5-1980

Extreme Value Distribution in Hydrology

Bill (Tzeng-Lwen) Chen
Utah State University

EXTREME VALUE DISTRIBUTION IN HYDROLOGY

by

Bill (Tzeng-Lwen) Chen

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Applied Statistics

UTAH STATE UNIVERSITY
Logan, Utah
1980

ACKNOWLEDGEMENTS

I would like to especially thank my major professor, Dr. Ronald V. Canfield for the many hours of precious help he extended to me during my studies. As a friend and teacher he made this time at Utah State University both enjoyable and educational.

Thanks also to my committee members Dr. Rex L. Hurst and Dr. Gregory W. Jones.

I would also like to thank Dr. David L. Turner, Dr. Ronald V. Sisson, Dr. David White and Mary V. Kolesar who made my education at Utah State a pleasant and successful one.

I dedicate this work to my parents who helped me financially and for their love and encouragement throughout my studies. Also to my wife Li-Chuan, whose love guided me through the past two years.

I also thank Stephen Kan, Joanna Chien and Mei-Eing Chiu for being my best friends during the past two years.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ii

LIST OF TABLES iv

LIST OF FIGURES v

ABSTRACT vi

Chapter

I. INTRODUCTION 1

Significance of Flood-Frequency-Analysis 1
Brief History of Flood-Frequency-Analysis 2
Objective of Study 4
Data and Methods. 4

II. EXTREME VALUE APPLICATION - HOMOGENEOUS DATA 5

Extreme Value Distributions 7
Determining Extreme Value Type 9
Estimation of Parameters 15
Goodness-of-fit Comparisons 15
A Graphical Technique 18

III. CONCLUSION 40

IV. FUTURE STUDIES 42

LITERATURE CITED 43

APPENDICES 45

LIST OF TABLES

Table Page

1. Selected Stations Exhibiting Homogeneous Sources. . . 6
2. Mean of the Absolute Relative Deviations. 22
3. Mean of the Quadratic Deviations. 23
4. Computed Flood Discharges (m^3/s) for Selected Return Periods. 24
5. Data Values D(T) (m^3/s) as Interpolated Between Adjacent Observations 25

LIST OF FIGURES

Figure Page

1. Straight Line Plot 10
2. Straight Line with Negative Slope 10
3. Straight Line with Positive Slope 10
4. Verification for the Kymijoki River in Finland 10
5. Station bB24--Mali River 26
6. Station HE60--Susquehanna River 27
7. Station IB06--Krishna River 28
8. Station BF40--Elbe River 29
9. Station BE38--Danube River 30
10. Station BF19--Gloma River 31
11. Station CF25--Neman River 32
12. Station ME19--Fraser River 33
13. Station JF792--Headingly River 34
14. Station IF00--Medicine Hat River 35
15. Station KF62--Saskatoon River 36
16. Station DF53--Prince Albert River 37
17. Station hE88a--Amos River. 38
18. Station jF50a--Slave Falls River 39

ABSTRACT

Extreme Value Distribution in Hydrology

by

Bill (Tzeng-Lwen) Chen, Master of Science

Utah State University, 1980

Major Professor: Dr. Ronald V. Canfield
Department: Applied Statistics

The problems encountered when empirical fit is used as the sole criterion for choosing a distribution to represent annual flood data are discussed. Some theoretical direction is needed for this choice. Extreme value theory is established as a viable tool for analyzing annual flood data. Extreme value distributions have been used in previous analyses of flood data. However, no systematic investigation of the theory has previously been applied. Properties of the extreme value distributions are examined. The most appropriate distribution for flood data has not previously been fit to such data. The fit of the chosen extreme value distribution compares favorably with that of the Pearson and log Pearson Type III distributions.

(59 pages)

CHAPTER I

INTRODUCTION

Significance of Flood-Frequency-Analysis

With continuing development of flood plains and rural watersheds for urban use, flood control becomes increasingly important. Construction of dams, water needed for irrigational purposes, keeping a river within its embankments, etc., all require estimation of flood frequency and severity. The design of structures related to water resources management and control is heavily dependent on the extreme hydrologic event.

The central hydrologic information to flood control and floodplain management planning is the relationship between peak flow and return period. (Note that the flood is defined to be the maximum annual flow.) The relationship is established by selecting an appropriate distribution to represent the population of peak flows from each year of record (the annual flood series) and estimating parameters for that distribution that best fit the recorded data. Selecting a distribution to describe floods has been essentially one of curve fitting. It is very necessary in the application of these distributions for design and management decisions to extrapolated, i.e., to estimate return periods beyond the range of the data. Thus, the hydrologist is forced to make decisions in regions in which he has no data.

Brief History of Flood-Frequency-Analysis

In the past, empirical fit has been the only criterion for choosing from among several candidates, the distribution to describe floods. It is sometimes suggested that no distribution is perfect; therefore, several may do an adequate job, and certainly the "best" fit will be close. This argument may be valid when the distributions are used to estimate probabilities or return periods of common events. However, when estimates are needed for extreme or rare events, a distribution selected on the basis of empirical fit can cause serious problems. The problem arises because the probabilities of rare events are computed from the tails of a distribution whereas empirical fit is dominated by the body of the data set. Complete reliance on empirical fit for choosing a distribution for homogeneous runoff is potentially dangerous because many distribution can provide a good empirical fit in the range of data set and yet have very different tail characteristics. It is the tail characteristic of the estimated distribution that is used in extrapolating return periods. Thus, in addition to empirical fit, the right hand tail of a distribution is an extremely important consideration. Since there is no data in this region, a theoretical motivation is needed. In the studies on rivers with homogeneous sources of runoff by Benson (1968), Beard (1974), Bobee and Robitaille (1977) and others; the characteristics of the right tail of the distributions examined were not even considered.

Methods of flood-frequency-analysis, which started about 1914, have developed along divergent lines, with resulting nonuniformity in methods of analysis and, hence, in results. This and the need for the upmost possible uniformity have induced the U.S. Water Resources Council to form the Work Group on Flow Frequency Methods with the object of developing an uniform technique to determine flood frequency. As was reported by Benson (1968) and his work-group, the main conclusions of the work-group is that after fitting several distributions to many different data sets representing a wide variety of condition choose log Pearson Type III distribution as a base method. It has been chosen from among several candidate distributions by first estimating the parameters of each distribution for each of the large number of gaged records. Then a goodness-of-fit criterion which emphasizes selected flood flows from 2 to 100 years (U.S. Water Resources Council 1976, Appendix 14) was used to select the best overall fit. Although selection of the log Pearson Type III is based upon fit in the right tail, estimation of parameters for each distribution is by standard methods which emphasized fit in the body of the data. In certain cases, the fit in the right tail is poor. Even if the fit is good, blind application of a distribution selected on the basis of empirical fit can lead to serious error. According to the report by B. B. Bobee and R. Robintaille in 1977 the main objective of their study has been the comparison between the Pearson Type III and the log Pearson Type III distributions. Different methods of fitting have been applied to a group of long-term records of annual flood peaks previously tested for independence and homogeneity. The conclusion has been that Pearson Type III distribution conforms generally better to annual flood data than the log Pearson Type III distribution.

Objective of Study

One theoretical basis for selection of the distribution for annual floods is evaluated in this paper. The annual flood event is the maximum or extreme value of all the events occurring during the year; therefore, extreme value theory would seem to provide a reasonable theoretical base and is the one examined here. Although extreme value distributions have been used in hydrology, no systematic application of the theory is reported in the literature. The application of extreme value theory for homogeneous runoff is suggested here as a possible solution which has never been tested.

Data and Methods

The data selected by B. B. Bobee and R. Robitaille is used here to estimate parameters of the extreme value distribution. The goodness-of-fit statistics used by them is used in this study. These statistics have the same basis as those used by the Work Group on Flow Frequency Methods (Benson, 1968). The statistics are essentially the average absolute deviation and the average guadratic deviation expressed as a percent between the predicted flow over selected recurrence intervals and the observed flow. By comparing the values computed from the same data set by Bobee and Robitaille for the distributions selected in his study, the usefulness of this distributions can be established.

CHAPTER II

EXTREME VALUE APPLICATION - HOMOGENEOUS DATA

The purpose of the research reported was to evaluate extreme value theory as a tool in identifying a distribution for annual floods. It should be understood that in all likelihood no single distribution is correct for all situations. For example, the systems with large carry-over storage or rivers which flow only intermittently may violate the assumptions of extreme value theory. In the first case, flood peaks become dependent on flows in the previous year; and in the second, having zero flows for all events is not really an extreme value situation.

However, if the theory is shown to apply in several cases, the hypotheses of the theory are sufficiently general to expect it to apply in a wide variety of cases. In this section a theoretical distribution is selected by matching physical characteristics of stream flow with the mathematical characteristics of the various extreme value forms. If the theory applies to stream flows, this distribution should provide good (but not necessarily best) fit over a wide variety of streams. This extreme value distribution is therefore fit to data for selected stations with long periods of record from around the world (Table 1) used in the study of Bobee

Table 1. Selected Stations Exhibiting Homogeneous Sources.

Station	Country	River	Location	Drainage Area, Km^2	Record	Missing Years	Years of Record
bB24	Senegal	Mali	Bakel	218,000	1903-1966		64
hE60	USA	Susquehanna	Harrisburg, PA	62,400	1891-1967	1906, 1922, 1927 1935, 1938, 1951	70
IB06	India	Krishna	Vijayawada	251,355	1901-1960		60
BF40	Czech.	Elbe	Decin	51,104	1851-1968	1857, 1863, 1866, 1873 1874, 1879, 1884, 1898	108
BE38	Germany	Danube	Hofkirchen	47,495	1901-1968		68
BF19	Norway	Gloma	Langnes	40,170	1902-1968	1964	66
CF25	USSR	Neman	Smalininkai	81,200	1812-1969	1944, 1945, 1946	155
mE19	Canada	Fraser	Hope	203,000	1912-1970		59
jE792	Canada	Headingly	Assinibione	162,000	1914-1970		57
iF00	Canada	Medicine Hat	S.Saskatchewan	58,400	1913-1970		58
KF62	Canada	Saskatoon	S.Saskatchewan	139,500	1912-1970		59
KF53	Canada	Prince Albert	N.Saskatchewan	119,500			59
hE88a	Canada	Amos	Hurricana	3,680	1915-1969	1932, 1933	53
jF50a	Canada	Slave Falls Power Plant	Winnipeg	126,000	1908-1970	1909, 1911-1912, 1917 1922-1926, 1931, 1934 1939-1942, 1949, 1958 1961, 1962, 1964, 1965 1967	50

and Robitaille (1977). The same measure of goodness-of-fit is used in order to compare these results with those obtained from the distributions of their study.

Extreme Value Distributions

As a beginning point for this application, some basic elements of extreme value theory need to be reviewed. Extreme value random variables are defined as follows. Let $X_1, X_2, X_3, \ldots\ldots, X_n$ be a sample of independent, identically distributed, continuous random variables. Let

$$Z_n = \max(X_1, X_2, \ldots, X_n) \quad \ldots\ldots\ldots\ldots \quad (1)$$

and

$$Y_n = \min(X_1, X_2, \ldots, X_n) \quad \ldots\ldots\ldots\ldots \quad (2)$$

extreme value theory is concerned with the asymptotic distribution of sequences $(Z_n - b_n)/a_n$ and $(Y_n - b_n')/a_n'$, $n=1, 2, \ldots\ldots, \infty$ The norming values a_n, b_n, a_n', b_n' are dictated by the theory. The interesting result of the theory is that if an asymptotic distribution exists, there are only three types for Z_n and three types for Y_n. The mathematical characteristics for the random variables X_i which determine the resulting distribution for Z_n and Y_n are given by Gnedenko (1943). These results are difficult to use because the distribution function must be known. A less mathematical but more workable approach is suggested here.

The term "flood" by nature suggests application of extreme value theory. Since the primary interest here is in the annual maximum flows, only the distribution of Z_n is considered. Under very general conditions, it has been shown by Gnedenko (1943) that the maximum of a sufficiently long sequence of independent random variables Z_n from a given distribution must be closely approximated by one of the following three types.

$$F_1(X) = \exp\left\{-\exp\left[-\left(\frac{x-b}{c}\right)\right]\right\} \qquad -\infty<x<\infty,\ c>0 \ \ldots\ldots(3)$$

$$F_2(X) = \begin{cases} 0 & x<b \\ \exp\left\{-\left(\frac{x-b}{c}\right)^{-a}\right\} & x\geq b,\ c>0,\ a>0 \end{cases} \qquad \ldots\ldots(4)$$

$$F_3(X) = \begin{cases} 1 & x\geq b \\ \exp\left\{-\left(\frac{b-x}{c}\right)^{a}\right\} & x<b,\ c>0,\ a>0 \end{cases} \qquad \ldots\ldots(5)$$

The assumption of independence of the $X_1, X_2, \ldots, X_n$ random variables is violated in many applications. However, Watson (1952) has shown that independence is not a necessary assumption. If the randomized sequence of X_i's satisfies the assumption for all n, the theory holds.

The advantage of the theory is that once an extreme value situation is recognized one can legitimately confine the search for best fit to three extreme value distributions. The mathematical characteristics of the three distributions are very different, thus it is relatively easy to determine the correct one for a given set of data. A graphical procedure is given below for use in identifying which of the extreme value distributions should be used with a given set of data.

Determining Extreme Value Type

The three distributions (3), (4), and (5) for the maximum have some easily observed characteristics.

The function of $F_1(X)$ has no bound on X, so it is not appropriate in flood analysis.

The form $F_2(X)$ is referred to as a "Cauchy type" because the extreme values for the Cauchy distribution follow distribution (4). Cauchy type distributions are "heavy tailed" and seldom occur in nature. Thus, distribution (4) has limited usefulness compared with the other two types. There is, however, reference to its use in Gumbel (1954).

