, 20, 201-215.

Holland, D.M. and Fitz-Simons, T. (1982). Fitting statistical distributions to air quality data by the maximum likelihood method. Atmospheric Environment, 16(5), 1071-1076.

Jakeman, A.J. and Taylor, J.A. (1985). A hybrid ATDL-gamma distribution model for predicting urban area source acid gas concentrations. Atmospheric Environment, 19, 1959-1967.

Johnson, N.L. and Kotz, S. (1970). Distributions in statistics: continuous univariate distributions-1. Boston: Houghton Mifflin Company.

Johnson, N.L. and Kotz, S., and Balakrishnan, N. (1994). Distributions in statistics: continuous univariate distributions. Volume 1, Second Edition. New York: John Wiley & Sons.

Kalpasanov, Y. and Kurchatova, G. (1976). A study of the statistical distribution of chemical pollutants in air. Journal of The Air Pollution Control Association, 26(10), 981-985.

Kendall, M.G. and Stuart, A. (1979). The advanced theory of statistics. Vol. 2 (4th edition). New York: Macmillan.

Kushner, E.J. (1976). On determining the statistical parameters for pollution concentration from a truncated data set. Atmospheric Environment, 10, 975-979.

Larsen, R.I. (1969). A new mathematical model of air pollutant concentration averaging time and frequency. Journal of The Air Pollution Control Association, 19(1), 24-30.

Larsen, R.I. (1973). An air quality data analysis system for interrelating effects, standards and needed source reductions. Journal of The Air Pollution Control Association, 23(11), 933-940.

Larsen, R.I. (1974). An air quality data analysis system for interrelating effects, standards and needed source reductions-part 2. Journal of The Air Pollution Control Association, 24(6), 551-558.

Larsen, R.I. (1977). An air quality data analysis system for interrelating effects, standards, and needed source reductions: part 4. A three-parameter averaging-time model. Journal of The Air Pollution Control Association, 27(5), 454-459.

Lee, L. and Krutchkoff, R.G. (1980). Mean and variance of partially-truncated distributions. Biometrics, 36, 531-536.

Mage, D.T. (1981). A review of the application of probability models for aerometric data. In Environmetrics 81: Selected Papers, SIAM-SIMS Conference Series No. 8. Society for Industrial and Applied Mathematics, Philadelphia.

Mage, D.T. and Ott, W.R. (1978). Refinements of the lognormal probability model for analysis of aerometric data. Journal of The Air Pollution Control Association, 28(8), 796-798.

Mage, D.T. and Ott, W.R. (1984). An evaluation of the methods of fractiles, moments and maximum likelihood for estimating parameters when sampling air quality data from a stationary lognormal distribution. Atmospheric Environment, 18(1), 163-171.

Marani, A., Lavagni, I., and Buttazzoni, C. (1986). Statistical study of air pollutant concentrations via generalized gamma distributions. Journal of The Air Pollution Control Association, 36(11), 1250-1254.

McDonald, J.B. (1984). Some generalized functions for the size distribution of income. Econometrica, 52(3), 647-663.

McDonald, J.B. and Richards, D.O. (1987). Model selection: some generalized distributions. Commun. Statist.-Theory Meth., 16(4), 1049-1074.

Millard, S.P. and Deverel, S.J. (1988). Nonparametric statistical methods for comparing two sites based on data with multiple nondetect limits. Water Resources Research, 24(12), 2087-2098.

Newman, M.C., Dixon, P.M., Looney, B.B., and Pinder, J.E. (1989). Estimating mean and variance for environmental samples with below detection limit observations. Water Resources Bulletin, 25(4), 905-916.

Okur, M.C. (1988). On fitting the generalized γ-distribution to air pollution data. Atmospheric Environment, 22(11), 2569-2572.

Ott, W.R. (1995). Environmental statistics and data analysis. Boca Raton: Lewis.

Owen, W.J. and DeRouen, T.A. (1980). Estimation of the mean for lognormal data containing zeroes and left-censored values, with applications to the measurement of worker exposure to air contaminants. Biometrics, 36, 707-719.

Roberts, E.M. (1979). Review of statistics of extreme values with applications to air quality data. Part II. Applications. Journal of The Air Pollution Control Association, 29(7), 733-740.

Schubert, J., Brodsky, A., and Tyler, S. (1967). The log-normal function as a stochastic model of the distribution of strontium-90 and other fission products in humans. Health Physics , 13, 1187-204.

Shimizu, K. and Crow, E.L. (1988). History, genesis, and properties. In E.L. Crow and K. Shimizu (Eds.), Lognormal distributions: theory and applications (pp. 1-25). New York: Marcel Dekker.

Simpson, R.W., Butt, J., Jakeman, A.J. (1984). An averaging time model of SO2 frequency distributions from a single point source. Atmospheric Environment, 18(6), 1115-1123.

Singh, K.P. (1989). A generalized log-logistic regression model for survival analysis: hazard rate characteristics. Biometrie-Praximetrie, 29, 63-74.

Singh, K.P., Bartolucci, A.A., and Burgard, S.L. (1994). Two-step procedure for survival data. Biometrie-Praximetrie, 34, 1-12.

Stoline, M.R. (1991). An examination of the lognormal and box and cox family of transformations in fitting environmental data. Environmetrics, 2(1), 85-106.

Taylor, J.A., Jakeman, A.J., and Simpson, R.W. (1986). Modeling distributions of air pollutant concentratios--I. Identification of statistical models. Atmospheric Environment, 20(9), 1781-1789.

Tiku, M.L. (1968). Estimating the parameters of log-normal distribution from censored samples. Journal of the American Statistical Association, 63, 134-140.

U.S. Environmental Protection Agency. (1979). Surface water monitoring data. EPA-905/4-79-029.

GRADUATE SCHOOL
UNIVERSITY OF ALABAMA AT BIRMINGHAM
DISSERTATION APPROVAL FORM

Name of Candidate Warsono

Major Subject Biostatistics

Title of Dissertation Analysis of Environmental Pollutant Data Using Generalized Log-Logistic Distribution

Dissertation Committee:

_______________, Chairman

Elwin Bradley Jr

Charles R Katholi

Pauline E. Jolly

Deodutta Roy

Alfred G. Bartolucci

Director of Graduate Program Elwin Bradley Jr

Dean, UAB Graduate School Joan Lorden

Date 8/28/96