NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Nonparametric Flood Frequency Analysis with Historic Information and Hydroclimatically Defined Mixed Distributions

by

Younes Alila

A thesis
Submitted to the School of Graduate Studies
in partial fulfillment of the requirements for the
degree of Master of Applied Sciences
in
Civil Engineering

at the
University of Ottawa
December, 1987

ISBN 0-315-53762-0

Abstract

In parametric flood frequency analysis extensive research efforts have been devoted to the model selection and parameters estimation problems for annual flood series. Furthermore, most of the parametric distributions currently in use are unimodal and consequently consider annual peakflows to be drawn from a single population. Yet, diverse flood generating mechanisms give rise to different flood characteristics. Very little attention has been devoted to the treatment of mixed distributions in parametric analysis. The nonparametric kernel density estimator is proposed in this study as an alternative to the existing methods. In nonparametric approach to flood frequency analysis there is no need for the apriori distribution choice and the parameters estimation problem is much less complex. In addition, the multimodal nature of nonparametric densities allows for differences in flood characteristics and hence making this approach more suitable for analysis of annual flood series of mixed distributions.

It is recognized, however, that the fixed kernel method has difficulties in fitting long - tailed distributed flood data. In an attempt to alleviate this problem three other alternatives, namely, transformed, variable and adaptive kernels, are investigated and introduced for applications in flood frequency analysis. A comparison of the asymptotic behaviour of the four kernel approaches shows that no real preference of one method over the others can be made based on standard error of estimates. The fixed kernel, on the other hand, exhibits the best results in terms of overall accuracy of flood estimates.

Peakflows of an annual flood series of the Turtle river near Mine Centre in Ontario are investigated numerically and proved to be caused by two different flood mechanisms (snowmelt and rainfall). The nonparametric approach was shown to be suitable for treatment of such hydroclimatically defined mixed distributions.

The effect of sample length variation on the accuracy of flood estimates is also assessed. It was found that a 25 years of flood data is required to keep to a minimum the bias and variance of quantile estimates.

Finally, a methodology of incorporating historical information is adopted to nonparametric flood frequency analysis for the first time in the literature. The inclusion of historical data, where only threshold exceedances are available, decreased the variability of flood estimates around their population mean values and hence increased their precision.

Acknowledgements

Grateful thanks are given to Dr. K. Adamowski who gave up countless hours in the aid and supervision of this research and whose advice and positive criticisms proved most helpful. Thanks are also expressed to Dr. W. Feluch, Ottawa visiting professor in statistical hydrology (1986 - 1987), for his valuable advice in the earlier stages of this study and for providing computer programs for the analysis.

I would like to express also my gratitude to P. Pilon and D. Harvey of Environment Canada for their enthusiasm and understanding. Further grateful acknowledgement is extended to Environment Canada (Inland Waters Branch) for providing the necessary hydrologic data required for the analysis.

Finally, sincere appreciation is expressed to the Tunisian Government and the National Science and Engineering Research Council of Canada (Grant # 7467) for providing financial support to my education program at the University of Ottawa.

Contents

List of Figures

List of Tables

Notations

a_k	=	parameter related to the variable kernel estimator
$B^*(\cdot)$	=	parameter related to the transformed kernel estimator
$d_{i,k}$	=	distance from flood observation x_i to its k-th nearest neighbour
$\bar{d}_k$	=	mean of the k-th nearest neighbour distances
$f(\cdot)$	=	true density function of flood data
$\hat{f}(\cdot)$	=	estimate of the density function of flood data
$\tilde{f}(\cdot)$	=	pilot estimate of the adaptive kernel
$F(\cdot)$	=	true probability distribution function
$\hat{F}(\cdot)$	=	estimate of the probability distribution function
$g(\cdot)$	=	fixed kernel density of the transformed data
$g'(\cdot)$	=	first derivative of density estimate $g(\cdot)$
h	=	smoothing factor of the fixed kernel method
H	=	historic time span (year)
k_*	=	number of floods exceeding Q_c
k_c	=	number of floods whose magnitudes are less than Q_c
$K(\cdot)$	=	kernel function
N	=	Number of repititions in a simulation experiment
n	=	sample length of a flood annual series
$\hat{p}(\cdot)$	=	exceedance probability
Q_T	=	T-year flood estimate (cms)
Q_c	=	censoring threshold (cms)
$T(\cdot)$	=	transformation
T	=	return period (year)
t	=	standard normal deviate
X	=	random variable of flood data in real space
x_i	=	flood observation (cms)
$W(\cdot)$	=	integral of the kernel function $K(\cdot)$

μ	=	population mean in real space
μ_y	=	population mean in log space
μ'_r	=	r-th absolute moment of the distribution
μ'_1	=	first absolute moment of the distribution (mean)
μ_r	=	r-th central moment of the distribution
σ	=	population standard deviation in real space
σ_y	=	population standard deviation in log space
$\sigma(d_k)$	=	standard deviation of the k-th nearest neighbour distances
λ_i	=	local smoothing factor at flood x_i
α	=	sensitivity parameter of the adaptive kernel
$\Gamma(\cdot)$	=	Gamma function

Abbreviations

AFS	=	annual flood series
AK	=	adaptive kernel method
APC	=	antecedent precipitation concept
ADAKER	=	adaptive kernel program
CFA1	=	consolidated frequency analysis package version 1
CITKER	=	circular transformed kernel program
CK	=	coefficient of kurtosis
CS	=	coefficient of skewness
CTK	=	circular transformed kernel method
CV	=	coefficient of variance
eff.	=	efficiency of kernel function $K(\cdot)$
FK	=	fixed kernel method
IMSE	=	integrated mean square error
L	=	likelihood function
LP (3)	=	Logpearson type (3) distribution
MSE	=	mean square error
RMSE	=	root mean square error
SD	=	standard deviation
SEE	=	standard error of flood estimate
SEE_0	=	standard error of flood estimate (historical analysis)
Sup	=	supremum
TCEV	=	two component extreme value distribution
TTK	=	triangular transformed kernel method
TRTKER	=	triangular transformed kernel program
USWRC	=	United States water resources council
$Var.$	=	variance
VARKER	=	variable kernel program
VK	=	variable kernel method
WSC	=	water survey of Canada

Chapter 1

INTRODUCTION

1.1 Motivation

Floods are considered one of the most destructive natural hazards. Floods cause each year many deaths and displace thousands from their homes (Thomas et al, 1985). In addition, from an economic standpoint several billion dollars are spent yearly in the design of civil engineering structures. These enormous expenditures have made flood frequency analysis an important and urgent field of research. In the first three decades of the twentieth century, hydrologists relied mainly on experience and empirical methods. According to Chow (1964), flood frequency analysis was performed graphically by means of probability plots and envelope curves. Hazen (1919) was the first to introduce probability papers. From the early thirty's, together

with the general sciences development, hydrologists began theorizing flood frequency analysis making use of more sophisticated statistical techniques. Since then, and especially with the advent of computers, increasingly complex hydrologic models have been developed including the parametric flood frequency analysis models.

The estimation of flood magnitudes corresponding to specified return periods in the parametric analysis is based on the assumption that the flood data sample is drawn from a population of known distribution. A variety of probability distributions have been used in flood frequency analysis. The selected distribution for an annual flood series should be able to abstract the maximum information from the data. Most of the distributions exhibit similar fits at the central part. At the tails, however, they all behave differently. Therefore, several distributions can be fitted successfully to a flood data sample leading to different quantile estimates. Goodness of fit tests, such as the Chi-Square test often used to select the distribution which best fit the data, are deemed necessary but not sufficient. Many debates have been written on the subject with no general agreements as to the best distributional choice. Since the *true* frequency distribution is not known and a selection has to be made, this constitutes the major source of error in parametric flood frequency analysis.

The second source of error is due to the estimation of the distribution parameters. Statistical parameters, such as shape, scale and location, are estimated from sample data. Often, flood data are not long enough to yield

accurate parameter estimates. Another difficulty in parametric analysis is the treatment of mixed distributions. Indeed, most of the frequency distributions used are unimodal. That is, floods in a data sample are assumed to be drawn from one and the same population. Yet, floodings in some watersheds are generated by different types of mechanisms. In Canada, for instance, floods may be caused by snowmelt, rainfall or a combination of both. Moreover, in other climatic conditions, floods can be generated by different types of storms. Recently, several methods have been suggested for treatment of annual flood series of mixed distributions. These techniques, however, are not freed from the apriori distributional assumption and added statistical complexity to the parametric analysis.

The sample length requirement is another critical issue related to frequency analysis. The fitted distribution is usually used to extrapolate from the recorded events to the design event of interest. Hydrologic flood data are usually not long enough to yield accurate extreme event estimates. One of the techniques applied to overcome this difficulty is the use of historical data. In parametric flood frequency analysis historical information was shown (Stedinger, 1986) to be valuable in improving the accuracy of high return period flood events.

To overcome some of the stated problems related to parametric analysis, a nonparametric approach was recently suggested. In nonparametric analysis the distribution of annual flood series is directly estimated from data without any assumption on its functional form. Besides, the problem of

parameter estimation is greatly simplified. A problem associated with the nonparametric technique is the difficulty in fitting highly skewed data, i.e. long-tailed distributions. This thesis introduces other alternative nonparametric approaches, such as the transformed, variable and adaptive kernel estimators. These new techniques are specially designed for fitting long-tailed data and explicitly estimating mixed distributions. Furthermore, this study addresses the multimodal nature of nonparametric probability distributions, which allow for the difference in flood characteristics arising from various flood causative factors. In addition, sample length requirements of nonparametric method and the effect of record length variations on quantile estimates are assessed. Finally, a method for incorporating historical information in nonparametric analysis is developed and investigated. Then, the effect of the use of historical data on the accuracy of flood estimates is evaluated.

This thesis extends the application of nonparametric approach in flood frequency analysis. To the best of the author's knowledge the topics *historical information* and *mixed distributions* have not been reported elsewhere in the literature as far as the nonparametric technique is concerned.

1.2 Objectives

The objectives of this study are

1. To investigate the various kernel approaches in flood frequency analysis and compare their performance with the parametric analysis.

2. To examine the performance of the different kernel methods in estimating flood density functions in terms of tail behaviour and bimodality.

3. To introduce and investigate the nonparametric technique to flood data of mixed distributions.

4. To compare the asymptotic behaviour of the various kernel estimators by a Monte-Carlo simulation study.

5. To determine the effect of sample length variations on the accuracy of flood estimates by the nonparametric approach.

6. To develop and evaluate the nonparametric flood frequency analysis with historical information.

Chapter 2

LITERATURE REVIEW

2.1 Parametric Analysis

2.1.1 Selection of a probability density function

In practice, the appropriate distribution of an annual flood series is not known apriori. The unknown probability density function will then be estimated by the form

$$f(x) = f(x, \theta) \tag{2.1}$$

where θ is a vector of p parameters characterizing the density function and estimated from the flood data sample. No explicit guidance as to the distribution choice can be formulated. An administrative imposition of choice is

often advocated (USWRC ,1982). Experience encountered from the application of other distributions can help in the selection. Properties of various distributions such as range, shapes, tail and boundary conditions, skewness etc, can also assist in selecting the distribution to be used in a given application. Nevertheless, the choice will affect flood quantile estimation. A large number of distributions are available for use in flood frequency analysis, each with its own parameters and characteristics. The following is a description of the most commonly used distributions.

Normal distribution and its transformations

The Normal probability density function is defined as (Kite, 1977):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\} \tag{2.2}$$

where μ and σ are the population mean and standard deviation. The normal distribution has a wide application in statistical hydrology for several reasons: (a) It fits symmetrical frequency distributions to random variables. (b) It can be used as a reference for comparing different distributions. (c) It is applied in the analysis of random errors. (d) It is employed in statistical inferences of distribution parameters, since all characteristics, such as mean and standard deviation, are either approximately or exactly normally distributed.

To account for skewness, sometimes a logarithmic transformation is neces-

sary, in which case the probability density function is expressed as:

$$f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} \exp\{-\frac{(\ln(x) - \mu_y)^2}{2{\sigma_y}^2}\} \tag{2.3}$$

in which μ_y and σ_y are the mean and standard deviation of the natural logarithms of x. This is known as the probability density function of the two parameter Lognormal distribution. It can not have negative skew coefficients and its variates take on values only in the interval of $[0, +\infty]$. This distribution found a wide range of applicability in hydrology because of not only its simplicity but also hydrologic variables such as discharge and precipitation are bounded from left by zero and are unbounded from right.

Another transform of the Normal distribution is the shifted Lognormal or three parameters Lognormal. Besides the scale (μ) and shape (σ) parameters, this distribution introduces a third term (c) called the location parameter. Using the transformation $z = \ln(x - c)$, its probability density function is given by

$$f(x) = \frac{1}{(x - c)\sigma_z\sqrt{2\pi}} \exp\{-\frac{(\ln(x - c) - \mu_z)^2}{2{\sigma_z}^2}\} \tag{2.4}$$

where μ_z and σ_z are the mean and variance of the variable Z, respectively. The three parameter Lognormal distribution is bounded by the location parameter c. For negatively skewed data c is an upper bound and for a positively skewed data c is a lower bound. The former case is obviously of limited applicability in hydrology. This distribution was found to be more efficient than the two parameter Lognormal in transforming the flood data to normality.

Often a transformation is needed to arrive at a probability distribution that will describe the data. Transformations that are common for most of the distributions are : logarithmics, translations along the x-axis and n-th power. Logarithmic transformations are usually made to decrease skews of the real space data. Translations along the x-axis, by introducing an extra location parameter, are useful in the case of bounded distributions. For this type of transformation, the addition of a third parameter make the estimation procedure of the parameters more complex.

Pearson and Log-Pearson type III

The three - parameters Gamma or Pearson distribution, introduced in 1895 by Karl Pearson, has the form of (Kite, 1977)

$$f(x) = \frac{1}{|a|\Gamma(b)}\left(\frac{x-c}{a}\right)^{b-1} \exp\left\{-\left(\frac{x-c}{a}\right)\right\} \qquad (2.5)$$

where a, b and c are its parameters and $\Gamma(b)$ is the Gamma function. If the transformed data $Y = \ln X$ act in accordance with the Pearson type III than the random variable Y is said to follow a Log-Pearson type III distribution given by (Kite, 1977)

$$f(x) = \frac{1}{x|a|\Gamma(b)}\left(\frac{\ln(x)-c}{a}\right)^{b-1} \exp\left\{-\left(\frac{\ln(x)-c}{a}\right)\right\} \qquad (2.6)$$

in which a, b and c are the scale, shape and location parameters respectively. Depending on the values of these parameters, the distribution can take a variety of shapes (Bobee, 1975). When a is negative and b less than 1, the distribution is negatively skewed, upper bounded at exp(c) and has

shapes such as U,J or S. These are of almost no use in hydrology. In case a is positive, however, the distribution is positively skewed, unbounded and has either a unimodal bell-shape ($b > 1$) or a reverse J-shape ($0 < b < 1$).

To promote uniform and consistent flood frequency analysis, the U.S. Water Resources Council (USWRC) recommended in 1977 the use of the Log-Pearson type III distribution by all federal agencies. The same recommendation was adopted also in Australia and South-Africa.

General Extreme Value distribution

The maximum or minimum values, chosen from N samples of n events, were found to follow an asymptotic form as n increases. This is given by the functional equation (Kite, 1977)

$$f^n(x) = F(a_n x + b_n) \tag{2.7}$$

in which a_n and b_n are functions of n. Fisher and Tippet found that there are three possible solutions in this case: the unbounded type I or Gumbel, the lower bounded type II and the upper bounded type III distribution. Jenkinson (1969) generalized the cumulative distribution function of the three types in one form such as:

$$F(x) = \exp\{-\left[\frac{1-k(x-c)}{a}\right]^{\frac{1}{k}}\} \tag{2.8}$$

where a and c are the scale and location parameters respectively. As k approaches zero the extreme value type I is obtained. The extreme value type

II, lower bounded at $c + a/k$, corresponds to negative values of k. And the extreme value type III, upper bounded at $c + a/k$, corresponds to positive values of k. The general extreme value distribution was recommended for use in great Britain (Hall, 1984).

Wakeby distribution

The Wakeby distribution, introduced by Houghton (1978), is a five parameter distribution with a very thick right-hand tail. It is defined as an inverse function

$$x = -a(1 - F)^b + c(1 - F)^{-d} + e \tag{2.9}$$

where a, b, c, d and e are shape parameters and F is the cumulative distribution function. The Wakeby is one of the very few distributions that account for the *separation effect* (Matalas et al, 1975). That is, it produces skews that are at least as stable as the natural skews. Furthermore, the Wakeby was found effective in generating synthetic flow data for Monte-Carlo experiments in the patterns of any of the more traditional distributions. For this reason, the Wakeby is often called the parent distribution. Nevertheless, as would be expected, the increase in number of parameters will make their estimation more complex and less accurate.

In general, there is no agreement between hydrologists as to which distribution should be used. In 1976 at the International Hydrological Decade meeting in Leningrad, it was found that over one hundred distributions

had been tried. Acceptance of a certain distribution for analysis of flood peaks depends on the goals and conditions to be satisfied. The selection of a model is based on two important criteria: (1) There should be a sound theory describing the phenomenon, and (2) the model should abstract the maximum information from the data, using proper estimation techniques. Although goodness of fit tests can by themselves determine the best frequency distribution, they in some cases provide reasons why certain models may not be suitable.

In practical applications, a specific distribution is recommended for use, not necessarily because of its statistical superiority but to promote a uniform and consistent flood frequency analysis. These types of recommendations are often followed without proper search for a statistically better fitting distribution. Benson's report (1968) states, however, that *the range of uncertainty in flood analysis, regardless of the method used, is still quite large and many questions concerning it remain unresolved.* Often, commonly used distributions provide a good fit in the central part of the density function. Differences, if any, are pronounced in the tails, where the T-year return period event needs to be estimated. Thus, the true shape of the distribution is not known and its estimation is a major source of error in parametric flood frequency analysis.

2.1.2 Parameter estimation

Problems faced by hydrologists, when applying parametric flood frequency analysis, do not end at the stage of selecting an appropriate density function. Parameter estimation techniques also have significant influence on flood estimation. The most commonly used techniques are the moments and the maximum likelihood methods.

The method of moments is widely used because of its computational simplicity. It is based on the relations of the parameters and the absolute or central moments of the distribution. In case the absolute or moment about the origin is used, the r-th absolute moment is given by

$$\mu'_r = \int_{-\infty}^{+\infty} x^r f(x) dx \tag{2.10}$$

When the central or the moment about the mean is used, on the other hand, the k-th central moment is expressed as

$$\mu_r = \int_{-\infty}^{+\infty} (x - \mu'_1)^r f(x) dx \tag{2.11}$$

Most of the distributions, in practice, are adequately described with the first two or three moments. The location parameter is related to the first order moment (mean), the scale parameter to the second order moment (standard deviation) and the shape parameter to the third order moment (skewness coefficient). The accuracy in estimating higher order moments from the sample data decreases rapidly as its order r increases. Wallis, et al (1974) showed that only the mean of the sample is an unbiased estimate

of the population mean. However, the standard deviation and coefficient of skewness are quite biased. Thus, if a distribution has more than two parameters, the use of the method of moments leads to biased parameter estimates and consequently will influence the accuracy of the flood quantile estimates. Rao (1983) provided a remedy to avoid using the biased third moment order in estimating the parameters of the Log-Pearson type III. He proposed the mixed moments method. This technique, which uses the mean and variance of real data and the mean of the log data, eliminates the use of the biased sample skewness coefficient. Despite being tedious, the estimation procedure was found to have superior statistical properties. To correct for the sampling error of the third order moment, the USWRC (1982) recommended the use of a linear combination of the sample skew and a regional skew obtained from skew maps. Matalas, et al (1974), however, found that estimates of skewness obtained from observed flood sequences are more variable than those obtained by any of the frequency distributions (phenomenon called *separation effect*). Beard, et al (1982), in an investigation on the uncertain nature of the bias in the method of moment estimate of skew for the Log-Pearson type III, criticized the use of the bias correction factor of skew. They state that *the nature of the bias correction factors for skew is probably markedly different from what is currently used.*

The maximum likelihood method, on the other hand, was introduced in 1910 by Fisher. It is based on the fact that the joint probability distribution of the n random observations in a flood data sample of density function

f(X,a,b,...) is proportional to the product

$$L = \prod_{i=1}^{n} f(x_i; a, b, c, ...) \tag{2.12}$$

where L is known as the likelihood function. The values of the parameters (a,b,..) that maximizes this product are called the maximum likelihood estimates. Since some probability distributions involve exponential terms, it is often easier to maximize the natural logarithm of the likelihood function. Parameters estimated by this method are usually considered the most reliable.