The function $F_3(X)$ is limited to some maximum value b (i.e., $F_3(X) = 1$ for $X \geq b$), thus random variables which are limited have extreme value form $F_3(X)$. The converse of this statement is not necessarily true, however, and variables which are not limited may have this form too (Gnedenko 1943).

Three simple plots constitute the easiest method of determining which of the extreme value distribution is appropriate. Let $X_{(1)}$, $X_{(2)}, \ldots, X_{(N)}$ represent the ordered extreme value date for the observed maximums.

For any random variable, the expected value of its distribution function evaluated at the ith order statistic is i/(N+1) where the sample size is N, (i.e., $E(F(X_{(i)}) = i/(N+1))$ (Lindgren 1976).

Define $E_i = i/(N+1)$. Note that from equation (3)

$$\ln(-\ln F_1(X_{(i)})) = -X_{(i)}/c + b/c \quad \ldots\ldots\ldots\ldots (6)$$

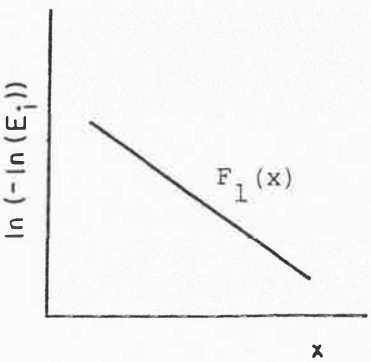

Figure 1. Straight Line Plot.

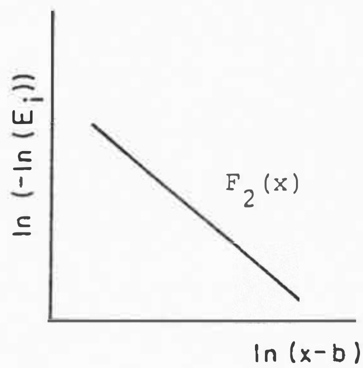

Figure 2. Straight Line with Negative Slope.

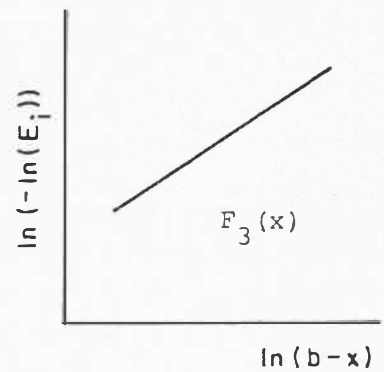

Figure 3. Straight Line with Positive Slope.

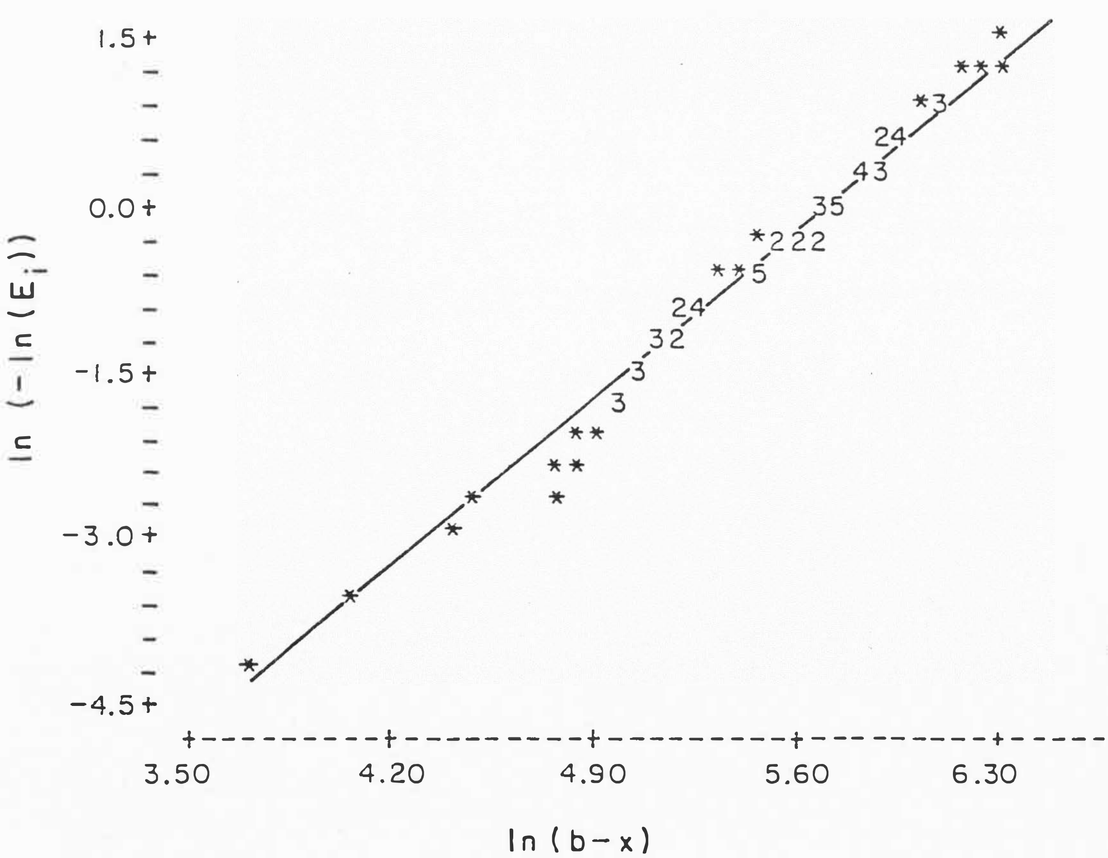

Figure 4. Verification for the Kymijoki River in Finland.

Note that the relationship in (6) is linear in $X_{(i)}$. Substituting E_i for $F(X_{(i)})$ in (6) and plotting $X_{(i)}$ vs. $\ln(-\ln F(X_{(i)})$ identifies data from a population with distribution function $F_1(X)$. If (3) is appropriate the plot will be a straight line as illustrated in Figure 1. If the data are from any other distribution, the plot will not be a straight line.

The plot which identifies data from an $F_2(X)$ population is similar. From (4) it follows that

$$\ln(-\ln F_2(X_{(i)}) = -a \ln (X_{(i)}-b) + a \ln c \quad \ldots(7)$$

Thus if data are from a population with distribution $F_2(X)$, the plot of $\ln(X_{(i)} -b)$ vs. $\ln(-\ln E_i)$ will be a straight line with negative slope as illustrated in Figure 2. The parameter b must be estimated before the plot can be made. Estimation of parameters is considered later.

The third plot which identifies $F_3(X)$ is motivated from (5) in the same manner, $\ln(-\ln F_3(X_{(i)})) = a \ln(b-X_{(i)}) - a \ln c$, i.e., the plot of $\ln(b-X_{(i)})$ vs. $\ln(-\ln E_i)$ is a straight line with positive slope as illustrated in Figure 3.

Prior to the observations of Ashkanasy and Weeks (1975), Potter (1958) noted the effect of mixture random variables in the statistical distribution of floods. He used the standard mixture distribution for the case of two components, i.e.,

$$F(X) = p_1F_1(X) + p_2F_2(X) \quad \ldots\ldots\ldots\ldots(8)$$

where $F_i(X)$, i = 1, 2 are the distribution functions of the 1st and 2nd components respectively, $p_i \geq 0$, i = 1, 2 and $p_1+p_2 = 1$. Estimation for mixtures is very difficult. Note that p_1 and p_2 must be estimated in addition to all of the parameters of both $F_1(X)$ and $F_2(X)$. Additional work in this area has been done by Hawkins (1971), (1972) which documents some of the problems associated with mixed distributions.

Without some theoretical guidance as to the choice of distributions for $F_1(X)$ and $F_2(X)$, it is an impossible task to select the best fitting forms. The mixture distributions contain so many parameters that they can fit almost any data set no matter what is used for $F_1(X)$ and $F_2(X)$. If the important tail characteristics of the distributions were not different it would matter little what choice is made. Potter (1958) chose to use extreme value forms in his analysis of such data. This seems a good choice relative to the tail characteristics since the data is observed extremes. However, it should be noted at this point that although the random variables governing stream flow may be mixtures, it does not follow that the flood (extreme event) should also be a mixture.

In fact, the classical extreme value theory suggests it should be one of the three forms given previously. However, it can be shown that for the case of mixtures, extremely large sample sizes are required for an adequate approximation of the distribution of the maximum of a sequence of mixtures.

Work by Canfield and Borgman (1975) on the distribution of the extreme in a sequence of mixture random variables in the context of reliability theory has provided a much more adequate approximating distribution. The results have direct application to the problem of choosing a distribution of maximum yearly river flow in hydrology. The results have merit because they provide a theoretical foundation which gives primary consideration to the shape of the right tails of the distributions involved. The form of the distribution of the extreme in a sequence of mixture random variables has been shown to be (Canfield and Borgman, 1975)

$$F(x) = \Phi_i(x)^{p_1} \, \Phi_{i'}(x)^{p_2} \quad \text{..........(9)}$$

where the components $\Phi_i(x)$ and $\Phi_{i'}(x)$ are extreme value forms (3), (4), or (5). Note that the parameters p_1 and p_2 can be absorbed by reparameterization so that (9) can be written

$$F(x) = \Phi_i(x)\Phi_{i'}(x) \quad \text{..........(10)}$$

thereby reducing the number of parameters in the distribution. Since it is theoretically motivated, it seems that if extreme value theory applies to floods, a distribution of this form should have the correct tail characteristics. Note that the tail shape in (8) is a weighted average of the tails of $F_1(X)$ and $F_2(X)$, whereas the shape of (10) is a produce of the tails of $\Phi_i(x)$ and $\Phi_{i'}(x)$. Even if extreme value distributions are used in (8), the tail shape is not necessarily correct.

As discussed by Bobee and Robitaille (1977) physical limitations of meteorological phenomena and basic characteristics which control river flow seems to indicate that flows are bounded above. Thus it seems that the most logical distribution for the statistical description of flood peaks is $F_3(x)$. Figure 4 verifies this choice for the Kymijoki River in Finland. It is very evident from a glance that the data are linear in this case. In less obvious cases, standard analysis techniques can be used to test for the existence of higher order polynomial effects.

In order to interpret the plot for $F_3(x)$ it is useful to examine the shape of this plot if the data were to originate from a Pearson or log Pearson Type III distribution. Relative to these distributions, if floods are bounded above the general shape of $\ln(b - X_{(i)})$ plotted against $\ln(-\ln E_i)$ is a curve, concave as viewed from the left. If floods are bounded below, the plot will appear as a curve convex as viewed from the left. Note that for this plot an upper bound is estimated as if the distribution were $F_3(x)$ even though it is not.

It is interesting to note that in the work of Bobee and Robitaille (1977), both the Pearson Type III and log Pearson Type III distributions introduce an apparent inconsistency. In some cases an upper bound for annual floods is appropriate and in others a lower bound is used. The Pearson and log Pearson distributions are not even consistent for a given data set. In some cases the Pearson distribution calls for an upper bound while the log Pearson calls for a lower bound. It seems that if an upper bound is valid due to meteorological and geographical limitations, it would be valid for all systems. The switch in boundedness is due to the inability of the Pearson and log Pearson Type III distribution to accommodate both positive and negative skewness for a given bound (upper or lower).

Estimation of Parameters

Although the existence of a limiting flood is easily justified, it is difficult to determine from geographical considerations. It was found that percentile estimates were very insensitive to the actual choice of b as long as it is relatively large. Therefore, ordinary maximum likelihood estimates of all of the parameters were used.

The distribution $F_3(x)$ is a transformed Weibull, i.e., if the $F_3(x)$ is transformed by $y = -x$ the distribution of y is Weibull with the same parameters as $F_3(x)$ (b is negative). Therefore, a program available for maximum likelihood (ML) estimation of Weibull parameters (Harter and Moore 1965) was used.

Some difficulties were experienced in applying ML methods. In general, the computer program was expensive to run and, in addition, required several passes to find acceptable scale factors and initial values. The resulting estimates were highly dependent on these values even when the convergence criterion for the computation was met. In some cases, a better fit was obtained using a less stringent convergence measure. These problems motivated additional research not directly connected with this project.

Goodness-of-fit Comparisons

The result of fitting $F_3(x)$ to the same data used by Bobee and Robitaille (1977) (Table 1) to evaluate the Pearson and log Pearson Type III distributions is given in this section. Maximum likelihood estimation with the accompanying difficulties described previously was used. The same goodness-of-fit statistics used by Bobee and Robitaille (1977) are used herein. Since classical tests of goodness-of-fit (Chi-square and Kolmogorov-Smirnov) are not powerful enough to discriminate between distribution functions or parameter estimation methods; Bobee and Robitaille used another procedure for purposes of comparison which has the same origin as the one used by the Work Group on Flow Frequency Methods (Benson, 1968). These statistics are essentially the average absolute deviation and the average quadratic deviation expressed as a percent between the predicted flow over selected recurrence intervals and the observed flow. The recurrence intervals or return periods are T = 2, 5, 10, 20, 50, 100 years (probability of being equaled or exceeded of 0.50, 0.20, 0.10, 0.05, 0.02, and 0.01).

The predicted flood discharges (value estimated from the fitted distribution), Q(T), for these return periods are calculated using program FLOOD. (See Table 4.)

The observed flood data values (the empirical for recurrence interval T), D(T) (Table 5), are obtained from the sample I, ranked in decreasing order, by using a formula of plotting position and by interpolating between the specified probability (or the selected recurrence interval). Linear interpolations are done graphically using normal probability paper. Three formulae of expected probabilities are used to obtain the data values given in Table 5:

Hazen $Pm = \frac{m-0.5}{N}$ ....(11)

Chegodayev $Pm = \frac{m-0.3}{N+0.4}$ ....(12)

Weibull $Pm = \frac{m}{N+1}$ ....(13)

where m is the rank of the observation in the sample of size N, varying from 1 for the lowest flow to N for the highest. (Bobee and Robitaille, 1977.)

