5-1979

Estimation of Floods When Runoff Originates from Nonhomogeneous Sources

David Ray Olson
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd
Part of the Applied Statistics Commons

Recommended Citation
Olson, David Ray, "Estimation of Floods When Runoff Originates from Nonhomogeneous Sources" (1979). All Graduate Theses and Dissertations. 7018.
https://digitalcommons.usu.edu/etd/7018

This Thesis is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact dylan.burns@usu.edu.
ESTIMATION OF FLOODS WHEN RUNOFF ORIGINATES FROM NONHOMOGENEOUS SOURCES

by

David Ray Olson

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

MASTER OF SCIENCE

in

Applied Statistics

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1979
ACKNOWLEDGMENTS

I must give special thanks to my major professor, Dr. Ronald V. Canfield, for providing the initial idea that motivated this study and patient guidance toward its completion.

I wish to express my gratitude to committee members, Dr. Rex L. Hurst and Dr. Gregory W. Jones, for their suggestions and assistance.

Sincere thanks are also extended to Dr. David L. Turner for the exceptional instruction and guidance I received through his outstanding leadership and teaching abilities.

Finally, I wish to express my appreciation to my family for assisting me in accomplishing this goal— to my father for his counsel, understanding and constant guidance— to my mother for typing the preliminary draft of this thesis, and to Carole Hibshman for typing the final copy.

David R. Olson
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF COMPUTER OUTPUTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Relevance of research</td>
<td>1</td>
</tr>
<tr>
<td>Mixture distributions</td>
<td>3</td>
</tr>
<tr>
<td>Objective of study</td>
<td>4</td>
</tr>
<tr>
<td>II. BACKGROUND</td>
<td>5</td>
</tr>
<tr>
<td>Extreme value theory</td>
<td>5</td>
</tr>
<tr>
<td>Mixture distributions in hydrology</td>
<td>6</td>
</tr>
<tr>
<td>Application to stream flow</td>
<td>9</td>
</tr>
<tr>
<td>III. METHODOLOGY</td>
<td>12</td>
</tr>
<tr>
<td>Parameter estimation technique</td>
<td>12</td>
</tr>
<tr>
<td>Scale factor</td>
<td>16</td>
</tr>
<tr>
<td>Example</td>
<td>17</td>
</tr>
<tr>
<td>IV. RESULTS AND DISCUSSION</td>
<td>23</td>
</tr>
<tr>
<td>Goodness-of-fit statistics</td>
<td>23</td>
</tr>
<tr>
<td>Example</td>
<td>25</td>
</tr>
<tr>
<td>Results</td>
<td>29</td>
</tr>
<tr>
<td>A graphical technique</td>
<td>39</td>
</tr>
<tr>
<td>V. CONCLUSION</td>
<td>55</td>
</tr>
<tr>
<td>LITERATURE CITED</td>
<td>56</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS (Continued)

<table>
<thead>
<tr>
<th>APPENDICES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A. Program ORDER</td>
<td>58</td>
</tr>
<tr>
<td>Appendix B. Program INTEGRATE</td>
<td>59</td>
</tr>
<tr>
<td>Appendix C. Program FLOOD</td>
<td>60</td>
</tr>
<tr>
<td>Appendix D. Program PFHTS</td>
<td>62</td>
</tr>
<tr>
<td>Appendix E. Program EPROB</td>
<td>66</td>
</tr>
<tr>
<td>Appendix F. Program DEVIATION</td>
<td>67</td>
</tr>
<tr>
<td>VITA</td>
<td>68</td>
</tr>
</tbody>
</table>

Page  70
LIST OF COMPUTER OUTPUTS

<table>
<thead>
<tr>
<th>Computer Output</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ordered data from the Saquenay River, 1913-1970</td>
<td>18</td>
</tr>
<tr>
<td>2. $E_{L_1}$, $E_{LSQ_1}$, and $Z_1$ for the Saquenay River</td>
<td>19</td>
</tr>
<tr>
<td>3. Parameter estimates $\hat{\alpha}$ and $\hat{\theta}$ for the Saquenay River</td>
<td>21</td>
</tr>
<tr>
<td>4. PFHTS finds the $P(T)$ for selected $T$ of Saquenay River</td>
<td>25</td>
</tr>
<tr>
<td>5. EPROB finds the $D(T)$ for selected $T$ of Saquenay River</td>
<td>28</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Selected stations exhibiting non-homogeneity in source</td>
<td>31</td>
</tr>
<tr>
<td>2.</td>
<td>Maximum flood height b, scale factor sf, and parameters estimated from equation (20)</td>
<td>32</td>
</tr>
<tr>
<td>3.</td>
<td>Parameter estimates of equation (8) for each station</td>
<td>33</td>
</tr>
<tr>
<td>4.</td>
<td>Computed flood discharges P(T) for selected return periods</td>
<td>34</td>
</tr>
<tr>
<td>5.1.</td>
<td>Data values D(T) as interpolated between adjacent observations by the Hazen method</td>
<td>35</td>
</tr>
<tr>
<td>5.2.</td>
<td>Data values D(T) as interpolated between adjacent observations by the Chegodayev method</td>
<td>36</td>
</tr>
<tr>
<td>5.3.</td>
<td>Data values D(T) as interpolated between adjacent observations by the Wilbull method</td>
<td>37</td>
</tr>
<tr>
<td>6.</td>
<td>Goodness-of-fit statistics</td>
<td>38</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1.</td>
<td>Basin cross section</td>
<td>7</td>
</tr>
<tr>
<td>2.</td>
<td>Station No. 1--Saquenay River</td>
<td>44</td>
</tr>
<tr>
<td>3.</td>
<td>Station No. 2--Niger River, Dire</td>
<td>45</td>
</tr>
<tr>
<td>4.</td>
<td>Station No. 3--Niger River, Koulikoro</td>
<td>46</td>
</tr>
<tr>
<td>5.</td>
<td>Station No. 4--Penobscot River</td>
<td>47</td>
</tr>
<tr>
<td>6.</td>
<td>Station No. 5--Kymijoki River</td>
<td>48</td>
</tr>
<tr>
<td>7.</td>
<td>Station No. 6--Vuoksi River</td>
<td>49</td>
</tr>
<tr>
<td>8.</td>
<td>Station No. 7--Oder River</td>
<td>50</td>
</tr>
<tr>
<td>9.</td>
<td>Station No. 8--Vanerngota River</td>
<td>51</td>
</tr>
<tr>
<td>10.</td>
<td>Station No. 9--Neva River</td>
<td>52</td>
</tr>
<tr>
<td>11.</td>
<td>Station No. 10--Assiniboine River</td>
<td>53</td>
</tr>
<tr>
<td>12.</td>
<td>Station No. 11--Red River</td>
<td>54</td>
</tr>
</tbody>
</table>
ABSTRACT

Estimation of Floods when Runoff Originates from Nonhomogeneous Sources

by

David Ray Olson, Master of Science

Utah State University, 1979

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

Extreme value theory is used as a basis for deriving a distribution function for flood frequency analysis when runoff originates from nonhomogeneous sources. A modified least squares technique is used to estimate the parameters of the distribution function for eleven rivers. Goodness-of-fit statistics are computed and the distribution function is found to fit the data very well.

The derived distribution function is recommended as a base method for flood frequency analysis for rivers exhibiting nonhomogeneous sources of runoff if further investigation also proves to be positive.

(70 pages)
CHAPTER I
INTRODUCTION

In June 1972, Hurricane Agnes brought torrential rains and high winds to Pennsylvania. The study of Reich (1973) illustrates the complete failure of previous historical data to predict the flood magnitude witnessed during this storm. Even including the hurricane data, existing estimation techniques are not satisfactory in this case. What is wrong with these estimation methods? Why don't they work? Is there a new technique which will work?

Relevance of research

With the continuing development of flood plains and rural water-sheds for urban use, flood control becomes increasingly important. Construction of dams--water needed for irrigational purposes--keeping a river within its embankments--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. Design parameters usually include the yearly maximum event with n-year return period. Prediction of flood frequency by height is the basis of most of the specifications for flood control. Considerable effort has been expanded to determine a distribution of maximum yearly river height which can be applied uniformly to all streams with reasonable accuracy. After fitting several distributions to many different data sets representing a wide variety of conditions, the log Pearson Type III distributions have been judged to give the best
overall fit (Benson, 1968). This has been reinforced by the work of Beard (1974). There has been disagreement that the log Pearson Type III distributions are "best" (Bobee and Robitaille, 1977), and Reich (1977) has questioned the advisability of even suggesting that a uniform method may apply.

The standard technique for estimating flood frequency is based upon a homogeneous source of runoff. It is designed for regions which exhibit a unimodal distribution of flood heights (a flood is defined here to be the maximum annual flow). In many applications flow characteristics may differ considerably, depending on origin (spring melt, hurricane, etc.). These cases usually result in a bimodal (or multimodal) distribution for flood heights, and are the result of a mixture random variable. It is widely accepted that the present methods of estimating return periods and probabilities of flood events do not work well when runoff originates from nonhomogeneous sources (see for example, Ashkanasy and Weeks, 1975). Additional problems arise when one of the sources is a rare event. An example of such an occurrence is Hurricane Agnes which struck Pennsylvania in June of 1972.

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 extrapolate, 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. A serious difficulty is inherent when one uses empirical fit to select a distribution for maximum river height. Many different distributions can provide a good empirical fit in the range of the data and yet have very different right tail characteristics. The most
important consideration in selecting a distribution for use in des-
cribing maximum yearly river height is the behavior of the right tail
of the distribution. It is from the right tail that return periods and
probabilities of rare events are determined. Considering the region
where greatest accuracy is needed, empirical fit of a distribution
over the data set is not adequate as a sole criterion for choosing a
distribution. Some theoretical principle is needed to assist in the
choice of a distribution due to the absence of data in the right tail.
In the studies on rivers with homogeneous sources of runoff by Benson
(1968), Beard (1974), Bobbee and Robitaille (1977) and others; the
characteristics of the right tail of the distributions examined were
not even considered.

Mixture distributions

The problem of empirical fit is magnified when the data consti-
tutes observations from a mixture of populations. Because mixture
distributions contain so many parameters, almost any distributions
can be used for the components of the mixture with essentially the
same fit to almost any data set. Again it is not possible to reason
that since all distributions fit well, any form will do. Even though
the empirical fit of several distributions may be excellent in the
range of the data, the more important behavior of the right tails of
these distributions may be very different. Since there is no data in
this region, theoretical considerations are invaluable.

The random variable "daily river height" is in many cases a mix-
ture random variable (see Ashkanasy and Weeks, 1975), while the ran-
dom event of interest is the random variable "yearly maximum river
height." Since these yearly floods are extreme events, it is natural to propose extreme value theory as the avenue for selecting distributions for floods. Extreme value distributions have not been fully studied in the previous work comparing distributions of floods. Notable omissions from these comparisons are the general type II and type III extreme value distributions. These distributions are fully described in Chapter II. An extension of extreme value theory in the context of reliability by Canfield and Borgman (1975) to the distribution of the maximum in a sequence of mixture random variables provides a theoretical base for selection of the distribution of floods in a region where runoff originates from nonhomogeneous sources. This extension is also discussed in the next chapter.

Objective of study

The primary objective of this thesis is to develop and evaluate a method of flood frequency analysis based on the theory given in Canfield and Borgman (1975); to explore its application to the yearly maximum height of a river whose daily heights are mixture random variables.
Extreme value theory

The term "flood" by nature suggests application of extreme value theory. Under very general conditions, it has been shown by Gnedenko (1943) that the maximum of a sufficiently long sequence of independent random variables from a given distribution must be closely approximated by one of the following three types:

\[
\phi_1(x) = \exp[-e^{-(x-b)/c}] \quad c > 0, \quad -\infty < x < \infty \quad (1)
\]

\[
\phi_2(x) = \begin{cases}
\exp[-\left(\frac{x-b}{c}\right)^a] & x \geq b, \quad c > 0, \quad a > 0 \\
0 & x < b
\end{cases} \quad (2)
\]

\[
\phi_3(x) = \begin{cases}
\exp[-\left(\frac{b-x}{c}\right)^a] & c > 0, \quad a > 0, \quad x \leq b \\
1 & x > b
\end{cases} \quad (3)
\]

In the work of Gnedenko (1943) independence of the members of the sequence of random variables was assumed. Watson (1952) has shown that this assumption may be relaxed. Estimation of the parameters of the extreme value distributions is given by Mann, Schafer, and Singpurwalla (1974).