Greenwood, et al (1979) introduced a new estimation technique namely the probability weighted moments. This procedure is of potential interest for distributions that can be written in the inverse form, such as Wakeby and Weibull. Landwehr, et al (1979) investigated by a Monte-Carlo experiment the performance of the three methods, that is: moments, maximum likelihood and probability weighted moment. They showed that, unlike the two former techniques, the latter produces unbiased parameter estimates in the case where the samples are drawn from a purely random process. When the process, regarded as purely random, is serially correlated, however, all methods produced biased estimates. Thus, the probability weighted moment procedure can not reduce the bias and the variance of the quantile estimates, since the assumption that the flood data sample is drawn from a random process is usually questionable. Nozdryn - Plotnicki and Watt (1979) conducted another study, whereby three different estimation procedures (moments, maximum likelihood and a method proposed by Bobee (1975), which preserve the moment of the untransformed flows) are com-

pared over a range of parameter values of the Log-Pearson type III distribution. It was found that all methods exhibited considerable bias in estimating the distribution parameters. In quantile estimates, however, no bias was detected except for the method proposed by Bobee for low return periods. In terms of standard error of estimate (i.e. precision), the three methods did not show a great deal of variations and no preference was indicated. Nevertheless, Nozdryn - Plotnicki and Watt concluded that the method of moment is the best to use because of its computational simplicity.

As a summary, population statistics, such as mean, standard deviation and skewness, are unknown and can only be estimated from sample data. Regardless of the method used, the annual flood series are most of the time not long enough to yield unbiased parameter estimates. As a result, this will affect the accuracy of flood estimates.

2.1.3 Treatment of mixed distributions

The number of modes in a density function often determines the number of populations involved in the corresponding flood data sample. Most of the parametric distributions used in hydrology are unimodal. Thus, parametric flood frequency analysis assumes automatically that the floods in a data sample are drawn from one and the same population. Different types of flood generating mechanisms, however, can give rise to different flood char-

acteristics. In Canada and other countries of similar climatic conditions, floods may be caused by different and complex mechanisms. Hazen (1930) classified flood causative factors into three main categories: melting snow, great storms of wide distributions (convective storms) and cloudbursts covering small areas (frontal storms). Hizer (1956), on the other hand, categorized precipitations into six types: (1)cold front, (2)warm front, (3)stationary front, (4)squall line, (5)warm air mass and (6)cold air mass. If more than one source of the aforementioned flood generating processes are present in a data sample, which is the most probable situation, the flood data in question should be treated with a mixed distribution. Almost none of the parametric distributions can handle this problem. Therefore, several attempts have been suggested in the literature. Wood, et al (1975) consider the use of a linear combination of individual distributions each weighted by a parameter. Waylen and Woo (1982) subdivided the historic streamflow into separate subsets (snowmelt and rainfall) using the antecedent precipitation criterion (APC). This segregation criterion is only feasible when floods are generated by either snowmelt or rainfall. The two annual maximum series, one for each generating process, are fitted by distinct Gumbel distributions. These components were then combined to obtain flood estimates produced by mixed mechanisms. A similar model was introduced by Rouselle, et al (1971), but this time with seasonally segregated components. Using the partial duration series approach, they assumed that exceedances which occurred during the same season are identically distributed and mutually independent. Rossi, et al (1984) criticized this separation from a physical stand point. They state that *the assumption of identically dis-*

tributed flood peaks within a season is not realistic, since the separation of floods by season only partially corresponds to separation by meteorological mechanisms. Floods caused by different type of storms might coexist in the same season (Gupta, et al, 1976). Rossi, et al (1984) introduced the two component extreme value model (TCEV) as another alternative for handling mixed distributions. They dealt with data from regions where floods are caused by different types of storms. Thus, an apriori separation of flood processes was not feasible: neither by season, because they can coexist in the same season, nor by storm characteristics, because the available meteorological information is insufficient. The TCEV distribution is based on the assumption that floods arise from a mixture of two exponential components, extreme value type I, whereby the four parameters are estimated jointly by the maximum likelihood method.

To summarize, in parametric flood frequency analysis very little concern is devoted to the treatment of mixed distributions. Yet, it is obvious that historic streamflow peaks do not all arise from one and the same distribution. Different flood generating mechanisms might be involved in one annual flood series. In this thesis a new technique *the nonparametric analysis* is investigated numerically to address this issue and proposed as a viable alternative for the treatment of a sample of flood data drawn from mixed populations.

2.1.4 Use of historical data

Once a distribution is selected, its characteristics (parameters) are usually inferred from the flood data. The longer the sample length the more information, related to the distribution, can be extracted and consequently the more accurate is the quantile prediction. This fact is more stressed when the flood quantiles to be estimated are of relatively large return periods, i.e. lying far in the tail of the distribution. Usually, extreme floods are the ones which most influence the behaviour of the distribution tails. Klemes (1986) shows how changing the smallest flood influences the estimates of the upper quantiles. Thus, it is obvious that hydrologists are interested in any source of flood information that can increase the accuracy of the estimate of the assumed distribution parameters. Several techniques can be used for this purpose including: record extension by regression on correlated rainfall or streamflow sequences (Matalas, et al 1964), regional flood estimation (Wood, et al, 1981) and inclusion of information on flood events that preceded the gauged records (Leese, 1973). Another approach, suggested by Costa (1978), is paleohydrology, in which the dates and magnitudes of flood events are deduced from physical evidence, such as deposited slack water sediments and age of vegetation at the flood site.

Hosking et al (1986b) found that many factors, such as choice and nature of flood distribution, method of fitting, sample length of systematic records and flood measurement errors, affect the utility of historical information in parametric flood frequency analysis. The number of parameters of a dis-

tribution, however, seems to be the critical issue. Hosking, et al (1986b) showed that where the distribution has only two parameters to be estimated, the historical information is valuable only if accurately measured. In such case, the accuracy of extreme flood events can be increased by up to 50%. If the parent distribution has more than two parameters, however, the historical information, even when subject to measurements errors, contributes to a considerable increase in the accuracy of the quantile estimates. Depending on the type of information available, different procedures have been followed for its incorporation in parametric flood frequency analysis. Benson (1950) used a graphical method to display flood data with historical information. He had, apart from systematic gauged records, measurements of floods that exceeded a particular stage (5.5 m) during a historic time span (88 years) prior to the installation of a gauge. Plotting positions, for the various historical and systematic flows, were assigned so as to estimate the flood frequency relationship. Karpuk and Gerard (1979) emphasized the fact that historical floods are observed because their values exceed a threshold of perception called *censoring threshold* Q_c. In statistical terminology, a censored sample is defined as a record of historical flood peaks above the censoring threshold. Stedinger (1986) classified historical information into censored data, whereby the magnitudes of historical flood peaks are known, and binomial data where only threshold exceedance information is available.

The incorporation of historical information by the method of moments is not possible for binomial data, because one does not have values of the his-

torical floods. Nevertheless, USWRC (1982) proposed the adjusted moment estimator, in which the procedure weighs the moments of the historical and systematic data to obtain adjusted moment estimates of the parameters of the distribution.

The maximum likelihood method for incorporating historical information was initially used by Hald (1949). This technique is more attractive in the sense that it can be used for either binomial, censored or any combination. In the case of censored data, the likelihood function has three components: one for the systematic records $\{x_1, x_2, ..., x_s\}$, one recognizing that only $k_\star$ floods $\{y_1, y_2, ..., y_{k_\star}\}$ exceeded the threshold Q_c during the historical time span and one representing the actual values of the historical flows. Therefore, the likelihood function in this case is given by (Stedinger, 1986):

$$L(\mu, \sigma) = \prod_{i=1}^{s} f(x_i) \begin{bmatrix} h_o \\ k_\star \end{bmatrix} F(X_c)^{(h_o - k_\star)} \prod_{j=1}^{k_\star} f(y_j) \qquad (2.13)$$

where $h_o = H - s$ (H being the historic period), $X_c = \ln(Q_c)$, $f(\cdot)$ and $F(\cdot)$ are the probability and cumulative probability density functions ,respectively, and x_i and y_j are flood values in log-space.

In the case of binomial data, on the other hand, the likelihood function is similar to eq.(2.13) without incorporating the likelihood of having observed the values of $\{y_1, y_2, ..., y_{k_\star}\}$ .This is given by the form (Stedinger ,1986)

$$L(\mu, \sigma) = \prod_{i=1}^{s} f(x_i) \left(\begin{bmatrix} h_o \\ k_\star \end{bmatrix} F(X_c)^{(h_o - k_\star)} [1 - F(X_c)]^{k_\star} \right) \qquad (2.14)$$

Condie and Lee (1982), using a Monte-Carlo experiment, compared the

performance of both the maximum likelihood and the weighted moments methods for fitting the three parameter Log-normal distribution to flood series with historic information. They concluded that the maximum likelihood method was preferable. A similar investigation was done by Stedinger, et al (1986), first, to quantify the improvement in the accuracy of flood estimates when using historical information and second, to compare the two methods of incorporation. With the two parameter Log-normal being the selected distribution, they showed that historical information can be of tremendous value in flood frequency analysis. In addition, they found that the adjusted weighted moment technique performed poorly in comparison with the maximum likelihood procedure.

Finally, it is clear that historic flood information, when added to systematic records in flood frequency analysis, increases the effective record length and as a result improves the accuracy of the quantile flood estimates. The term *use of historical information* is new as far as the nonparametric approach is concerned. This thesis includes an investigation to adopt the historical information to the nonparametric flood frequency analysis.

2.2 Nonparametric Analysis

2.2.1 Density estimation by the kernel method

Since a knowledge of the density function $f(\cdot)$ is fundamental to any prediction, density estimation is the most important subject in applied statistics. The usual aim is to infer the distribution characteristics from a data sample $x_1, x_2, ..., x_n$. Histogram analysis was used extensively for univariate data as the first elementary nonparametric density estimator. This method, however, presented some inconveniences, such as the sensitivity of the frequency histogram to the subjective choice of the class interval width and location of the first class mark. In fact the appropriate class interval depends upon, not only the range of the data, but also its unknown behaviour. Furthermore, the histogram method was found to be very tedious and time consuming in the bivariate data analysis case. It is mainly because of these two reasons that the continuous nonparametric density estimation becomes a very active area of research. Rozenblatt (1956) introduced the shiftable or naive histogram estimator of the density function. It consisted of subjectively choosing an interval h, known as the smoothing factor. Then, the density at any point is estimated as being proportional to the number of observations within that interval, centered at the point under consideration. That is based on a random sample $[X_i]_{i=1}^n$ from a continuous but unknown density function $f(\cdot)$, this estimation is given by (Silverman, 1986)

$$\hat{f}(x) = \frac{1}{2hn}[no.\, of\, x_1, x_2, ..., x_n\, falling\, in\, (x-h, x+h)] \qquad (2.15)$$

By defining a weight function $W(\cdot)$ as

$$W(x) = \begin{cases} 1/2 & \text{if} \mid x \mid < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.16}$$

the naive histogram estimator can be expressed more transparently by the form

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} W\left(\frac{x - x_i}{h}\right) \tag{2.17}$$

The only difference between the conventional and naive histograms is that the latter is freed from a choice of bin positions. Every point is at the center of a sampling interval. The choice of bin width, however, still remains common for both estimators. For the naive histogram this is governed by the parameter h, which controls the smoothness of the estimated density function. Parzen (1962) generalized Rosenblatt's estimator to include series of kernel function $K(\cdot)$ of which eq.(2.16) is only an example. The kernel estimator is defined as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{2.18}$$

Fig.(2.1) is a qualitative graph showing how a density can be estimated using this technique. the nonparametric density is the sum of kernels placed at the observations. The shapes of these kernels are controlled by the so called kernel function $K(\cdot)$ and their widths are determined by the smoothing factor h. In this section only the general concept of the kernel estimator is presented. Further discussions are devoted for the choice of the smoothing factor and kernel function in chapter (3).

Since the introduction of the kernel method by Rosenblatt (1956), many

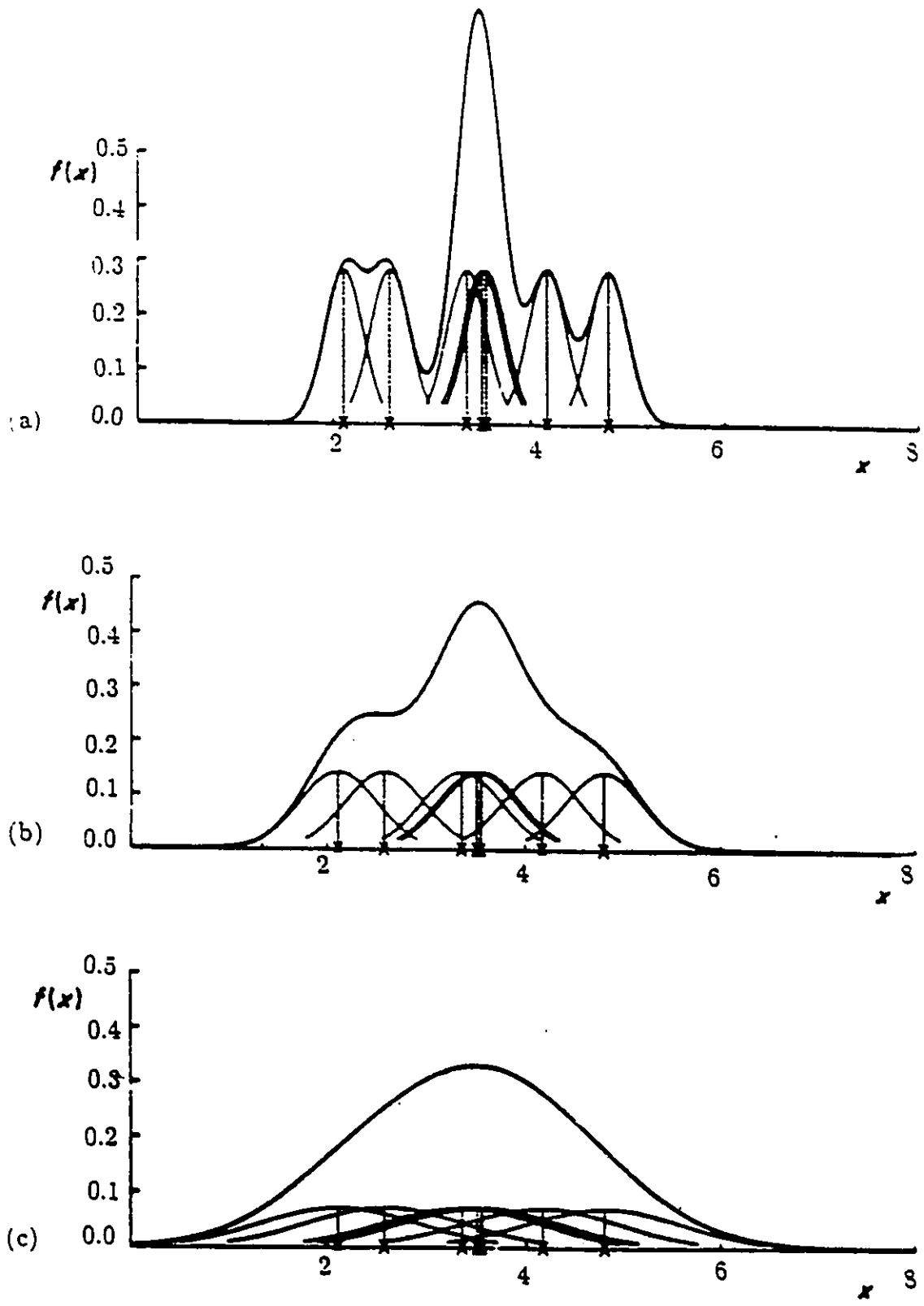


Figure 2.1: Fixed Kernel Estimates of Density Function (Gaussian Kernel , Smoothing Factors (a) h=0.2, (b) h=0.4 and (c) h=0.8. After Silverman, 1986 p. 14)

other important nonparametric approaches of density estimation, though not as popular as the kernel method, came into application. Among these are the series methods (Watson, 1969 and Brunk, 1978), the maximum likelihood method (Wegman, 1970) and the nearest neighbourhood method (Loftsgaarden et al, 1965). Some of these estimators are of limited use in flood frequency analysis because either the estimate is not usually a probability density function or it has finite supports. The kernel method, on the other hand, producing a probability density function which can have infinite supports, is the basis of the nonparametric approach to flood frequency analysis investigated in this thesis.

2.2.2 Introduction of the nonparametric approach to flood frequency analysis

Limitations associated with the conventional parametric flood frequency analysis, such as the apriori distributional assumptions, parameter estimation and treatment of mixed distributions, brought about a recent trend to apply nonparametric statistical models in flood frequency analysis. Such new technique should not be viewed as a magic solution for all problems of parametric methods. The nonparametric procedure, however, does not require the apriori distribution selection, is particularly efficient in treating mixed distributions and greatly simplifies the problem of parameter estimation

Adamowski (1985) was the first to introduce the application of nonparametric method in flood frequency analysis. The emphasis was made on the fixed kernel estimator using several data-based algorithms for the determination of the smoothing factor h. The comparison of parametric versus nonparametric methods was made based on a simulation study performed by Labatiuk (1985). Despite the limitation of the sampling experiment, the results of the investigation have led to the following major conclusions.

- (a) From real data analysis, it was shown that flood estimates obtained by the fixed kernel and those obtained by the ordinary parametric methods have the same order of magnitude and lead to similar designs.
- (b) From synthetic data analysis, it was found that the flood magnitude estimates by the fixed kernel method can be equally or even more accurate than those of the parent distribution (LP III) used to generate the data.
- (c) The estimation of the smoothing factor h, using four data- based algorithms, might be less critical than the computation of the distribution parameters in parametric analysis.
- (d) The nonparametric procedures, the fixed kernel in particular, usually give more weight to the tails of the distribution, the region of interest. The parametric procedure, however, allows a great weight to the central part of the distribution.

It is clear that the nonparametric approach is a viable alternative to the determination of design flood estimates, yet, its potentials are not fully realized. It is the purpose of this thesis to expand the work already done on the subject. Some of the issues related to flood frequency analysis, to which the response of the nonparametric approach is not known, are to be investigated. First, other alternative kernel approaches, such as the transformed, variable and adaptive kernels, are suggested to achieve better overall smoothness and tail behaviour of the density estimate and to avoid the critical estimation of the smoothing parameter of the fixed kernel. Second, it is shown how the nonparametric method can easily handle mixed distribution of a streamflow data sample, where floods are generated by more than one mechanism. Third, the effect of sample length variation on quantile estimates of the nonparametric approach is assessed. And finally, the use of historical information in nonparametric flood frequency analysis is investigated.

Chapter 3

THEORETICAL DEVELOPMENT

3.1 Introduction

The fixed kernel method does not explicitly address the estimation of mixed distributions and highly skewed data (i.e. long tails). This is due to the fact that the smoothing factor h is not allowed to vary from one observation to another. To remedy the above two shortcomings of the fixed kernel method other alternative kernels namely the transformed, variable and adaptive kernels are investigated. The transformed kernel (Devroye and Gyorfi, 1985) consider the estimation of the density of a transformed random variable and then taking the inverse transform. This approach,

besides allowing for an analytical expression of h, is mainly applied to improve the tail behaviour of the density function estimate. The variable kernel (Breiman, et al 1977), on the other hand, is an estimator which adapts the amount of smoothing by allowing h to vary depending on the local density. Hence, the smoothing factor will be larger in low density regions and smaller in high density regions. Finally, the third approach called the adaptive kernel (Silverman, 1986), permits also the variation of the smoothing factor from one observation to another. This technique, however, requires an apriori knowledge of the density function to decide whether or not an observation is indeed in a low density region. It is to be noted that the transformed, variable and adaptive kernel approaches are applied for the first time in flood frequency analysis. The main purpose in introducing these new techniques is to achieve better tail fit to heavy-tailed distributed flood data. This chapter presents the concepts of the different kernel techniques along with the incorporation methodology of historical information in nonparametric analysis.