For each data set the relative deviation in percent, q(T), is computed between Q(T) and D(T) corresponding to each return period T.

$$q(T) = \frac{Q(T) - D(T)}{D(T)} * 100 \quad ....(14)$$

To evaluate the fit for the data set, the following quantities are computed:

$$A = \frac{1}{L} \sum_T |q(T)|$$

...........(15)

$$B = \frac{1}{L} \sum_T q^2(T)$$

where "A" represents the average of the absolute values of the relative deviations over the "L" selected recurrence intervals and "B" represents the quadratic deviation averaged over the "L" selected recurrence intervals. The goodness-of-fit values for the log Pearson Type III distribution and for the distribution and method of fitting judged best by Bobee and Robitaille (1977) (Pearson Type III) are also tabulated in Table 2 and 3 for comparative purposes.

A Graphical Technique

It is impossible to interpret the information on Table 2 and 3 without viewing plots of these data sets. The plots are shown in Figures 5-18.

Given N years of maximum yearly river flows, the observations are ordered low to high, producing the order statistics $X_{(i)}$, i = 1, 2, ...N. This is done using the subroutine ORDER of program PLOT.

From a previous discussion we know

$$E[F(X_{(i)})] = \frac{i}{N+1} \quad \text{..........(16)}$$

Let

$$Z_i = \ln(b-X_{(i)}) \quad \text{..........(17)}$$

$$Y_i = \ln(-\ln(E[F(X_{(i)})])$$

$$= \ln(-\ln(\frac{i}{N+1})) \quad \text{........(18)}$$

Then

$$Y_i = \ln(-\ln(F(X_{(i)})))$$

$$= \ln(-\ln(\exp[-(\frac{b-X_{(i)}}{c})^a]))$$

$$= a \ln(b-X_{(i)}) - a \ln(c)$$

$$= aZ_i - a \ln(c)$$

This implies that

$$Y_i = \ln(-\ln(\frac{i}{N+1}))$$

$$= aZ_i - a \ln(c) \quad ....(19)$$

Therefore, Y_i is a linear function of Z_i where "a" is the slope and "-a ln(c)" is the intercept. Hence, by plotting Z_i against Y_i on a graph, the largest maximum yearly floods should form a straight line whose slope approximates "a" with an intercept of approximately "-a ln(c)".

The Z_i and Y_i of equation (19) are plotted for each of the fourteen stations in this study (see Figures 5-18) using the command PLOT from the MINITAB II Reference Manual (1978). The Z_i are along the C1 axis while the Y_i are along the C2 axis in each plot.

As can be observed from the figures, the data for Mali River (Figure 5) manifest almost a straight line indicating a highly linear relationship. Therefore the extreme value distribution should provide the best fit compared with the Pearson Type III and log Pearson Type III distributions. This is evidenced by the deviations tabulated in Table 2 and 3. Date from Gloma River (Figure 10) also show a linear relationship, and the goodness-of-fit also demonstrates the $F_3(X)$ to be the best one.

The linear relationship between ln(b-X) and ln(-ln(E_i)) is also found for the Amos River (Figure 17). A good fit was shown for this case though not all the deviations from extreme value distribution are smaller than that from the Pearson Type III and log Pearson Type III distribution, it is apparent that data of the above mentioned rivers are homogeneous.

For the plots of Danube River (Figure 9) and Fraser River (Figure 12), relationships of roughly linearity are observed. However, for both plots, there is a data point far apart from the others, located at the lower left corner of the graph. This may be an indication of non-homogeneous sources. Since there is only one observation, it is unmature to advance a more conclusive argument. The deviations of the two rivers, nonetheless are not too much far off comparing to the deviations from the Pearson Type III and log Pearson Type III distributions. But, according to the plots, the data do not show a curved relationship and therefore Pearson Type III distribution cannot be the correct distribution. In other words, the data seem to be nonhomogeneous and none of the distribution (F_3(X), Pearson Type III, log Pearson Type III) are appropriate.

The rest of the rivers are found to have a poor fit by the extreme value distribution. The plot for Susquehanna River (Figure 6) reveals a pattern of two straight lines with the breaking point approximately at the position (10.25, -3.0). The plot for Krishna River (Figure 7) manifests a "S" shape. This may be a result of at least three sources affecting the data. The graph for Elbe River

(Figure 8) shows (though not quite apparent) two straight lines. For the Neman River (Figure 11) and Slave Falls River (Figure 18), both the graphs show a slightly "S" shape while the plot of Headingly River (Figure 13) shows a rather clear "S" shape. Finally the plot of Prince Albert River (Figure 16) shows a clear curvilinear relationship and the plots of Medicine Hat River (Figure 14) and Saskatoon River (Figure 15) indicate relationships with several breaking points. For the last three rivers, the deviations from extreme value distribution are much larger than that from the two alternative distributions. For all the rivers discussed in this paragraph, it is clear that the data are from nonhomogeneous sources. Moreover, although Pearson Type III and log Pearson Type III distributions provide deviations of relatively small magnitude comparing to $F_3(X)$ to these rivers, it does not imply that the distributions are appropriate. In other words, none of the distributions considered adequately describe the data. Analysis and estimation for nonhomogeneous sources have been considered by Olson (1979).

These plots underscore their importance in fitting data. Whenever several distributions are fit to a given data, one will always have a "best" fit. However, none of these tried may be appropriate. The plots very easily point this out.

Table 2. Mean of the Absolute Relative Deviations.

Station	Pearson Type III			log Pearson Type III			$F_3(x)$		
	H[a]	C[a]	W[a]	H	C	W	H	C	W
bB24	1.4	1.7	2.1	1.8	1.7	2.1	1.6	1.4	1.6
hE60	3.6	4.0	4.9	3.7	3.5	4.3	7.5	5.4	5.4
IB06	3.4	2.9	3.4	3.3	3.8	4.7	7.4	7.4	8.3
BF40	3.6	4.2	4.2	3.8	4.7	4.8	7.7	7.8	8.4
BE38	3.1	2.9	2.4	2.5	2.4	2.4	2.7	2.1	3.9
BF19	3.5	4.0	4.0	3.5	4.1	4.1	3.4	3.9	4.0
CF25	2.8	2.9	3.3	3.3	3.3	3.6	7.4	6.1	6.5
mE19	2.7	2.2	3.4	2.5	2.1	3.3	3.4	2.8	3.8
jE792	7.6	5.8	6.1	6.2	5.1	4.8	6.4	6.3	6.8
iF00	2.9	4.1	5.9	4.2	5.9	7.7	15.8	17.1	15.5
KF62	4.8	4.5	4.5	4.8	5.8	5.8	10.4	11.3	11.3
KF53	6.6	4.6	6.8	6.6	4.8	8.5	13.7	11.2	14.5
hE88a	1.4	1.8	2.8	1.7	2.5	3.5	1.8	2.3	2.5
jF50a	4.4	3.6	4.4	3.8	3.4	4.2	4.2	4.4	5.4

[a]H = Hazen Formula
C = Chegodayev Formula
W = Weibull Formula

Table 3. Mean of the Quadratic Deviations.

Station	Pearson Type III			log Pearson Type III			$F_3(x)$		
	H[a]	C[a]	W[a]	H	C	W	H	C	W
bB24	2.9	4.1	9.4	4.3	5.1	11.2	5.0	3.4	4.6
hE60	13.4	17.6	32.3	18.9	20.8	41.3	101.0	56.9	56.9
IB06	20.4	21.2	28.2	24.0	32.1	43.6	87.7	95.8	121.1
BF40	18.0	21.9	23.7	21.9	27.7	30.9	75.7	80.1	91.4
BE38	16.2	10.2	7.0	11.0	7.1	8.7	9.6	8.1	20.9
BF19	14.5	17.7	19.7	15.7	19.6	22.2	14.0	17.2	19.3
CF25	14.2	15.2	16.0	17.6	18.4	20.1	95.1	72.45	77.2
mE19	10.7	6.6	20.7	10.6	5.8	22.7	14.2	10.7	19.8
jE792	81.4	47.8	49.5	47.6	33.1	33.7	59.4	63.6	72.9
iF00	11.4	19.2	40.9	29.2	45.1	72.8	297.0	351.0	228.9
KF62	23.9	20.7	21.7	26.0	34.5	35.8	122.7	157.2	163.6
KF53	81.3	41.3	82.0	55.6	26.8	122.8	312.0	192.4	380.4
hE88a	2.6	4.5	11.5	4.4	7.6	16.5	4.2	6.9	8.2
jF50a	31.7	13.8	21.7	21.7	13.3	22.2	22.7	24.1	37.1

[a]H = Hazen Formula
C = Chegodayev Formula
W = Weibull Formula

Table 4. Computed Flood Discharges (m^3/s) for Selected Return Periods.

Station	T, year					
	2	5	10	20	50	100
bB24	4655.23635066	6272.89996590	7247.95542250	8117.35581700	9154.20721560	9870.63355980
hE60	8027.23079960	10363.71015940	11821.54383950	13156.76624720	14797.73328050	15965.91768930
IB06	13990.32594110	17615.19406840	19854.97081710	21890.88496210	24371.84126000	26123.19515320
BF40	1802.21725839	2502.53573263	2931.43685430	3318.62328482	3786.82135218	4114.81009459
BE38	1745.40536416	2233.58595037	2534.79673630	2808.28677428	3141.14507955	3375.82534808
BF19	2034.96169358	2486.24995434	2762.55134177	3011.92067710	3313.38592850	3524.51698148
CF25	2510.38990915	3439.16060698	4010.87103999	4529.01771528	5158.35140159	5601.15915400
mE19	8638.36425050	10231.50331440	11221.44454570	12125.22535810	13231.97626070	14017.05808160
jE792	261.12998649	390.79758282	466.94913456	533.49377880	611.08609254	663.50999960
iF00	1157.47069317	1744.43547696	2105.72248554	2433.14438486	2830.80467355	3110.58867341
KF62	1383.69398129	1980.61646688	2345.32852489	2673.96335685	3070.54083848	3347.78891778
KF53	1241.82383925	1765.84557378	2096.54420352	2402.11957186	2781.40712678	3054.08241218
hE88a	183.52505580	228.39903791	255.81949763	380.52965035	310.35129306	331.20164779
jF50a	1448.55935091	1931.61701706	2220.09326349	2475.46596313	2777.58237005	2984.66871659

Table 5. Data Values D(T) (m^3/s) as Interpolated Between Adjacent Observations

Station	Hazen						Chegodayev					Weibull				
	T=2	T=5	T=10	T=20	T=50	T=100	T=2	T=5	T=10	T=20	T=50	T=2	T=5	T=10	T=20	T=50
bB24	4650	6378	7190	7757	9145	9655	4650	6391	7208	7847	9265	4650	6410	7235	8004	9538
hE60	7730	10715	11790	13705	17530	19460	7730	10720	11790	13750	17920	7734	10724	11798	13817	18870
IB06	13555	17520	20379	26307	27410	29094	13555	17557	20559	26492	27841	13555	17614	20845	26809	28796
BF40	1630	2596	2964	3771	4215	4522	1630	2600	2970	3772	4271	1630	2605	2986	3775	4374
BE38	1780	2267	2660	2819	3045	3531	1780	2271	2689	2837	3149	1780	2278	2698	2867	3395
BF19	2119	2368	2700	3170	3437	3502	2119	2370	2722	3180	3451	2119	2373	2751	3197	3485
CF25	2500	3400	4200	4873	6200	6568	2500	3400	4215	4892	6200	2500	3400	4238	4954	6200
mE19	8580	9960	10800	11460	12900	14700	8580	9960	10800	11510	13400	8580	9970	10800	11600	14300
jE792	248	392	540	581	598	611	248	397	545	586	602	248	404	550	592	609
iF00	987	1839	2333	2891	3764	4003	987	1840	2367	2962	3828	987	1847	2413	3082	3967
KF62	1260	1932	2669	3140	3447	3810	1260	1947	2681	3140	3726	1260	1970	2700	3140	3750
KF53	1180	1634	2129	2855	3286	4767	1180	1636	2141	2883	3673	1180	1640	2160	2930	4523
hE88a	179	230	262	279	321	335	179	230	262	284	325	179	230	262	291	334
jF50a	1390	2004	2397	2559	2788	2807*	1390	2010	2400	2621	2795	1390	2019	2405	2733	2808*

*Obtained by extrapolation.