The first type (1) is sometimes called the Gumbel distribution and has been used in the previous works comparing distributions for...
floods. If daily heights were normally distributed the yearly maxima should be closely approximated by this type.

The second type (2) is often denoted the Cauchy type since the maximum in a sufficiently long sequence of Cauchy random variables has this approximating distribution. It is therefore characterized as having a density function with a thick tail as compared with the other types. Previous comparisons of distributions for floods have not included this type.

The third type (3) is a "limited" distribution since x is restricted from above, i.e., $x \leq b$. This type seems to have received no attention in recent hydrologic literature. There is a general reluctance to consider a distribution which puts a bound on the height of floods (probably in memory of Noah). This is the parameter b. When one considers the physical aspects of a flood basin it seems natural to place a bound on flood height at any given position along the river as illustrated in Figure 1. Note that this bound on flood height in no way restricts the volume of flood water. Thus it seems that the type III (3) extreme value form with b equal to the basin ridge elevation at an appropriate cross section may be a very natural distribution and should be considered.

Mixture distributions in hydrology

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 mixed distribution for the case of two components in his analysis, i.e.,
where \( F_i(x), i = 1, 2 \) are the distribution functions of the first and second components of the mixture respectively. The parameters \( p_i, i = 1, 2 \) are such that \( p_i > 0, i = 1, 2 \) 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 has been done in this area by Hawkins (1972) and (1974), 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 for the $F_i(x)$ in his analysis of flood data. This seems a good choice relative to the tail characteristics since the data is observed extremes. However, it should be noted that although the random variable describing "daily river height" may be a mixture, it does not follow that the flood (an extreme event, i.e., "maximum yearly river height") 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 yearly flood height 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)

$$F(x) = \phi_1(x)^{p_1} \phi_2(x)^{p_2}$$  \hspace{1cm} (5)

where the components $\phi_1(x)$ and $\phi_2(x)$ are extreme value forms (1),
(2), or (3). Note that the parameters $p_1$ and $p_2$ can be absorbed by reparameterization so that (5) can be written

$$F(x) = \phi_1(x)\phi_1'(x)$$

(6)

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 (4) is a weighted average of the tails of $F_1(x)$ and $F_2(x)$ whereas the shape of (6) is a product of the tails of $\phi_1(x)$ and $\phi_1'(x)$. Even if extreme value distributions are used in (4), the tail shape is not necessarily correct.

Application to stream flow

The primary benefit of the foregoing theory when applied to stream flow is that an appropriate distribution for the components in (6) can be selected from only three types. These three types are very different in their tail characteristics which further simplifies the problem of choice.

The type (1) distribution has been applied to flood data exhibiting a homogeneous source of runoff. For nonhomogeneous data, the form of (6) for this case is

$$F(h) = \exp \left[ -e^{-(h-b)/c} - e^{-(h-b')/c'} \right]$$

(7)

where $c$ and $c' > 0$ and $h$ is the height. This form will not be considered in this thesis.
The type (2) distributions as stated previously are "heavy
tailed," indicating a tendency toward extreme events which are larger
and occur with greater frequency than with types (1) or (3). This
characteristic seems inappropriate for flood height if the relationship
between flow volume and river height is considered. A large increase
in water volume will in most basins produce a very small increase in
flood height. The type (2) distributions will not be considered here.

The most distinguishing feature of the type (3) distributions is
that they are limited above, i.e., \( x \leq b \) in (3). This means that there
is a maximum flood height attainable if used to describe flood height.
However, the relationship between flow volume and river height indi­
cates a possible limiting potential for height. A very large increase
in flow volume produces a small increase in height for most basins.
In addition, considering spill over into neighboring basins, it seems
that topographic features of a basin provide a natural bound for flood
height (See Figure 1). Thus the type (3) distribution should not be
ignored. It seems to have found little use in hydrology, probably
because of reluctance to assign a maximum height parameter. The dis­
tribution was not used even in the comparisons of distributions by
Beard (1974) and Benson (1968) for homogeneous flood data. The
appropriate form of (6) for this case (for nonhomogeneous data)
after reparameterization is given by

\[
F(h) = \begin{cases} 
\exp \left[ - \left( \frac{h-h}{c} \right)^a - \left( \frac{h-h}{c'} \right)^{a'} \right] & h \leq b \\
1 & h > b 
\end{cases}
\]
This extreme value type (8) is examined in detail in the chapters which follow. The parameters of $F(h)$ are estimated for several data sets with nonhomogeneous sources of runoff. The goodness-of-fit statistics used by Bobee and Robitaille (1977) which have the same basis as those used by the Work Group on Flow Frequency Methods (Benson, 1968) are calculated. A comparison is then made of this new method with the standard technique.
CHAPTER III
METHODOLOGY

The usefulness of the model developed in the last chapter depends upon three points: the goodness-of-fit to a given data set, the right tail, and the availability of estimation techniques for the parameters in the distribution function. The goodness-of-fit to data sets exhibiting nonhomogeneous sources of runoff is considered in Chapter IV. The theoretical basis for having the correct right tail has been explained in the last chapter. Although this theory behind extreme value form (8) is exciting and convincing, it isn't worth anything in practice unless the parameters equation (8) contains can be estimated. Estimation is considered here.

Parameter estimation technique

Parameter estimation for a case similar to the one being examined in this study has been given by Canfield and Borgman (1975). It is based upon a least squares technique reported by Bain and Antle (1967) for the Weibull distribution. A slight extension provides a method of estimation compatible with extreme value form (8).

Let \( h_{(i)}, i = 1, 2, \ldots, n \) be the \( i \)th order statistic of \( n \) yearly maximum flood heights. Then for any distribution \( F(h) \) of maximum flood heights

\[
E[F(h_{(i)})] = \frac{i}{n+1}
\]  
(9)

where \( E[ ] \) means expected value (see Lindgren, 1976).
Estimates of the parameters of (8) are taken to be those values which minimize the expression

$$\Psi = \sum_{i=1}^{n} \left\{ \ln E[F(h(i))] - \ln F(h(i)) \right\}^2$$

$$= \sum_{i=1}^{n} \left\{ \ln E[F(h(i))] + \left( \frac{b-h(i)}{c} \right)^a + \left( \frac{b-h(i)}{c'} \right)^{a'} \right\}^2$$

White (1969) has suggested that $\ln E[F(h(i))]$ be replaced by $E[\ln F(h(i))]$ which removes many biases. Thus (10) becomes

$$\Psi = \sum_{i=1}^{n} \left\{ E[\ln F(h(i))] + \left( \frac{b-h(i)}{c} \right)^a + \left( \frac{b-h(i)}{c'} \right)^{a'} \right\}^2$$

(11)

In order to speed the convergence of the minimization process in the numerical solution and to place greater emphasis on the larger floods, a weight factor is also included. Let

$$z_i^{-1} = E\left[ E[\ln F(h(i))] - \ln F(h(i)) \right]^2$$

(12)

The weight factor for the minimization of (11) is given by

$$w_i^{-1} = \frac{z_i^{-1}}{z_n^{-1}}$$

(13)

Equation (11) can now be written

$$\Psi = \sum_{i=1}^{n} \left\{ E[\ln F(h(i))] + \left( \frac{b-h(i)}{c} \right)^a + \left( \frac{b-h(i)}{c'} \right)^{a'} \right\}^2 w_i$$

(14)
are approximated using numerical integration by the trapezoid rule.

\[
E[\ln F(h_{(i)})] = \frac{n!}{(i-1)!(n-1)!} \int_0^1 \ln F(h_{(i)}) \left[ F(h_{(i)}) \right]^{i-1} \left[ 1-F(h_{(i)}) \right]^{n-1} dF(h_{(i)})
\]

(9-2)

\[
E\left[ \left\{ E[\ln F(h_{(i)})] - \ln F(h_{(i)}) \right\}^2 \right] = E\left[ \left( \ln F(h_{(i)}) \right)^2 \right] - \left\{ E[\ln F(h_{(i)})] \right\}^2 = \frac{n!}{(i-1)!(n-1)!} \int_0^1 \left[ \ln F(h_{(i)}) \right]^2 \left[ F(h_{(i)}) \right]^{i-1} \left[ 1-F(h_{(i)}) \right]^{n-1} dF(h_{(i)}) - \left\{ \frac{n!}{(i-1)!(n-1)!} \int_0^1 \ln F(h_{(i)}) \left[ F(h_{(i)}) \right]^{i-1} \left[ 1-F(h_{(i)}) \right]^{n-1} dF(h_{(i)}) \right\}^2
\]

(9-3)

(See Statistical Theory by Lindgren (1976), page 218 for the density function of the ith order statistic, page 113 for the expectation of a function of a random variable, and page 484 for equation (9-1)).

For convenience let

\[
EL_{i} = E[\ln F(h_{(i)})]
\]

\[
ELSQ_{i} = E\left[ \left\{ E[\ln F(h_{(i)})] - \ln F(h_{(i)}) \right\}^2 \right]
\]

\[
Y_{i} = b-h_{(i)}
\]

\[
\alpha' = (\alpha_1, \alpha_2) = (a, a')
\]
\[ \theta' = (\theta_1, \theta_2) = \left( \frac{1}{c}, \frac{1}{(c')^a} \right) \]

Equation (14) can be written in final form as

\[ \psi = \sum_{i=1}^{n} \left\{ (EL_i + \theta_1 Y_{i1} + \theta_2 Y_{i2}) \right\}^2 w_i \]

(15)

\[ = \sum_{i=1}^{n} \left\{ \sqrt{w_i EL_i + \theta_1 Y_{i1} \sqrt{w_i} + \theta_2 Y_{i2} \sqrt{w_i}} \right\}^2 \]

Estimation of \(a, a', c\) and \(c'\) is accomplished by estimating \(a\) and \(\theta\)
and then solving for \(a, a', c\) and \(c'\) respectively.

In order to minimize equation (15) appropriate partial derivatives
of \(\psi\) are evaluated and set equal to zero.

\[ \frac{\partial \psi}{\partial \theta_1} = \sum_{i=1}^{n} w_i EL_i Y_{i1} + \theta_1 \sum_{i=1}^{n} w_i Y_{i1} + \theta_2 \sum_{i=1}^{n} w_i Y_{i1} = 0 \]  

(16-1)

\[ \frac{\partial \psi}{\partial \theta_2} = \sum_{i=1}^{n} w_i EL_i Y_{i1} + \theta_1 \sum_{i=1}^{n} w_i Y_{i1} + \theta_2 \sum_{i=1}^{n} w_i Y_{i1} = 0 \]  

(16-2)

\[ \frac{\partial \psi}{\partial \alpha_1} = \sum_{i=1}^{n} w_i EL_i Y_{i1} \ln Y_{i1} + \theta_1 \sum_{i=1}^{n} w_i Y_{i1} \ln Y_{i1} + \theta_2 \sum_{i=1}^{n} w_i Y_{i1} \ln Y_{i1} = 0 \]  

(16-3)

\[ \frac{\partial \psi}{\partial \alpha_2} = \sum_{i=1}^{n} w_i EL_i Y_{i1} \ln Y_{i1} + \theta_1 \sum_{i=1}^{n} w_i Y_{i1} \ln Y_{i1} + \theta_2 \sum_{i=1}^{n} w_i Y_{i1} \ln Y_{i1} = 0 \]  

(16-4)

Solving (16-1) for \(\theta_2\) yields
Substituting for $\theta_2$ in equation (16-2) and solving for $\theta_1$ gives

$$
\theta_1 = \frac{\left( \sum_{i=1}^{n} \frac{\alpha_1}{W_{i1}Y_{i1}} \right) - \left( \sum_{i=1}^{n} \frac{2\alpha_2}{W_{i1}Y_{i1}} \right)}{\left( \sum_{i=1}^{n} \frac{\alpha_1 + \alpha_2}{W_{i1}Y_{i1}} \right)^2} - \left( \sum_{i=1}^{n} \frac{2\alpha_1}{W_{i1}Y_{i1}} \right) \left( \sum_{i=1}^{n} \frac{2\alpha_2}{W_{i1}Y_{i1}} \right)
$$

The result of equation (18) is substituted into equation (17) to yield equations for both $\theta_1$ and $\theta_2$ which involve the parameters $\alpha_1$ and $\alpha_2$ as the only unknowns. These equations are substituted for $\theta_1$ and $\theta_2$ in equations (16-3) and (16-4) giving two equations in two unknowns---$\alpha_1$ and $\alpha_2$. This system of equations can be solved numerically using the IMSL (1977) library subroutine ZSYSTM. Given this solution as $\hat{\alpha}$, the estimate $\hat{\theta}$ of $\theta$ is computed from equations (17) and (18).