3.2 Fixed Kernel

The fixed kernel concept of estimating the density function was introduced in section (2.2.1). Methodologies for computing the smoothing factor h and kernel function $K(\cdot)$ are provided in this section. The choice of h and $K(\cdot)$ usually affects the general performance of the estimated density (i.e. how well the distribution will fit the data). A global measure of accuracy of

the estimated density is given by the so called Mean Square Error (MSE) defined by (Silverman, 1986 p.35)

$$MSE\left(\hat{f}(x)\right) = Var\left[\hat{f}(x)\right] + Bias^2\left[\hat{f}(x)\right] \tag{3.1}$$

where $\hat{f}(\cdot)$ is the estimate of the density function. According to Tapia and Thompson (1978), the variance and bias terms in eq.(3.1) are given by

$$Var\left[\hat{f}(x)\right] \sim \frac{\hat{f}(x)}{nh}\int_{-\infty}^{+\infty} K^2(x)dx \tag{3.2}$$

and

$$Bias^2\left[\hat{f}(x)\right] \sim h^4 {k_2}^2 |\, \hat{f}^{(2)}(x)\,|^2 \tag{3.3}$$

respectively, in which n is the number of observations in a data sample, h is the smoothing factor, $K(\cdot)$ the kernel function and k_2 is a coefficient defined as

$$k_2 = -\frac{1}{2}\int_{-\infty}^{+\infty} x^2 K(x)dx \tag{3.4}$$

The MSE, which measures the closeness of the estimator $\hat{f}(\cdot)$ to the true density $f(\cdot)$ in terms of precision and bias, will be used as the criterion for the derivation of the optimal smoothing factor and kernel function.

3.2.1 Choice of the kernel function

From the definition of the fixed kernel (eq.2.18), it is clear that for $\hat{f}(\cdot)$ to be a density function, the kernel $K(\cdot)$ should also be a density function itself. This is the only condition that insures that the estimate $\hat{f}(\cdot)$ is everywhere non-negative. The estimate $\hat{f}(\cdot)$ will inherit all the characteristics

of the kernel function, such as continuity and differentiability. The above conditions can be expressed mathematically as (Wegman, 1972)

$$\int_{-\infty}^{+\infty} K(x)dx = 1 \tag{3.5}$$

$$\int_{-\infty}^{+\infty} |K(x)|^2 dx < 0 \tag{3.6}$$

$$Sup_{-\infty<x<+\infty} |K(x)| < \infty \tag{3.7}$$

$$\lim_{n\to\infty} |xK(x)| = 0 \tag{3.8}$$

$$K(x) = K(-x) \tag{3.9}$$

The choice of the kernel function is empirical. Anderson (1969) and others have found that this choice is not very critical. Epanechnikov (1969) showed that there is an optimal kernel, which gives better results than any other kernel by only four percent. For the derivation of this optimal $K(\cdot)$, he considered the minimization of the Integrated Mean Square Error (IMSE) defined as

$$IMSE\left[\hat{f}(x)\right] = \int_{-\infty}^{+\infty} MSE\left[\hat{f}(x)\right] dx \tag{3.10}$$

where MSE is given by eq.(3.1). Epanechnikov derived the optimal kernel

$$K(x) = \begin{cases} \frac{3}{4}(5)^{-1/2}(1-\frac{x^2}{5}) & \text{if } |x| \leq (5)^{1/2} \\ 0 & \text{otherwise} \end{cases} \tag{3.11}$$

In order to compare the performance of various kernels, the efficiency of $K(\cdot)$ can be used, and is defined as (Silverman, 1986)

$$eff(K) = \frac{3}{5\sqrt{5}}\left[\int x^2 K(x)dx\right]^{\frac{-1}{2}}\left[\int K(x)^2 dx\right]^{-1} \tag{3.12}$$

Some kernels and their efficiencies are displayed in Table (3.1). It can be seen that the efficiencies are so close to each other that there is a little to

Table 3.1: Some Kernels and their Efficiency (After Silverman, 1986 p. 43)

Kernel	$K(\cdot)$	Efficiency
Epanechnikov	$\frac{3}{4\sqrt{5}}(1-\frac{x^2}{5})\ for\ \mid x \mid < \sqrt{5},$	1
	$0 \quad otherwise$	
Biweight	$\frac{15}{16}(1-x^2)^2\ for\ \mid x \mid < 1,$	0.9939
	$0 \quad otherwise$	
Triangular	$1-\mid x \mid\ for\ \mid x \mid < 1,$	0.9859
	$0 \quad otherwise$	
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp\{-(1/2)x^2\}$	0.9512
Rectangular	$\frac{1}{2} \quad for\ \mid x \mid < 1;$	0.9295
	$0 \quad otherwise$	

choose between the various kernels. Nevertheless, in this study the optimal Epanechnikov kernel (eq.3.11) will be used for the fixed, transformed, variable and adaptive kernels.

3.2.2 Choice of the smoothing factor

The smoothness properties of the fixed kernel density estimate are very sensitive to the choice of the smoothing factor h. Extensive research efforts were devoted to the determination of the optimal value of h. One approach (Tapia and Thompson, 1978 p.66) consists of drawing graphs, called *test graph*, of the density function estimate. Starting with too-large values and decreasing until too-rough density estimates are obtained. Silverman (1978) considered graphs of the second derivatives of the density function as being more sensitive to any change of the smoothing factor. The applicability of this technique is very limited, subjective, time - consuming and requires an interactive environment.

Another approach, investigated first by Woodroofe (1970), is a data - based experimental procedure. It is theoretically based on the minimization of the Mean Square Error. The MSE, however, is a function of the density (refer to eq.3.1). Thus, to determine h this approach requires an apriori knowledge of density. Besides being computationally impractical, this technique has sometimes a slow convergence rate (Bowman, 1982 and Hall, 1983).

Adamowski (1985) suggested selecting the smoothing factor by minimizing

the distance between kernel and plotting position estimates of the distribution function. The optimal choice of h is obtained by optimizing the Least Square function

$$L = \sum_{i=1}^{n} (p_i - p_n(x_i))^2 \tag{3.13}$$

where $x_1, x_2, ..., x_i$ are the ranked sample observations, $p_i = \frac{(i-.025)}{(n+0.5)}$ and $p_n(x_i)$ is the fixed kernel estimate of the cumulative distribution function given by

$$p_n(x_i) = \frac{1}{nh} \sum_{i=1}^{n} \int_{x}^{+\infty} K\left(\frac{x - x_i}{h}\right) dx \tag{3.14}$$

The Adamowski approach was implemented using the rectangular kernel given in Table (3.1).

Another data - based algorithm for the determination of the smoothing factor considers the minimization of the Integrated Mean Square Error (eq.3.10). Using the Epanechnikov optimal kernel (eq.3.11), Flanagan (1980) derived an optimal value of h as being

$$h = \frac{W_n}{n(n - \frac{10}{3})\sqrt{5}} \tag{3.15}$$

where

$$W_n = (n-1)(x_n - x_1) + \sum_{i=1}^{n-1} (2i - n - 1)x_i \tag{3.16}$$

This optimal value of h will be used in coordination with the fixed kernel in the numerical analysis of this thesis.

3.3 Transformed Kernel

To overcome the inefficiency of the fixed kernel in estimating long - tailed distributions and the problems encountered in choosing the smoothing factor h, Devroye and Gyorfi (1985) suggested the use of the transformed kernel estimator. They introduced a new parameter called $B^*(f)$ and defined as

$$B^*(f) = \left[1/2\left(\int \sqrt{f}\right)^4 \int \mid f^{(2)} \mid\right]^{1/5} \tag{3.17}$$

where $f(\cdot)$ is the density function and $f^{(2)}(\cdot)$ is its second derivative. This parameter measures the difficulty posed by $f(\cdot)$ when the fixed kernel estimator is used. Two major components appear in the expression of $B^*(f)$: $\left(\int \sqrt{f}\right)$ which measures how heavy the tail of $f(\cdot)$ is, and $\left(\int \mid f^{(2)} \mid\right)$ which measures how oscillatory $f(\cdot)$ is. The authors claimed that by minimizing $B^*(f)$ the density function will have better tail estimates. The overall infimum of $B^*(f)$s is reached when the density to be estimated is isosceles triangular. This is the basis for the development of the transformed kernel estimator. Therefore, the attempt is to estimate the density of a transformed random variable, take the inverse transform to minimize $B^*(f)$ and hence improve the performance of the original estimate, especially in the tails.

In mathematical terms, let $x_1, x_2, ..., x_n$ be a random sample of flood events and $K^*(\cdot)$ be the optimal Epanechnikov kernel (eq.3.11). For the following development, h_* will be denoting the smoothing factor of the fixed kernel and h that 's of the transformed kernel estimator. Let $T(\cdot)$ be the

transformation from $R \rightarrow [0,1]$, which is strictly monotonically increasing, continuously differentiable, one-to-one and onto, and which has a continuously differentiable inverse. The transformed kernel density estimate of the random variable x_i is given by

$$\hat{f}(x) = g\left(T(x)\right) T'(x) \tag{3.18}$$

where $g(\cdot)$ is the fixed kernel estimate of the density function of the new random variable $Y_i = T(X_i)$, i.e. the transformed data. Thus,

$$g(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - y_i}{h}\right) \tag{3.19}$$

Since $g(\cdot)$ has to be a real density on $[0,1]$, the following normalization will be required

$$\hat{g}(y) = \frac{g(y)}{\int_0^1 g(y)dy} \tag{3.20}$$

The only unknown at this stage is the transformation $T(\cdot)$, which is chosen in a way to minimize $B^*(f)$. This objective is attained when $g(\cdot)$ is the isosceles triangular density on $[0,1]$. If the distribution function $F(\cdot)$ of $f(\cdot)$ was known, this optimal transformation would be (Devroye and Gyorfi, 1985)

$$T(x) = \begin{cases} \sqrt{\frac{F(x)}{2}} & \text{if } F(x) \leq 1/2 \\ 1 - \sqrt{\frac{1-F(x)}{2}} & \text{if } F(x) > 1/2 \end{cases} \tag{3.21}$$

$f(\cdot)$, however, is not known and can only be approximated:

$$\hat{f}(x) = \frac{1}{nh_*} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_*}\right) \tag{3.22}$$

The corresponding distribution function $F(\cdot)$ is given by the form:

$$\hat{F}(x) = 1 - \hat{p}(x) \tag{3.23}$$

where $\hat{p}(\cdot)$ is the exceedance probability ,

$$\begin{aligned} \hat{p}(x) &= \int_x^{+\infty} \hat{f}(x)dx \\ &= 1/2 - \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x - x_i}{h_\star}\right) \end{aligned} \tag{3.24}$$

and $W(\cdot)$ is the integral of the kernel function $K(\cdot)$. The whole procedure of the transformed kernel is summarized in a more convenient way in the flowchart (Fig.3.1). The optimal Epanechnikov kernel (eq.3.11) is used. The optimal smoothing factor (eq.3.15) is employed in the preliminary estimation of the density function (eq.3.22). Devroye and Gyorfi suggested to use the conventional parametric technique for the first estimation of $f(\cdot)$ in eq.(3.22). This, however, was not considered in the forgoing analysis because the overall attempt is to free the flood frequency analysis from any distributional assumption.

3.3.1 Choice of the transformation

Devroye and Gyorfi (1985 p.110) have made an investigation on the choice of the transformation to be used. They suggested several transforms, such as uniform on $[0,1]$, isosceles triangular on $[0,1]$, normal and Laplace transforms. Their study showed that the isosceles triangular on $[0,1]$ is the transformation which gave a minimum value of $B^*(g)$

$$B^*(g) = 1.446 \tag{3.25}$$

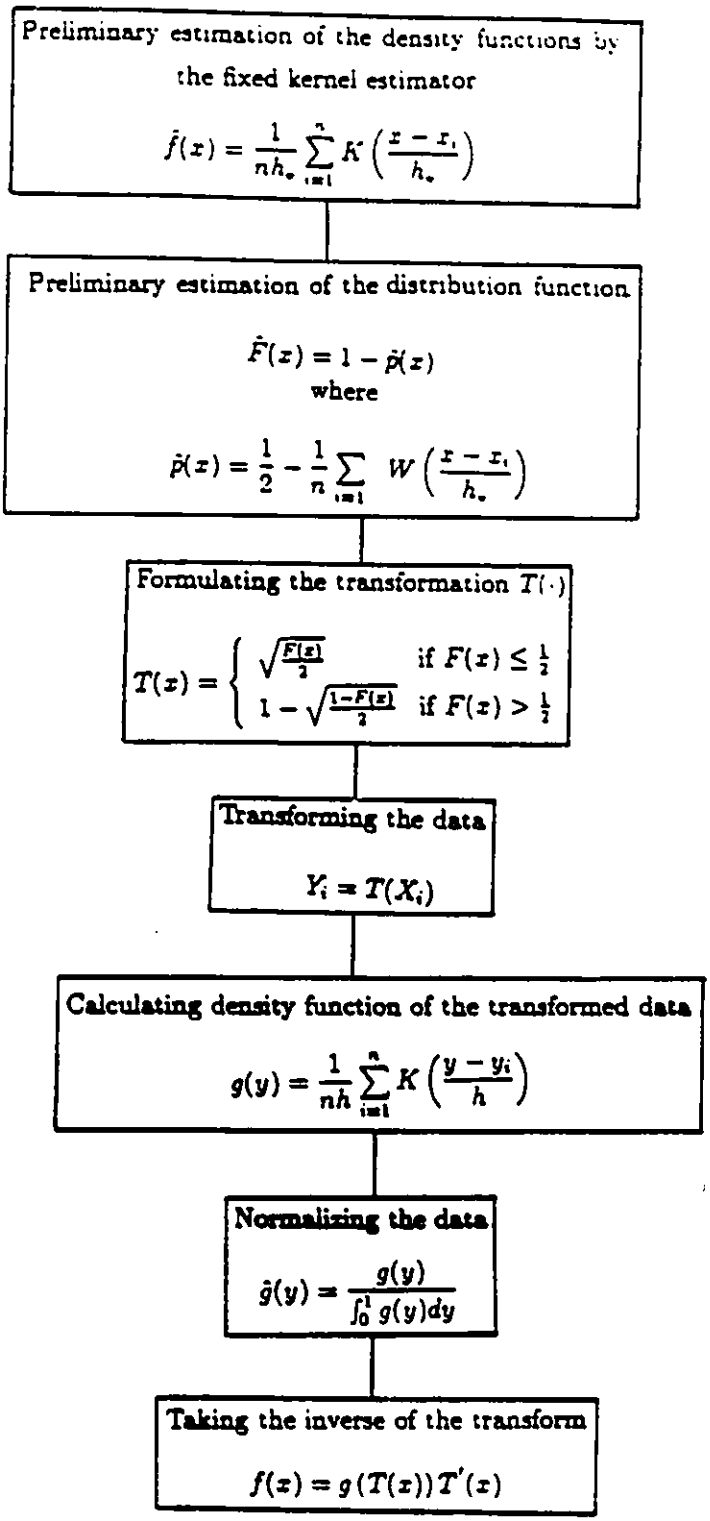


Figure 3.1: Flowchart Summarizing the Transformed Kernel Procedure.

and

$$h = \left(\frac{5\pi}{192n}\right)^{\frac{1}{5}} \tag{3.26}$$

In an attempt to find other transforms, Adamowski and Feluch (1986) demonstrated that a circular transform, theoretically should perform better than the isosceles triangular. This circular density function is given by

$$g(y) = \begin{cases} \frac{3}{4}(a)^{-1/2}(1 - \frac{y^2}{a}) & \text{if } \mid y \mid \leq (a)^{1/2} \\ 0 & \text{otherwise} \end{cases} \tag{3.27}$$

Notice that when $a = 5$ eq.(3.27) yields the Epanechnikov optimal kernel function (eq.3.11). For the circular density transform, in the case $a = 1$, the corresponding $B^*(g)$ will be having a value of (Adamowski and Feluch, 1986)

$$B^*(g) = 1.3872 \tag{3.28}$$

When comparing eq.(3.25) with eq.(3.28), it is evident that the circular transform should be more efficient. Using the optimal form of h recommended by Devroye and Gyorfi (1985, p.106) and defined as

$$h = \left(\frac{8}{\pi}\left(\frac{\int \sqrt{g}}{\int \mid g' \mid}\right)^2 \frac{1}{n}\right)^{\frac{1}{3}} \tag{3.29}$$

Adamowski and Feluch showed that the smoothing factor of the circular density will be

$$h = \left(\frac{5\pi}{16n}\right)^{\frac{1}{5}} \tag{3.30}$$

The transformation associated with the circular density function (eq.3.27) has the form of

$$T(x) = 2\cos\left(\frac{4\pi + \arccos(1 - 2F(x))}{3}\right) \tag{3.31}$$

In the numerical analysis both transforms, the isosceles triangular and the circular are used and a comparison of their performance is conducted.

3.4 Variable Kernel

A limitation in the application of fixed, and to some extent the transformed kernels in flood frequency analysis, is that the peakedness of the kernels is not data responsive. For instance, in regions of low densities the estimate will have a peak at a point and too low over the remaining of the region. On the other hand, in regions of high densities the observations are densely packed together and the fixed and transformed kernels can not respond appropriately to variations in magnitude of $f(\cdot)$. The main cause of this problem is that the smoothing factor is constant for all observations in a flood data sample. By allowing h to vary from one observation to another, the kernel will be more data responsive. To accomplish this task, Breiman et al (1977) proposed the use of the variable kernel given by

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} a_k d_{i,k} K\left(\frac{x - x_i}{a_k d_{i,k}}\right) \tag{3.32}$$

where $d_{i,k}$ is the distance from the point x_i to its k-th nearest neighbour, and a_k a constant multiplicative factor. Obviously, in low density regions, $d_{i,k}$ will be large and the kernel will be spread out. In regions of high density, however, the reverse phenomenon will occur. A common characteristic of both fixed and variable kernels is that $f(\cdot)$ is an average of some contributions, one from each sample point $x_1, x_2, x_3, ..., x_n$. For the variable kernel,

however, these contributions depend not only on the distance between x and x_i, but also on the local density near x_i. Regarding the symmetric properties of the estimator, to some extent the fixed kernel possesses more symmetry than the variable kernel. For the former, the contribution to the estimate at x_i from a point x_j is the same as that at x_j from a point x_i. For the latter, however, these two contributions are no more the same.

The usual question which arises is: how to estimate the a_k and k values which give the best estimate of $f(\cdot)$? Breiman, et al (1977) used a minimization search procedure, which was proved to be successful in locating the region of parameter (a_k, k) values. They found that good density fits can be obtained for a wide range of k, as long as the corresponding value of a_k keeps the following ratio constant

$$\lambda = \frac{a_k(\bar{d}_k)^2}{\sigma(d_k)} \tag{3.33}$$

where $\bar{d}_k$ is the mean of the k-th nearest neighbour distances and $\sigma(d_k)$ their standard deviation. Hence, the objective here is to find the single parameter λ to calibrate the variable kernel estimator. In order to make possible the incorporation of historical information in nonparametric flood frequency analysis, the aforementioned algorithm for the determination of the λ and a_k parameters was not used. Rather, the maximum likelihood method, which will be introduced in section (3.6), was applied. The performance of the variable kernel in improving the tail behaviour of the density function estimate will depend on the efficiency of the maximum likelihood method in locating the optimal parameters (a_k *and* k) region. In this thesis, the variable kernel is used in the comparison of the asymptotic behaviour

of the different kernel approaches, and for the incorporation of historical information in nonparametric flood frequency analysis.

3.5 Adaptive Kernel

The adaptive kernel estimator concept was introduced by Silverman (1986, p.100). It has the same features as the variable kernel estimator. The only difference being the introduction of a criterion to decide whether a flood observation is located in a low density region. The adaptive kernel deals with this aspect in two stages. First, a pilot estimate of the density function is obtained by means of the fixed kernel estimator.

$$\tilde{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{3.34}$$

where $K(\cdot)$ is the optimal Epanechnikov kernel (eq.3.11) and h is the optimal smoothing factor (eq.3.15). Second, a local smoothing factor for each observation is determined in the following way

$$\lambda_i = \left(\frac{\tilde{f}(x_i)}{g}\right)^{-\alpha} \tag{3.35}$$

where g is the geometric mean of the $\tilde{f}(x_i)$'s i.e.

$$\log g = \frac{1}{n} \sum_{i=1}^{n} \log\left(\tilde{f}(x_i)\right) \tag{3.36}$$

and α is called the sensitivity parameter satisfying the condition $0 \leq \alpha \leq 1$. Then, the adaptive kernel estimate of the density function is given by

(Silverman, 1986)

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h\lambda_i}K\left(\frac{x-x_i}{h\lambda_i}\right) \tag{3.37}$$

The sensitivity parameter controls the smoothness properties of the tails. From practical experience, Abramson (1982) and others suggested a value for α equal to 0.5. As will be seen later in section (4.4.1), however, this might not be usually the case. Furthermore, it can be noticed from eq.(3.37) that as α approaches 0 the method converges to the fixed kernel. Theoretically, the adaptive kernel estimator is the most promising in improving the tail behaviour of the density estimate, and it will be further investigated numerically using observed and synthetic data.