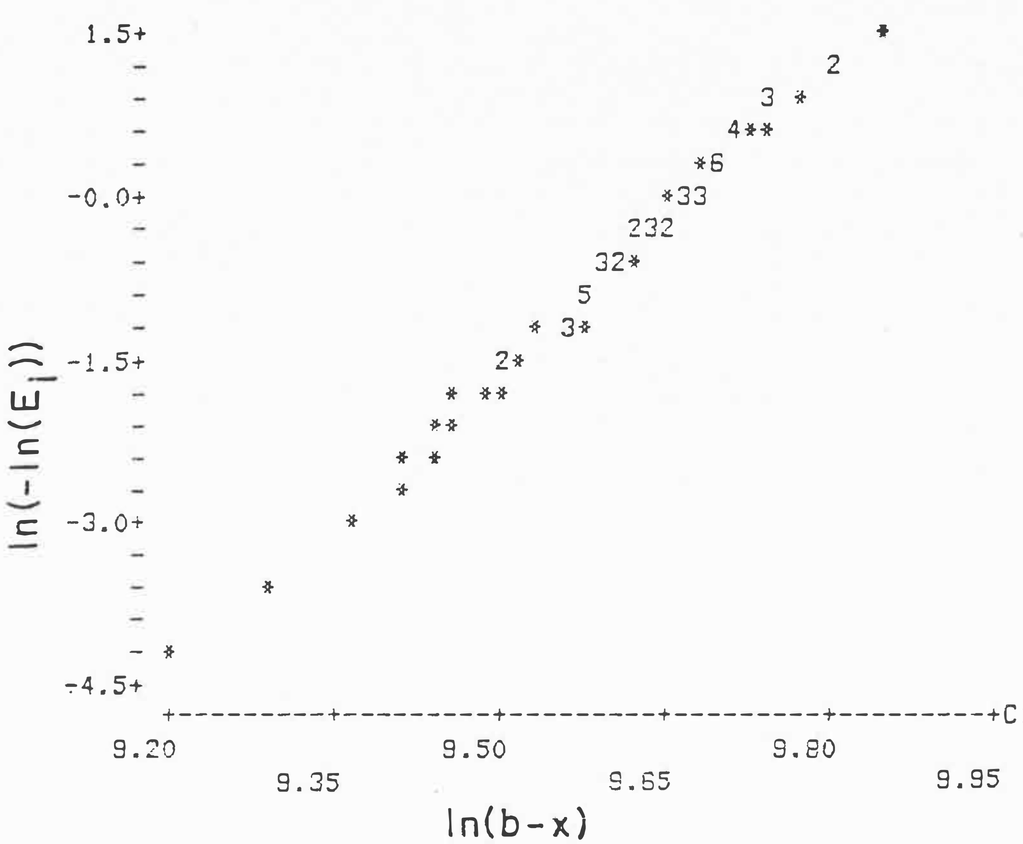

Figure 5. Station bB24--Mali River.

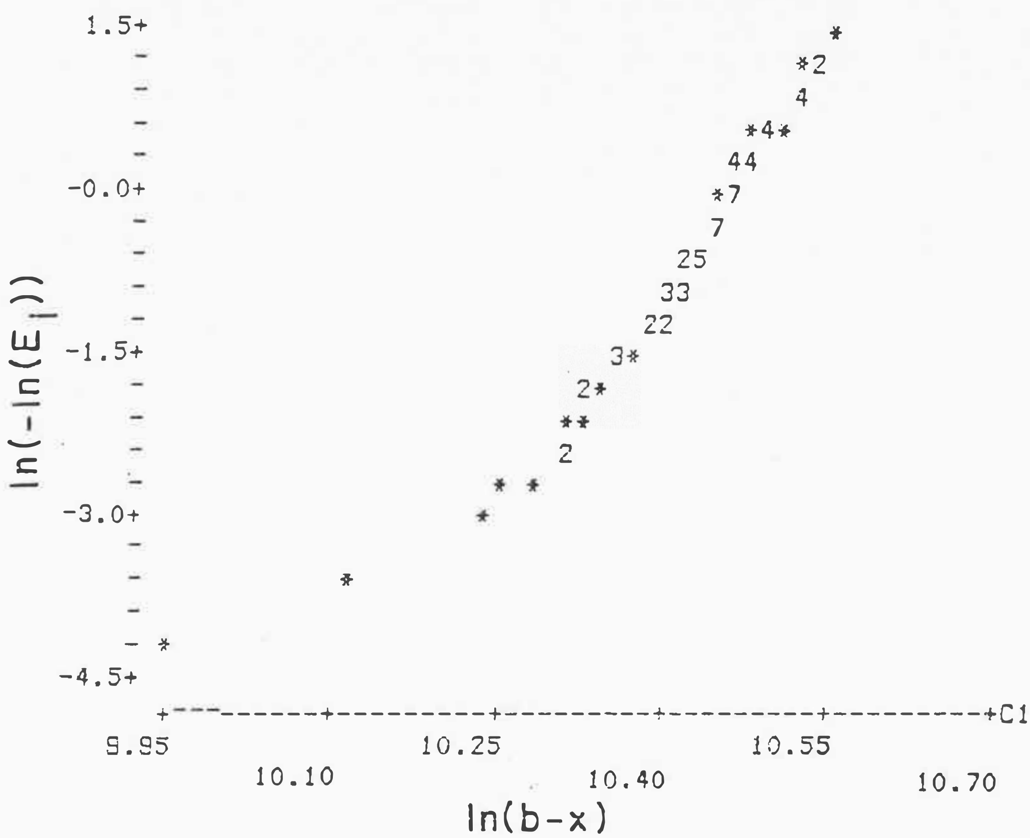

Figure 6. Station HE60--Susquehanna River.

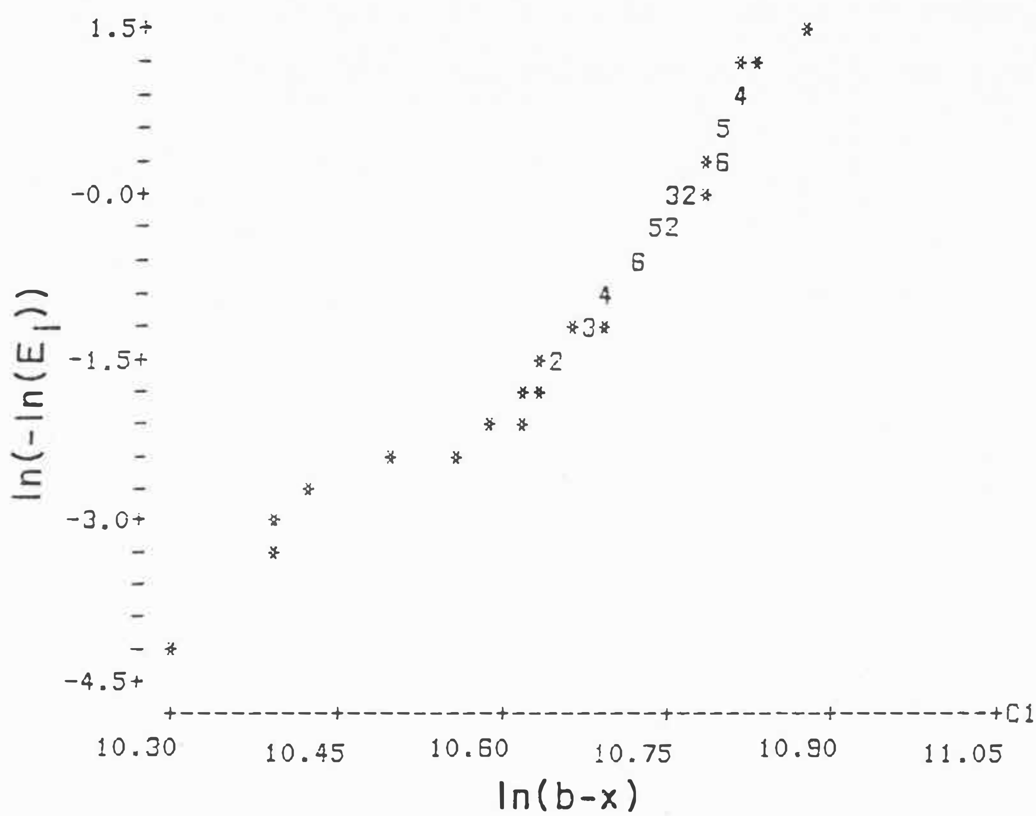

Figure 7. Station IB06--Krishna River.

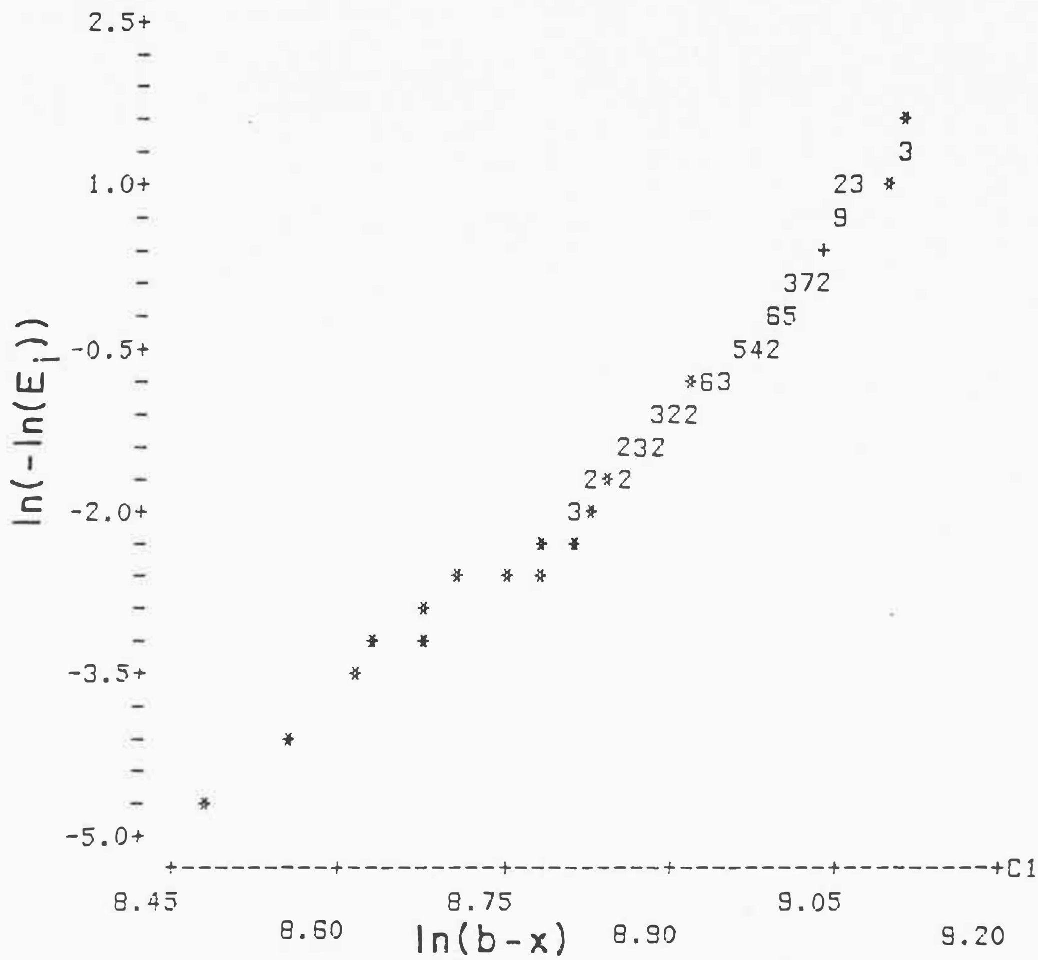

Figure 8. Station BF40--Elbe River.

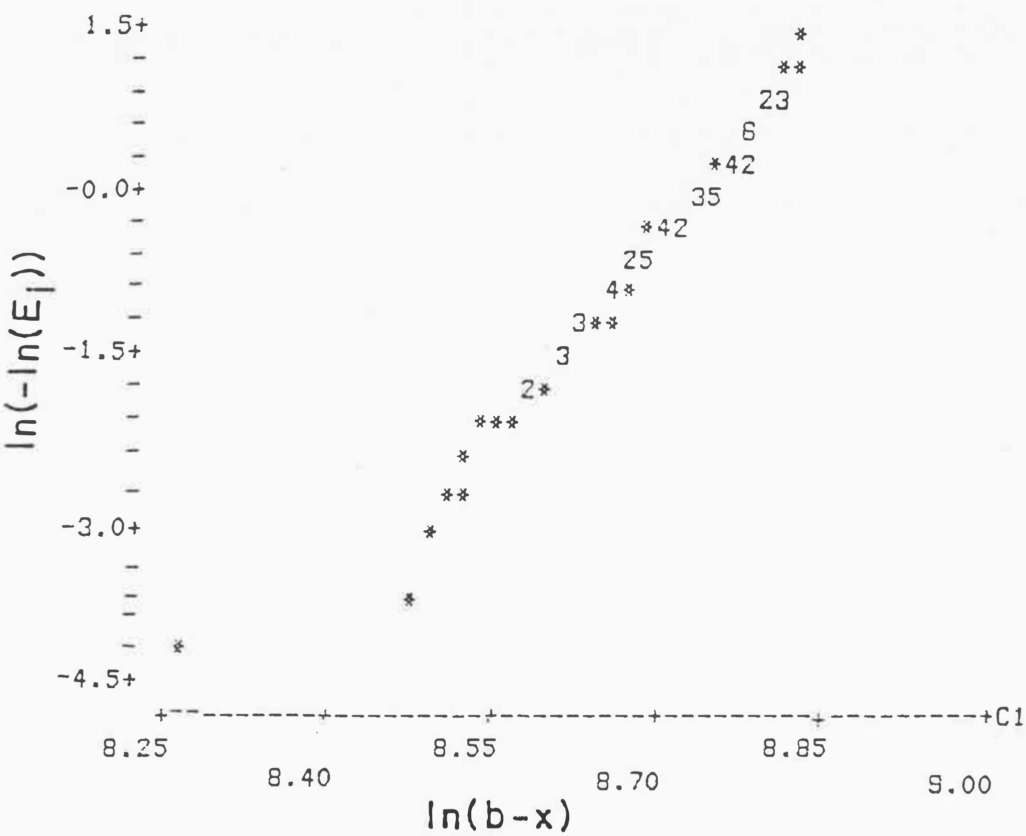

Figure 9. Station BE38--Danube River.