**Scale factor**

A Burroughs 6700 computer was used to solve for $\hat{\alpha}$ and since the Burroughs or any other computer system is finite, a scaling factor was found to be a computational necessity, i.e., equation (15) becomes

$$\psi = \sum_{i=1}^{n} \left\{ \frac{\sqrt{W_{i1}}}{\sqrt{\psi}} + (sf) \theta_1 \left( \frac{Y_{i1}}{sf} \right)^{\alpha_1} + (sf) \theta_2 \left( \frac{Y_{i1}}{sf} \right)^{\alpha_2} \right\} 2
$$

For convenience $\theta_1$ and $\theta_2$ are changed so that
\[ \psi = \sum_{i=1}^{n} \left\{ \sqrt{W_1} \cdot EL_i + \theta_1 \left( \frac{Y_i}{s_f} \right)^{\alpha_1} \sqrt{W_1} + \theta_2 \left( \frac{Y_i}{s_f} \right)^{\alpha_2} \sqrt{W_1} \right\}^2 \]  \quad (20)

For eight of the eleven data sets used in this study, an acceptable scale factor was the difference between the specified maximum possible flood height and the first order statistic or smallest of the maximum yearly floods:

\[ sf = b - h_{(1)} \]  \quad (21)

The other three data sets required manipulation of the scale factor to insure that no numbers became too large or too close to zero for the computer to handle. Of course, larger and more powerful computer facilities would lessen the importance of the scale factor.

Example

In order to clearly demonstrate the estimation procedure described in this chapter, one example is worked in detail. The parameters of equation (20) are estimated for the Saquenay River in Canada. All computer programs mentioned are found and explained in detail in the appendix. The output for runs on the Saquenay River is found and explained in this chapter.

There are fifty-eight years of data for the Saquenay River, from 1913 to 1970. Each observation is the maximum flood occurring in that particular year. Using the program ORDER (see appendix), the data is ordered into ascending order of magnitude, is written to disk, and is printed out for observation (see Computer Output 1). Input is the number of observations and the observations themselfed.

The number of years of record: 58

The ordered maximum yearly floods:

2370.0  2380.0  2410.0  2730.0  2830.0
3400.0  3510.0  3600.0  3650.0  3770.0
3820.0  3850.0  3850.0  3960.0  4050.0
4050.0  4080.0  4110.0  4190.0  4190.0
4250.0  4420.0  4420.0  4420.0  4450.0
4530.0  4530.0  4590.0  4640.0  4670.0
4670.0  4870.0  4930.0  4930.0  4950.0
5010.0  5070.0  5150.0  5180.0  5270.0
5550.0  5660.0  5720.0  5830.0  5920.0
6030.0  6120.0  6370.0  6460.0  6460.0
6480.0  6740.0  6770.0  6820.0  7390.0
7930.0  9060.0  9260.0
Note that the largest flood in fifty-eight years is 9260 and that the smallest is 2370. The parameter \( b \) can be any height which equals or exceeds the basin ridge height. Several representative "maximum possible flood heights" are examined for the data set and the one exhibiting the "best" fit is chosen as an appropriate bound. (Determining goodness-of-fit is explained in Chapter IV.) The maximum possible flood height (\( b \)) is chosen in this case to be 25,000. The scaling factor is found using equation (21):

\[
sf = b - h(1) = 25,000 - 2370 = 22,630
\]

Using the program INTEGRATE \( EL_1 \), \( ELSQ_1 \), and \( Z_1 \) are found with numerical integration by the trapezoid rule (see Computer Output 2) and are written to disk. Input is the number of years of data. The \( EL_1 \) are the expected values of the log of the distribution of the fifty-eight order statistics. The \( ELSQ_1 \) and \( Z_1 \) are intermediate results in forming the weight factor for the Saquenay River.

Computer Output 2. \( EL_1 \), \( ELSQ_1 \), and \( Z_1 \) for Saquenay River.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( EL_1 )</td>
<td>( ELSQ_1 )</td>
<td>( Z_1 )</td>
</tr>
<tr>
<td>4.462284161774</td>
<td>21.15399255548</td>
<td>0.80514995802</td>
</tr>
<tr>
<td>3.472463664175</td>
<td>12.82640122958</td>
<td>1.30109712449</td>
</tr>
<tr>
<td>3.141561459905</td>
<td>10.235183726183</td>
<td>2.72456398253</td>
</tr>
<tr>
<td>2.818547297039</td>
<td>8.181957615097</td>
<td>3.76076148499</td>
</tr>
<tr>
<td>2.562950659591</td>
<td>6.773053716392</td>
<td>4.89496688550</td>
</tr>
<tr>
<td>2.362917958813</td>
<td>5.747515382778</td>
<td>6.08869385619</td>
</tr>
<tr>
<td>2.196254308278</td>
<td>4.959984108114</td>
<td>7.32802948222</td>
</tr>
<tr>
<td>2.005397407690</td>
<td>4.332484444571</td>
<td>8.61746703513</td>
</tr>
<tr>
<td>1.923947852830</td>
<td>3.819135152521</td>
<td>9.95835298750</td>
</tr>
<tr>
<td>1.917238259697</td>
<td>3.396802375420</td>
<td>11.35425169088</td>
</tr>
<tr>
<td>1.717288333973</td>
<td>3.02745107512</td>
<td>12.808585686717</td>
</tr>
<tr>
<td>$EL_1$</td>
<td>$ELSQ_1$</td>
<td>$Z_1$</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>-1.625377248466</td>
<td>2.714091122769</td>
<td>14.249449180795</td>
</tr>
<tr>
<td>-1.543043015352</td>
<td>2.443848352798</td>
<td>15.90739201833</td>
</tr>
<tr>
<td>-1.46512838609</td>
<td>2.206456981174</td>
<td>17.56028948525</td>
</tr>
<tr>
<td>-1.39492267180</td>
<td>1.997011147991</td>
<td>19.28401878315</td>
</tr>
<tr>
<td>-1.328025408514</td>
<td>1.811052178133</td>
<td>21.09695179638</td>
</tr>
<tr>
<td>-1.26552504014</td>
<td>1.645048974869</td>
<td>22.99172683797</td>
</tr>
<tr>
<td>-1.20670871102</td>
<td>1.49815343303</td>
<td>24.97893993143</td>
</tr>
<tr>
<td>-1.15146515546</td>
<td>1.362045605403</td>
<td>27.06553415708</td>
</tr>
<tr>
<td>-1.09801048049</td>
<td>1.24012287977</td>
<td>29.25253369490</td>
</tr>
<tr>
<td>-1.054143936594</td>
<td>1.131060794317</td>
<td>31.56834273618</td>
</tr>
<tr>
<td>-1.00982568820</td>
<td>1.031202288927</td>
<td>34.00244664011</td>
</tr>
<tr>
<td>-0.951441343525</td>
<td>0.940211693565</td>
<td>36.57135726767</td>
</tr>
<tr>
<td>-0.91903022656</td>
<td>0.85712483766</td>
<td>39.25731562421</td>
</tr>
<tr>
<td>-0.87329615949</td>
<td>0.781132414051</td>
<td>42.16373738974</td>
</tr>
<tr>
<td>-0.82994345989</td>
<td>0.71150920772</td>
<td>45.21396121128</td>
</tr>
<tr>
<td>-0.787134477528</td>
<td>0.647640245696</td>
<td>48.45484149459</td>
</tr>
<tr>
<td>-0.75479744491</td>
<td>0.588945810324</td>
<td>51.90428534081</td>
</tr>
<tr>
<td>-0.71603554776</td>
<td>0.5356176886</td>
<td>55.58429892043</td>
</tr>
<tr>
<td>-0.684000796156</td>
<td>0.485479079555</td>
<td>59.18638954711</td>
</tr>
<tr>
<td>-0.651267462822</td>
<td>0.43983954648</td>
<td>63.73347928720</td>
</tr>
<tr>
<td>-0.61509938306</td>
<td>0.397822400823</td>
<td>68.26047708943</td>
</tr>
<tr>
<td>-0.57753393406</td>
<td>0.35913413428</td>
<td>73.13575296077</td>
</tr>
<tr>
<td>-0.55715536934</td>
<td>0.323512531712</td>
<td>78.04186637701</td>
</tr>
<tr>
<td>-0.52904503247</td>
<td>0.29720840653</td>
<td>84.10555254530</td>
</tr>
<tr>
<td>-0.49947317426</td>
<td>0.26054700322</td>
<td>90.30526841392</td>
</tr>
<tr>
<td>-0.47109539548</td>
<td>0.232798439581</td>
<td>96.10983122513</td>
</tr>
<tr>
<td>-0.44466369921</td>
<td>0.207301445132</td>
<td>102.47685915224</td>
</tr>
<tr>
<td>-0.41835250447</td>
<td>0.18389784715</td>
<td>108.62573772956</td>
</tr>
<tr>
<td>-0.39271154306</td>
<td>0.16244346823</td>
<td>114.65322157930</td>
</tr>
<tr>
<td>-0.36771515840</td>
<td>0.14280824049</td>
<td>121.133951970419</td>
</tr>
<tr>
<td>-0.34281319043</td>
<td>0.12481141475</td>
<td>127.59244116590</td>
</tr>
<tr>
<td>-0.31951178794</td>
<td>0.10852257622</td>
<td>134.55784639138</td>
</tr>
<tr>
<td>-0.29625597314</td>
<td>0.09366449269</td>
<td>142.169571554405</td>
</tr>
<tr>
<td>-0.27526700414</td>
<td>0.08019353167</td>
<td>149.88647510816</td>
</tr>
<tr>
<td>-0.25156479191</td>
<td>0.06455842542</td>
<td>158.11949351087</td>
</tr>
<tr>
<td>-0.22956347757</td>
<td>0.05711213925</td>
<td>166.7069424305</td>
</tr>
<tr>
<td>-0.20829375212</td>
<td>0.04734279016</td>
<td>175.2363694697</td>
</tr>
<tr>
<td>-0.18745741677</td>
<td>0.03886459283</td>
<td>184.7502963113</td>
</tr>
<tr>
<td>-0.16704925548</td>
<td>0.03101346701</td>
<td>194.77803330215</td>
</tr>
<tr>
<td>-0.14769255405</td>
<td>0.02433121584</td>
<td>205.36931274235</td>
</tr>
<tr>
<td>-0.12744112920</td>
<td>0.01856473895</td>
<td>216.94027921776</td>
</tr>
<tr>
<td>-0.10821038216</td>
<td>0.01366295868</td>
<td>229.16530312170</td>
</tr>
<tr>
<td>-0.09934263026</td>
<td>0.00957956456</td>
<td>242.59411401621</td>
</tr>
<tr>
<td>-0.0809245367</td>
<td>0.00627046744</td>
<td>257.20154178983</td>
</tr>
<tr>
<td>-0.052834572763</td>
<td>0.00369516988</td>
<td>273.04319412607</td>
</tr>
<tr>
<td>-0.047709275142</td>
<td>0.001815325703</td>
<td>290.24160824696</td>
</tr>
<tr>
<td>-0.043758526344</td>
<td>0.000594256942</td>
<td>317.02427517514</td>
</tr>
</tbody>
</table>
Program FLOOD reads the ordered data, $EL_i$ and $Z_i$ from disk and requires as input $b$, $sf$, and the number of years of data. Also required as input are starting values for $\alpha_1$ and $\alpha_2$. (Method for obtaining starting values is described in Chapter IV.) FLOOD minimizes equation (20) and prints out $\hat{a}$ and $\hat{\theta}$ from which $a$, $a'$, $c$ and $c'$ can be computed (see Computer Output 3).