3.6 Nonparametric Flood Frequency with Historical Information

To evaluate the improvement in the accuracy of high return period flood estimates by the use of historical data, the procedure adopted by Condie and Lee (1982) in parametric analysis is followed here. It is assumed that one has both threshold exceedances information and also systematic gauged records. In this case, the maximum likelihood method, used to estimate the parameters of the variable kernel, is expressed in two terms. The likelihood function for the n systematic gauged records $\{x_1, x_2, ..., x_n\}$ and the likelihood function that recognizes the fact that there were exactly k_c flows whose magnitude are less than the censoring threshold Q_c during the his-

torical period. The likelihood function is written in the form (Condie and Lee, 1982)

$$L(a_k) = \prod_{i=1}^{n} f(x_i) \left(\frac{(n+k_c)!}{k_c!} \right) (F(Q_c))^{k_c} \qquad (3.38)$$

where a_k Parameter of the variable kernel

k_c Number of censored floods

Q_c Censoring threshold

n Number of fully defined floods

x_i Gauged streamflow in real space

Maximizing $L(a_k)$ by taking the partial derivative will lead to the maximum likelihood estimate of the parameter a_k. Notice here that when k_c is equal to zero, eq.(3.28) yields the conventional likelihood function, where no historical information is included (method used to find a_k for section 3.4). The use of censored data, whereby the magnitude of historical flood peaks are all known is beyond the scope of this thesis. The application of historical information is an original contribution to the nonparametric flood frequency analysis.

Chapter 4

RESULTS AND DISCUSSION

4.1 Introduction

For data analysis, computer programs were developed for triangular transformed (TRTKER), circular transformed (CITKER), variable (VARKER) and adaptive (ADAKER) kernel estimators. Flowcharts and listings of the programs are given in Appendix (B). Epanechnikov 's optimal kernel (eq.3.11) and the optimal smoothing factor h (eq.3.15) were used in all programs. The main difference between program TRTKER and CITKER is subroutine TR (refer to Appendix B). While the triangular transformation (eq.3.21) is used for the former program, the latter is based on a circular transformation (eq.3.31). Both programs provide the option of the fixed kernel estimator. Parameter KL is controlling this option (Kl=0 for trans-

formed kernel and KL=1 for fixed kernel). Program VARKER includes also more than one option. Parameters IPAL and KL control these options in the following way

$$IPAL = \begin{cases} 0 & \text{Fixed kernel (KL=1) ,circular transformed kernel (KL=0)} \\ 1 & \text{Variable kernel with or without historical information} \end{cases} \tag{4.1}$$

Parameters of the variable kernel are estimated using the maximum likelihood method (section 3.6, eq.3.38).

In this chapter the results of five main issues related to nonparametric flood frequency analysis are presented and discussed. These are, in order, (1)tail behaviour, smoothness and discontinuity of flood densities as estimated by the four kernel approaches, (2)mixed distributions estimated by the kernel method, (3)relative performance of the various kernels in quantile estimation, (4)effect of record length on flood estimates and (5)use of historical data in nonparametric flood frequency analysis.

4.2 Data Selection

In section (4.3) flood quantile estimates determined by the four nonparametric kernel approaches and by the Logpearson type (3) distribution are compared. For this purpose, the Northeast Margaree river at Margaree valley (station WSC 01FB001) in Canada is selected. This data set has relatively high skewness and long period of records (1917 - 1983).

In section (4.4), the ability of the four kernel approaches in explicitly estimating mixed distributions and fitting highly skewed data is compared. To perform this task, the flood data of Warta river at Poznan in Poland is selected. This data set is a 140 year set of records (1828 - 1965), with a skewness of 1.770.

Section (4.5), on the other hand, is devoted to an investigation on mixed distributions as estimated by the nonparametric technique. To address this issue the annual flood series of the Turtle river (station 05PB014) at Mine Centre are studied numerically. This station is located in Manitoba, Northwestern Ontario region, (latitude 48 51 00 N, longitude 92 43 30 W). The Turtle river basin represents a common type of geography of Canadian shield. The basin elevation differences are very small and its population density is quite low. The basin has a length of 115 km and an area of 4870 km^2. The soil cover is mainly gravel and loam over rock. Vegetation in the basin is mostly coniferous forest. Main tributaries above the station are the Little Turtle, the Big Turtle and Little Heron rivers which are not gauged.

Lakes comprise approximately 30% of surface area and have a large impact on runoff response. The winter temperature across the basin differs substantially. Peakflows of the Turtle river data set are drawn from two distinct populations (snowmelt and rainfall). The independence, randomness, trend and homogeneity of the analyzed data are checked by statistical hypothesis testing using the Consolidated Frequency Analysis Program (CFA1). This package is documented by Pilon, et al (1985). Results of the testing along with the data are given in Appendix (A).

The three stations studied in this thesis are for natural unregulated watersheds. All annual floods consisted of maximum daily flows and statistical characteristics of the analyzed river station data are given in Table (4.1).

Table 4.1: Statistical Characteristics of the Analyzed Flood Data.

River	Mean	SD	CV	CS	CK
Margaree					
Real space	171.83	57.37	0.334	1.073	4.500
Log space	5.10	0.32	0.063	0.166	3.135
Warta					
Real space	436.46	319.69	0.732	1.770	6.046
Log space	5.86	0.65	0.110	0.272	2.760
Turtle					
Real space	128.87	58.11	0.541	0.668	3.468
Log space	4.75	0.48	0.101	-0.310	2.550

Note: SD is the standard deviation; CV, CS and CK are the coefficientsof variation, skewness and kurtosis respectively.

4.3 Relative Performance of the Four Kernel Methods

For the Margaree river data, flood quantiles of arbitrarily selected return periods of 10, 50, 100 and 200 years are estimated by the four nonparametric approaches and are presented in Table (4.2). Flood quantiles for the same return periods are also estimated by the Log-Pearson type III distribution (using the parameter estimation method recommended by the USWRC) and are included in the table for comparison purposes.

Table 4.2: Flood Estimates for the Northeast Margaree River at Margaree Valley.

Event	LP3	FK	TTK	CTK	VK	AK
Q10	246	247	251	251	247	243
Q50	322	343	349	352	341	337
Q100	355	354	360	362	356	370
Q200	389	361	367	369	365	393

Note: LP3 is the Logpearson type (3); FK, TTK, CTK, VK and AK are the fixed, triangular transformed, circular transformed, variable and adaptive kernel estimators respectively.

From Table (4.2), it is clear that the four kernel approaches gave flood es-

timates in the same order of magnitude when compared with the LP III estimates. The triangular and circular transformed kernels gave approximately the same flood estimates. Furthermore, among the nonparametric kernel methods, the adaptive kernel is the most conservative for return periods larger than 100 years and the least conservative for return periods less than 100 years. No preference of one method over the others can be made at this stage because the true value of the estimates are not known. However, the fact that nonparametric and parametric analysis give flood estimates in the same range make further investigation on the various kernel methods worthy.

4.4 Overall Smoothness and Tail Behaviour of Flood Densities

The application of other alternative methods, such as transformed, variable and adaptive kernels has been proposed in order to explicitly take into account mixed distributions and skewed data. The amount of smoothness in a nonparametric density estimate affects its multimodal nature and tail characteristics. Fig.(4.1-A to 4.1-D) show density plots of the Warta river data by the four different kernel estimators. Fig.(4.1-A) give the density plot of the fixed kernel, with the optimal choice of the smoothing factor h. It can be seen that the bimodality nature of the data is perhaps obscured by an oversmoothness around the peak region and the tail of the density

is oscillatory. A marginal improvement in the tail is brought by the application of the transformed kernel estimator as shown in Fig.(4.1-B). The transformed kernel has a less oscillatory and less discontinuous tail. On the other hand, again bimodality was not well pronounced. The variable kernel estimator shows (Fig.4.1-C) a highly oscillatory and discontinuous density. This is most probably caused by an inefficiency of the maximum likelihood method in locating the optimal region of the a_k and k parameters. In fact, if the true density is multimodal the likelihood procedure might lead to a local maximum rather than an optimum one. Fig.(4.1-D) displays the density function plot estimated by the adaptive kernel. This method, compared to the fixed, transformed and variable kernels, exhibits a pronounced bimodality with the least oscillatory and least discontinuous upper tail. Similar findings were observed for other data, therefore, it was decided to further investigate the adaptive kernel method.

4.4.1 Effect of varying the sensitivity parameter α on the adaptive kernel density estimate

The amount of smoothing in the adaptive kernel method is controlled by the parameter α. Abramson (1982) suggested to use a value of $\alpha = 0.5$. It will be shown, however, that this choice is not usually satisfactory. Fig.(4.2 A-D) display density function plots of the Warta river data for values of $\alpha = 0.2, 0.4, 0.6 \, and \, 0.8$. Unlike the fixed and variable kernels, which are very sensitive to the choice of their smoothing parameter, the overall shape

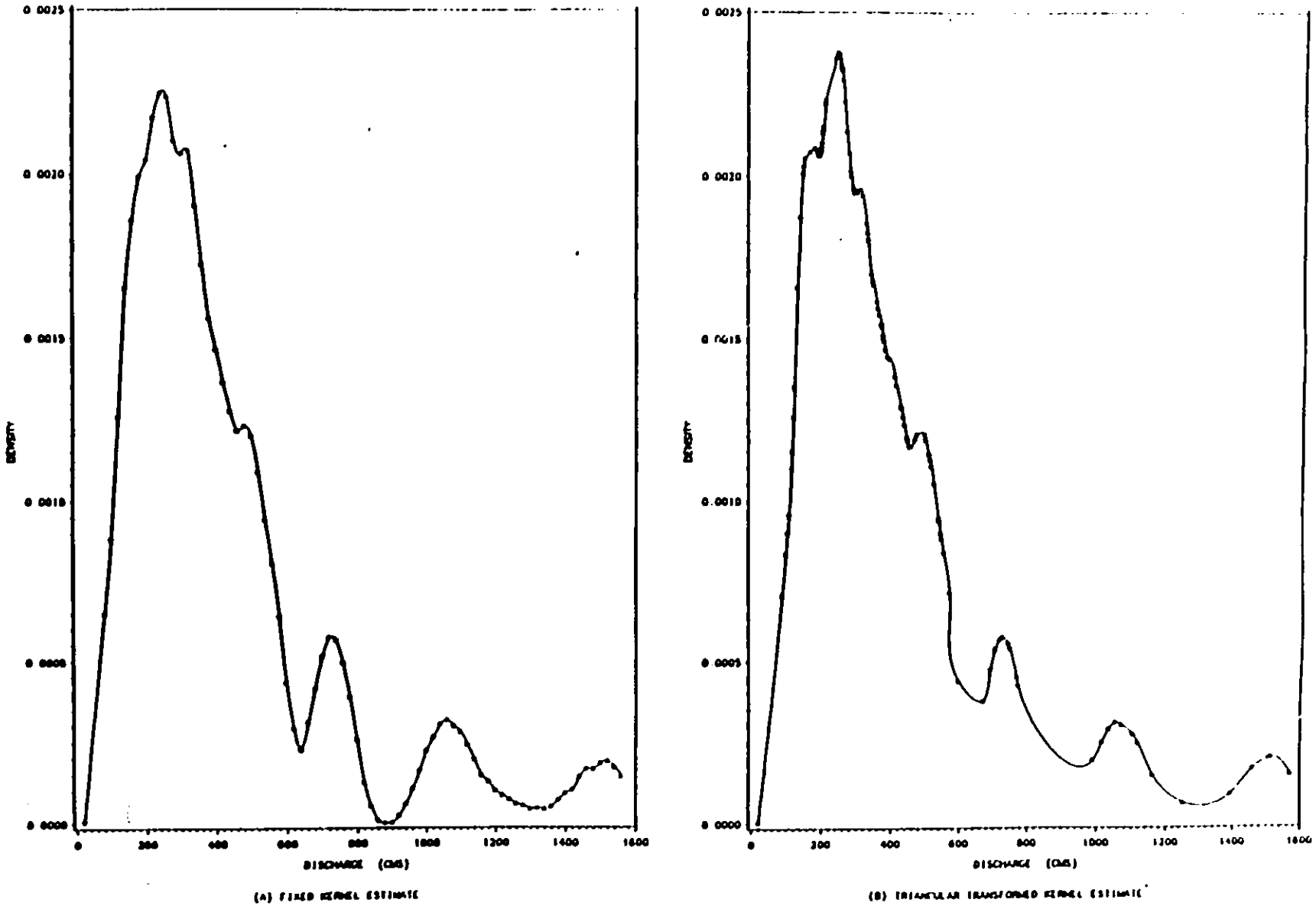


Figure 4.1: Smoothness and Tail Behaviour of the Nonparametric Density Functions (Warta River at Poznan)

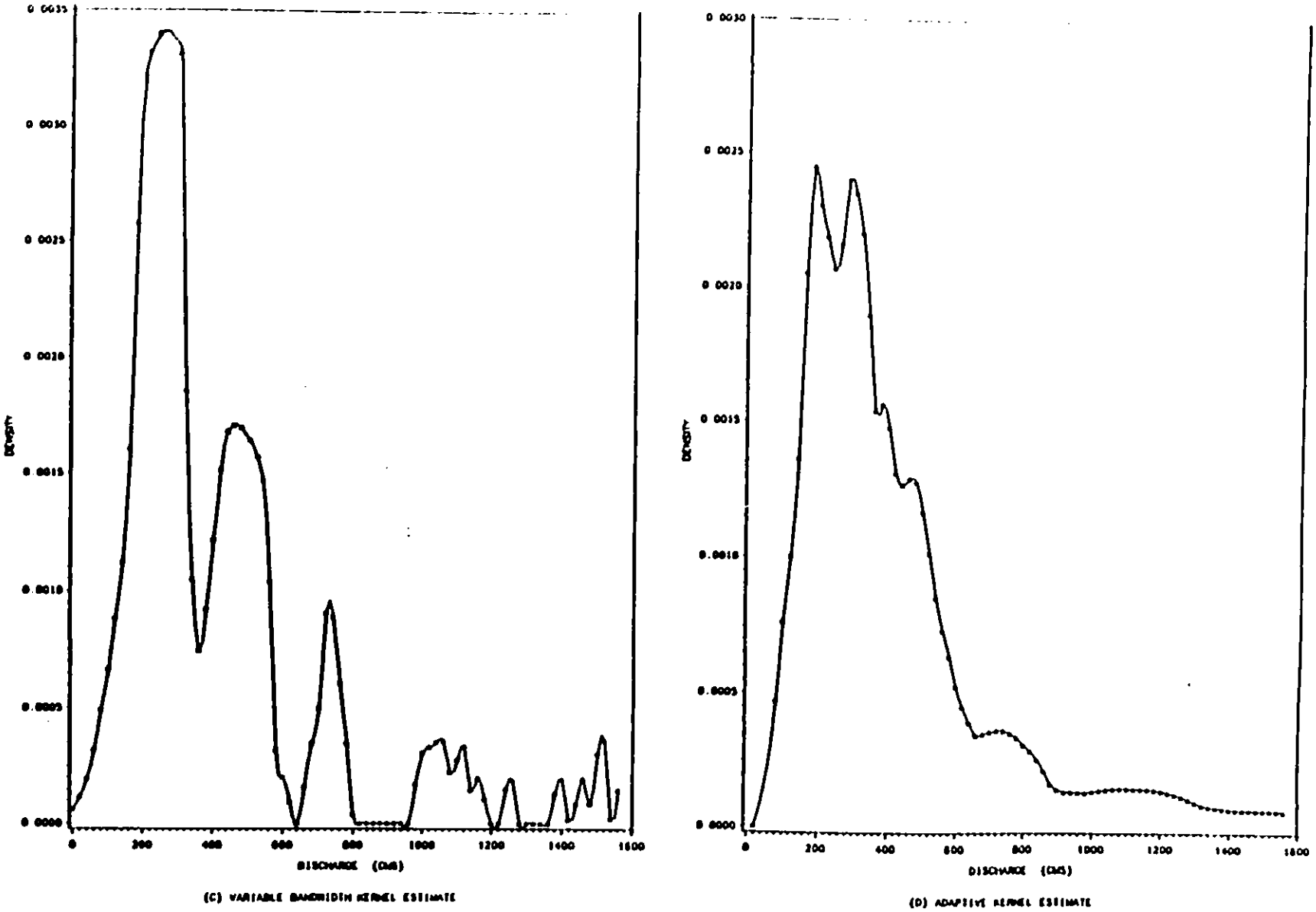


Figure 4.1: Smoothness and Tail Behaviour of the Nonparametric Density Functions (Warta River at Poznan)

and smoothness of the adaptive kernel is hardly affected by a change in the α value. In fact, it can be observed that any variability in the sensitivity parameter α affects mainly the smoothness of the tail only, while the bimodality characteristic at the central part of the distribution is preserved. Moreover, as α increases the tail of the density becomes less oscillatory. Fig(4.2-D) shows the density plot estimated by the adaptive kernel technique with a value of $\alpha = 0.8$. For the Warta river data set, this seems to be an optimal choice.

Another issue related to the adaptive kernel is the choice of the pilot estimate $\tilde{f}(\cdot)$ (eq.3.34) as an intermediate step for the determination of the local bandwidth factor λ_i (eq.3.35). Silverman (1986, p.101) confirmed that the method is insensitive to the final detail of the pilot estimate, and hence any convenient estimate can be used. To free this procedure from any apriori distributional assumption, parametric estimates have been avoided and the fixed kernel is used as the pilot. It can be concluded that the adaptive kernel is the best method to achieve better tail behaviour without obscuring the general characteristics of the distribution at the central part of the density.

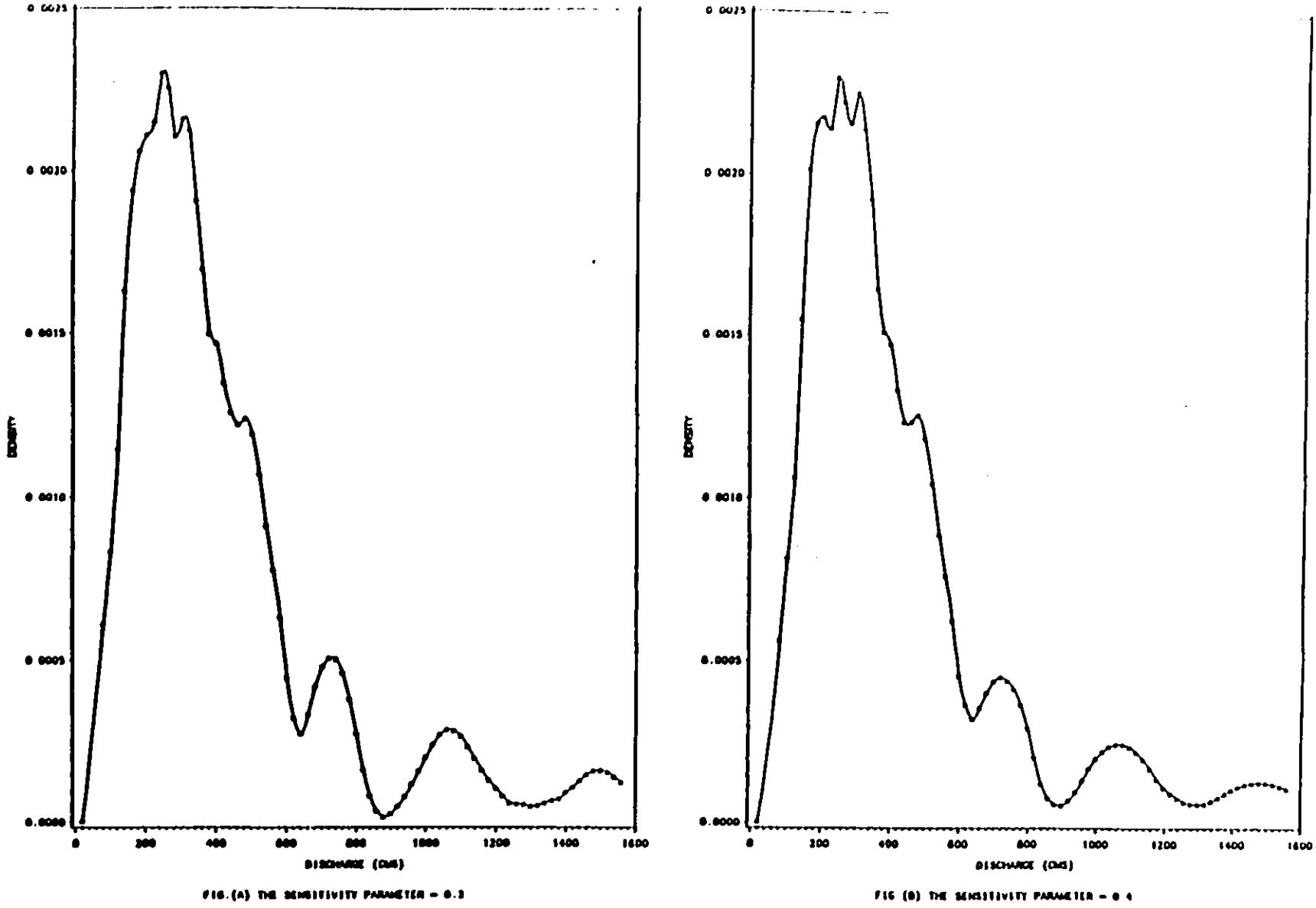


Figure 4.2: Effect of Varying the Sensitivity Parameter (α) on the Adaptive Kernel Density Function (Warta River at Poznan)