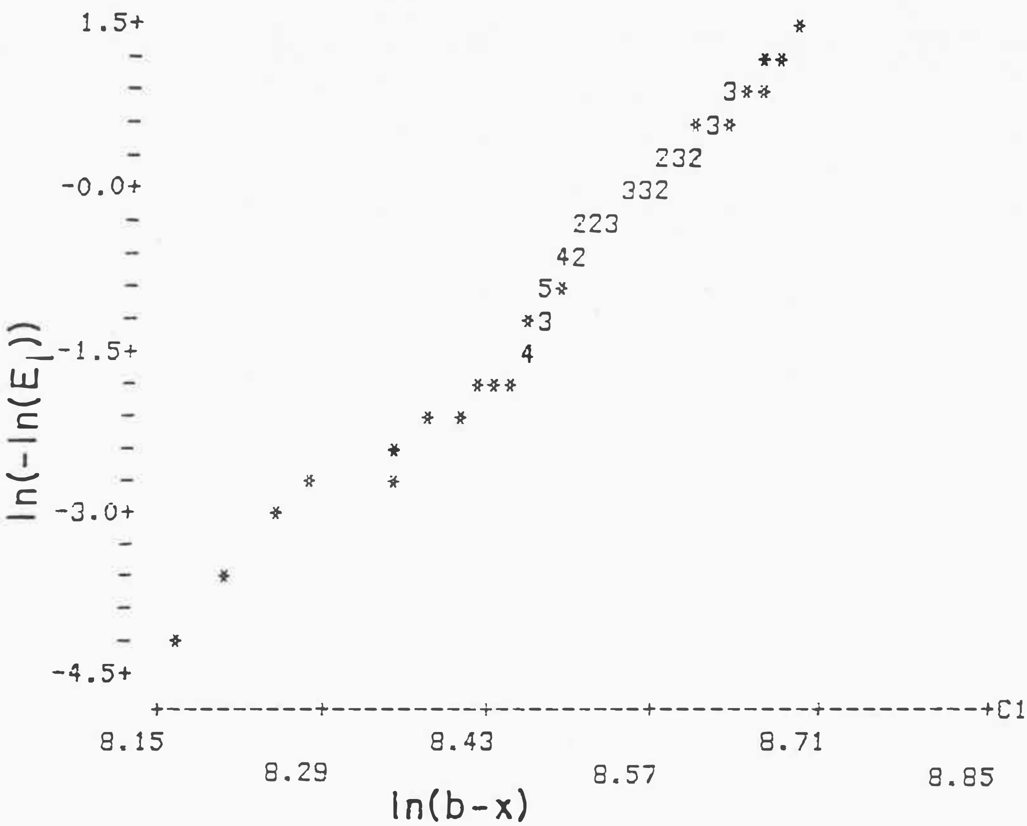

Figure 10. Station BF19--Gloma River.

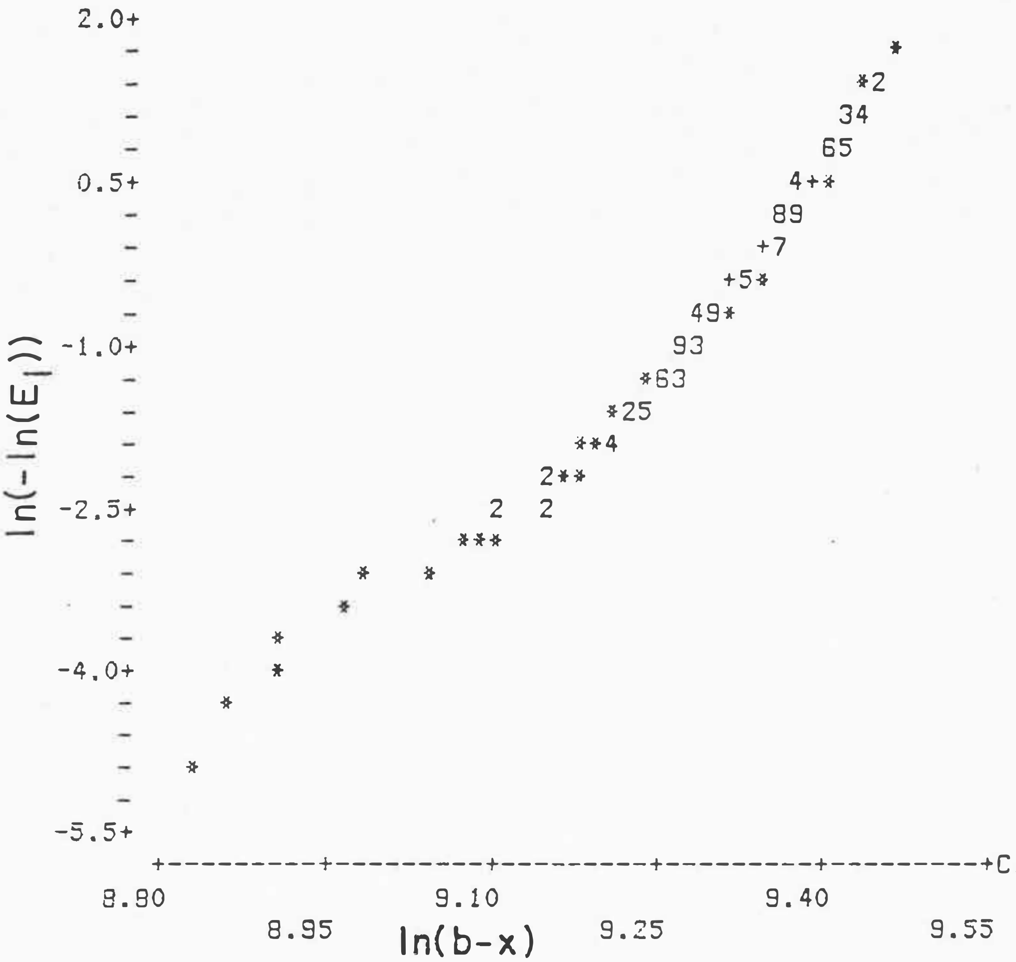

Figure 11. Station CF25--Neman River.

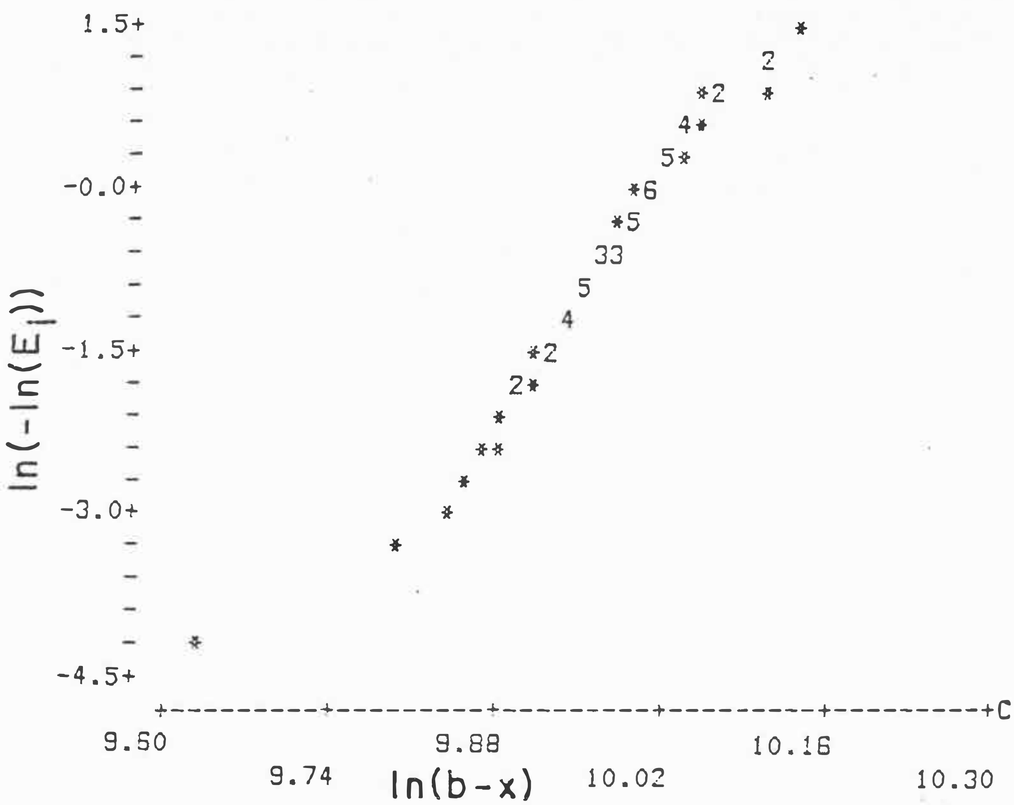

Figure 12. Station ME19--Fraser River.

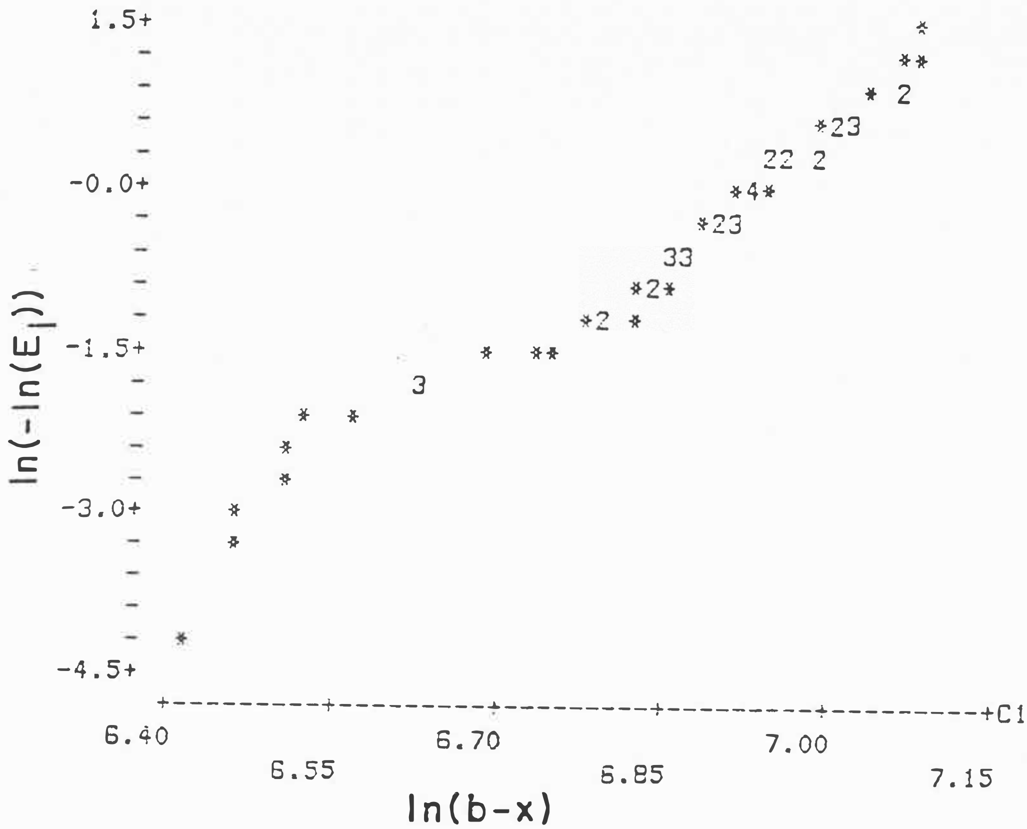

Figure 13. Station JE792--Headingly River.

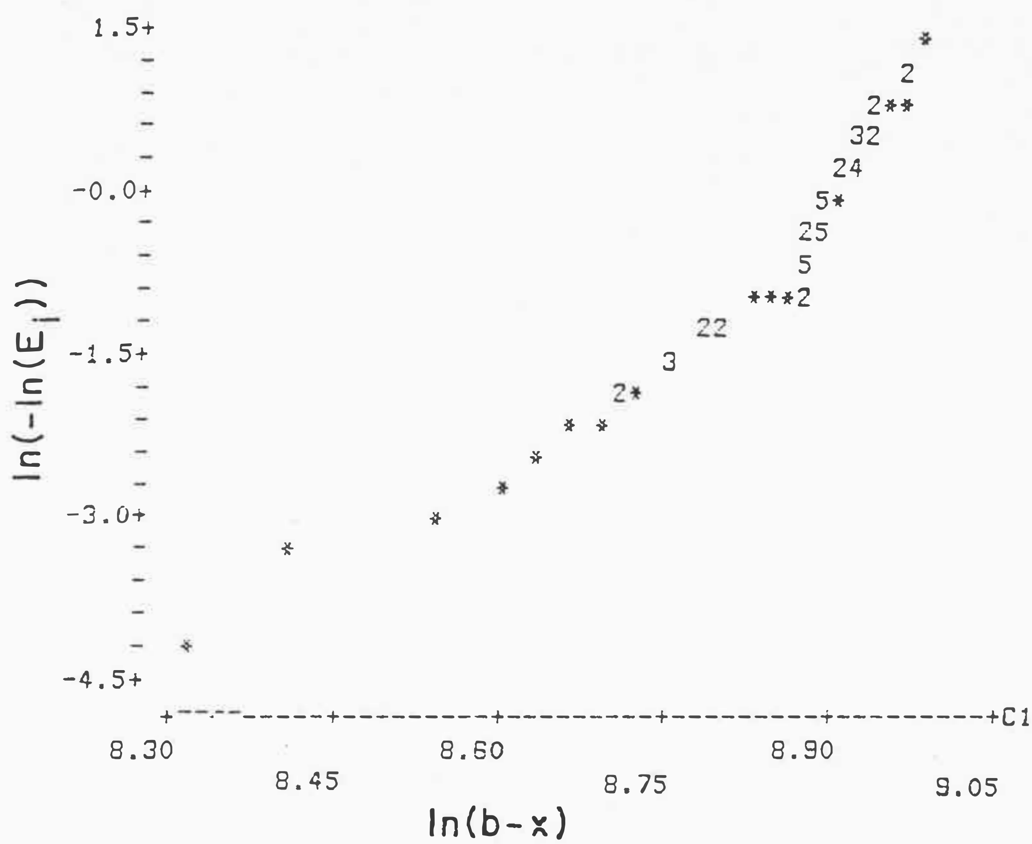

Figure 14. Station IF00--Medicine Hat River.