Computer Output 3. Parameter estimates $\hat{\alpha}$ and $\hat{\theta}$ for the Saquenay River.

<table>
<thead>
<tr>
<th>The Number of Years of the Data Record</th>
<th>58</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Maximum Possible Flood Height</td>
<td>25000</td>
</tr>
<tr>
<td>The Scale Factor</td>
<td>22830</td>
</tr>
</tbody>
</table>

The Starting Values Are

<table>
<thead>
<tr>
<th>Alpha(1)</th>
<th>Alpha(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.62162</td>
<td>10.48941</td>
</tr>
</tbody>
</table>

Number of Iterations of External Function: 32

Error Message: 0

Parameter Estimates Are

<table>
<thead>
<tr>
<th>Alpha(1)</th>
<th>10.3261278196</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theta(1)</td>
<td>4.264827193</td>
</tr>
<tr>
<td>Alpha(2)</td>
<td>8.0873938840</td>
</tr>
<tr>
<td>Theta(2)</td>
<td>0.0010937330</td>
</tr>
</tbody>
</table>
\( a = \alpha_1 = 16.33 \)

\( a' = \alpha_2 = 8.67 \)

\( c = \frac{sf}{1/a'} = 20706.33 \)

\( c' = \frac{sf}{1/a'} = 49696.00 \)

Thus, the estimated distribution function for maximum yearly flood height on the Saquenay River is

\[
F(h) = \exp \left[ - \left( \frac{25000-h}{20706.33} \right)^{16.33} - \left( \frac{25000-h}{49696} \right)^{8.67} \right]
\]

A distribution function can be obtained in like manner for any river manifesting nonhomogeneous sources of runoff, which in theory has the proper right tail for predicting floods. But how well does such a distribution function fit the data?
CHAPTER IV
RESULTS AND DISCUSSION

Goodness-of-fit statistics

The goodness-of-fit statistics used by Bobee and Robitaille (1977) are used in this study. 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; they 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 squared deviation expressed as a percent between the predicted flow over selected recurrence intervals and the observed flow. The recurrence intervals or return periods are \( T = 2, 5, 10, 20, 50, \) and 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, \( P(T) \), for these return periods are calculated using program PFHTS which is an interactive (terminal) program. For a given distribution function \( F(h) \) (equation (8) with parameters \( \alpha \) and \( \theta \) which are found by minimizing equation (20) for a set of data, PFHTS calculates \( F(h) \) for any given \( h \). The six heights \( h \) which give \( F(h) = 0.5, 0.8, 0.9, 0.95, 0.98, \) and 0.99 respectively are the predicted flood heights for the return periods.

The observed flow data values, \( D(T) \), are obtained from the ordered data with program EPROB which uses a formula of plotting position; and then by interpolating between the two adjacent floods of the sample that bracket the specified probability (or the selected recurrence interval).
The three formulae of expected probabilities used to obtain the data values are:

Hazen \[ P_0 \left( \frac{m - i + 1}{m} \right) - 0.5 \]

Chagodayev \[ P_1 \left( \frac{m - i + 1}{m + 0.4} \right) \]

Weibull \[ P_2 \left( \frac{m - i + 1}{m + 1} \right) \]

where \( i \) is the rank of the observation in a sample of size \( m \), varying from 1 for the lowest flow to \( m \) for the highest.

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

\[ q(T) = \frac{P(T) - D(T)}{D(T)} \times 100 \]

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

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

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

where \( A \) represents the average of the absolute value of the relative deviations over the \( L \) selected recurrence intervals and \( B \) represents the quadratic deviation averaged over the \( L \) selected recurrence intervals.
Example

To clearly demonstrate the goodness-of-fit procedures the Saquenay River is examined in detail. The predicted flood discharges, \( P(T) \), are found using PFHTS for the desired return periods (see Computer Output 4). Note that the parameters \( \hat{\alpha} \) and \( \hat{\theta} \) are the estimated obtained from the program FLOOD (Computer Output 3).

Computer Output 4. PFHTS finds the \( P(T) \) for selected \( T \) of Saquenay River.

```
ENTER BB AND CC#?
25330,22330
ENTER ALPHA(1) AND ALPHA(2)
16.3261278195,8.667393884

ENTER THETA(1) AND THETA(2)
4.2546271929,0.00109373298461

ENTER X
4754
F(X) = 0.499975620849000

ENTER X
4754.5
F(X) = 0.500115313902000

ENTER X
6112
F(X) = 0.799930535040000

ENTER X
6112.5
F(X) = 0.800007568872000

ENTER X
6961
F(X) = 0.899962275258000
```
Thus, the predicted flood heights \( P(T) \) for the selected recurrence intervals are:
Note that BB is the maximum possible flood height and CC is the scale factor in Computer Output 4.

The observed flood heights or data values, D(T), for the given return periods are found using program EPROB. It utilizes the three plotting position formulas Hazen, Chegodayev and Weibull of equation (22) to find expected probabilities of being equaled or exceeded for over half of the largest ordered data observations. Then the D(T) for T = 2, 5, 10, 20, 50, and 100 years (or equivalently for the probabilities of being equaled or exceeded of $P_m = 0.5$, 0.2, 0.1, 0.05, 0.02, and 0.01 respectively) are computed by linear interpolation between the two adjacent floods which bracket the specified probability (see Computer Output 5). For example, for a twenty year recurrence interval or a probability of being equaled or exceeded of 0.05 ($T = 20$ is equivalent to $P_m = 0.05$) both the Hazen and Chegodayev methods bracket $P_m = 0.05$ with the yearly maximum flood heights 7390 and 7930. The Weibull method yields 7930 and 9060 as the bracketing heights. The data values D(T) for the Saquenay River as interpolated between adjacent observations using the three different methods are:

<table>
<thead>
<tr>
<th>T</th>
<th>D(T)</th>
<th>Hazen</th>
<th>Chegodayev</th>
<th>Weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4754</td>
<td>4655</td>
<td>4655</td>
<td>4655</td>
</tr>
<tr>
<td>5</td>
<td>6112</td>
<td>6111</td>
<td>6125</td>
<td>6170</td>
</tr>
<tr>
<td>10</td>
<td>6961</td>
<td>6761</td>
<td>6766</td>
<td>6775</td>
</tr>
<tr>
<td>20</td>
<td>7740</td>
<td>7714</td>
<td>7811</td>
<td>7987</td>
</tr>
<tr>
<td>50</td>
<td>8699</td>
<td>9128</td>
<td>9166</td>
<td>9224</td>
</tr>
<tr>
<td>100</td>
<td>9382</td>
<td>9244</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
Computer Output 5. EPROB finds the D(T) for selected T of Saquenay River.

<table>
<thead>
<tr>
<th>T DATA</th>
<th>T HAZEN</th>
<th>T CHEGODAYEV</th>
<th>T WELSHULL</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.590</td>
<td>0.52568969366</td>
<td>0.52568931506</td>
<td>0.525287278614</td>
</tr>
<tr>
<td>4.640</td>
<td>0.508620689656</td>
<td>0.50859164335</td>
<td>0.508474576271</td>
</tr>
<tr>
<td>4.670</td>
<td>0.491379310344</td>
<td>0.491438356135</td>
<td>0.491524237129</td>
</tr>
<tr>
<td>4.670</td>
<td>0.474137931034</td>
<td>0.47431506642</td>
<td>0.474576271186</td>
</tr>
<tr>
<td>4.670</td>
<td>0.45849551725</td>
<td>0.45719178082</td>
<td>0.457627110643</td>
</tr>
<tr>
<td>4.930</td>
<td>0.439655172413</td>
<td>0.44006491350</td>
<td>0.440677966102</td>
</tr>
<tr>
<td>4.930</td>
<td>0.422413793103</td>
<td>0.42294520547</td>
<td>0.425728013559</td>
</tr>
<tr>
<td>4.950</td>
<td>0.405172413733</td>
<td>0.405821917808</td>
<td>0.408779661102</td>
</tr>
<tr>
<td>5.010</td>
<td>0.38793103442</td>
<td>0.38869430137</td>
<td>0.3895098473</td>
</tr>
<tr>
<td>5.070</td>
<td>0.370669655172</td>
<td>0.371575342464</td>
<td>0.37208135932</td>
</tr>
<tr>
<td>5.100</td>
<td>0.353448275862</td>
<td>0.354452354795</td>
<td>0.355342233391</td>
</tr>
<tr>
<td>5.160</td>
<td>0.33620682653</td>
<td>0.337328767122</td>
<td>0.338263050848</td>
</tr>
<tr>
<td>5.270</td>
<td>0.318965517241</td>
<td>0.320205479451</td>
<td>0.32033898405</td>
</tr>
<tr>
<td>5.550</td>
<td>0.301724137931</td>
<td>0.303082191780</td>
<td>0.305064157505</td>
</tr>
<tr>
<td>5.600</td>
<td>0.28468735621</td>
<td>0.285958404109</td>
<td>0.28613559322</td>
</tr>
<tr>
<td>5.720</td>
<td>0.267241379310</td>
<td>0.268350166458</td>
<td>0.269186440177</td>
</tr>
<tr>
<td>5.830</td>
<td>0.250109000000</td>
<td>0.251712328767</td>
<td>0.25423728136</td>
</tr>
<tr>
<td>5.920</td>
<td>0.23275866090</td>
<td>0.234589041094</td>
<td>0.23726813593</td>
</tr>
<tr>
<td>6.030</td>
<td>0.21551723137</td>
<td>0.217465753423</td>
<td>0.22033883050</td>
</tr>
<tr>
<td>6.120</td>
<td>0.19872586269</td>
<td>0.20034245752</td>
<td>0.203589635097</td>
</tr>
<tr>
<td>6.370</td>
<td>0.18103462759</td>
<td>0.183219178041</td>
<td>0.186440677966</td>
</tr>
<tr>
<td>6.460</td>
<td>0.163793103447</td>
<td>0.16609589041</td>
<td>0.169481525425</td>
</tr>
<tr>
<td>6.460</td>
<td>0.14655174134</td>
<td>0.14972052739</td>
<td>0.152542372802</td>
</tr>
<tr>
<td>6.480</td>
<td>0.12931034462</td>
<td>0.131849315966</td>
<td>0.13559320339</td>
</tr>
<tr>
<td>6.740</td>
<td>0.11208965517</td>
<td>0.11472602397</td>
<td>0.118440677977</td>
</tr>
<tr>
<td>6.770</td>
<td>0.094627586206</td>
<td>0.097602739725</td>
<td>0.10164415265</td>
</tr>
<tr>
<td>6.820</td>
<td>0.077506206963</td>
<td>0.080479452054</td>
<td>0.084745766711</td>
</tr>
<tr>
<td>7.390</td>
<td>0.060344627586</td>
<td>0.063356164343</td>
<td>0.067796610169</td>
</tr>
<tr>
<td>7.950</td>
<td>0.043103446275</td>
<td>0.0462328576712</td>
<td>0.050847457625</td>
</tr>
<tr>
<td>9.460</td>
<td>0.025862068945</td>
<td>0.02910458941</td>
<td>0.03384830504</td>
</tr>
<tr>
<td>9.260</td>
<td>0.008620696895</td>
<td>0.011986301389</td>
<td>0.016449152542</td>
</tr>
</tbody>
</table>
Note that for the Chegodayev and Weibull methods the sample size is too small to obtain the observed flood height for the 100 year return period. Extrapolation is not used for we know nothing beyond the range of the data.

The goodness-of-fit statistics $A$ and $B$ from equations (23) and (24) are computed for all three expected probability methods—Hazen, Chegodayev and Weibull—using the program DEVIATION. Of course, the smaller these statistics are, the better the fit (see Computer Output 6).

Results

Bobee and Robitaille (1977) evaluated 28 long-term river records of annual flood peaks from stations across the world for the condition that runoff occurs due to a single source. The records of the selected stations of Canada are from the National Committee of IHD (1972) and those of the world excluding Canada are from the Unesco Inventory (1971). Of these samples, eleven were considered nonhomogeneous in source and were eliminated from their study. These eleven stations are listed in Table 1.