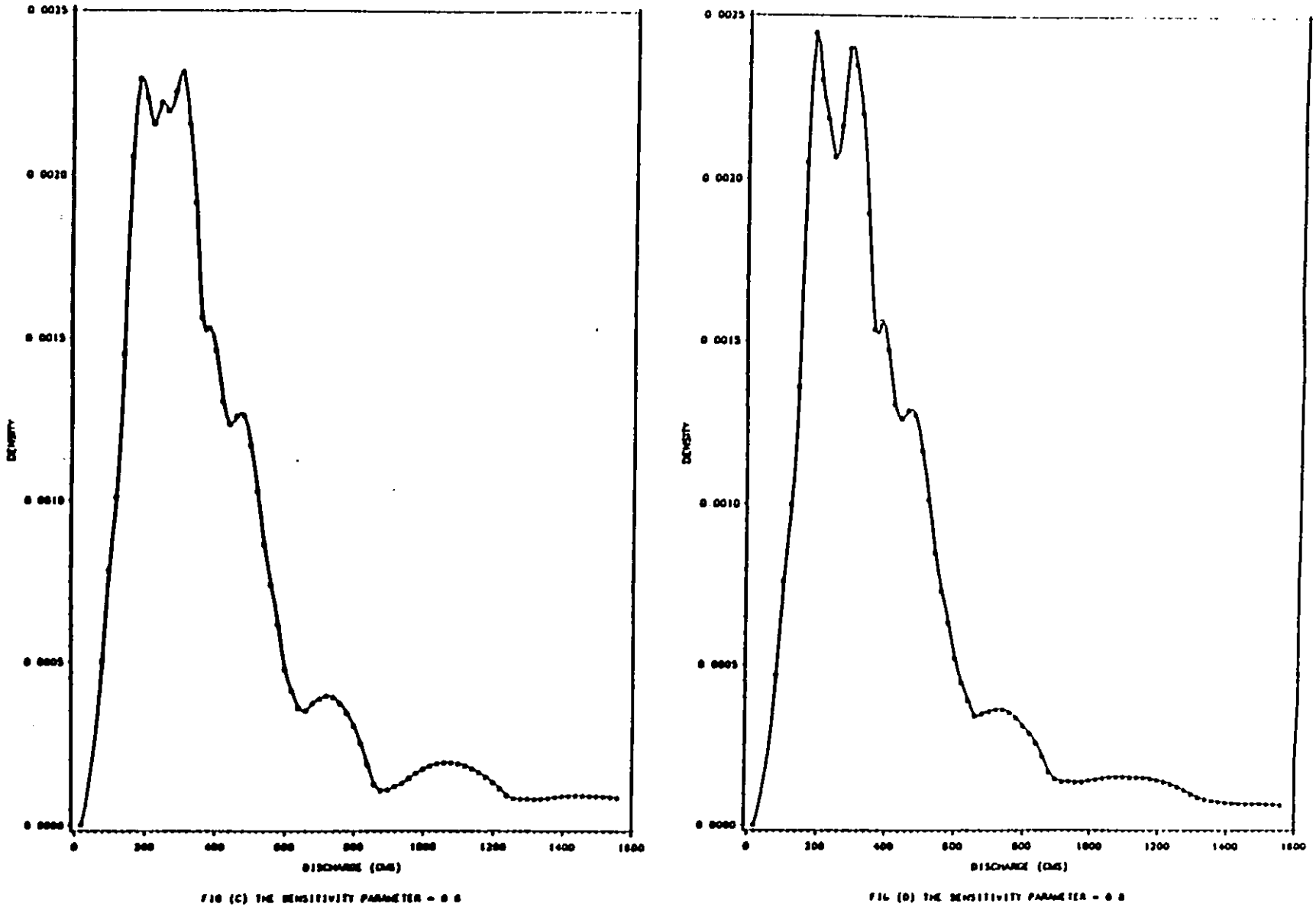


Figure 4.2: Effect of Varying the Sensitivity Parameter (α) on the Adaptive Kernel Density Function (Warta River at Poznan)

4.5 Mixed Distributions Estimated by the Nonparametric Technique

In the traditional flood frequency analysis , it is usually assumed that the annual flood series is a sample from a single population. A number of investigators, including (Hazen, 1930 and Hizer, 1956), have recognized that this assumption is not always valid. Therefore, mixed distribution approaches were introduced (Rouselle et al 1971, Waylen and Woo 1982 and Rossi et al 1984). Although physically justifiable, these approaches add an additional statistical complexity to the analysis. The nonparametric method has been found to be especially suitable for mixed distribution (Labatiuk, 1985). In this study, it is hypothesized that each mode in a nonparametric density function corresponds to a flood generating mechanism (i.e. snowmelt, rainfall or a combination of the two components)

Station 05PB014 of the Turtle river data is an example where annual flood series should not be considered as a sample from a single population. Annual flood maxima recorded at this station occurred in summer and winter time as shown in Appendix (A). Thus, a seasonal based separation is possible. Two different attempts were used to identify the snowmelt from the rainfall component.

In the first method, the data were separated arbitrarily into summer (Jun. to Nov.) and winter (Dec. to May) floods. A distinct histogram analysis for each set is performed, summer and winter modes are located and su-

perimposed on the combined density. Choosing the number of classes in the range of 5 to 8, the modes were found to be 75 and 155 cms for summer and winter floods, respectively. Fig.(4.3) displays the results of this technique. As it can be noticed, there is a quite good agreement between the modes found by histogram analysis for both seasons and those of the nonparametric density function.

The second method used to separate the snowmelt from the rainfall component considers the concept of antecedent precipitation criterion (APC) (Waylen and Woo, 1982). Daily precipitation and temperature data were collected from monthly meteorological records available at Environment Canada Library, and are shown in Appendix (A) (Table A.6). Fig.(4.4) shows the annual flood discharge versus the one day APC for the Turtle river. The data suggest that an arbitrary value of 30 mm of APC will distinguish the two mechanisms satisfactorily. The choice of the 30 mm threshold is also physically reasonable, since below this level rain - induced annual floods can unlikely be generated. Furthermore, floods above this demarcation occurred in the period from June to October suggesting a rainfall component. Floods below the 30 mm APC, however, happen strictly in the spring time (April to June), where a snowmelt is more likely to occur. Fig.(4.4) suggests also that these annual maxima can perhaps be further subdivided into two different clusters: one around 70 cms and the other around 155 cms. For the former, floods were associated with antecedent temperatures below 20 C, possibly suggesting a simultaneous snowmelt and rainfall mechanism. For the latter, on the other hand, peakflows occurred

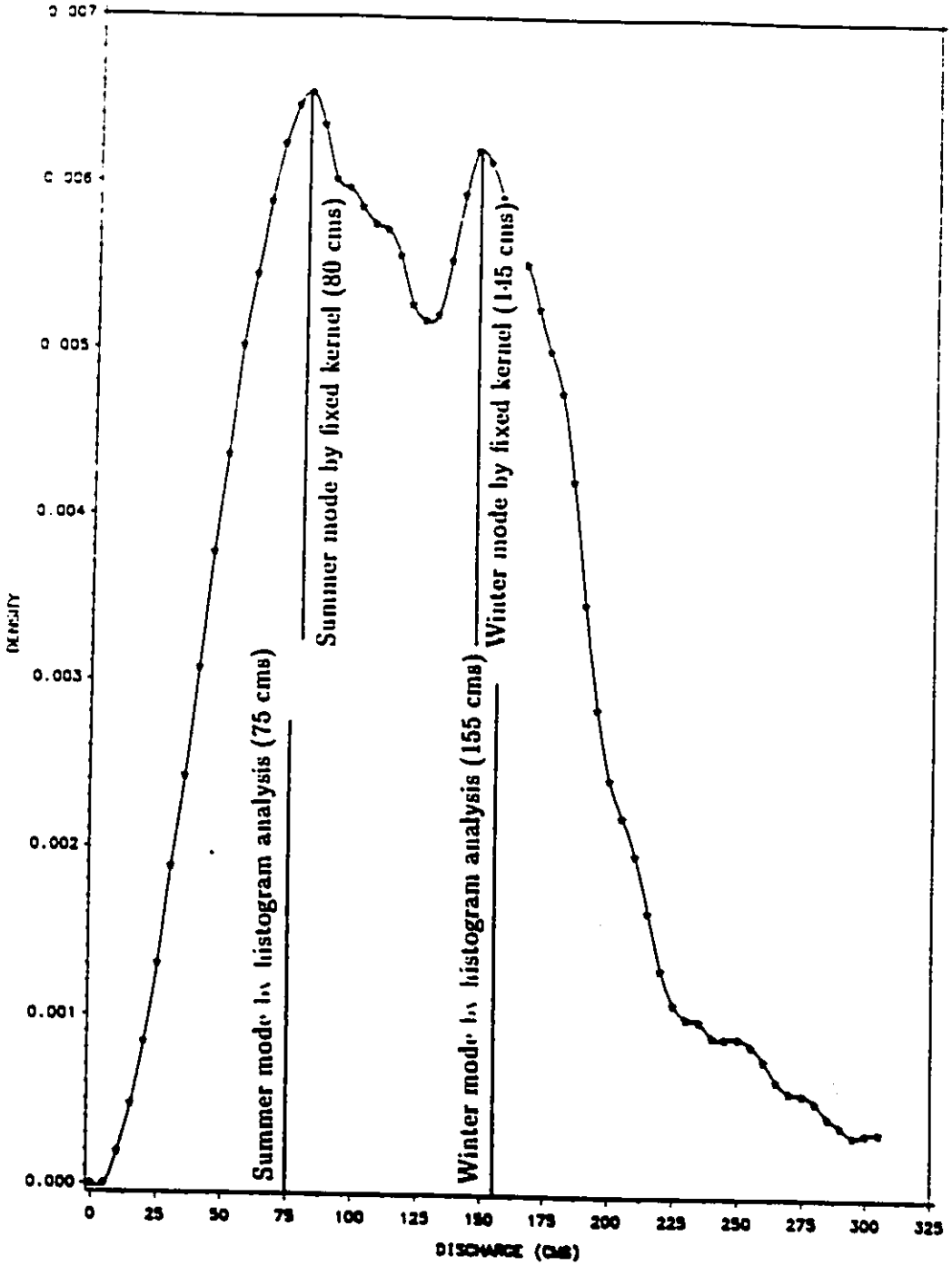


Figure 4.3: Summer and Winter Flood Modes as Determined by Histogram Analysis and as Superimposed on the Combined Nonparametric Fixed Kernel Density (Turtle River near Mine Center)

on the average at temperatures above 20 C, implying a purely snowmelt component. This is well explained by Fig.(4.5-A), which is a plot of the density function for floods with an APC less then 30 mm. This density exhibits two modes, 68 cms for floods caused by a simultaneous snowmelt and rainfall mechanism and 148 cms for purely snowmelt induced floods. The density plot on Fig.(4.5-B) (floods with an $APC < 30\,mm$) displays only one mode, 104 cms for purely rainfall generated floods.

Hence annual flood series of the Turtle river data appear to be generated three different runoff processes. Consequently, one would expect the combined density to be trimodal. Fig.(4.5-C) shows only two modes, 80 and 145 cms, for the two more dominant components, rainfall and snowmelt respectively.

Therefore, each mode in the nonparametric density function is associated with a flood mechanism. This shows that annual flood series such as the Turtle river data are not necessarily identically distributed and can not be treated with the conventional unimodal parametric distributions. The parametric analysis assumption that the AFS is a sample from a single population is questionable. This assumption would not take into consideration different flood causative factors. The nonparametric method, compared to other parametric techniques suggested for treatment of AFS of mixed distributions, involves much less computational efforts and most important of all does not require the apriori distributional choice.

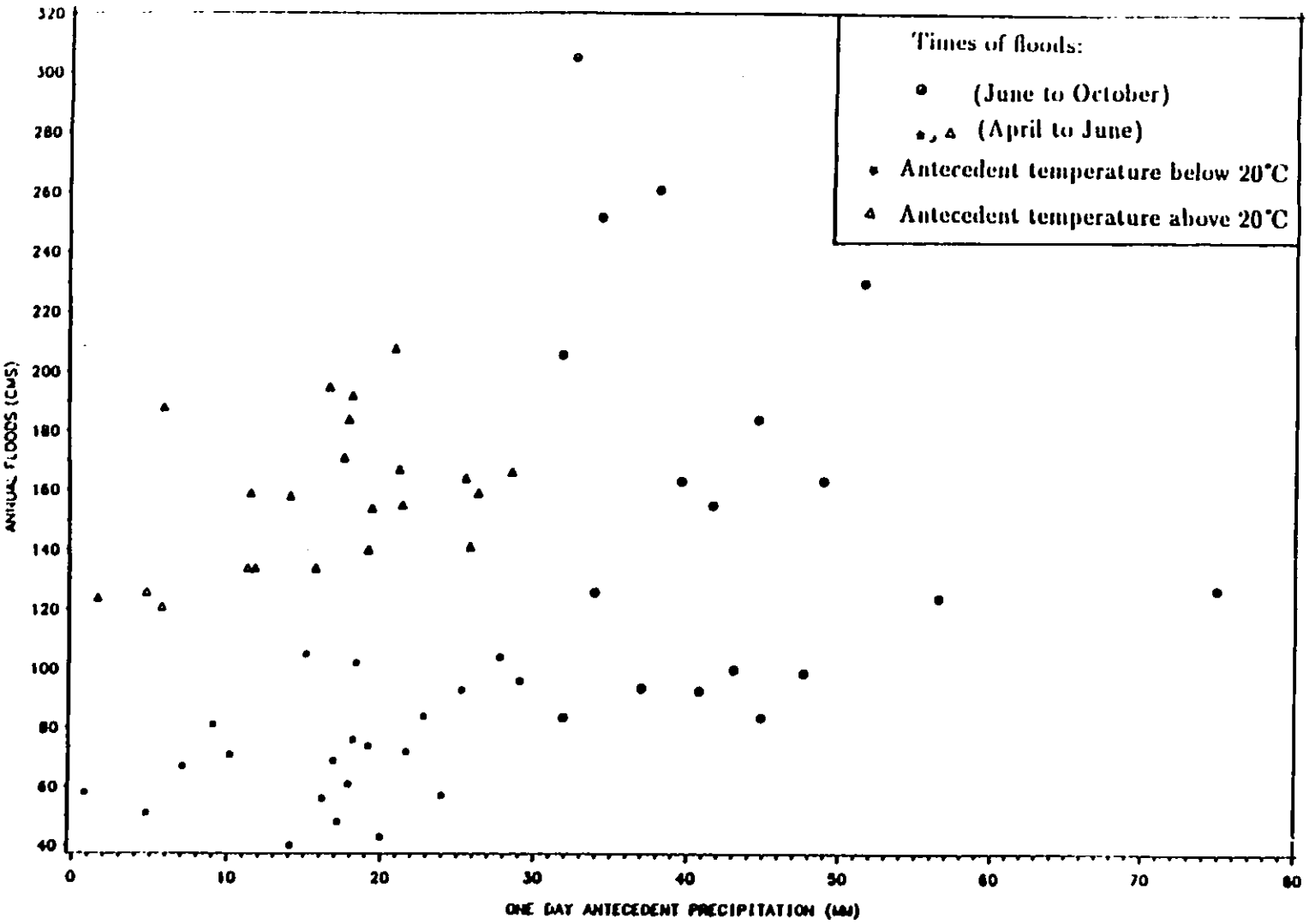


Figure 4.4: Plot of the Annual Floods versus the One Day Antecedent Precipitation (Turtle River near Mine Center)

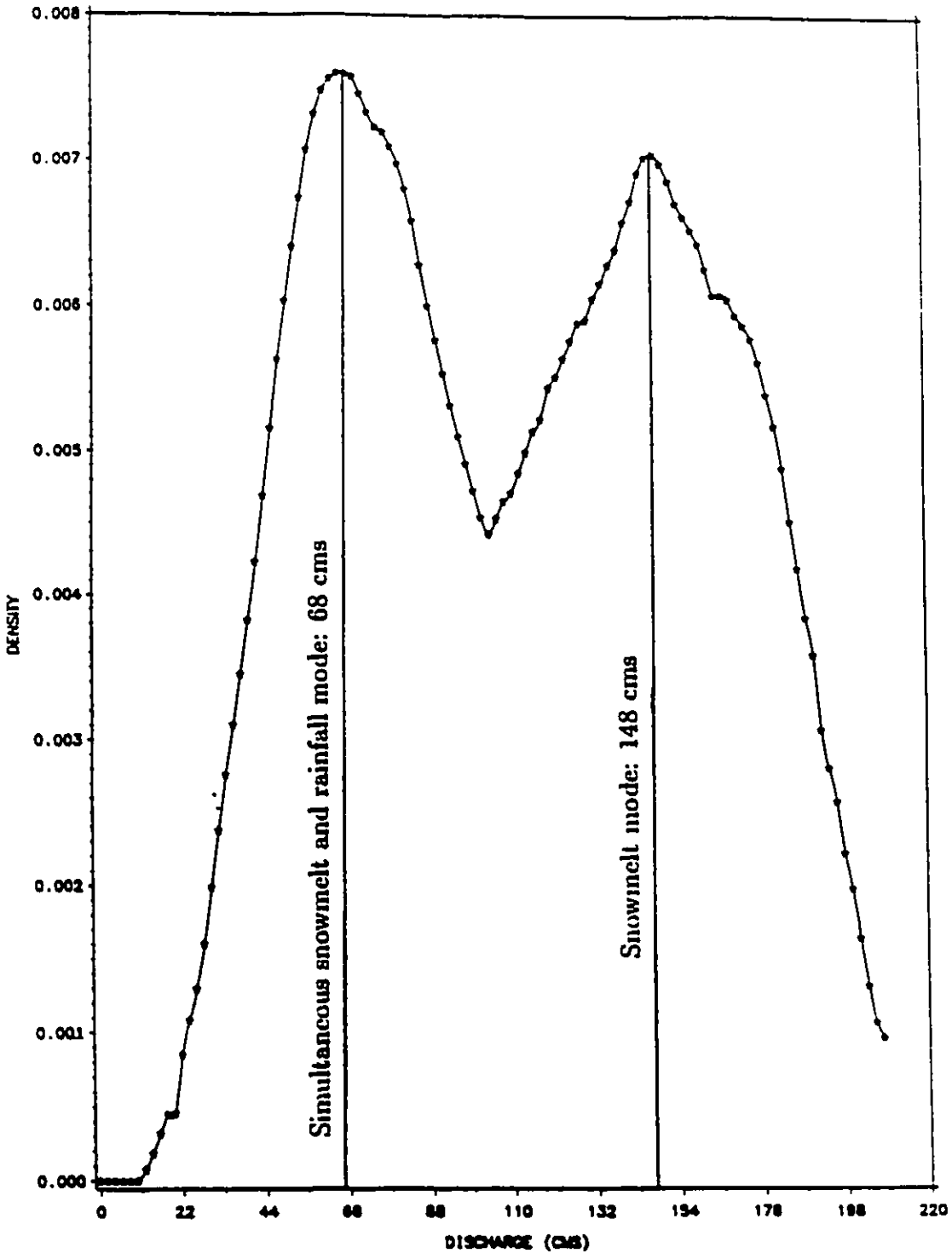


Figure 4.5-A: Fixed Kernel Density (floods of APC < 30 mm, Turtle river data)