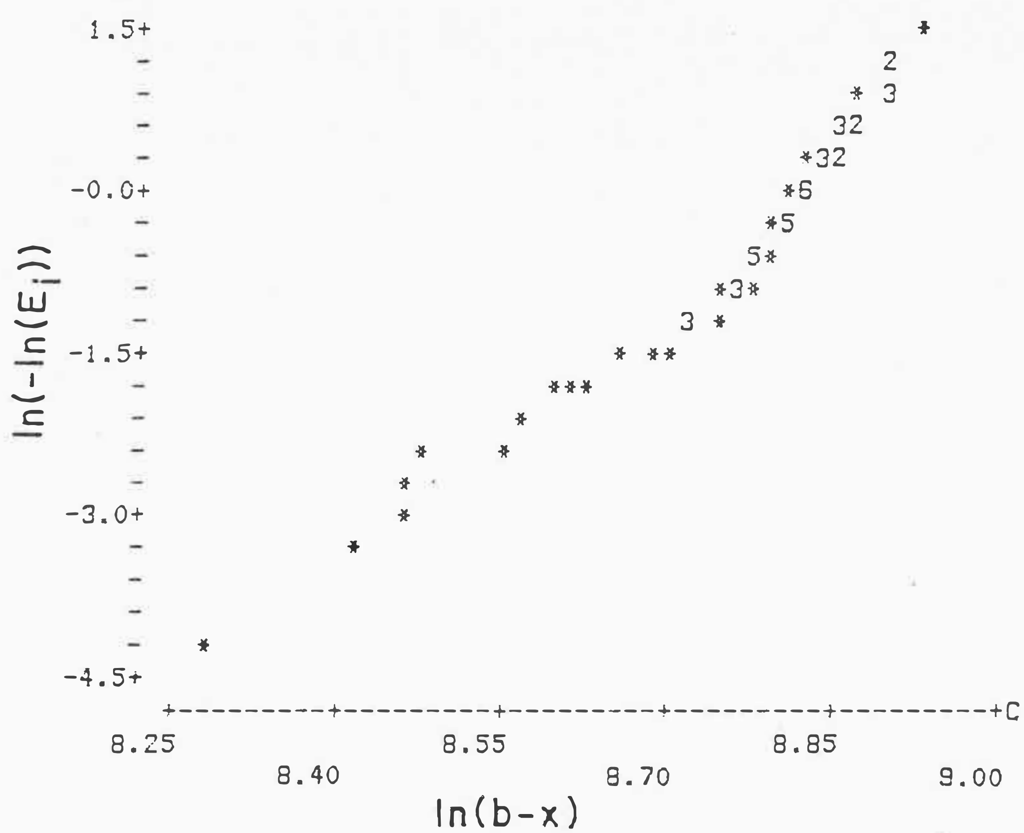

Figure 15. Station KF62--Saskatoon River.

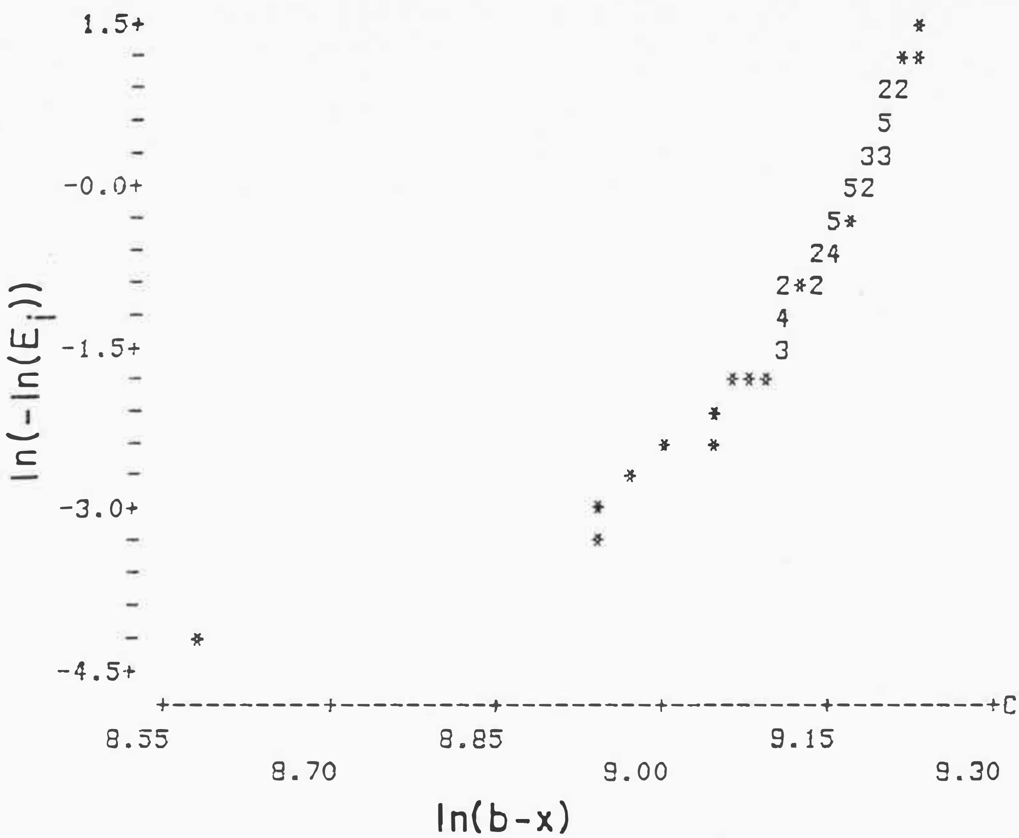

Figure 16. Station DF53--Prince Albert River.

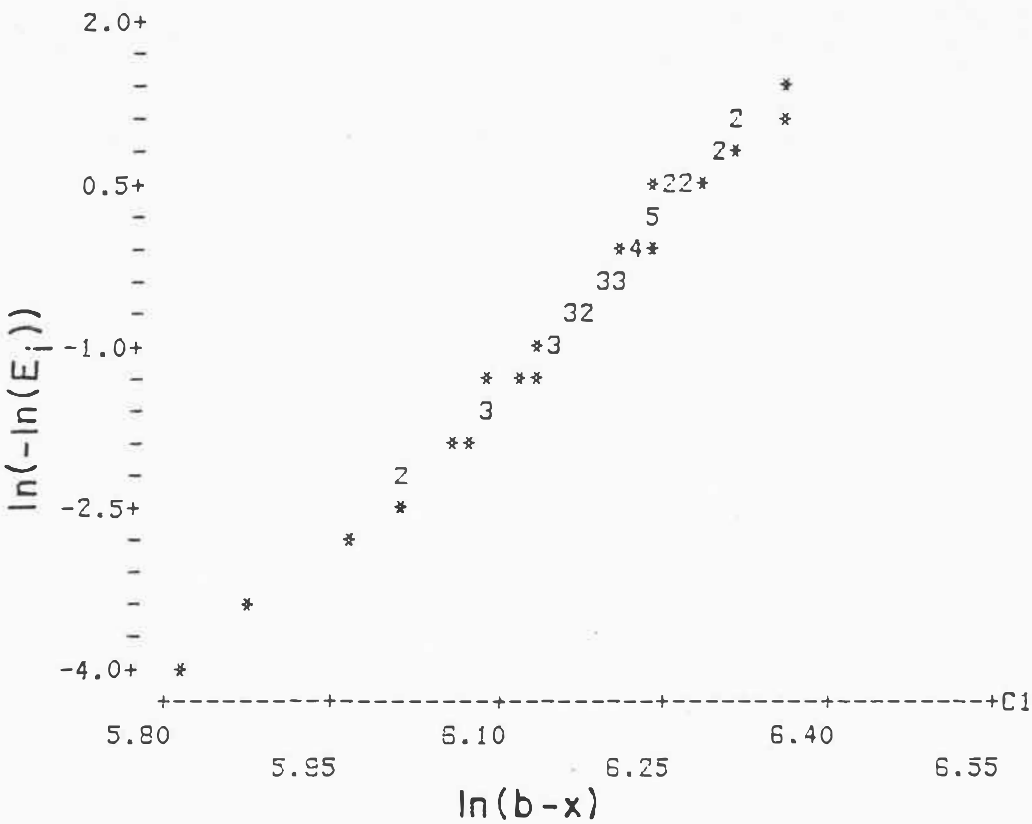

Figure 17. Station hE88a--Amos River.

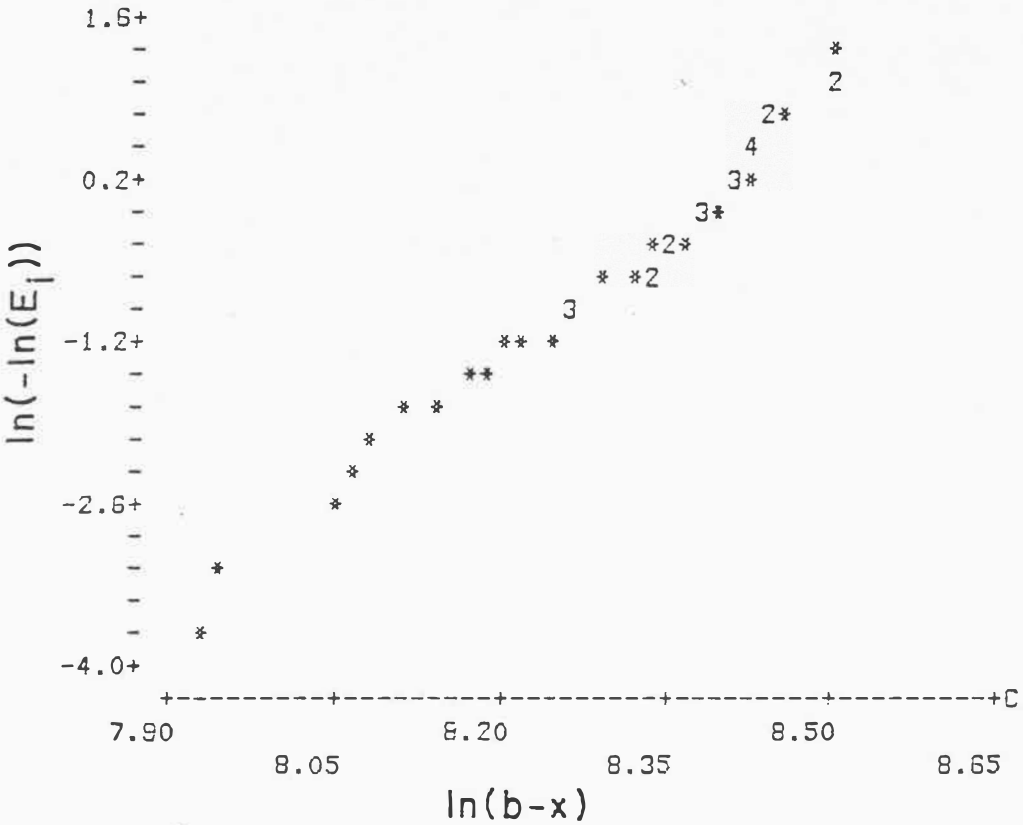

Figure 18. Station jK50a--Slave Falls River.

CHAPTER III

CONCLUSION

There is extensive literature describing distribution function which provide the "best" fit for the random variable "maximum yearly river flow" to rivers which exhibit a single homogeneous source of runoff. But in estimating n-year return periods, it is often necessary to extrapolate. Some theoretical guideline should be used when working beyond the range of the data to ensure the proper right tail characteristics of the estimated distribution function. In this research, extreme value theory has been applied to the estimation of the flood frequency.

The following steps are offered as guidelines for flood frequency analysis based on extreme value theory as presented in this research.

1. Select a value b in the order of two or three times the magnitude of the largest flood of record and plot the date in the form of Figure 3.

2. If the plot in Step One is linear, estimate parameters a, b, and c (Equation 5) and apply the results for estimating flood frequency.

3. If the plot in Step One is curved, some other distributions are probably more applicable; and alternatives should be considered.

4. If the plot in Step One is two straight lines, it means the data value are from nonhomogeneous sources (more than one source).

Finally, two major points can be concluded from the results of the study. First, all of the data sets in this study do not belong to Pearson Type III or log Pearson Type III distribution. Even though these two distributions provide deviations of smaller magnitude, it does not imply that they appropriate for the data. This observation is easily confirmed by plotting the data. Straight line plots as described in Chapter II indicate our extreme value form with homogeneous sources. A broken line plot indicates an extreme value form with nonhomogeneous sources. Plots other than those considered could be Pearson Type III or log Pearson Type III distributions if they are either concave or convex but not "S" shaped. Very few of the data sets observed could possibly be from a Pearson Type III or log Pearson Type III distribution. Secondly, a three parameter extreme value distribution is preferable to the two alternative distributions, i.e., Pearson Type III and log Pearson Type III, if the data are homogeneous. For the nonhomogeneous data, the three-parameters model is not so useful. However, a study (Olson, 1979) indicates that an extreme value distribution for nonhomogeneous sources provided excellent food-of-fit for this type of data.

CHAPTER IV

FUTURE STUDIES

Some difficulties were experienced in applying Maximum-Likelihood methods. In general, the computer program was expensive to run and, in addition, required several passes to find acceptable scale factors and initial values. The resulting estimates were highly dependent on these values even when the convergence criterion for the computation was met. In some cases, a better fit was obtained using a less stringent convergence measure. These problems motivated additional research which will result in a computationally more efficient method of estimation developed for all extreme value distributions. This method of estimation should not depend upon sensitive convergence criteria.

LITERATURE CITED

Ashkanasy, N. M. and Weeks, W. D. (1975): "Flood Frequency Distribution in a Catchment Subject to Two Storm Rainfall Producing Mechanisms", Hydrology Symposium, Armidale, NSW, Australia, May 18-21, Institute of Eng., Sydney, Australia, pp. 153-157.

Beard, L. R. (1974): "Flood Flow Frequency Techniques", Report No. CR WR-119 to the Office of Water Research and Technology, Department of the Interior, Washington, D.C., pp. 224.

Benson, M. A. (1968): "Uniform Flood-Frequency Estimating Methods for Federal Agencies", Water Resour., Vol. 4, No. 5, pp. 891-908.