Using the flood frequency analysis technique described in the previous chapter, the parameters $\alpha$ and $\theta$ are estimated for each of the data records. These parameters are listed in Table 2. Using $\hat{\alpha}$ and $\hat{\theta}$ for each station $a$, $a'$, $c$, and $c'$ are calculated, which comprise the parameters of the estimated distribution function $F(h)$ of maximum yearly flood height for each of the eleven rivers (see Table 3). These distribution functions all have the form of equation (8).

<table>
<thead>
<tr>
<th></th>
<th>Mean of the Absolute Relative Deviations</th>
<th>Mean Quadratic Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hazen</td>
<td>1.94</td>
<td>6.28</td>
</tr>
<tr>
<td>Chegodayev</td>
<td>2.25</td>
<td>7.94</td>
</tr>
<tr>
<td>Weibull</td>
<td>2.92</td>
<td>10.97</td>
</tr>
</tbody>
</table>

Program PFHTS yields predicted flood discharges for each river (see Table 4) and in Tables 5.1, 5.2, and 5.3 are found the data values as interpolated between adjacent observations for each station using the Hazen, Chegodayev, and Weibull methods and program EPROB. Finally, in Table 6 are located the goodness-of-fit statistics computed for each of the eleven data sets.

Bobee and Robitaille (1977) computed these goodness-of-fit statistics for the seventeen stations which exhibited a single source of runoff. They compared the Pearson Type III and log Pearson Type III distributions using eleven different methods of parameter estimation. Their methods resulted in mean absolute relative deviations ranging from 0.9 to 10.4 and mean quadratic deviations ranging from 1.1 to 192.9.
<table>
<thead>
<tr>
<th>No.</th>
<th>Station</th>
<th>Country</th>
<th>River</th>
<th>Location</th>
<th>Drainage area, km²</th>
<th>Record</th>
<th>Missing years</th>
<th>Years of</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hE1833</td>
<td>Canada</td>
<td>Saquenay</td>
<td>Isle-Maligne</td>
<td>73,000</td>
<td>1913-1970</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>2</td>
<td>aB36</td>
<td>Mali</td>
<td>Niger</td>
<td>Dire</td>
<td>340,000</td>
<td>1924-1968</td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>aB72</td>
<td>Mali</td>
<td>Niger</td>
<td>Koulikoro</td>
<td>120,000</td>
<td>1907-1968</td>
<td></td>
<td>62</td>
</tr>
<tr>
<td>4</td>
<td>aE85</td>
<td>U.S.A.</td>
<td>Penobscot</td>
<td>West Enfield, Me.</td>
<td>17,090</td>
<td>1902-1967</td>
<td>1913, 1928, 1944</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>CG60</td>
<td>Finland</td>
<td>Kymijoki</td>
<td>Pernoo</td>
<td>36,535</td>
<td>1900-1968</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td>6</td>
<td>CG81</td>
<td>Finland</td>
<td>Vuoksi</td>
<td>Imatra</td>
<td>61,280</td>
<td>1847-1968</td>
<td></td>
<td>122</td>
</tr>
<tr>
<td>7</td>
<td>BF42</td>
<td>Poland</td>
<td>Oder</td>
<td>Gozdowice</td>
<td>109,365</td>
<td>1901-1968</td>
<td>1945</td>
<td>67</td>
</tr>
<tr>
<td>8</td>
<td>BF28</td>
<td>Sweden</td>
<td>Vanerngota</td>
<td>Vanesborg</td>
<td>46,830</td>
<td>1807-1968</td>
<td></td>
<td>162</td>
</tr>
<tr>
<td>9</td>
<td>DF09</td>
<td>USSR</td>
<td>Neva</td>
<td>Novosaratovka</td>
<td>281,000</td>
<td>1859-1969</td>
<td>1942</td>
<td>90</td>
</tr>
<tr>
<td>10</td>
<td>JE9955</td>
<td>Canada</td>
<td>Assiniboine</td>
<td>Brandon</td>
<td>92,000</td>
<td>1902-1970</td>
<td></td>
<td>65</td>
</tr>
<tr>
<td>11</td>
<td>JE791</td>
<td>Canada</td>
<td>Red</td>
<td>Emerson</td>
<td>104,000</td>
<td>1913-1970</td>
<td></td>
<td>58</td>
</tr>
</tbody>
</table>
Table 2. Maximum flood height $b$, scale factor $sf$, and parameters estimated from equation (20).

<table>
<thead>
<tr>
<th>No.</th>
<th>$b$</th>
<th>$sf$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25000</td>
<td>22630</td>
<td>16.33</td>
<td>8.67</td>
<td>4.26</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>3000</td>
<td>1053</td>
<td>3.53</td>
<td>719.68</td>
<td>3.74</td>
<td>0.52</td>
</tr>
<tr>
<td>3</td>
<td>21000</td>
<td>17354</td>
<td>14.05</td>
<td>14.05</td>
<td>0.14</td>
<td>6.06</td>
</tr>
<tr>
<td>4</td>
<td>18000</td>
<td>17179</td>
<td>0.91</td>
<td>27.57</td>
<td>0.005</td>
<td>2.92</td>
</tr>
<tr>
<td>5</td>
<td>700</td>
<td>562</td>
<td>2.30</td>
<td>2.30</td>
<td>2.34</td>
<td>2.06</td>
</tr>
<tr>
<td>6</td>
<td>2500</td>
<td>2167</td>
<td>14.55</td>
<td>2898.16</td>
<td>9.82</td>
<td>4.90</td>
</tr>
<tr>
<td>7</td>
<td>6000</td>
<td>5293</td>
<td>22.89</td>
<td>6.69</td>
<td>2.06</td>
<td>1.40</td>
</tr>
<tr>
<td>8</td>
<td>1300</td>
<td>1047</td>
<td>6.55</td>
<td>6.59</td>
<td>0.25</td>
<td>11.32</td>
</tr>
<tr>
<td>9</td>
<td>6000</td>
<td>4000</td>
<td>4.89</td>
<td>6.46</td>
<td>0.02</td>
<td>8.56</td>
</tr>
<tr>
<td>10</td>
<td>670</td>
<td>347.9</td>
<td>8.98</td>
<td>1.03</td>
<td>0.013</td>
<td>0.14</td>
</tr>
<tr>
<td>11</td>
<td>3100</td>
<td>1200</td>
<td>9.31</td>
<td>0.87</td>
<td>5.6 E-4</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Table 3. Parameter estimated of equation (8) for each station.

<table>
<thead>
<tr>
<th>No.</th>
<th>a</th>
<th>a'</th>
<th>c</th>
<th>c'</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.33</td>
<td>8.67</td>
<td>20706.33</td>
<td>49696.00</td>
</tr>
<tr>
<td>2</td>
<td>3.53</td>
<td>719.68</td>
<td>724.52</td>
<td>1053.96</td>
</tr>
<tr>
<td>3</td>
<td>14.05</td>
<td>14.05</td>
<td>19932.25</td>
<td>15265.06</td>
</tr>
<tr>
<td>4</td>
<td>0.91</td>
<td>27.57</td>
<td>5605113.20</td>
<td>16523.166</td>
</tr>
<tr>
<td>5</td>
<td>2.30</td>
<td>2.30</td>
<td>388.32</td>
<td>410.86</td>
</tr>
<tr>
<td>6</td>
<td>14.55</td>
<td>2898.16</td>
<td>1852.06</td>
<td>2165.81</td>
</tr>
<tr>
<td>7</td>
<td>22.89</td>
<td>6.69</td>
<td>5128.49</td>
<td>5033.41</td>
</tr>
<tr>
<td>8</td>
<td>6.55</td>
<td>6.59</td>
<td>1297.56</td>
<td>724.48</td>
</tr>
<tr>
<td>9</td>
<td>4.89</td>
<td>6.46</td>
<td>8904.89</td>
<td>2869.71</td>
</tr>
<tr>
<td>10</td>
<td>8.98</td>
<td>1.03</td>
<td>563.86</td>
<td>2303.15</td>
</tr>
<tr>
<td>11</td>
<td>9.31</td>
<td>0.87</td>
<td>2683.34</td>
<td>49293.56</td>
</tr>
</tbody>
</table>
Table 4. Computed flood discharges $P(T)$ for selected return periods.

<table>
<thead>
<tr>
<th>No.</th>
<th>T in years</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>20</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>4754</td>
<td>6112</td>
<td>6961</td>
<td>7740</td>
<td>8699</td>
<td>9382</td>
</tr>
<tr>
<td>2</td>
<td>2347</td>
<td>2526</td>
<td>2617</td>
<td>2688</td>
<td>2760</td>
<td>2803</td>
</tr>
<tr>
<td>3</td>
<td>6153</td>
<td>7303</td>
<td>8015</td>
<td>8664</td>
<td>9455</td>
<td>10015</td>
</tr>
<tr>
<td>4</td>
<td>1699</td>
<td>2364</td>
<td>2796</td>
<td>3212</td>
<td>3779</td>
<td>4278</td>
</tr>
<tr>
<td>5</td>
<td>448</td>
<td>546</td>
<td>589</td>
<td>619</td>
<td>646</td>
<td>660</td>
</tr>
<tr>
<td>6</td>
<td>694</td>
<td>829</td>
<td>913</td>
<td>990</td>
<td>1084</td>
<td>1150</td>
</tr>
<tr>
<td>7</td>
<td>1351</td>
<td>1988</td>
<td>2407</td>
<td>2772</td>
<td>3192</td>
<td>3470</td>
</tr>
<tr>
<td>8</td>
<td>617</td>
<td>725</td>
<td>787</td>
<td>840</td>
<td>901</td>
<td>941</td>
</tr>
<tr>
<td>9</td>
<td>3290</td>
<td>3727</td>
<td>3976</td>
<td>4190</td>
<td>4433</td>
<td>4594</td>
</tr>
<tr>
<td>10</td>
<td>151</td>
<td>262</td>
<td>413</td>
<td>542</td>
<td>618</td>
<td>644</td>
</tr>
<tr>
<td>11</td>
<td>552</td>
<td>902</td>
<td>1174</td>
<td>1622</td>
<td>2546</td>
<td>2852</td>
</tr>
</tbody>
</table>
Table 5.1. Data values $D(T)$ as interpolated between adjacent observations by the Hazen method.

<table>
<thead>
<tr>
<th>No.</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4655</td>
<td>6111</td>
<td>6761</td>
<td>7714</td>
<td>9128</td>
<td>9244</td>
</tr>
<tr>
<td>2</td>
<td>2335</td>
<td>2561</td>
<td>2640</td>
<td>2661</td>
<td>2677</td>
<td>*</td>
</tr>
<tr>
<td>3</td>
<td>6250</td>
<td>7041</td>
<td>7649</td>
<td>8964</td>
<td>9552</td>
<td>9676</td>
</tr>
<tr>
<td>4</td>
<td>1738</td>
<td>2339</td>
<td>2650</td>
<td>3081</td>
<td>3777</td>
<td>4251</td>
</tr>
<tr>
<td>5</td>
<td>454</td>
<td>545</td>
<td>577</td>
<td>614</td>
<td>646</td>
<td>655</td>
</tr>
<tr>
<td>6</td>
<td>703</td>
<td>794</td>
<td>880</td>
<td>931</td>
<td>1138</td>
<td>1153</td>
</tr>
<tr>
<td>7</td>
<td>1350</td>
<td>1867</td>
<td>2412</td>
<td>2700</td>
<td>3401</td>
<td>3655</td>
</tr>
<tr>
<td>8</td>
<td>627</td>
<td>726</td>
<td>773</td>
<td>806</td>
<td>927</td>
<td>937</td>
</tr>
<tr>
<td>9</td>
<td>3300</td>
<td>3750</td>
<td>4000</td>
<td>4100</td>
<td>4500</td>
<td>4540</td>
</tr>
<tr>
<td>10</td>
<td>154</td>
<td>251</td>
<td>422</td>
<td>498</td>
<td>613</td>
<td>644</td>
</tr>
<tr>
<td>11</td>
<td>540</td>
<td>835</td>
<td>1253</td>
<td>1518</td>
<td>2149</td>
<td>2607</td>
</tr>
</tbody>
</table>

*Beyond the range of the data.
Table 5.2. Data values $D(T)$ as interpolated between adjacent observations by the Chegodayev method.