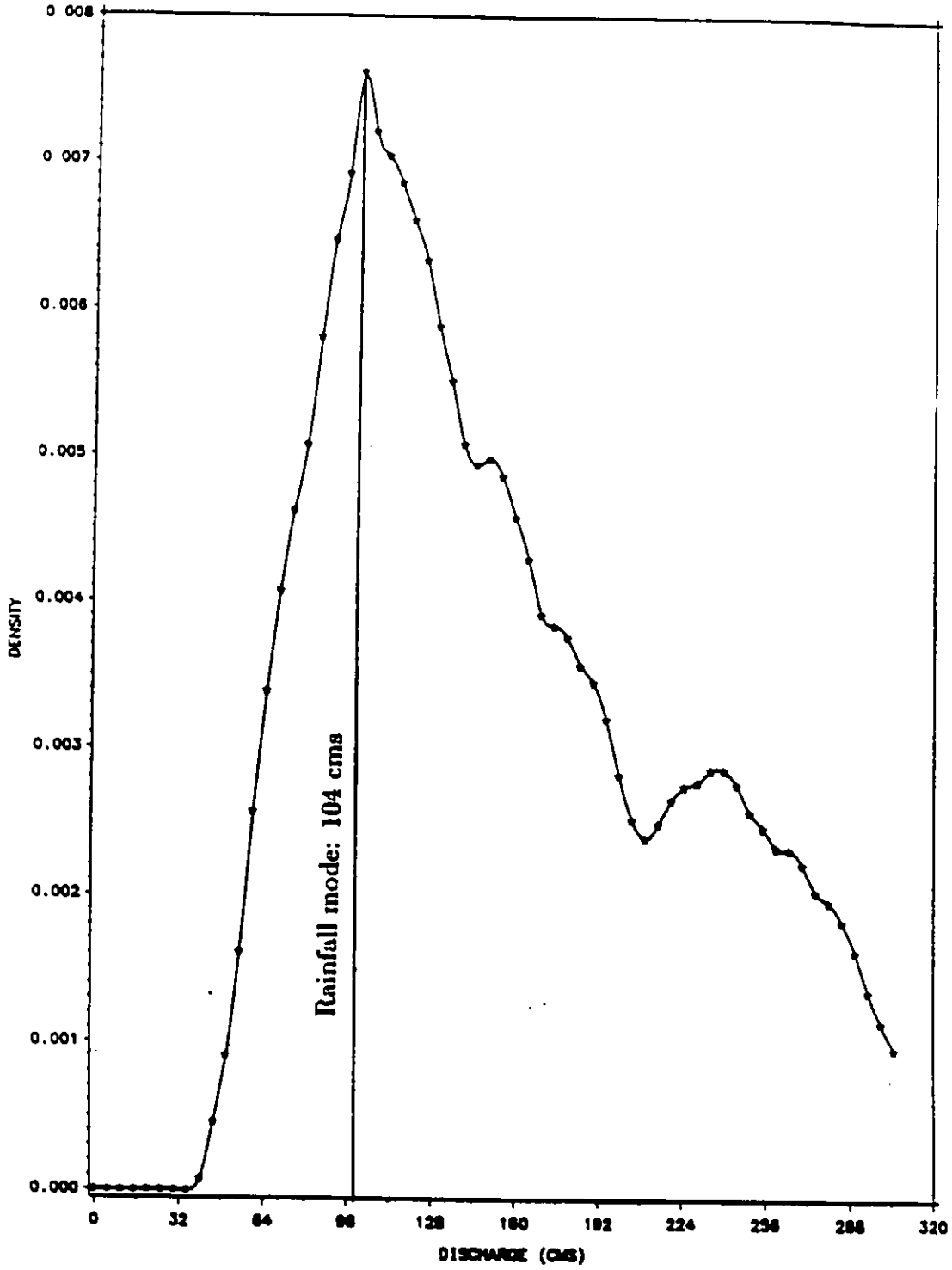


Figure 4.5-B: Fixed Kernel Density (floods of APC > 30 mm, Turtle river data)

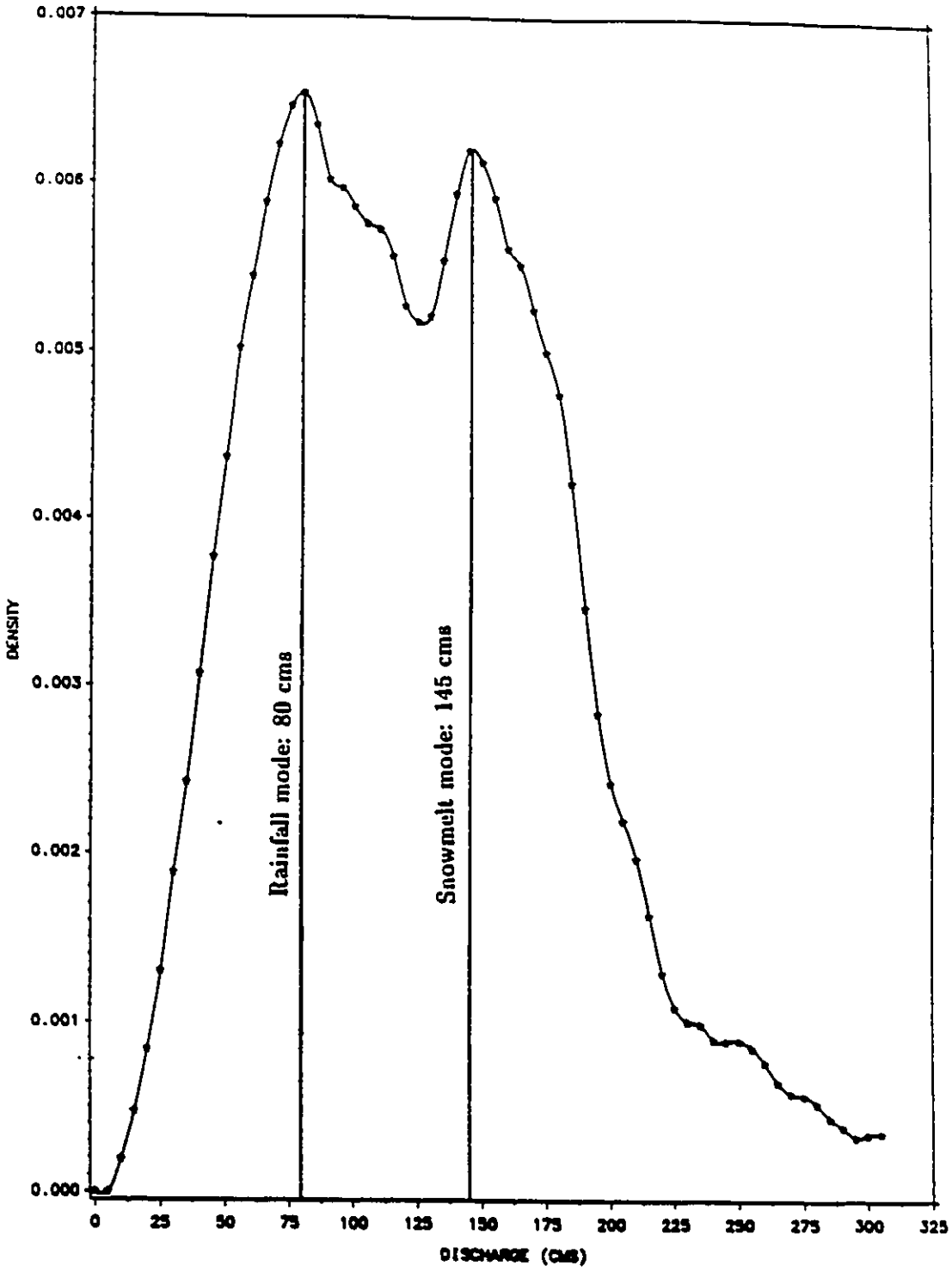


Figure 4.5-C: Fixed Kernel Density of the Combined Floods (Turtle River Data)

4.6 Data Generation for Simulation Experiment

Since the shape of the underlying distribution is not known, a Monte-Carlo experiment is considered for the analysis whereby the population values of quantile flood estimates are known. To conduct this experiment synthetic data with the same statistical properties as the observed data are used. The synthetic data are supplied by Environment Canada in already generated form. These data have been used extensively for the comparison of performance of different parametric distributions in the Consolidated Frequency Analysis Package (Pilon, et al 1985). The parent distribution considered in data generation is the Log-Pearson type III. This choice of LP III is an arbitrary one. The generation was based on the Wilson-Hilferty transformation (Pilon, et al 1987)

$$\ln(X_T) = C + A\left[\frac{t}{3B^{1/6}} - \frac{1}{9B^{2/3}} + B^{1/3}\right]^3 \tag{4.2}$$

Where A,B and C are the scale, shape and location parameters respectively, T is the return period, X_T is the corresponding LP III variate and t is the standard normal deviate defined as:

$$t = \frac{\left(\sum_{i=1}^{100} u_i - \frac{100}{2}\right)}{\sqrt{\frac{n}{12}}} \tag{4.3}$$

in which u_i's are random numbers for a uniform distribution on the interval [0,1].

Typical Canadian river values for A,B and C are used, such as $A = -0.06$,

$B = 25$ and $C = 7$. The synthetic data, which comprise fifty thousand observations, will be used as the basis of the sampling experiment in the following three sections (4.6,4.7 and 4.8), which deal respectively with (1) a comparison of the performance of the kernel methods in quantile estimation, (2) effect of sample length variation on the accuracy of floods estimated by the nonparametric approach and (3) the use of historical information in nonparametric flood frequency analysis.

To evaluate the performance of different flood quantile estimators in the simulation experiment, various statistical measures are used, namely, bias, precision and accuracy. Bias is the difference between the estimated and the true value of a statistic obtained by random sampling. So an algorithm is said to be unbiased when it yields estimates relatively close to the population values. Precision or Standard Error of Estimates (SEE), however, is the ability of an algorithm to give consistent results. This would imply that a technique is said to be precise if it gives results of different analysis sufficiently close to each other. The accuracy comprises both bias and precision. It is represented in this thesis by the Root Mean Square Error (RMSE) defined as:

$$MSE = (Bias)^2 + Variance \tag{4.4}$$

$$RMSE = (MSE)^{0.5} \tag{4.5}$$

Different statistical measures will be used in different simulation experiments, depending on the objectives to be met.

4.7 Comparison of the Asymptotic Behaviour of the Kernel Methods

Tables (4.3) and (4.4) summarize the results of the simulation analysis of the four different nonparametric approaches; fixed, transformed, variable and adaptive kernel estimators. The simulation is based on one thousand repetitions each of fifty years of record. The population means of the design floods listed in column two, were computed using eq.(4.2) with the LP III parent parameters. The bias in percent,the standard error of estimate (SEE) in percent and the RMSE are also listed.

From table (4.3) several observations related to bias can be made. Generally, the bias increases with an increasing value of the return period. Furthermore, the nonparametric procedures tend to underestimate the flood event magnitudes of a return period beyond 100 years (negative bias), except for the adaptive and triangular transformed kernels. For flood events of return periods less than 100 years, on the other hand, the circular transformed and the variable kernel result in the least bias of all algorithms; Nevertheless, no method displays excessive bias (a maximum of 5 % and a minimum of 0%) Finally, when the transformed kernels are compared with each other, it is noted that the circular transformed exhibits less bias than the triangular transformed kernel.

The SEE values (Table 4.4) indicates that almost all algorithms have the same order of precision. The variation of the SEE is so small that it can

hardly be differentiated. In terms of RMSE, however, all procedures display a decreasing accuracy as the return period increases. The circular transformed kernel is more accurate than the triangular transformed in estimating quantiles. Among the four kernel estimators, on the other hand, the fixed kernel exhibits the lowest RMSEs.

By combining all of the above observations, it can be concluded that the fixed kernel gave the best estimates in terms of accuracy. The variable kernel did not improve the results probably because of the inefficiency of the maximum likelihood method in estimating the parameters. In fact, the multimodal nature of the nonparametric density might cause the likelihood method to yield a local maximum rather than an optimal one. Furthermore, the adaptive kernel did not also improve the accuracy of flood estimates. This is, perhaps, due to the fact that its sensitivity parameter α can not be controlled in a simulation experiment. It can only, however, be fixed at a constant value (i.e. $\alpha = 0.8$). Among the transformed kernels the circular transformed performed better than the triangular transformed. This is congruous with the theoretical results mentioned in section (3.3). The SEE, which is a measure of precision, did not exhibit a great deal of variation across the kernel methods. For more reliable conclusions based on this measure of precision, the number of repititions (N) in the simulation experiment might have to be increased. In terms of computer execution time requirements, the four methods used are ranked in ascending order: fixed, transformed, adaptive and variable kernel estimators. To illustrate this, simulation time requirements of one thousand samples, each of fifty

Table 4.3: Simulation Results (Mean Estimates and Bias).

Flood Event	Population Mean	Mean Estimate					Bias (%)				
		FK	TTK	CTK	VK	AK	FK	TTK	CTK	VK	AK
Q10	354	357	358	355	356	351	0.8	1.1	0.2	0.6	-0.8
Q50	424	429	444	428	425	430	1.2	4.7	0.9	0.2	1.4
Q100	450	449	463	446	446	465	-0.2	2.9	-0.9	-0.9	3.3
Q200	475	461	474	456	456	500	-2.7	0.0	-4.0	-1.9	5.0

Table 4.4: Simulation Results (Standard Error of Estimates and Root Mean Square Error).

Flood Event	Population Mean	SEE (%)					RMSE				
		FK	TTK	CTK	VK	AK	FK	TTK	CTK	VK	AK
Q10	354	1.2	1.2	1.4	1.2	1.2	19.0	19.7	21.6	19.4	19.6
Q50	424	2.2	2.5	2.3	2.1	2.2	34.3	39.8	36.1	33.3	34.5
Q100	450	2.6	2.6	2.6	2.7	2.8	41.0	43.0	42.7	43.2	44.8
Q200	475	2.7	2.8	2.7	3.7	3.4	47.0	44.9	43.5	59.2	54.8

Note: FK, TTK, CTK, VK and AK are the fixed, triangular transformed, circular transformed, variable and adaptive kernel estimators respectively.

years of record, for the four algorithms are given in table (4.5).

Table 4.5: Computer Time Requirement for Simulation Based on 1000 Repititions.

Kernel Method	CPU Seconds
Fixed	885
Transformed	2803
Adaptive	5560
Variable Bandwidth	8495

4.8 Effect of Sample Size on Quantile Estimation

The sample size plays an important role in parametric analysis, since for a distribution to be known, a larger length of record is needed. The sample size available to hydrologists for flood frequency analysis supports the use of nonparametric techniques even for small and intermediate sample sizes. This section addresses the issue of sample length variation and its effect on the accuracy of quantile estimates. For this purpose, a limited experiment is designed. It consists of simulations with 500 sequences each, for different record lengths (n) ranging from 10 to 100 with an increment of 10 years.

Each time, quantiles (Q_T) corresponding to return periods of 5, 10, 15, 100 and 200 years are computed and compared with the population mean values. Plots of the discharge Q_T versus the record length n are displayed in Fig (4.6 - 4.10). RMSE's are also calculated and two illustrative plots of RMSE versus n are presented in Fig (4.11).

It can be observed from Fig.(4.6 to 4.10) that a sample size of $N_c = 20$ to 30 years is a remarkable threshold. Below this level, the nonparametric method exhibits considerable bias. Furthermore, as the sample size increases beyond N_c, the variation in the bias is much less accentuated and the quantile estimate reaches asymptotically the population mean value. On the other hand, the mild slope in Fig (4.11-A) suggests that for relatively low return periods the rate of decrease of the RMSE is small above N_c. For larger return periods, however, Fig (4.11-B) shows that the rate of convergence of the flood estimate to its true value is relatively slow and any additional records even beyond 100 years still make a considerable decrease in the RMSE.

It would therefore appear that the nonparametric kernel approach in flood frequency analysis requires a minimum record length of about 25 years to yield reliable estimates.

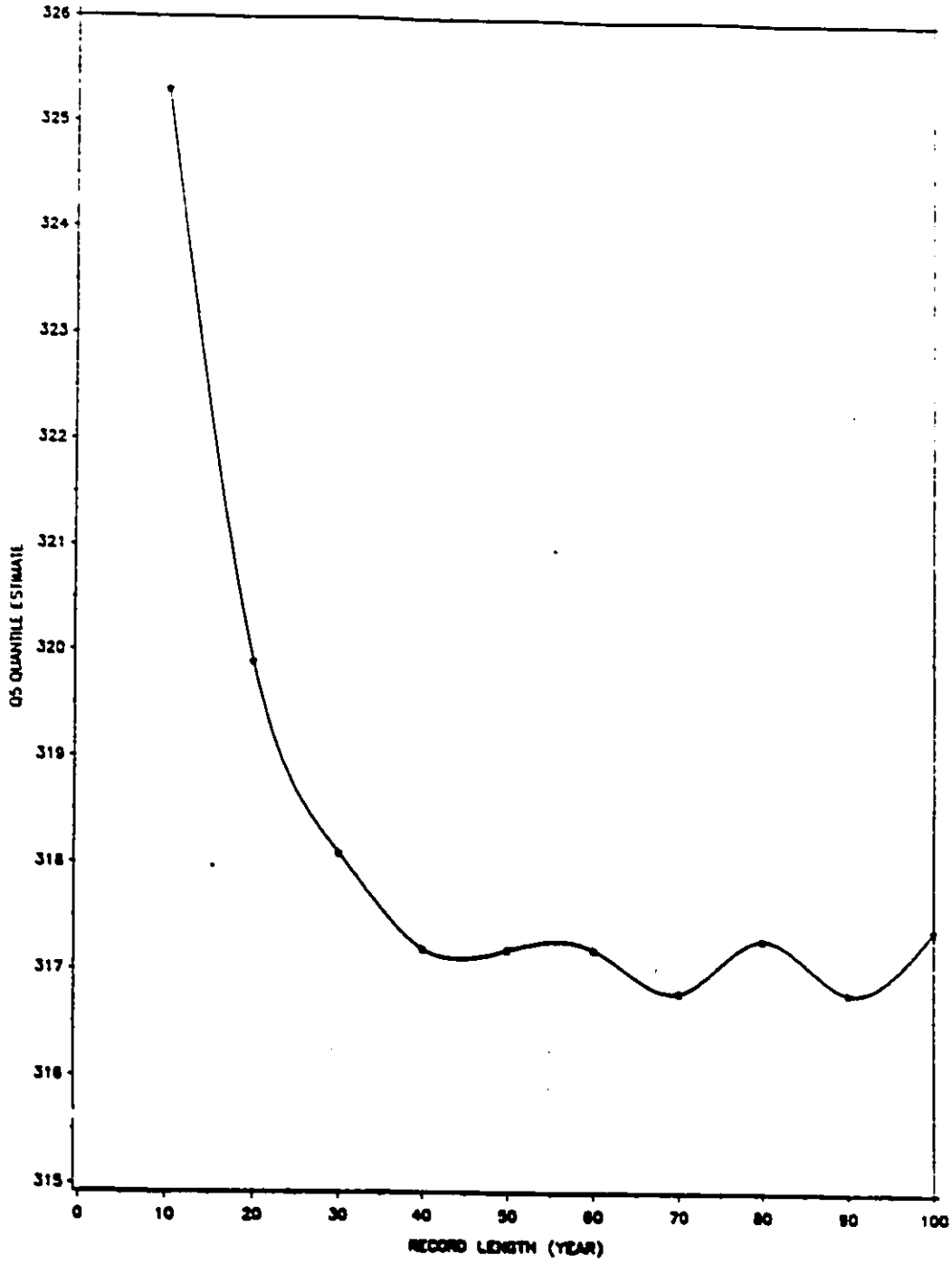


Figure 4.6: Effect of Record Length Variation on the 5 Year Quantile Estimate (Population Mean Value Q5= 316 cms).