Bobee, B. B. and Robitaille, R. (1977): "The Use of the Pearson Type 3 and Log Pearson Type 3 Distributions Revisited", Water Resources Research, Vol. 13, No. 2, pp. 427-443.

Canfield, R. V. and Borgman, L. E. (1975): "Some Distributions of Time to Failure for Reliability Applications", Technometrics, Vol. 17, No. 2, pp. 263-268.

Gnedenko, B. V. (1943): "Sur la Distribution Limite du Terms Maximum d'une Serie Aleatorie", Annals of Mathematics, Vol. 44, pp. 423-453.

Gumbel, E. J. (1958): Statistics of Extremes, Columbia University Press, New York, p. 375.

Hager, H. W., Bain, L. J. and Antle, C. E. (1971): "Reliability Estimation for the Generalized Gamma Distribution and Robustness of the Weibull Model", Technometrics, Vol. 13, No. 3, pp. 547-558.

Harter, H. L. and Moore, A. H. (1965): "Maximum Likelihood Estimation of Gamma and Weibull Populations from Complete and Censored Samples", Technometrics, Vol. 7, No. 6, pp. 639-643.

Hawkins, R. H. (1972): "A Note on Multiple Solutions to the Mixed Distribution Problem", Technometrics, Vol. 14, No. 4, pp. 973-976.

Hawkins, R. H. (1974): "A Note on Mixed Distributions in Hydrology", Proceeding of Symposium on Statistical Hydrology, USDA., Ag. Res. Service, Misc. Pub. No. 1275, USGPO, Washington, D.C., pp. 336-344.

Lindgren, B. W. (1976): Statistical Theory, Macmillan Publishing Co., Inc., New York.

Mann, N., Schafer, R. and Singpurwalla, N. (1974): Methods for Statistical Analysis of Reliability and Life Data, John Wiley and Sons, New York, p. 168.

Olson, D. R. (1979): "Estimation of Floods When Runoff Originates from Nonhomogeneous Sources", M.S. Thesis, Department of Applied Statistics and Computer Science, Utah State University, Logan, Utah, 84322.

Potter, W. D. (1958): "Upper and Lower Frequency Curves for Peak Rates of Runoff", Transactions of the American Geophysics Union, Vol. 39, pp. 100-105.

Ryan, T. A., Joiner, B. L., and Ryan, B. F. (1978): Minitab II Reference Manual, Statistics Department, Pennsylvania State University.

U.S. Water Resources Council. (1976): Guidelines for Determining Flood Flow Frequency, Bulletin No. 17 of the Hydrology Committee, Washington, D.C.

Watson, G. S. (1952): "Extreme Value Theory for m Dependent Stationary Stochastic Processes", Annals of Mathematical Statistics, Vol. 25, pp. 798-803.

White, J. S. (1969): "The Moments of Log-Weibull Order Statistics", Technometrics, Vol. 11, No. 2, pp. 373-385.

APPENDICES

Appendix A

Data Used in Analysis

STATION		COUNTRY		RIVER		LOCATION	
bB24		SENEGAL		SENEGAL		BAKEL	
1040	1740	1880	2290	2750	2850	2850	2890
3140	3290	3320	3400	3480	3550	3560	3560
3600	3600	3600	3760	3770	3840	3840	4180
4200	4200	4300	4350	4400	4460	4620	4620
4680	4790	4850	4970	5070	5260	5330	5330
5430	5450	5450	5450	5450	5450	5590	5590
5620	6030	6310	6410	6430	6570	6640	7000
7030	7180	7300	7600	7630	8170	9070	9940

STATION		COUNTRY		RIVER		LOCATION	
hE60		U.S.A		SUSQUEHANNA		HARRIABURA, PA.	
3850	4330	4390	5010	5012	5040	5100	5150
5250	6000	6060	6116	6230	6460	6500	6513
6540	6650	6850	6853	6910	6940	6990	7050
7050	7051	7079	7140	7150	7390	7500	7500
7620	7646	7650	7820	7870	7957	8100	8160
8210	8330	8410	8410	8440	8670	8720	8920
9160	9170	9170	9175	9400	9571	10100	10700
10730	10817	11100	11400	11600	11700	11780	11800
12000	12700	13705	14000	17400	21000		

STATION		COUNTRY		RIVER		LOCATION	
I806		INDIA		KRISHNA		VIJAYAWADA	
7190	9058	9915	10017	10204	10212	10360	10458
10478	10495	10613	10793	10813	10878	10882	10916
11105	11122	11374	11500	12091	12393	12560	12912
12979	13069	13113	13260	13465	13528	13582	13686
14033	14132	14220	14242	14503	14520	15396	15514
15647	15816	15872	16009	16380	16524	16782	17372
17680	17908	17970	18511	18888	19879	20970	23501
25902	26873	27073	29768				

STATION		COUNTRY		RIVER		LOCATION	
BF40		CZECHOSLOVAKIA		ELBE		DECIN	
543	587	595	610	726	1038	1046	1058
1112	1117	1138	1138	1149	1160	1166	1172
1175	1181	1181	1198	1205	1207	1234	1246
1265	1265	1269	1270	1282	1293	1300	1312
1317	1350	1354	1360	1372	1396	1429	1454
1462	1474	1492	1498	1522	1527	1546	1561
1565	1565	1575	1601	1610	1618	1643	1702
1717	1742	1768	1845	1848	1853	1874	1915
1930	1930	1940	2038	2040	2040	2083	2109
2124	2146	2158	2250	2284	2301	2373	2379
2385	2400	2410	2515	2540	2565	2600	2626
2643	2666	2725	2815	2850	2876	2937	2937
2940	2975	3100	3172	3343	3600	3770	3779
4058	4143	4450	4822				

STATION		COUNTRY		RIVER		LOCATION	
BE38		GERMANY		DANUBE		HOFKIRCHEN	
947	956	1090	1090	1100	1120	1230	1230
1250	1250	1260	1260	1310	1310	1320	1320
1340	1350	1380	1400	1440	1450	1450	1460
1460	1480	1540	1580	1600	1640	1650	1720
1730	1760	1800	1810	1810	1850	1850	1880
1890	1900	1920	1930	1980	2020	2030	2040
2050	2070	2150	2170	2180	2240	2270	2310
2390	2400	2450	2540	2600	2690	2780	2780
2810	2930	3000	3880				

STATION		COUNTRY		RIVER		LOCATION	
BF19		NORWAY		GLOMA		LANGNES	
1157	1267	1351	1358	1413	1504	1504	1518
1533	1557	1568	1580	1643	1650	1675	1707
1734	1738	1770	1783	1817	1822	1839	1872
1878	1910	1916	1953	2031	2050	2050	2100
2106	2133	2168	2172	2180	2195	2232	2240
2255	2256	2258	2260	2288	2299	2302	2311
2312	2321	2346	2359	2363	2380	2385	2390
2515	2582	2585	2715	2850	2877	3160	3224
3429	3543						

STATION	COUNTRY	RIVER	LOCATION
CF25	USSR	NEMAN	SMALININKAI

810	870	980	1050	1100	1150	1150	1200
1240	1250	1300	1350	1400	1400	1400	1400
1450	1500	1550	1550	1600	1600	1600	1650
1650	1700	1700	1700	1700	1700	1750	1750
1750	1800	1800	1800	1800	1850	1850	1900
1900	1950	1950	1950	1950	1950	2000	2000
2000	2000	2100	2100	2100	2100	2100	2100
2100	2100	2100	2100	2100	2200	2200	2200
2300	2300	2300	2300	2300	2300	2300	2300
2400	2400	2400	2400	2400	2500	2500	2500
2500	2500	2600	2600	2600	2600	2600	2600
2700	2700	2700	2700	2700	2700	2700	2700
2700	2800	2800	2800	2800	2900	2900	2900
3000	3000	3000	3000	3000	3000	3000	3000
3100	3100	3100	3100	3200	3200	3200	3200
3200	3200	3300	3400	3400	3400	3400	3400
3500	3500	3600	3600	3600	3700	3700	3800
3900	3900	4100	4200	4300	4300	4300	4600
4600	4700	4800	4900	5200	5600	5800	6200
6200	6600	6800					

STATION	COUNTRY	RIVER	LOCATION
mE19	CANADA	FRASER	HOPE

5130	5810	6000	6060	6830	7080	7220	7220
7420	7480	7560	7620	7700	7820	7820	7820
7840	7900	8040	8040	8040	8160	8210	8330
8470	8500	8500	8520	8550	8580	8670	8670
8720	8840	8980	9010	9060	9250	9290	9350
9520	9540	9690	9690	9770	9770	9910	9970
10300	10300	10500	10600	10800	10800	11100	11300
11600	12500	15200					

STATION	COUNTRY	RIVER	LOCATION
JE792	CANADA	ASSINIBOINE	HEADINGLEY

48	54	61	62	65	92	114	116
117	129	139	146	146	153	174	185
191	202	204	206	216	216	217	222
228	230	233	236	248	264	269	275
276	281	286	289	292	300	306	317
320	340	346	360	382	388	430	473
473	481	518	547	564	566	592	595
615							

STATION		COUNTRY		RIVER		LOCATION	
iFOO		CANADA		S. SASKATCHEWAN		MEDICINE HAT	
230	317	379	391	524	572	575	581
649	683	683	688	722	725	731	733
821	824	827	889	912	940	940	952
957	960	963	974	983	991	991	1030
1040	1040	1070	1090	1090	1090	1130	1290
1370	1520	1550	1630	1690	1830	1840	1880
2080	2090	2170	2200	2400	2550	2710	3060
3710	4080						

STATION		COUNTRY		RIVER		LOCATION	
KF62		CANADA		S. SASKATCHEWAN		SASKATOON	
399	541	583	583	595	632	793	816
852	855	855	861	901	926	980	994
1050	1070	1070	1080	1110	1120	1140	1150
1170	1180	1190	1210	1250	1260	1270	1280
1370	1370	1420	1420	1420	1420	1530	1540
1540	1570	1630	1760	1780	1820	1850	1970
2180	2330	2420	2490	2530	2700	3060	3140
3140	3370	3940					

STATION		COUNTRY		RIVER		LOCATION	
KF53		CANADA		N. SASKATCHEWAN		PRINCE ALBERT	
487	527	589	620	623	683	685	756
759	762	765	770	790	796	799	875
926	940	952	954	991	1010	1010	1050
1070	1110	1120	1130	1140	1180	1190	1200
1230	1250	1250	1270	1290	1340	1350	1510
1540	1560	1570	1570	1570	1620	1620	1640
1650	1790	1800	1980	2090	2160	2460	2790
2930	2970	5300					

STATION		COUNTRY		RIVER		LOCATION	
hE88a		CANADA		HURRICANA		AMOS	
99	99	117	118	125	132	132	135
142	146	150	154	158	158	161	161
161	164	164	166	167	172	172	173
173	174	179	183	183	185	192	194
195	195	201	202	204	205	213	213
216	229	230	230	235	240	244	262
262	264	283	317	337			

STATION		COUNTRY		RIVER		LOCATION	
jF50a		CANADA		WINNIPEG		SLAVE FALLS	
666	668	668	901	986	1000	1020	1030
1050	1060	1060	1090	1100	1140	1200	1250
1250	1270	1290	1370	1390	1420	1450	1450
1510	1590	1720	1720	1750	1790	1920	1970
1990	2040	2190	2260	2390	2410	2450	2780
2800							