<table>
<thead>
<tr>
<th>No.</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4655</td>
<td>6125</td>
<td>6766</td>
<td>7811</td>
<td>9166</td>
<td>*</td>
</tr>
<tr>
<td>2</td>
<td>2335</td>
<td>2562</td>
<td>2641</td>
<td>2664</td>
<td>2677</td>
<td>*</td>
</tr>
<tr>
<td>3</td>
<td>6250</td>
<td>7066</td>
<td>7670</td>
<td>9064</td>
<td>9590</td>
<td>*</td>
</tr>
<tr>
<td>4</td>
<td>1738</td>
<td>2342</td>
<td>2659</td>
<td>3124</td>
<td>3929</td>
<td>*</td>
</tr>
<tr>
<td>5</td>
<td>454</td>
<td>545</td>
<td>578</td>
<td>614</td>
<td>648</td>
<td>*</td>
</tr>
<tr>
<td>6</td>
<td>703</td>
<td>794</td>
<td>881</td>
<td>933</td>
<td>1139</td>
<td>1157</td>
</tr>
<tr>
<td>7</td>
<td>1350</td>
<td>1875</td>
<td>2418</td>
<td>2759</td>
<td>3474</td>
<td>*</td>
</tr>
<tr>
<td>8</td>
<td>627</td>
<td>726</td>
<td>773</td>
<td>809</td>
<td>927</td>
<td>945</td>
</tr>
<tr>
<td>9</td>
<td>3300</td>
<td>3762</td>
<td>4000</td>
<td>4118</td>
<td>4500</td>
<td>4560</td>
</tr>
<tr>
<td>10</td>
<td>154</td>
<td>252</td>
<td>423</td>
<td>509</td>
<td>622</td>
<td>*</td>
</tr>
<tr>
<td>11</td>
<td>540</td>
<td>836</td>
<td>1283</td>
<td>1532</td>
<td>2300</td>
<td>*</td>
</tr>
</tbody>
</table>

*Beyond the range of the data.
Table 5.3. Data values $D(T)$ as interpolated between adjacent observations by the Weibull method.

<table>
<thead>
<tr>
<th>No.</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4655</td>
<td>6170</td>
<td>6775</td>
<td>7987</td>
<td>9224</td>
<td>*</td>
</tr>
<tr>
<td>2</td>
<td>2335</td>
<td>2563</td>
<td>2643</td>
<td>2670</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>3</td>
<td>6250</td>
<td>7103</td>
<td>7701</td>
<td>9216</td>
<td>9648</td>
<td>*</td>
</tr>
<tr>
<td>4</td>
<td>1738</td>
<td>2346</td>
<td>2673</td>
<td>3188</td>
<td>4156</td>
<td>*</td>
</tr>
<tr>
<td>5</td>
<td>454</td>
<td>546</td>
<td>579</td>
<td>615</td>
<td>652</td>
<td>*</td>
</tr>
<tr>
<td>6</td>
<td>703</td>
<td>794</td>
<td>881</td>
<td>935</td>
<td>1142</td>
<td>1164</td>
</tr>
<tr>
<td>7</td>
<td>1350</td>
<td>1888</td>
<td>2426</td>
<td>2848</td>
<td>3583</td>
<td>*</td>
</tr>
<tr>
<td>8</td>
<td>624</td>
<td>727</td>
<td>773</td>
<td>814</td>
<td>927</td>
<td>973</td>
</tr>
<tr>
<td>9</td>
<td>3300</td>
<td>3780</td>
<td>4000</td>
<td>4145</td>
<td>4500</td>
<td>4589</td>
</tr>
<tr>
<td>10</td>
<td>154</td>
<td>255</td>
<td>425</td>
<td>524</td>
<td>636</td>
<td>*</td>
</tr>
<tr>
<td>11</td>
<td>540</td>
<td>841</td>
<td>1310</td>
<td>1567</td>
<td>2528</td>
<td>*</td>
</tr>
</tbody>
</table>

*Beyond the range of the data.

<table>
<thead>
<tr>
<th>No.</th>
<th>Mean of the absolute deviations</th>
<th>Mean of the quadratic deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hazen</td>
<td>Chegodayev</td>
</tr>
<tr>
<td>1</td>
<td>1.9</td>
<td>2.2</td>
</tr>
<tr>
<td>2</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>3</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>4</td>
<td>2.3</td>
<td>3.0</td>
</tr>
<tr>
<td>5</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>6</td>
<td>3.4</td>
<td>3.5</td>
</tr>
<tr>
<td>7</td>
<td>3.4</td>
<td>3.0</td>
</tr>
<tr>
<td>8</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>9</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>10</td>
<td>3.1</td>
<td>3.1</td>
</tr>
<tr>
<td>11</td>
<td>8.6</td>
<td>7.1</td>
</tr>
</tbody>
</table>
Realizing that the rivers examined in their study are not the same as those considered here, but just comparing the magnitude of the deviations, the new distribution (equation (8)) evaluated in this thesis provides at least as good if not a better fit to rivers with nonhomogeneous sources of runoff than the Pearson and log Pearson Type III distributions do to rivers having a homogeneous source (see Table 6).

A graphical technique

In order to facilitate the analysis, interpretation, and discussion of the results, a simple graphical technique proves invaluable.

Consider a river where the random variable "daily river height" is a mixture random variable with two nonhomogeneous components. These components could conceivably be two sources of runoff—precipitation and spring melt, or perhaps one component produces runoff whenever either source is acting alone, since precipitation and snow melt may be homogeneous sources (not significantly different), and the second nonhomogeneous component is some combination of the two sources acting simultaneously to produce significantly different flood heights. Another possibility, which the author feels is most likely, is that precipitation and spring melt are always homogeneous sources whether present in combination or individually, and a nonhomogeneous component appears, forming a mixture random variable, only during some rare event such as a hurricane. Whatever the case, the random variable "yearly maximum river height" may or may not be a mixture.

Observe this river for \( n \) years, obtaining the maximum annual flow for each year. Order these observations from low to high, producing
the order statistics \( h(i), i = 1, 2, \ldots, n \). Making a large (but useful) simplification, assume that the largest maximum yearly floods were caused solely by one component of the mixture and that the smallest maximum yearly floods were caused solely by the other. The intermediate floods were caused by some interaction of the components, or some separating point occurs where all smaller floods are caused by the other component.

Examine equation (8) again in detail and concentrate on the largest maximum yearly floods only. Since they were caused by solely one component of the mixture, the other component is negligible and disappears from the distribution function of the "largest" maximum yearly floods. Without loss of generality, suppose that the \( a', c' \) component vanishes. Then equation (8) becomes

\[
F(h) = \begin{cases} 
\exp \left[ - \left( \frac{b-h}{c} \right)^a \right] & h \leq b \\
1 & h > b 
\end{cases} \tag{25}
\]

For large \( n \), \( E \left[ F(h(i)) \right] \approx f(h(i)) \) where \( \approx \) means approximately equal. From Chapter III

\[
E \left[ F(h(i)) \right] = \frac{i}{n+1} \tag{9-1}
\]

Let

\[
X_i = \ln(b - h(i)) \tag{26}
\]

\[
Y_i = \ln(-\ln E[F(h(i)]) = \ln(-\ln \left( \frac{i}{n+1} \right)) \tag{27}
\]
Replace $E F(h(i))$ with $F(h(i))$ on the right-hand side of equation (27). Then

$$Y_i = \ln(-\ln(F(h(i))))$$

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

$$= a \ln(b - h(i)) - a \ln(c)$$

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

This implies that

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

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

Therefore, $Y_i$ is a linear function of $X_i$ where $a$ is the slope and $-a \ln(c)$ is the intercept. Hence, by plotting $X_i$ 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 $-a \ln(c)$. Similarly for the smallest maximum yearly floods, plotting $X_i$ against $Y_i$ should form a straight line whose slope approximates $a'$ with an intercept of approximately $-a' \ln(c')$. The intermediate maximum yearly floods in such a graph should follow one of the straight lines if caused by a single component or some gradual shift from one line to the other if there is some interaction between the components.

The $X_i$ and $Y_i$ of equation (28) are plotted for each of the eleven stations in this study (See Figures 2 through 12) using the command
PLOT from the MINITAB II Reference Manual (1978). The $X_1$ are along the $C1$ axis while the $Y_1$ are along the $C2$ axis in each plot. The slopes and intercepts as determined graphically very closely match $a$, $a'$, $c$, and $c'$ as determined by the parameter estimation technique of Chapter III. It should be noted that the starting values for $\alpha_1$ and $\alpha_2$ in program FLOOD are determined using equation (28) on the first four and the last four order statistics respectively. After viewing the graphs, a better procedure would be to use the first thirty and the last four order statistics.

The Saquenay River (Figure 2) manifests almost a straight line plot and may have nearly homogeneous sources, although the two largest floods could be from another source. The Niger River, location Dire (Figure 3) and location Koulikoro (Figure 4), exhibits two sharply different components with little or no interaction. The Penobscot River (Figure 5) also has two well defined components. The Kymijoki River (Figure 6) appears to have homogeneous sources with close to a straight line plot. Its estimated parameters indicate likewise—$a = 2.03$ and $a' = 2.03$ while $c = 388.32$ and $c' = 410.86$—very close to identical components. The Vuoksi River (Figure 7) has two nonhomogeneous sources with no interaction. The Oder River (Figure 8) has a lot of interaction between its two components. The Varangota River (Figure 9) and the Neva River (Figure 10) both have well defined components with little interaction. The Assiniboine River (Figure 11) and the Red River (Figure 12) have interaction between two different components. The non-homogeneity of source is very clear for most of these rivers.
The goodness-of-fit for the first ten stations is excellent. The fit for the Red River is not as good, although it does fit the Weibull method observed flood heights well. Perhaps the largest maximum yearly flood which is clearly an outlier (see Figure 12), much larger than any other on record, is the only observation from a particular source population. It is impossible to achieve a good estimate of the parameters of a population with only one observation. The outlier also may just have been an extremely rare event and in that case the river needs to be observed for a longer period of time for more data collection.
Figure 2. Station No. 1--Saquenay River.
Figure 3. Station No. 2—Niger River, Dire.

\[ X_i = (b - h(i))^\frac{n}{b} \]
$Y_i = \ln\left(\frac{\frac{t}{i}}{m+1}\right)$

$X_1 = (b-h_i)$

Figure 4. Station No. 3--Niger River, Koulikoro.
Figure 5. Station No. 4—Penobscot River.

\[ X_i = \left( \frac{b - \ln(i)}{i} \right) \]
Figure 6. Station No. 5--Kymijoki River.
$Y_i = \ln \left( \frac{(i)}{n+1} \right)$

$X_i = (b-h(i))$

Figure 7. Station No. 6--Vuoksi River.
Figure 8. Station No. 7--Oder River.

\[ Y_i = \ln \left( -\ln \left( \frac{X_i}{X_0} \right) \right) \]

\[ X_i = \left( b - h_i \right) \]
Figure 9. Station No. 8--Vanerngota River.
$Y_i = \ln(\frac{t}{m+1})$

$x_1 = (b-h)_i$

Figure 10. Station No. 9—Neva River.
Figure 11. Station No. --Assiniboine River.

\[ Y_i = \ln \left( 1 - \frac{1}{n+1} \right) \]

\[ X_i = (b - h_{(i)}) \]
Figure 12. Station No. 11--Red River.

$X_i = (b-h(i))$

$Y_i = \ln\left(\frac{i}{n+1}\right)$
CHAPTER V
CONCLUSION

There is extensive literature describing distribution functions which provide the "best" fit for the random variable "maximum yearly river height" 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. This is an area which requires more research.

Rivers which exhibit nonhomogeneous sources of runoff are examined in this study. The theoretical distribution function in equation (8) has the correct right tail characteristics for predicting maximum annual flow of such rivers. The modified least squares parameter estimation technique gives excellent fit to the eleven data sets considered herein.

Future research in this area should include testing the goodness-of-fit of other parameter estimation methods and applying the distribution function to many more sets of data from rivers with nonhomogeneous sources. If such studies result in positive findings, as this one has, the extreme value type distribution function (equation (8)) should be adopted as a base method for flood frequency analysis with non-homogeneity in runoff sources.