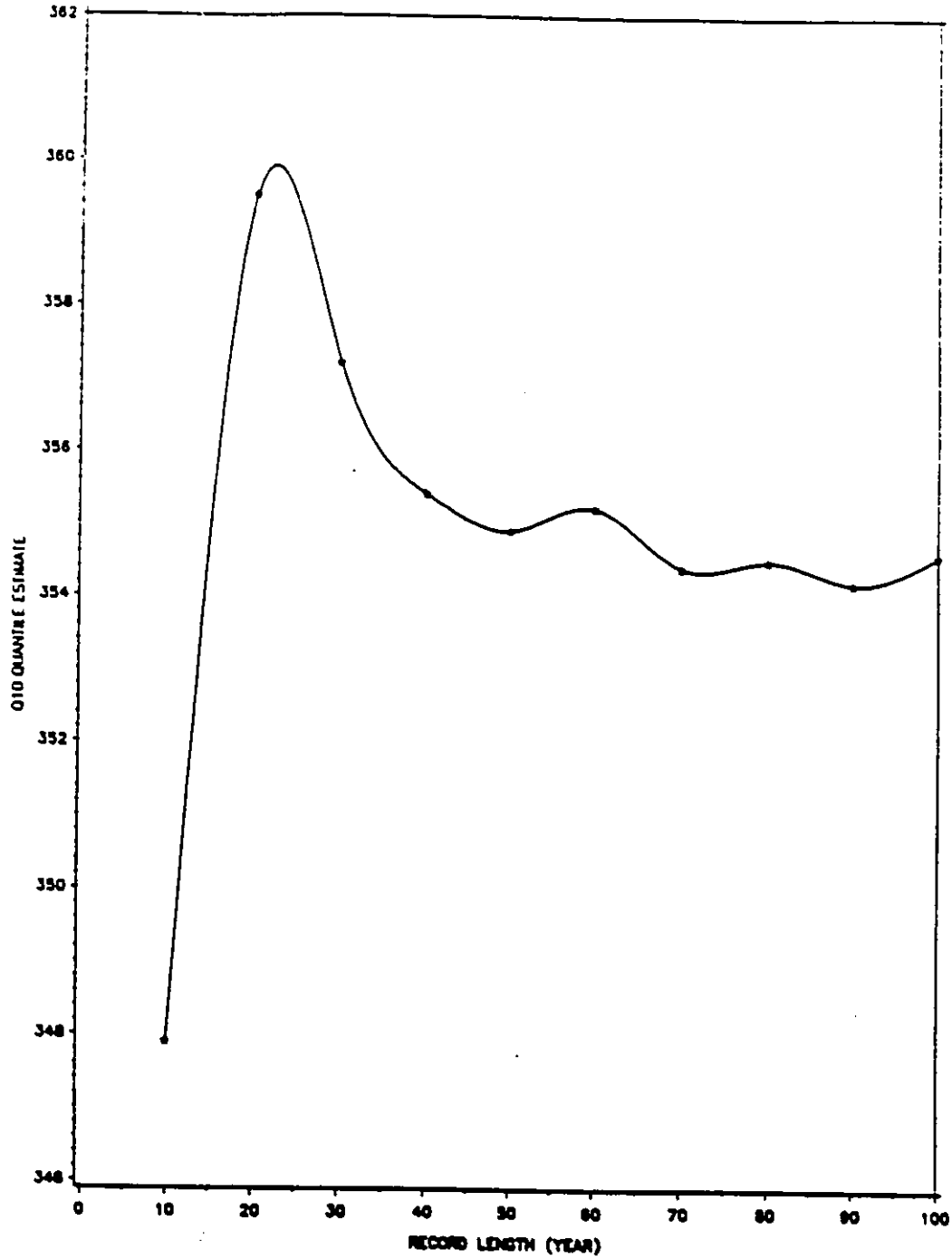


Figure 4.7: Effect of Record Length Variation on the 10 Year Quantile Estimate (Population Mean Value Q10= 354 cms).

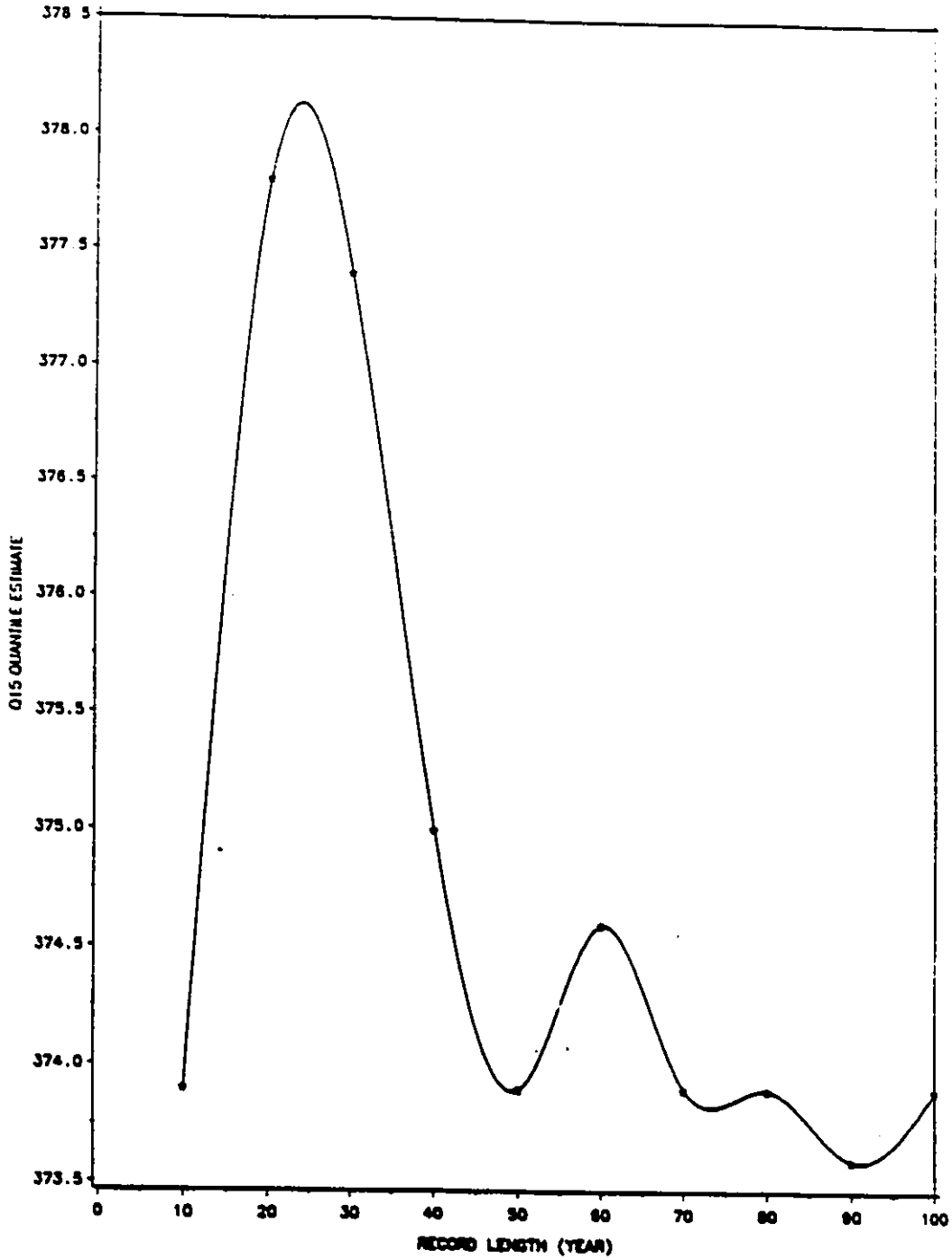


Figure 4.8: Effect of Record Length Variation on the 15 Year Quantile Estimate (Population Mean Value Q15= 374 cms).

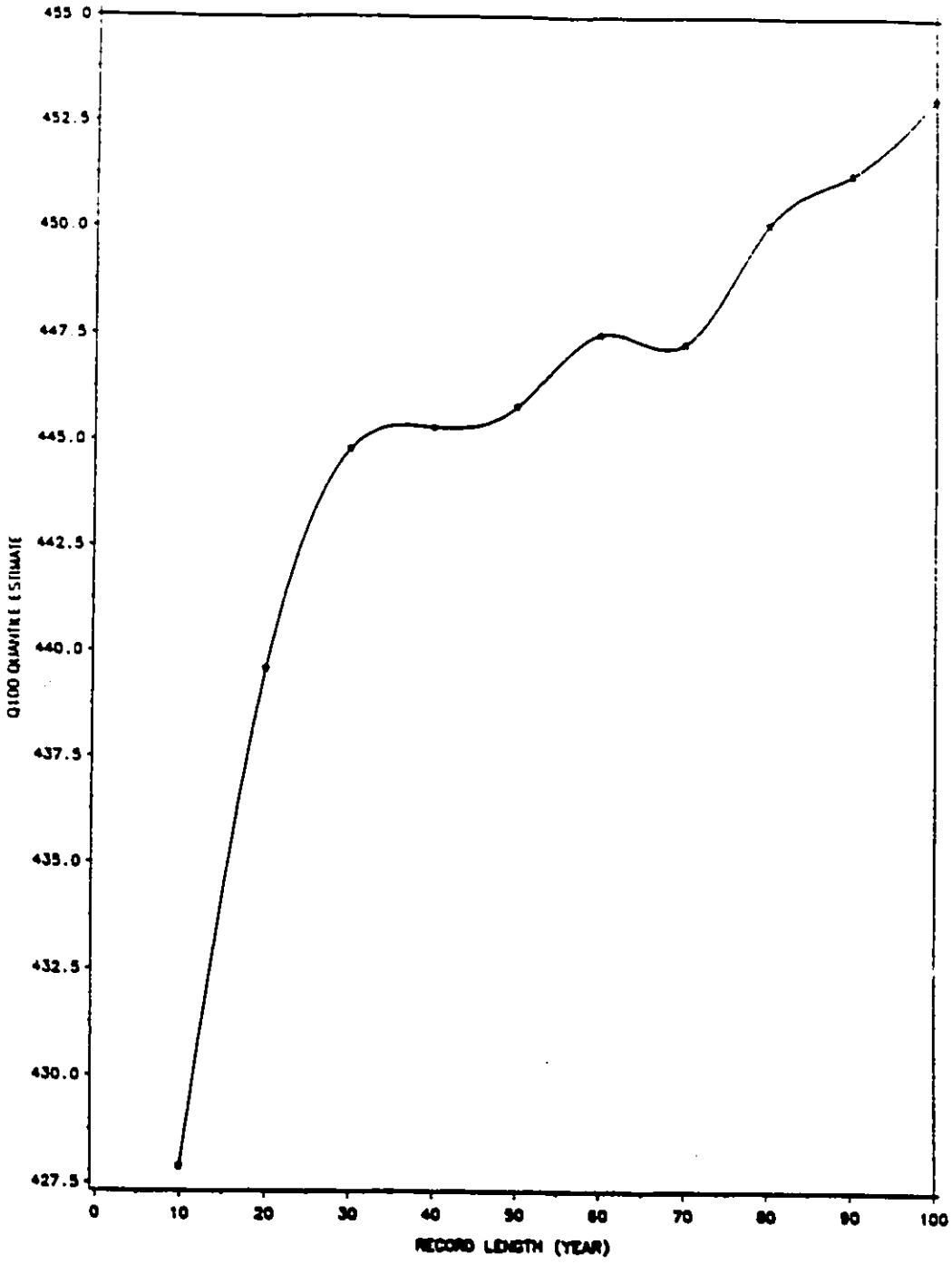


Figure 4.9: Effect of Record Length Variation on the 100 Year Quantile Estimate (Population Mean Q100= 450 cms).

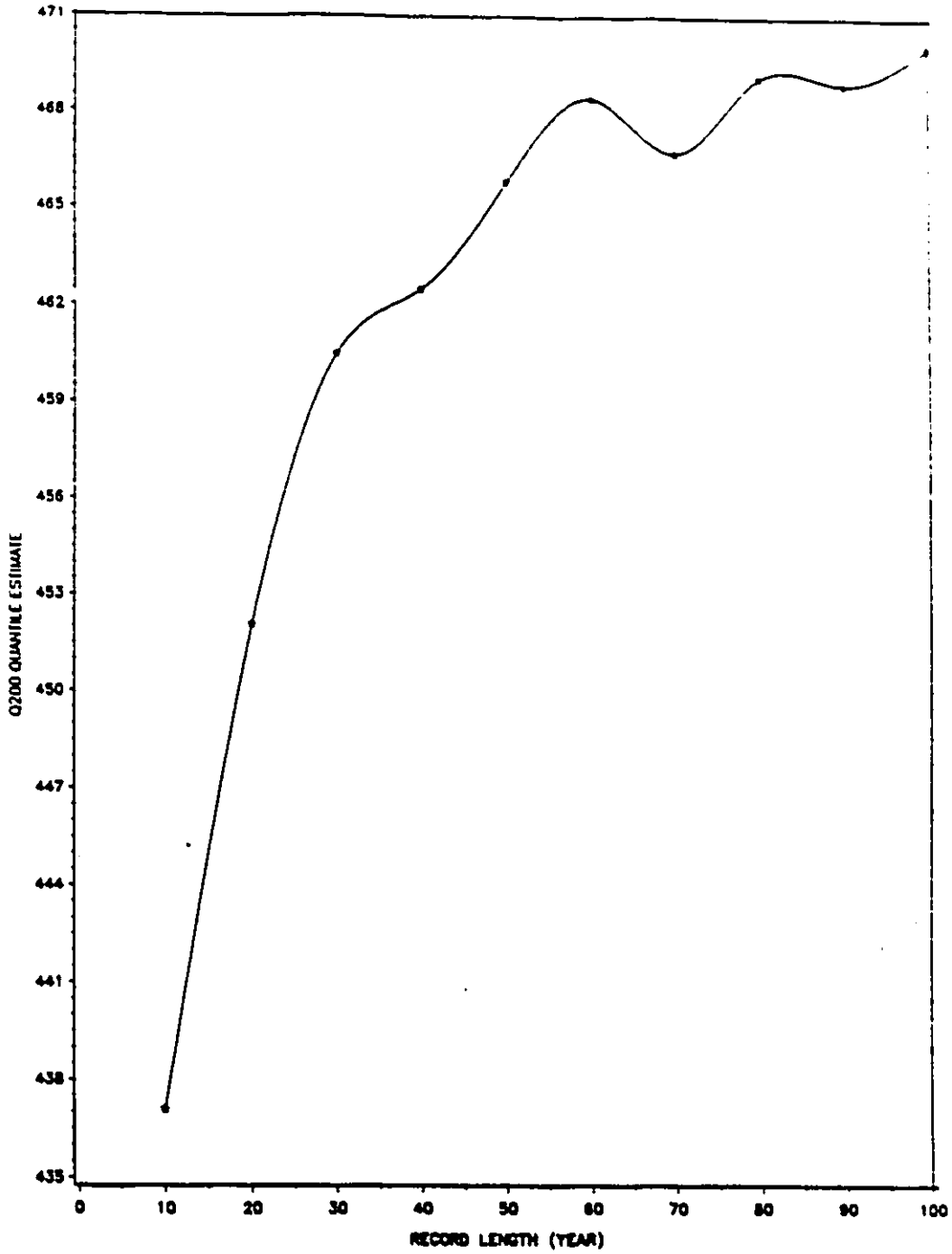


Figure 4.10: Effect of Record Length Variation on the 200 Year Quantile Estimate (Population Mean Q200= 475 cms).

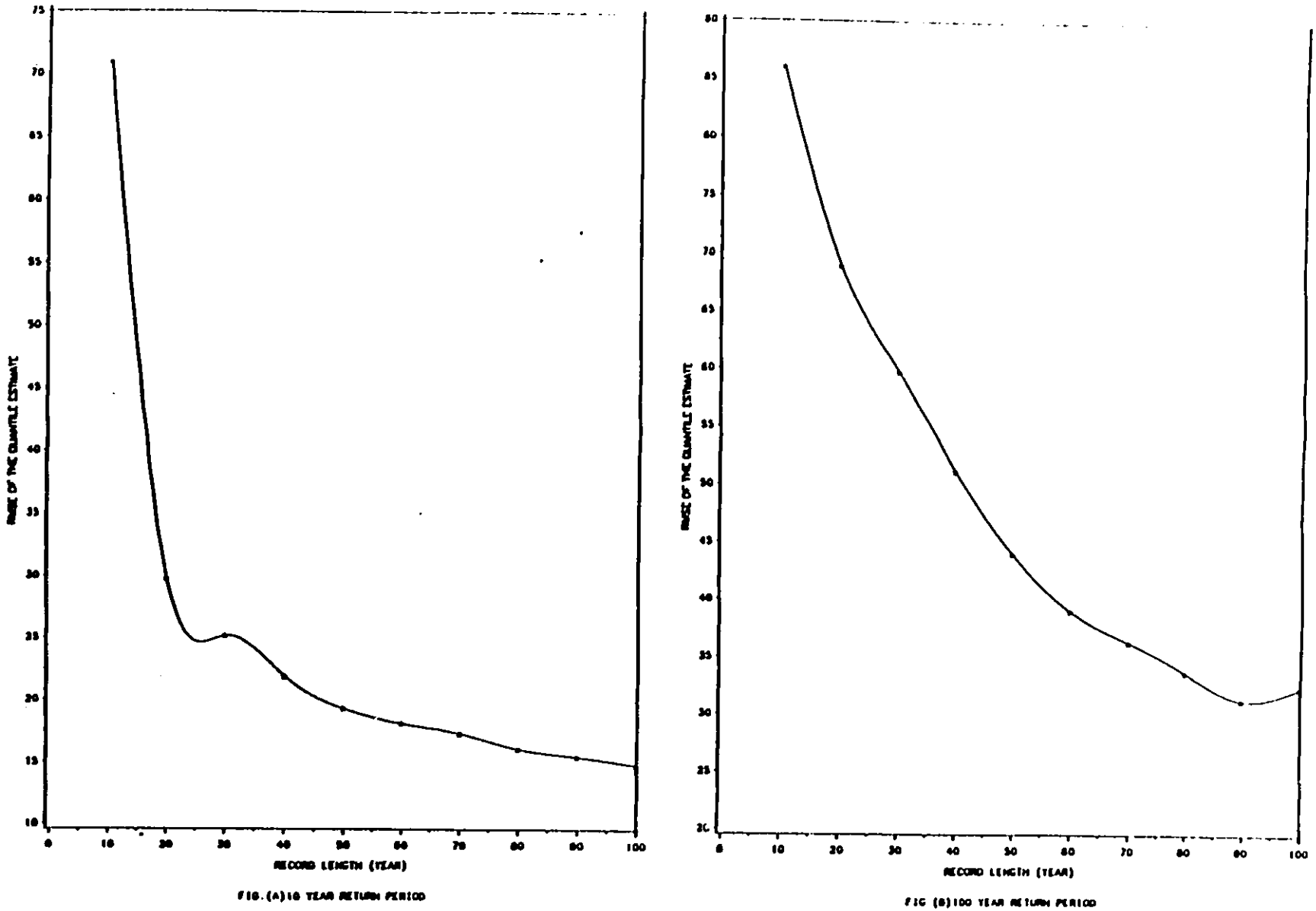


Figure 4.11: Variation of the Accuracy (RMSE) of Flood Estimates with the Record Length.

4.9 Use of Historical Data in Nonparametric Analysis

The shortage of flood data records is a common problem in performing flood frequency analysis, especially when estimating floods of high return periods. Hosking (1986), Stedinger (1986) and others showed that the inclusion of historical data in parametric analysis is an efficient method in improving the accuracy of high return period flood estimates. In this section, the historical data are incorporated into nonparametric analysis. The procedure is verified numerically by a specially designed simulation experiment. Historical sequences are obtained from the simulated 50,000 observation data file (LP III as a parent). Choosing a historic time span of 101 years, the maximum extreme event (Q_c) in each 101 year sample is considered to be the censoring threshold. The first n values generated were then censored excluding the maximum value Q_c, if it fell in that set. This means that the remaining $(101 - n)$ observations are disregarded, yet the fact that their values are less than Q_c is considered to be known. Three hundred such sequences were generated for n values ranging from 10 to 50 years, with an increment of 10 and simulation was performed accordingly.

The conducted experiment consisted of two separate analysis: conventional (without use of historical information) and historical analysis. Standard error of estimates were computed and given in Table (4.6) for conventional and historical analysis. Five arbitrarily selected sample lengths (n=10, 20,

30 and 50 years) were considered. Table (4.7) shows the percent change in the SEE values due to the use of historical information. In this table, a positive value of the percent change in the SEE implies that the incorporation of historical data has decreased the variability of the flood estimates around their population mean values. A negative value of the percent change in the SEE indicates that the reverse phenomenon had occurred. In general, the SEE values have been reduced by the use of historical information (a maximum reduction of 20% and a minimum of 4%). The only exception is for the 100 and 200 year return period quantiles with a sample length of 10 and 20 years. This is congruous with the results obtained in section (4.7), where it was found that a minimum of 20 to 30 year sample length is required for nonparametric method to give reliable flood estimates.

As a result and within the limitation of this experiment, it can be concluded that the historical data improves precision of flood quantiles estimated by the nonparametric approach. The historic time span, however, should include a minimum of 30 years of systematic gauged records (fully defined floods). The conducted experiment consisted of two separate analysis: conventional (without use of historical information) and historical analysis. Standard error of estimates were computed and given in Table (4.6) for conventional and historical analysis. Five arbitrarily selected sample lengths (n=10, 20, 30 and 50 years) were considered. Table (4.7) shows the percent change in the SEE values due to the use of historical information. In this table, a positive value of the percent change in the SEE implies that the incorporation of historical data has decreased the variability of the

flood estimates around their population mean values. A negative value of the percent change in the SEE indicates that the reverse phenomenon had occurred. In general, the SEE values have been reduced by the use of historical information (a maximum reduction of 20% and a minimum of 4%). The only exception is for the 100 and 200 year return period quantiles with a sample length of 10 and 20 years. This is congruous with the results obtained in section (4.7), where it was found that a minimum of 20 to 30 year sample length is required for nonparametric method to give reliable flood estimates.

As a result and within the limitation of this experiment, it can be concluded that the historical data improves precision of flood quantiles estimated by the nonparametric approach. The historic time span, however, should include a minimum of 30 years of systematic gauged records (fully defined floods).