Appendix B

Program Flood

```
C      THIS PROGRAM FINDS ESTIMATES FOR THE PARAMETERS
C      PA, PB, AND PC IN TRANSFORMED WEIBULL DISTRIBUTION
C      FUNCTION. WHICH PA IS SHAPE PARAMETER,
C      PB IS LOCATION PARAMETER AND PC IS SCALE PARAMETER.
C      REQUIRED INPUT INCLUDES U, PROBABILITIES OF
C      RECURRENCE INTERVALS.  F(K), THE OBSERVED
C      FLOOD DATA VALUE.
       DIMENSION T(200),X(200),AA(200),BB(200),F(200),
      *Q(200),U(6),XX(200),Y(200)
       DATA U/.5,.8,.9,.95,.98,.99/
       CALL ORDER(X,XLAR,N)
       D=2*XLAR
       WRITE(6,/) (X(I),I=1,N),XLAR,D,N
C      USE CC AS A SCALE FACTORS.
       CC=-X(1)+D
       DO 1 J=1,N
        NJ=N-J+1
       T(NJ)=-X(NJ)+D
       WRITE(6,4)T(NJ),X(J)
 4     FORMAT(2(X,E15.4))
   1   CONTINUE
       B=1
C      THE FOLLOWING 19 STATEMENTS ARE FOR FINDING
C      INITIAL VLAUES FOR THE THREE PARAMETERS FOR
C      TRANSFORMED WEIBULL DISTRIBUTION.
       XSUM=0
       YSUM=0
       XSUM2=0
       SUM1=0
       DO 1001 J=1,N
       XJ=J
       Y(J)=ALOG(-ALOG(XJ/(N+1.)))
       XSUM=XSUM+ALOG(T(J))
       YSUM=YSUM+Y(J)
       XSUM2=XSUM2+(ALOG(T(J)))**2
       SUM1=SUM1+ALOG(T(J))*Y(J)
 1001  CONTINUE
       XSUM3=(XSUM**2)/N
       SUM4=(XSUM*YSUM)/N
       A=(SUM1-SUM4)/(XSUM2-XSUM3)
       ALOGC=(-(YSUM/N)/A+(XSUM/N))
       C=EXP(ALOGC)
       B=B/CC
```

```
      C=C/CC
      DO 1002 J=1,N
      NJ=N-J+1
      T(NJ)=(-X(J)+D)/CC
 1002 CONTINUE
      CALL EST(T,N,A,N,C,D,PC,PA,PB)
      WRITE(6,/)PA,PB,PC
      PC=PC*CC
      PB=D-PB*CC
      WRITE(6,/)PA,PB,PC
      READ(5,/)(F(K),K=1,6)
      TEMPB=0
      TEMPA=0
C     THE FOLLOWING STATEMENTS ARE FOR FINDING AA(I),
C     THE AVERAGE OF THE RELATIVE DEVIATIONS AND
C     BB(I), THE AVERAGED QUADRATIC DEVIATION.
      DO 100 K=1,6
C     X(K) IS THE PREDICTED FLOOD DISCHARGES.
      X(K)=PC-PA*(-ALOG(U(K)))**(1.0/PB)
C     Q(K) IS THE RELATIVE DEVIATION IN PERCENT.
      Q(K)=(X(K)-F(K))/F(K)*100
      TEMPA=TEMPA+ABS(Q(K))
  100 TEMPB=TEMPB+Q(K)**2
      AA(I)=TEMPA/6
      BB(I)=TEMPB/6
 200  WRITE(6,/)(X(K),K=1,6),AA(I),BB(I)
      STOP
      END
```

```
C     THIS SUBROTINE READS THE YEARLY MAXIMUM FLOOD
C     DATA OF A RIVER, ORDERS THIS DATA INTO ASCENDING
C     ORDER, THE SMALLEST X(1) TO LARGEST X(N).
C     NECESSARY INPUT IS THE NUMBER OF YEARS OF THE
C     RECORD N, AND THE ACTUAL DATA IN ARRAY X.
      SUBROUTINE ORDER(X,XLAR,N)
      DIMENSION X(200)
C     N, THE NUMBER OF YEARS OF DATA IS READ.
      READ(5,/)N
C     THE DATA IS READ FREE FORMAT AND STORED IN ARRAY X.
      READ(5,/)(X(I),I=1,N)
      NM=N-1
      DO 30 I=1,NM
      JM=N-I
      DO 20 J=1,JM
      IF(X(J).LE.X(J+1))GO TO 20
      TEMP=X(J)
      X(J)=X(J+1)
      X(J+1)=TEMP
   20 CONTINUE
   30 CONTINUE
      XLAR=X(N)
      RETURN
      END
```

```
      SUBROUTINE EST(T,N,A,B,C1,D,PA,PB,PC)
C     INPUT
C     N=SAMPLE SIZE (BEFORE CENSORING),N=100 OR LESS
C     AS DIMENSIONED
C     SS1=0 IF SCALE PARAMETER THETA IS KNOWN
C     SS1=1 IF SCALE PARAMETER THETAIS TO BE ESTIMATED
C     SS2=0 IF SHAPE PARAMETER K IS KNOWN
C     SS2=1 IF SHAPE PARAMETER K IS TO BE ESTIMATED
C     SS3=0 IF LOCATION PARAMETER C IS KNOWN
C     SS3=1 IF LOCATION PARAMETER C IS TO BE ESTIMATED
C     T(I)=I-TH ORDER STATISTIC OF SAMPLE (I=1,N)
C      (SUBSTITUTE BLANK CARDS FOR UNKNOWN CENSORED
C     OBSERVATIONS)
C     M=NUMBER OF OBSERVATIONS REMAINING AFTER
C     CENSORING N-M FROM ABOVE
C     C(1)=INITIAL ESTIMATE (OR KNOWN VALUE) OF C
C     THETA(1)=INITIAL ESTIMATE (OR KNOWN VALUE) OF THETA
C     EK(1)=INITIAL ESTIMATE (OR KNOWN VALUE) OF K
C     MR=NUMBER OF OBSERVATIONS CENSORED FROM BELOW,
C     NORMALLY 0 INITIAL
C     OUTPUT
C     N,SS1,SS2,SS3,M,C(1),THETA(1),EK(1),MR-
C     SAME AS FOR INPUT
C     C(J)=ESTIMATE AFTER J-1 ITERATIONS (OR KNOWN
C     VALUE) OF C
C     THETA(J)=ESTIMATE AFTER J-1 ITERATIONS (OR KNOWN
C     VALUE) OF THETA
C     EK(J)= ESTIMATE AFTER J-1 ITERATIONS (OR KNOWN
C     VALUE) OF K
C      (MAXIMUM VALUE OF J AS PRESENTLY DIMENSIONED
C     IS 550)
C     EL=NATURAL LOGARITHM OF LIKELIHOOD FOR C(J),
C     THETA(J), EK(J)
C     REFERENCE
C     HARTER,H. LEON AND MOORE, ALBERT H.,
C     MAXIMUM-LIKELIHOOD ESTIMATION OF THE
C     PARAMETERS OF GAMMA AND WEIBULL POPULATIONS
C      FROMCOMPLETE AND FROM CENSORED SAMPLES,
C     TECHNOMETRICS, 7 (1965), 639-643. ERRATA,9 (1967)
C     195
      DOUBLE PRECISION SLK
      DIMENSION T(200),C(550),THETA(550),EK(550),
     *X(56),Y(55)
```

```
      SS1=1.
      SS2=1.
      SS3=1.
      C(1)=C1
      THETA(1)=A
      EK(1)=B
      M=N
      MR=0
      EN=N
32    WRITE(6,5)M,D-C(1),THETA(1),EK(1),MR
5     FORMAT (I4,3F10.4,I4)
      EM=M
31    ELNM=0.
      EMR=MR
      MRP=MR+1
33    NM=N-M+1
      DO 34 I=NM,N
      EI=I
34    ELNM=ELNM+ALOG(EI)
      IF(MR) 66,35,74
74    DO 75 I=1,MR
      EI=I
75    ELNM=ELNM-ALOG(EI)
35    DO 30 J=1,550
      IF(J-1) 66,25,37
37    JJ=J-1
      SK=0.
      SL=0.
      DO 6 I=MRP,M
6     SK=SK+(T(I)-C(JJ))**EK(JJ)
      IF(SS1)7,7,8
7     THETA(J)=THETA(JJ)
      GO TO 9
8     IF(MR) 66,19,20
   19 THETA(J)=((SK+(EN-EM)*(T(M)-C(JJ))**EK(JJ))
     */EM)**(1./EK(JJ))
      GO TO 9
20    X(1)=THETA(JJ)
      LS=0
      DO 21 L=1,55
      LL=L-1
      LP=L+1
      X(LP)=X(L)
      ZRK=((T(MRP)-C(JJ))/X(L))**EK(JJ)
```

```
      Y(L)=-EK(JJ)*(EM-EMR)/X(L)+EK(JJ)*SK/X(L)**(EK
     *(JJ)+1.)+EK(JJ)*(EN-EM)*(T(M)-C(JJ))**
     *EK(JJ)/X(L)**(EK(JJ)+1.)-EMR*EK(JJ)*ZRK*
     2EXP(-ZRK)/(X(L)*(1.-EXP(-ZRK)))
      IF(Y(L)) 53,73,54
53    LS=LS-1
      IF (LS+L) 58,55,58
54    LS=LS+1
      IF (LS-L) 58,56,58
55    X(LP)=.5*X(L)
      GO TO 61
56    X(LP)=1.5*X(L)
      GO TO 61
58    IF(Y(L)*Y(LL)) 60,73,59
59    LL=LL-1
      GO TO 58
60    X(LP)=X(L)+Y(L)*(X(L)-X(LL))/(Y(LL)-Y(L))
61    IF(ABS(X(LP)-X(L))-1.E-3) 73,73,21
21    CONTINUE
73    THETA(J)=X(LP)
9     EK(J)=EK(JJ)
10    IF(SS2) 12,12,11
11    DO 17 I=MRP,M
17    SL=SL+ALOG(T(I)-C(JJ))
      X(1)=EK(J)
      LS=0
      DO 51 L=1,55
      SLK=0.
      DO 18 I=MRP,M
18    SLK=SLK+(ALOG(T(I)-C(JJ))-ALOG(THETA(J)))*
     *(T(I)-C(JJ))**X(L)
      LL=L-1
      LP=L+1
      X(LP)=X(L)
      ZRK=((T(MRP)-C(JJ))/THETA(J))**X(L)
      Y(L)=(EM-EMR)*(1./X(L)-ALOG(THETA(J)))+SL-SLK/
     *THETA(J)**X(L)+(EN-EM)*(ALOG(THETA(J))-
     +ALOG(T(M)-C(JJ)))*(T(M)-C(JJ))**X(L)/
     2THETA(J)**X(L)+EMR*ZRK*(ALOG(ZRK)/X(L))*EXP(-ZRK)/
     3(1.-EXP(-ZRK))
      IF(Y(L)) 43,52,44
43    LS=LS-1
      IF(LS+L) 47,45,47
44    LS=LS+1
      IF(LS-L) 47,46,47
45    X(LP)=.5*X(L)
```

```
      GO TO 50
46    X(LP)=1.5*X(L)
      GO TO 50
47    IF(Y(L)*Y(LL)) 49,52,48
48    LL=LL-1
      GO TO 47
49    X(LP)=X(L)+Y(L)*(X(L)-X(LL))/(Y(LL)-Y(L))
50    IF(ABS(X(LP)-X(L))-1.E-3) 52,52,51
51    CONTINUE
52    EK(J)=X(LP)
12    C(J)=C(JJ)
62    IF(SS3) 25,25,14
14    IF(1.-EK(J)) 16,78,78
78    IF(SS1+SS2) 57,57,16
16    X(1)=C(J)
      LS=0
      DO 23 L=1,55
      SK1=0.
      SR=0.
      DO 15 I=MRP,M
      SK1=SK1+(T(I)-X(L))**(EK(J)-1.)
15    SR=SR+1./(T(I)-X(L))
      LL=L-1
      LP=L+1
      X(LP)=X(L)
      ZRK=((T(MRP)-X(L))/THETA(J))**EK(J)
      Y(L)=(1.-EK(J))*SR+EK(J)*(SK1+(EN-EM)*(T(M)-
     *X(L))**(EK(J)-1.))/THETA(J)**EK(J)-EMR*EK(J)
     **ZRK*EXP(-ZRK)/((T(MRP)-X(L))*(1.-EXP
     2(-ZRK)))
      IF(Y(L)) 39,24,40
39    LS=LS-1
      IF(LS+L) 70,41,70
40    LS=LS+1
      IF(LS-L) 70,42,70
41    X(LP)=.5*X(L)
      GO TO 22
42    X(LP)=.5*X(L)+.5*T(1)
      GO TO 22
70    IF(Y(L)*Y(LL)) 72,24,71
71    LL=LL-1
      GO TO 70
72    X(LP)=X(L)+Y(L)*(X(L)-X(LL))/(Y(LL)-Y(L))
```

```
 22    IF(ABS(X(LP)-X(L))-1.E-3) 24,24,23
 23    CONTINUE
 24    C(J)=X(LP)
       GO TO 25
 57    C(J)=T(1)
 25    IF(MR) 66,38,69
 38    DO 63 I=1,M
       IF(C(J)+1.E-4-T(I)) 68,67,67
 67    MR=MR+1
 63    C(1)=T(1)
 68    IF(MR) 66,69,31
 69    SK=0.
       SL=0.
       DO 36 I=MRP,M
       SK=SK+(T(I)-C(J))**EK(J)
 36    SL=SL+ALOG(T(I)-C(J))
       ZRK=((T(MRP)-C(J))/THETA(J))**EK(J)
       EL=ELNM+(EM-EMR)*(ALOG(EK(J))-EK(J)*ALOG
      *(THETA(J)))+(EK(J)-1.)*SL-
      1(SK+(EN-EM)*(T(M)-C(J))**EK(J))/(THETA(J)
      ***EK(J))+EMR*ALOG(1.-EXP
      2(-ZRK))
       WRITE(6,26)D-C(J),THETA(J),EK(J),EL
 26    FORMAT(4X,3F10.4,E18.8)
       IF(J-3) 30,27,27
 27    IF(ABS(C(J)-C(JJ))-1.E-3) 28,28,30
 28    IF(ABS(THETA(J)-THETA(JJ))-1.E-3) 29,29,30
 29    IF(ABS(EK(J)-EK(JJ))-1.E-3) 66,66,30
 30    CONTINUE
C      PC IS ESTIMATED LOCATION PARAMETER.
 66    PC=C(J)
C      PA IS ESTIMATED SCALE PARAMETER.
       PA=THETA(J)
C      PB IS ESTIMATED SHAPE PARAMETER.
       PB=EK(J)
       RETURN
       END
```

Appendix C

Program Plot

```
C     THIS PROGRAM GETS THE VALUE OF AXIS Z AND Y.
C     ORDERS THE FLOOD DATA IN ASCENDING ORDER FIRST,
C     THEN FROM THE EXTREME VALUE DISTRIBUTION
C     EXP(-((B-X)/C**A) WE KNOW THE EXPECTED VALUE OF
C     ITS DISTRIBUTION FUNCTION EVALUATED AT THE I-TH
C     ORDERED STSTISTICS IS I/(N+1) WHERE THE SAMPLE
C     SIZE IS N.  ALSO FROM THE DISTRIBUTION FUNCTION
C     WE KNOW LN(-LN F(X))=A LN(B-X)-A LN(C) WHICH IS
C     A LINEAR IN X AND F(X).  LET Z=LN(B-X(J))
C     AND Y=LN-LN(J/(N+1)). THEN WE CAN USE THESE
C     VALUES IN MINITAB AND PLOT THE DATA.
      DIMENSION X(200)
      CALL ORDER(X,XLAR,N)
      READ(5,/) B
      DO 10 J=1,N
         XJ=J
         Y=ALOG(-ALOG(XJ/(N+1.0)))
         Z=ALOG(B-X(J))
         WRITE(8,101) Z,Y
   10 CONTINUE
  101 FORMAT(2F10.5)
      WRITE(6,102) (X(I),I=1,N)
  102 FORMAT(5X, 8I7)
      STOP
      END
```