APPENDICES
Appendix A. Program ORDER

C THIS PROGRAM READS THE YEARLY MAXIMUM FLOOD DATA OF A RIVER,
C ORDERS THIS DATA INTO ASCENDING ORDER, AND THEN STORES THE
C DATA ON DISK FOR FUTURE ANALYSIS. NECESSARY INPUT IS THE
C NUMBER OF YEARS OF THE RECORD AND THE ACTUAL DATA. M IS THE
C NUMBER OF YEARS OF DATA RECORD. X IS AN ARRAY FOR THE DATA
C ITSELF.
C FILE 1(KIND=DISK,TITLE="SAGUENAY/DATA")

C DIMENSION X(200)
C M, THE NUMBER OF YEARS OF DATA IS READ.
C READ(5,/)M
C THE DATA IS READ FREE FORMAT AND STORED IN ARRAY X.
C READ(5,/) (X(I),I=1,M)
C THE DATA IS ORDERED IN ASCENDING ORDER, THUS X(1) IS THE
C SMALLEST AND X(M) IS THE LARGEST MAXIMUM YEARLY FLOOD.
C NESTED=M
C L=NESTED-1
C DO 20 J=1,L
C NESTED=NESTED-1
C DO 20 I=1,NESTED
C IF (X(I)>X(I+1)) 20,20,30
C SAVE=X(I)
C X(I)=X(I+1)
C X(I+1)=SAVE
C 20 CONTINUE
C WRITE(6,100)M
C 100 FORMAT(1X,' THE NUMBER OF YEARS OF RECORD=',I15,///)
C WRITE(6,200)
C 200 FORMAT(1X,' THE ORDERED MAXIMUM YEARLY FLOODS',///)
C WRITE(1,101)(X(I),I=1,M)
C WRITE(6,120)(X(I),I=1,M)
C 120 FORMAT(1X,5F10.1,///)
C 101 FORMAT(1X,F12.2)
C ORDERED DATA IS SAVED ON DISK.
C LOCK 1
C STOP
C END
Appendix B. Program INTEGRATE

```
C THIS PROGRAM CALCULATES THE EL(I), ELSQ(I), AND M(I) BY
C NUMERICAL INTEGRATION WITH THE TRAPEZOID RULE. M, THE
C NUMBER OF YEARS OF DATA, IS THE ONLY REQUIRED INPUT.
C C IS THE STEP SIZE.
FILE 2(KIND=DISK,TITLE="SAGUNAY/EL")
FILE 3(KIND=DISK,TITLE="SAGUNAY/M")

IMPLICIT DOUBLE PRECISION (A-H,O-Z)
DIMENSION EL(200),ELSQ(200),M(200)
C=0.01
C M, THE NUMBER OF YEARS OF DATA, IS READ.
READ(5,/)M
D=M
WRITE(6,110)
110 FORMAT(1X,15x,'EL(I)',25x,'ELSQ(I)',25x,'M(I)',/)
10 DO 1 I=1,M
20 EL(I)=0.
30 ELSQ(I)=0.
40 IF(I.EQ.1)GO TO 20
50 GO TO 14
11 T1A=LOG(F1)
21 T1ASQ=LOG(F1)**2.
31 T2A=F1**(I-1)
41 T3A=(1.-F1)**(M-I)
51 A=T1A*T2A*T3A
61 ASW=T1ASQ*T2A*T3A
70 GO TO 14
13 A=0.
80 ASW=0.
90 F2=C
14 T1B=LOG(F2)
10 T1BSQ=LOG(F2)**2.
11 T2B=F2**(I-1)
12 T3B=(1.-F2)**(M-I)
13 B=T1B*T2B*T3B
14 ASQ=T1BSQ*T2B*T3B
15 EL(I)=(A+B)*C/2.
16 ELSQ(I)=(ASQ+BSQ)*C/2.
20 GO TO 10
10 F1=C/4.
20 F2=C
GO TO 11
```
10  A=B
    ASQ=HSQ
    F2=F2+C
    IF(F2.GT.1)GO TO 15
    T1=LOG(F2)
    T1BSQ=(DLOG(F2))**2.
    T2=F2**4(I=1)
    T3=(I.-F2)**(M=1)
    B=T1*T2*T3
    BSQ=T1BSQ*T2*T3
    EL(I)=EL(I)+(4+B)*C/2.
    ELSQ(I)=ELSQ(I)+(ASQ+BSQ)*C/2.
    GO TO 10
15   FL(I)=EL(I)*D
    ELSQ(I)=ELSQ(I)*D
    w(I)=ELSQ(I)=EL(I)**2.
    w(I)=1./w(I)
    WRITE(8,100)EL(I),ELSQ(I),w(I)
100  FORMAT(1X,3F30.12)
    D=D*(N=I+1)
1    CONTINUE
    WRITE(2,200)(EL(I),I=1,M)
    WRITE(3,200)(w(I),I=1,M)
200  FORMAT(1X,3F40.15)
C    THE EL(I) AND W(I) ARE STORED ON DISK FOR FUTURE USE.
LOCK 2
LOCK 3
STOP
END
Appendix C. Program FLOOD

This program finds estimates for the parameters ALPHA(1), ALPHA(2), THETA(1), and THETA(2) by minimizing equation (20) of Chapter 3. Required input includes M, the number of years of the data record, BB, the maximum possible flood height, and CC, the scale factor. The ordered flood data, the EL(i), and the Z(i) are read into arrays X, EL, and W respectively (the W(i) are computed and stored in array W during execution). The minimization process is achieved with a subroutine from the IMSL (1977) library called ZSYSTM which solves the two equations in two unknowns discussed in Chapter 3. This subroutine requires an external function (F), two convergence criteria (EPS and NSIG), the number of unknowns (N), the maximum number of iterations of the external function F (ITMAX), a work area of computer storage (HA), an array for passing parameters (PAP, which is not used in this study), an error message variable (IER), and starting values for the ALPHAS, the starting values for ALPHA(1) and ALPHA(2) are computed from the ordered data. Output consists of ALPHA(1), ALPHA(2), THETA(1), THETA(2), ITMAX, and IER. The error message, IER=0 means there are no errors and minimization was completed to the accuracy specified by the convergence criteria.

For more detailed information on the subroutine ZSYSTEM, see the IMSL (1977) library.

FILE 1(KIND=DISK,TITLE="(R78073)SAGUENAY/DATA")
FILE 2(KIND=DISK,TITLE="(R78073)SAGUENAY/EL")
FILE 3(KIND=DISK,TITLE="(R78073)SAGUENAY/W")

EXTERNAL F
DIMENSION ALPHA(2),WA(20),PAP(2),XREG(200),YREG(200)
COMMON M,BB,CC,X(200),X(200),Y(200),THETA(2),Y(200)
EPS=1.0E-9
NSIG=5
N=2
ITMAX=100
IER=0
M=THE NUMBER OF YEARS OF DATA, BB=THE MAXIMUM POSSIBLE FLOOD HEIGHT, AND CC=THE SCALE FACTOR ARE READ.
READ(5,/)M
```fortran
READ(5,/BB,CC
C THE ORDERED DATA, THE FL(I), AND THE Z(I) ARE READ INTO ARRAYS
C X, EL, AND W RESPECTIVELY.
READ(1,101)(X(I),I=1,M)
101 FORMAT(1X,F12.2)
READ(2,200)(EL(I),I=1,M)
READ(3,200)(W(I),I=1,M)
200 FORMAT(1X,F4.15)
C THE W(I) ARE CALCULATED.
DO 23 I=1,M
  W(I)=W(I)/W(M)
23 CONTINUE
C STARTING VALUES ARE DETERMINED FOR ALPHA(1) AND ALPHA(2).
SUMX1=0.
SUMY1=0.
SUMXY1=0.
SUMXX1=0.
DO 15 I=1,4
  XREG(I)=ALOG(EE-X(I))
  YREG(I)=ALOG(-ALOG(I/(N+1.)))
  SUMX1=SUMX1+XREG(I)
  SUMY1=SUMY1+YREG(I)
  SUMXY1=SUMXY1+XREG(I)*YREG(I)
  SUMXX1=SUMXX1+XREG(I)**2
15 CONTINUE
XBAR=SUMX1/4.
YBAR=SUMY1/4.
ALPHA(1)=(SUMXY1-4.*XBAR*YBAR)/(SUMXX1-4.*XBAR**2)
SUMX2=0.
SUMY2=0.
SUMXY2=0.
SUMXX2=0.
DO 16 J=M-3,M
  XREG(J)=ALOG(EE-X(J))
  YREG(J)=ALOG(-ALOG(J/(N+1.)))
  SUMX2=SUMX2+XREG(J)
  SUMY2=SUMY2+YREG(J)
  SUMXY2=SUMXY2+XREG(J)*YREG(J)
  SUMXX2=SUMXX2+XREG(J)**2
16 CONTINUE
XBAR=SUMX2/4.
YBAR=SUMY2/4.
ALPHA(2)=(SUMXY2-4.*XBAR*YBAR)/(SUMXX2-4.*XBAR**2)
WRITE(6,50)M,ER,CC
```
FORMAT(1X,'THE NUMBER OF YEARS OF THE DATA RECORD=',I15,/, *1X,'THE MAXIMUM POSSIBLE FLOOD HEIGHT=',I15,/,1X, * 'THE SCALE FACTOR=',I15,/, *1X,'THE STARTING VALUES ARE',/,1X,'ALPHA(1)=' ,F15.5, *5X,'ALPHA(2)=',F15.5,///)

C ZSYSYM IS CALLED TO MINIMIZE EQUATION (20) AND OUTPUT THE
C ESTIMATED PARAMETERS.
CALL ZSYSYM(F,EPS,NSIGN,ALPHA,ITMAX,WA,PAR,IER)
WRITE(6,70)ITMAX,IER

FORMAT(1X,'NUMBER OF ITERATIONS OF EXTERNAL FUNCTION=',I5, *///,1X,'ERROR MESSAGE=',I5,///)
WRITE(6,80)

FORMAT(1X,'PARAMETER ESTIMATES ARE',//)
WRITE(6,90)ALPHA(1),ALPHA(2),THETA(1),THETA(2)

FORMAT(1X,'ALPHA(1)=' ,F20.10,5X,'ALPHA(2)=' ,F20.10)/

STOP

END

FUNCTION F(ALPHA,EK,PAR)
THIS FUNCTION EVALUATES THE TWO EQUATIONS IN TWO UNKNOWNS.
COMMON M,RA,CC,X(200),Y(200)
DIMENSION MEL(200),MLX(200),MELX(200),ALPHA(2),PAR(2)
COMMON M,RA,CC,X(200),Y(200),MEL(200),MLX(200),THETA(2),Y(200)

A3=2.*ALPHA(1)+4=2.*ALPHA(2); A5=ALPHA(1)+ALPHA(2)
Z1=0.; Z2=0.; Y1=0.; Y2=0.; Y3=0.; Y4=0.; Y5=0.; Y6=0.; Y7=0.; Y8=0.
B1=0.; B2=0.; B3=0.; B4=0.; B5=0.
CC 20 I=1,

YA1=Y(I)**ALPHA(1)
YA2=Y(I)**ALPHA(2)
YA3=Y(I)**A3
YA4=Y(I)**A4
YA5=Y(I)**A5
Z1=Z1=MEL(I)*YA1
Z2=Z2=MEL(I)*YA2
Y1=Y1+M(I)*YA3
Y2=Y2+M(I)*YA4
Y3=Y3+M(I)*YA5
Y4=Y4+M(I)*YA1
Y5=Y5+MLX(I)*YA3
Y6=Y6+MLX(I)*YA5
Y7=Y7+MLX(I)*YA2
Y8=Y8+MLX(I)*YA4
TH1 = (Z1 + Y2 + Z2 + Y3) / (Y1 + Y2 + Y3)
TH1 = ABS(TH1)
TH2 = (Z1 + Y1 + Y1) / Y3
TH2 = ABS(TH2)
ALPHA(1) = ABS(ALPHA(1))
ALPHA(2) = ABS(ALPHA(2))
GO TO (55, 56), KK
55 FY = Y4 + TH1 * Y5 + TH2 * Y6
THETA(1) = TH1
THETA(2) = TH2
RETURN
56 FY = Y7 + TH1 * Y6 + TH2 * YA
THETA(1) = TH1
THETA(2) = TH2
RETURN
END

BEGIN BINDING ZSYST OF "OUT FROM IMSL/ZSYSTM"
ZSYST
"IMSL"
UERTST
"UERTST"
PREC
"PREC"
DELTA
"DELTA"
ZERO
"ZERO"
PM1
"PM1"
PT1
"PT1"
P2
"P2"
<SEG DICT ITEM>
"<SEG DICT ITEM>
(<01, 0002>) CHANGED TO (<02, 0002>)
(<02, 0000>) CHANGED TO (<02, 0017>)
(<02, 0003>) CHANGED TO (<02, 0014>)
(<02, 0004) CHANGED TO (<02, 0019>)
(<02, 0005) CHANGED TO (<02, 001A>)
(<02, 0006) CHANGED TO (<02, 0018>)
(<02, 0007) CHANGED TO (<02, 001C>)
(<02, 0008) CHANGED TO (<02, 001D>)
(<01, 0002) CHANGED TO (<01, 0010>)
(<01, 0004) CHANGED TO (<01, 0012>)
(<01, 0005) CHANGED TO (<01, 0013>)
(<01, 0006) CHANGED TO (<01, 0014>)
(<01, 0007) CHANGED TO (<01, 0015>)
(<01, 0008) CHANGED TO (<01, 0016>)
(<01, 0009) CHANGED TO (<01, 0017>)
(<01, 000A) CHANGED TO (<01, 001B>)
END OF BINDING ZSYST

BEGIN BINDING UERTST OF "OUT FROM IMSL/UERTST"
UERTST
"UERTST"
FILED
"FILED"
ITYP
"ITYP"
TERM
"TERM"
WARF
"WARF"
WARN
"WARN"
IBIT
"IBIT"
PRINR
"PRINR"
<SEG DICT ITEM>
"<SEG DICT ITEM>
(<01, 0002) CHANGED TO (<01, 0017>)
(<02, 0003) CHANGED TO (<02, 0010>)
(<02, 0005) CHANGED TO (<02, 001F>)
(<02, 0006) CHANGED TO (<02, 0020>)
(<02, 0008) CHANGED TO (<02, 0020>)
(<02, 000A) CHANGED TO (<02, 0020>)
(<02, 0006) CHANGED TO (<02, 0020>)
(<02, 0007) CHANGED TO (<02, 0021>)
(<01, 0002) CHANGED TO (<01, 0019>)
(<01, 0003) CHANGED TO (<01, 0014>)
(<01, 000A) CHANGED TO (<01, 001B>)
END OF BINDING UERTST
Appendix D. Program PFHTS

1 C- THIS PROGRAM IS AN INTERACTIVE (TERMINAL) PROGRAM.
2 C- GIVEN A DISTRIBUTION FUNCTION F(X) OF THE FORM OF
3 C- EQUATION (8) OF CHAPTER 2 WHERE C AND C' HAVE BEEN
4 C- REPARAMETERIZED AS THETA(1) AND THETA(2) AS IN
5 C- EQUATION (20) OF CHAPTER 3, FOR ANY X (FLOOD HEIGHT)<
6 C- F(X) THE PROBABILITY THAT ANY POSSIBLE FLOOD IS
7 C- LESS THAN OR EQUAL TO X IS EVALUATED. THUS THE
8 C- SELECTED RETURN PERIODS OR RECURRANCE INTERVALS
9 C- CAN BE FOUND BY FINDING SOME X (TO THE NEAREST
10 C- INTEGER) WHICH YIELDS THE DESIRED F(X) PROBABILITY.
11 C- REQUIRED AS INPUT ARE BB--THE MAXIMUM POSSIBLE
12 C- FLOOD HEIGHT, CC--THE SCALE FACTOR, ALPHA(1), ALPHA(2),
13 C- THETA(1), AND THETA(2)--THE PARAMETERS OF THE
14 C- PARTICULAR DISTRIBUTION FUNCTION F(X), AND X--THE
15 C- FLOOD HEIGHT FOR WHICH F(X) IS DESIRED. F(X) MAY
16 C- BE FOUND FOR AS MANY X VALUES AS REQUIRED.
17 C- WHEN FINISHED SIMPLY ENTER 'END' AND A NEW
18 C- F(X) MAY BE EXAMINED OR ONE MAY LOG OFF THE
19 C- COMPUTER AS DESIRED.
20 DIMENSION ALPHA(2), THETA(2)
21 WRITE(6,160)
22 160 FORMAT(1X,"ENTER BB AND CC")
23 READ(5,/)BB,CC
24 WRITE(6,170)
25 170 FORMAT(1X,"ENTER ALPHA(1) AND ALPHA(2)",/)
26 READ(5,/)ALPHA(I),I=1,2
27 WRITE(6,180)
28 180 FORMAT(1X,"ENTER THETA(1) AND THETA(2)",/)
29 READ(5,/)THETA(I),I=1,2
30 WRITE(6,190,END=99)
31 190 FORMAT(1X,"ENTER X")
32 READ(5,/,END=99)X
33 Y=(BB-X)/CC
34 F=EXP(-THETA(1)*Y**ALPHA(1)-THETA(2)*Y**ALPHA(2))
35 WRITE(6,100)F
36 100 FORMAT(1X,"F(X)=",F20.15,/)
37 GO TO 1
38 99 STOP
39 END
Appendix E. Program EPROB

THIS PROGRAM CALCULATES THE EXPECTED PROBABILITY OF A
MAXIMUM YEARLY FLOOD BEING GREATER THAN OR EQUAL TO A GIVEN
HEIGHT USING THE OBSERVED DATA RECORD. THESE PROBABILITIES
ARE ESTIMATED USING THE THREE FORMULAE GIVEN IN CHAPTER 4 -
= HAZEN, CHEGDAYEV, AND WEIBULL. THE DESIRED RECURRANCE
INTERVALS OR RETURN PERIODS ARE FOUND BY LINEAR INTER-
PULATION BETWEEN THE TWO OBSERVED FLOOD HEIGHTS WHOSE
EXPECTED PROBABILITIES BRACKET THE DESIRED PROBABILITY.
REQUIRED INPUT IS THE NUMBER OF YEARS OF THE DATA RECORD
AND THE DATA ITSELF. M IS THE NUMBER OF YEARS OF DATA AND
X IS AN ARRAY FOR THE FLOOD RECORD.

FILE 1(KIND=DISK,TITLE="SAGUENAY/DATA")

DIMENSION X(200),HAZP(200),CHEGPR(200),WEIBPR(200)

M, THE NUMBER OF YEARS OF DATA, IS READ.
READ(5,/)M

THE FLOOD DATA IS READ INTO ARRAY X FROM DISK.
READ(1,101)(X(I),I=1,M)
101 FORMAT(1X,F12.2)

EXPECTED PROBABILITIES ARE CALCULATED, HAZP, CHEGPR,
AND WEIBPR ARE ARRAYS FOR THE PROBABILITIES FOUND USING THE
HAZEN, CHEGDAYEV, AND WEIBULL FORMULAE RESPECTIVELY.
105 CONTINUE
WRITE(6,200)

200 FORMAT(1X,'EXPECTED PROBABILITIES',/1/)  
WRITE(6,99)
99 FORMAT(1X,5X,'DATA',32X,'HAZEN',25X,'CHEGDAYEV',19X,'WEIBULL')
WRITE(6,100) (X(I),HAZP(I),CHEGPR(I),WEIBPR(I),I=IEST,M)
100 FORMAT(1X,F12.2,18X,3F30.16)
STOP
END
Appendix F. Program DEVIATION


ZAEN, CHEGODAYEV, AND WEIBULL ARE TREATED AS DIFFERENT METHODS. REQUIRED INPUT INCLUDES TH, TC, AND TW, THE NUMBER OF SELECTED RECURRENCE INTERVALS FOR THE ZAEN, CHEGODAYEV, AND WEIBULL METHODS RESPECTIVELY. THE PREDICTED FLOOD HEIGHTS OF THE ESTIMATED DISTRIBUTION FUNCTION ARE OBTAINED FROM THE INTERACTIVE PROGRAM PLOTS AND ARE INPUT AS ARRAY PF.

THE EXPECTED FLOOD HEIGHTS FOR THE SELECTED RETURN PERIODS ARE FOUND USING THE PROGRAM PLOT AND LINEAR INTERPOLATION AND ARE INPUT AS ARRAYS EFH, EFC, AND EFW FOR THE ZAEN, CHEGODAYEV AND WEIBULL METHODS RESPECTIVELY.

DIMENSION EFH(10), EFC(10), EFW(10), PF(10), H(10), C(10), W(10), PR(10)
PR(1)=.5; PR(2)=.8; PR(3)=.9; PR(4)=.95; PR(5)=.98; PR(6)=.99
TH, TC, AND TW, THE NUMBER OF RETURN PERIODS, ARE READ.
READ(5,/) TH, TC, TW

THE PREDICTED FLOOD HEIGHTS ARE READ INTO ARRAY PF.
READ(5,/) (PF(I), I=1,6)

THE EXPECTED FLOOD HEIGHTS ARE READ INTO ARRAYS EFH, EFC, AND EFW RESPECTIVELY.
READ(5,/) (EFH(I), I=1, TH)
READ(5,/) (EFC(I), I=1, TC)
READ(5,/) (EFW(I), I=1, TW)
WRITE(6,10)

10 FORMAT(1X,' PROBABILITY', 2X,' PREDICTED HEIGHT', 5X,' ZAEN', 5X,' CHEGODAYEV', 6X,' WEIBULL')
WRITE(6,20) (PR(I), PF(I), EFH(I), EFC(I), EFW(I), I=1,6)
20 FORMAT(1X,' F10.2,R8.2,A8', R8.2,F8.2,F8.2)

MEAN ABSOLUTE AND MEAN QUADRATIC DEVIATIONS ARE COMPUTED FOR EACH METHOD EMPLOYING THREE DO LOOPS USING EQUATIONS (23) AND (24) OF CHAPTER 4. THE SMALLER THE DEVIATIONS THE BETTER THE FIT.

DIFFH=0.; DIFFC=0.; DIFFW=0.; DHS=0.; DCS=0.; DSW=0.
DO 1 I=1, TH
   H(I)=ABS((PF(I)-EFH(I)) / EFH(I)*100.)
   DIFFH=DIFFH+H(I)
   DHS=DHS+H(I)*H(I)
1 CONTINUE
DO 2 I=1, TC
   C(I)=ABS((PF(I)-EFC(I)) / EFC(I)*100.)
   DIFFC=DIFFC+C(I)
   DCS=DCS+C(I)*C(I)
2 CONTINUE
DO 3 I=1,TH
\( w(I) = \text{ABS}(PF(I) - EFw(I)) / EFw(I) * 100.0 \)
DIFFN = DIFFN + W(I)
DWS = DWS + W(I) * K(I)
3 CONTINUE
ADIFFH = DIFFH / TH
ADIFFC = DIFFC / TC
ADIFFN = DIFFN / TH
ADHS = DHS / TH
ADCS = UCS / TC
ADWS = DWS / TH
WRITE (6, 100) ADIFFH, ADIFFC, ADIFFN
100 FORMAT (///, ' MEAN OF THE ABSOLUTE RELATIVE DEVIATIONS', ///, '10X, '
* ' HAZEN', F20.2, //, ' CHEGODAYEV', F20.2, //, '8X, ' WEIBULL',
* F20.2, ///)
WRITE (6, 200) ADHS, ADCS, ADWS
200 FORMAT (8X, ' MEAN QUADRATIC DEVIATION', ///, '10X, ' HAZEN',
STOP
END
VITA

David Ray Olson

Candidate for the Degree of

Master of Science

Thesis: Estimation of Floods when Runoff Originated from Nonhomogeneous Sources

Major Field: Applied Statistics

Biographical Information:

Personal Data: Born in Logan, Utah, May 7, 1955, son of Gerald Ray Olson and Effie Carolyn Brady.

Education: Graduated from Sky View High School, Smithfield, Utah in 1973; received the Bachelor of Science Degree in Mathematics from Utah State University in 1975; completed requirements for the Degree of Master of Science in 1979.

Professional Experience: Fall quarter 1974 and Winter and Spring quarters 1978, Instructor, Department of Mathematics, Utah State University; school year 1978-79, Statistical Consultant, Department of Applied Statistics and Computer Science, Utah State University; Summer 1979, data manipulation and statistical analysis for Dr. Walter R. Borg, Department of Psychology, Utah State University.