Table 4.6: Standard Error Values of Flood Estimates (SEE (Including) and SEE_0 (excluding) Historical Information)

Event	n = 10		n = 20		n = 30		n = 40		n = 50	
	SEE_0	SEE	SEE_0	SEE	SEE_0	SEE	SEE_0	SEE	SEE_0	SEE
Q10	3.75	2.25	1.67	1.56	1.38	1.33	1.27	1.15	1.15	1.04
Q50	4.04	4.45	3.00	2.77	2.35	1.96	2.14	1.78	1.78	1.67
Q100	4.66	5.95	3.86	4.62	3.46	3.35	2.88	2.65	2.65	2.25
Q200	5.08	6.93	4.44	5.83	4.10	4.73	3.86	3.70	3.86	3.46

Table 4.7: % Change in SEE Values Due to Use of Historical Information

Event	n = 10	n = 20	n = 30	n = 40	n = 50
Q10	66	8	4	10	10
Q50	-11	8	20	20	7
Q100	-22	-16	3	9	18
Q200	-27	-24	-13	4	12

Note: SEE is the standard error of flood estimates and n length of the systematic records.

Chapter 5

GENERAL CONCLUSIONS AND RECOMMENDATIONS

5.1 General conclusions

From real data analysis the following major conclusions can be made:

1. The fixed, transformed, variable, and adaptive kernel methods yield flood estimates in the same order of magnitude as those obtained by parametric analysis.

2. The transformed kernel did improve the tail behaviour of the density function when compared to the fixed kernel. This relative improve-

ment, however, obscured the bimodality nature of the data by an oversmoothness in the central part of the distribution.

3. The variable kernel exhibits the most oscillatory density function among the four kernels. This is perhaps due to the fact that the nonparametric density might be discontinuous and/or bimodal in which case the maximum likelihood procedure, used for parameter estimation, will lead to a local maximum rather than an optimal one.

4. Among the four kernel estimators, the adaptive kernel gave the best performance in terms of pronouncing the bimodal nature of the data and smoothing the right tail of the distribution.

5. It was hypothesized and verified that each flood generating mechanism is associated with a mode of the nonparametric density function of the corresponding annual flood series. The nonparametric method has successfully estimated mixed distribution of such mixed population flood data.

Based on the analysis of the synthetic data, the following major conclusions are made:

1. The standard error values of flood estimates were so close to each other that no preference of one kernel method over the others can be made decisively.

2. The fixed kernel approach gave the best flood estimates in terms of bias and accuracy.

3. The variable and adaptive kernels were less accurate in estimating flood quantiles. For the variable kernel this perhaps was due to the property of the maximum likelihood method in locating the optimal value of the smoothing parameters. For the adaptive kernel, however, this was caused by the subjectivity in choosing the sensitivity parameter α.

4. The circular transformed kernel exhibits smaller bias, standard error and root mean square error of flood estimates than the triangular transformed kernel. This was found congruous with the theoretical results.

5. The transformed, variable and adaptive kernel estimators are statistically more complex and are much more demanding in terms of computer time requirements than the fixed kernel. Therefore, when dealing with relatively low skewed distributions the fixed kernel is to be recommended.

6. From a limited numerical experiment, it can be suggested that the nonparametric kernel approach requires approximately 25 years of record length for its implementation.

7. The simulation experiment with historical information showed that the use of binomial data, with a minimum of 30 years of systematic records, did improve the precision of flood estimates.

5.2 Recommendations

1. Other alternative methods of estimating the variable kernel parameters should be tried to achieve better overall smoothness and tail behaviour of density estimates. The use of the minimization search procedure suggested by Breiman, et al (1977) is a possibility.

2. A simulation experiment might be designed to assess the effect of varying the sensitivity parameter α on the overall performance of the adaptive kernel estimator.

3. In a censored data sample the magnitude of historic flood peaks is known. This would imply that a censored data sample contains more information than a binomial sample, where only threshold exceedances are available. As a result the use of historical information with censored sample will improve more the accuracy of flood estimates. An extended investigation on historical data use in nonparametric approach along this line is recommended.

4. A comparison between parametric and nonparametric approaches based on a simulation experiment with bimodal data can be done to evaluate the merits and suitability of each method in reflecting various hydrologic regimes. Similarly, annual flood series where peakflows are generated by different type of storms (convective or frontal) can be investigated numerically to show how the nonparametric method can also handle such flood data of mixed distributions.

5. The transformed, variable and adaptive kernel approaches proposed for the treatment of heavy-tailed distributed flood data did not fulfill the required objective. Parametric / nonparametric mixture density estimation as an application in flood frequency analysis, introduced by Yakowitz, et al (1986), can overcome the inadequacy of the non-parametric approach in estimating long-tailed distributions. Research areas related to parametric / nonparametric mixtures are still open and highly recommended .

Chapter 6

BIBLIOGRAPHY

Bibliography

[1] Adamowski, K. (1985).*Nonparametric Kernel Estimation of Flood Frequencies*, Water Resources Research, Vol. 21, No. 11, 1585-1590.

[2] Adamowski, K. and V. Feluch (1986).*The Transformed Kernel Estimate of Flood Frequencies*, Department of Civil Engineering, University of Ottawa, Ottawa (Unpublished).

[3] Abramson, I.S. (1982).*On Bandwidth Variation in Kernel Estimates - a Square Root Law*, Ann. Statist., 10, 1217-1223.

[4] Anderson, G. D. (1969). *A Comparison of Probability Density Estimates*, Presented at IMS Annual Meeting, 19-22 August, New-York.

[5] Beard, L. R. and Lall U. (1982). *Estimation of Pearson Type 3 Moments*, Water Resources Research, 12(2), 1563-1569.

[6] Benson, M. A. (1950). *Use of Historical Data in Flood Frequency Analysis*. Transactions American Geophysical Union, 31, 419-424.

[7] Benson, M.A. (1968). *Uniform Flood Frequency Estimating Methods for Federal Agencies*, Water Resources Research, 4(5),891-908.

[8] Bobee, B. (1975). *The Log - Pearson Type III Distribution and its Application in Hydrology*, Water Resources Research, 11(5), 681- 690.

[9] Breiman, L., Meisel W. and Purcell E. (1977). *Variable Kernel Estimates of Multivariate Densities*, Technometrics, 19, 135-144.

[10] Brunk, H.D. (1978). *Univariate Density Estimation of Orthogonal Series*, Biometrika, 65, 521-528.

[11] Condie, R. and K. A. Lee (1982). *Flood Frequency Analysis with Historical Information*, Journal of Hydrology, 58, 47-61.

[12] Costa, J. E. (1978). *Holecene Stratigraphy in Flood Frequency Analysis*, Water Resources Research, 14(4), 626-632.

[13] Chow, V. T. (ed). (1964). *Handbook of Applied Hydrology* , McGraw -Hill NewYork.

[14] Devroye, L. and L.Gyorfi (1985). *Nonparametric Density Estimation: The L_1 View*, New York: Wiley.

[15] Epanechnikov, V. (1969). *Nonparametric Estimates of a Multivariate Probability Density*, Theory Probab. Appl. , 14, 153-158.

[16] Flanagan, P. (1980). *Computer Programs for a Probability Density Function*, M.S. Thesis, University of South Carolina.

[17] Greenwood, J. A., T. A. Landwehr, N. C. Matalas, and J. R. Wallis (1979). *Probability Weighted Moments*, Water Resources Research, 15(5), 1049-1054.

[18] Gupta, V. K., L. Duckstein and R. W. Peebles (1976). *On the Joint Distribution of the Largest Floods and its Time of Occurrence*, Water Resources Research, 12(2), 295-304.

[19] Haan, C. T. (1982).*Statistical Methods in Hydrology*, The Iowa State University Press /Ames.

[20] Hald, A. (1949). *Maximum Likelihood Estimation of the Parameters of a normal Distribution which is Truncated at a known Point*, Skand. Aktuarietidskrift, 32(1/2), 119-134.

[21] Hall, M. J. (1985). *Urban Hydrology*, Elsevier Applied Science Publishers, London.

[22] Hand, D. J. (1982). *Kernel Discriminant Analysis*, Chichester: Research Studies Press.

[23] Hazen, A.(1930). *Flood Flows, a Study of Frequencies and Magnitudes* , John Wiley and Sons, Inc. New York.

[24] Hizer, H. H. (1956). *Type Distributions of Precipitations at Selected Stations in Illinois*, Transactions, American Geophysical Union, 37, 421-424.

[25] Hosking, J. R. M. and J. R. Wallis (1986a). *The Value of Historical Data in Flood Frequency Analysis*, Water Resources Research, 22(11), 1606-1612.

[26] Hosking, J. R. M. and J. R. Wallis (1986b). *Paleoflood Hydrology and Flood Frequency Analysis*, Water Resources Research, 22(4), 543-550.

[27] Houghton, J. C. (1978). *Birth of a Parent: The Wakeby Distribution for Modeling Flood Flows*, Water Resources research, 14(6), 1105-1114.

[28] Karpuk, E. W. and R. Gerard (1979). *Probability Analysis of Historical Flood Data*, Journal of Hydraul. Div. Am. Soc. Civ. Eng., 105(HY9), 1153-1165.

[29] Klemes, V. (1986).*Dilettantism in Hydrology: Transition or destiny*, Water Resources Research, 22(9), 177S-188S.

[30] Leese, M. N. (1973). *The Use of Censored Data in Estimating T- Year Floods*, Proc. UNESCO -WMO- IAHS Symp., on Design of Water Resources Projects with Inadequate Data, Madrid, 1: 236-247.

[31] Loftsgaarden, D. O. and C. P. Quensenberry (1965). *A Nonparametric Estimate of a Multivariate Density Function*, Annals of Mathematical Statistics, 38, 1261-1266.

[32] Kite, G. W. (1985). *Frequency and Risk Analysis in Hydrology*, Water Resources Publications, Littleton, Colorado, U.S.A.

[33] Labatiuk, C. W. (1985). *A Nonparametric Approach to Flood Frequency Analysis*, M. A. Sc. Thesis, University Of Ottawa, Civil Eng. Dept.

[34] Landwehr, J. M., N. C. Matalas and J. R. Wallis (1979). *Probability Weighted Moments Compared with some Traditional Techniques in Estimating Gumble Parameters and Quantiles*, Water Resources Research, 15(5), 1055 -64.

[35] Matalas, N. C. and B. Jacobs (1964). *A Correlation Procedure for Augmenting Hydrologic Data*, U. S. Geol. Surv. Prof. Pap., 434-E.

[36] Matalas, N. C., J. R. Slack and J. R. Wallis (1974). *Regional Skew in Search of a Parent*, Water Resources Research, 11(6), 815-841.

[37] Nozdryn-Plotnicki, M. J. and W. E. Watt (1979). *Assessement of Fitting Techniques for the Logpearson Type 3 Distribution Using Monte-Carlo Simulation*, Water Resources Research, 15(3), 714-722.

[38] Parzen, E. (1962). *On Estimation of Probability Density and Mode*, Ann. Math. Statist., 33, 1065-1076.

[39] Parzen, E. A. (1979). *Nonparametric Statistical Modeling*, J. Amer. Statist. Assoc., 74, 105-131.

[40] Pilon, P. J., R. Condie and K. D. Harvey (1985). *Consolidated Frequency Analysis Package Version 1*, Water Resources Branch, Inland Waters Directorate, Environment Canada, Ottawa.

[41] Pilon, P. J., K. D. Harvey and R. Condie. *An assessment of Contemporary Flood Frequency Distributions*, Water Resources Branch, Inland Water Directorate, Environment Canada, Ottawa, Ontario, Canada.

[42] Potter, W. (1958). *Upper and Lower Frequency Curves for Peak Rates of Runoff*, Transaction, American Geophysics Union, 39(1), 100-105.

[43] Potter, W. K. and J. F. Walker (1985). *An Empirical Study of Flood Measurement Error*, Water Resources research, 21(3), 403-406.

[44] Potter, W. K. (1987). *Research on Flood Frequency Analysis 1983 - 1986*,Reviews of Geophysics, 25(2), 797-804.

[45] Prakasa, R. (1983). *Nonparametric Functional Estimation*, New York: Academic Press.

[46] Rao, D. V. (1983). *Estimating Logpearson Parameters by Mixed Moments*, Journal of Hydraulic Engineering, 109(8), 1118-1131.

[47] Rosenblatt, M. (1956). *Remarks on some Nonparametric Estimates of a Density Function*, Ann. Math. Statist., 27, 832-837.

[48] Rossi, F. M. and P. Versace (1984). *Two- Component Extreme Value Distribution for Flood Frequency Analysis*, Water Resources Research, 20(7), 847-856.

[49] Schuster, E. and S. Yakowitz (1986).*Parametric/Nonparametric Mixture Density Estimation with the Application to Flood Frequency Analysis*, Water Resources Bulletin, American Water Resources Association, 21(5), 797-804.

[50] Scott, D. W. and L. E. Factor (1981). *Monte-Carlo Study of Three Data-Based Nonparametric Probability Density Estimators*, Journal of the American Statistical Association, 76(373), 9-15.

[51] Silverman, B. W. (1978). *Choosing the Window Width when Estimating a Density*, Biometrika, 65, 1-11.

[52] Silverman, B. W. (1981). *Using Kernel Density Estimates to Investigate Multimodality*, J. Roy. Statist. Soc. B, 43, 97-99.

[53] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, School of Mathematics University of Bath, U.K., Chapman and Hall Ltd.

[54] Stedinger, J. R. and T. A. Cohn (1986). *Flood Frequency Analysis with Historical Information*, Water Resources Research, 22(5), 785-793.

[55] Stoddart, R. B. L. and W. E. Watt (1970). *Flood Frequency Prediction for Intermediate Drainage Basins in Southern Ontario*, C. E. Research Report No. 66 July,1970, Department of Civil Engineering Queen 's University at Kingston, Ontario.

[56] Tapia, R. A. and J. R. Thompson (1978). *Nonparametric Probability Density Estimation*, Baltimore: Johns Hopkins University Press.

[57] Titterington, D. M. (1983). *Kernel-Based Density Estimation Using Censored, Truncated or Grouped Data*, Comm. Statist. Theor. Meth., 12(18), 2151-2167.

[58] Thomas, W. O. (1985). *A Uniform Technique for Flood Frequency Analysis*, Journal of Water Resources Planning and Management, American Society of Civil Engineering, 111(3), 320-337.

[59] U. S. Water Resources Council (USWRC) (1982). *Guidelines for Determining Flood Flow Frequency*, Bulletin 17B, Hydrol. Comm., Washington D. C.

[60] Water Resources Council (1967). *A Uniform Technique for Determining Flood Frequencies*, Bulletin No. 15, Water Resources Council, Washington, D. C.

[61] Watson, G. S. (1969). *Density Estimation by Orthogonal Series*, Ann. Math. Statist., 40, 1496-1498.

[62] Waylen, P. and M. K. Woo (1982). *Prediction of Annual Floods Generated by Mixed Processes*, Water Resources Research, 18(4), 1283-1286.

[63] Waylen, P. and M. K. Woo (1983). *Annual Floods in Southwestern British Columbia, Canada*, Journal of Hydrology, 62, 95-105.

[64] Waylen, P. and M. K. Woo (1984). *Areal Prediction of Annual Floods Generated by Two Distinct Processes*, Journal of Hydrological Sciences, 29(1), 75-88.

[65] Wegman, E. J. (1970). *Maximum Likelihood Estimation of a Unimodal Density Function 2*, Ann. Math. Statist., 41, 2169-2174.

[66] Wegman, E. J. (1972). *Nonparametric Probability Density Estimation. A Summary of Available Methods*, Technometrics, 14, 533-546.

[67] Wood, E. F. and I. Rodriguez-Iturbe (1975). *A Baysian Approach to Analyzing Uncertainty among Flood Frequency Models*, Water Resources Research, 11(6), 839-882.

[68] Wood, E. F. and N. P. Greis (1981). *Regional Flood Frequency Estimate and Network Design*, Water Resources Research, !7, 1167-1177.

[69] Woodroofe, M. (1970). *On Choosing a Delta Sequence*, Ann. Math. Statist., 41, 1665-1671.

Appendix A

DATA AND STATISTICAL TESTINGS

Table A.1: Flood Data for . . rtheast Margaree River at Margaree Valley, Station 01FB001

Year	Month	Streamflow (cms)	Year	Month	Streamflow (cms)
1917	11	163	1951	8	129
1918	5	283	1952	1	125
1919	4	124	1953	2	174
1920	3	237	1954	2	77
1921	3	206	1955	9	118
1922	5	175	1956	1	187
1923	10	249	1957	5	123
1924	5	146	1958	4	131
1925	2	137	1959	11	121
1926	5	178	1960	4	131
1927	11	272	1961	5	167
1928	7	125	1962	4	343
1929	5	158	1963	5	144
1930	3	149	1964	5	137
1931	4	80	1965	5	108
1932	1	156	1966	5	94
1933	11	119	1967	11	263
1934	5	226	1968	12	206
1935	1	225	1969	5	292
1936	3	225	1970	5	121
1937	11	187	1971	11	206
1938	4	163	1972	5	224
1939	10	130	1973	5	138
1940	5	168	1974	4	123
1941	5	185	1975	12	216
1942	5	120	1976	5	166
1943	5	167	1977	6	201
1944	5	119	1978	1	199
1945	5	109	1979	1	162
1946	4	177	1980	12	186
1947	5	214	1981	4	189
1948	5	190	1982	4	355
1949	11	161	1983	3	150
1950	4	111			

Table A.2: Statistical Testings of Northeast Margaree,Station 01FB001

Test	Critical Value at 5 % level	Test Result
SPEARMAN TEST FOR INDEPENDENCE Spearman Rank Order Serial Coeff.=0.112 D.F = 64 Corresponds to students T = 0.898	1.670	Not significant
SPEARMAN TEST FOR TREND Spearman Rank Order Correlation Coef.=-0.018 D.F. = 65 Corresponds to students T = - 0.148	-1.998	Not significant
RUN TEST FOR GENERAL RANDOMNESS Runs above and below median = 28 Runs above median=32 Runs below median=33 For this test, Z=1.374	1.960	Not significant
MANN-WHITNEY HOMOGENEITY TEST Split by time span, Size of Sample 1 = 33 Size of Sample 2 = 34 For this test, Z=-0.684	-1.645	Not significant

Table A.3: Flood Data for Warta River At Poznan (Poland)

Year	Streamflow (cms)	Year	Streamflow (cms)	Year	Streamflow (cms)
1828	191	1874	275	1920	770
1829	275	1875	280	1921	241
1830	1035	1876	1105	1922	301
1831	393	1877	365	1923	518
1832	206	1878	380	1924	1573
1833	382	1879	333	1925	199
1834	692	1880	572	1926	413
1835	105	1881	475	1927	545
1836	303	1882	134	1928	539
1837	548	1883	269	1929	299
1838	725	1884	257	1930	109
1839	512	1885	200	1931	362
1840	525	1886	775	1932	192
1841	745	1887	260	1933	168
1842	118	1888	1510	1934	116
1843	207	1889	1460	1935	250
1844	300	1890	312	1936	150
1845	1015	1891	1162	1937	330
1846	728	1892	515	1938	450
1847	153	1893	500	1939	292
1848	418	1894	335	1940	990
1849	457	1895	728	1941	750
1850	1395	1896	199	1942	671
1851	264	1897	503	1943	148
1852	540	1898	247	1944	195
1853	718	1899	257	1945	455
1854	609	1900	418	1946	386
1855	1515	1901	448	1947	1055

Table A.4: Flood Data for Warta River at Poznan (continued)

Year	Streamflow (cms)	Year	Streamflow (cms)	Year	Streamflow (cms)
1856	399	1902	441	1948	511
1857	122	1903	297	1949	255
1858	198	1904	297	1950	193
1859	125	1905	89	1951	150
1860	555	1906	214	1952	148
1861	432	1907	362	1953	706
1862	353	1908	455	1954	207
1863	100	1909	1253	1955	182
1864	187	1910	185	1956	214
1865	285	1911	539	1957	299
1866	252	1912	294	1958	437
1867	375	1913	205	1959	185
1868	548	1914	287	1960	182
1869	295	1915	202	1961	293
1870	346	1916	431	1962	350
1871	1120	1917	1073	1963	350
1872	370	1918	376	1964	305
1873	213	1919	141	1965	469

Table A.5: Statistical Testings of Warta River at Poznan

Test	Critical Value at 5 % level	Test Result
SPEARMAN TEST FOR INDEPENDENCE Spearman Rank Order Serial Coeff.=0.152 D.F = 137 Corresponds to students T = 1.794	1.645	significant
SPEARMAN TEST FOR TREND Spearman Rank Order Correlation Coef.= 0.127 D.F. = 138 Corresponds to students T = 1.508	1.960	Not significant
RUN TEST FOR GENERAL RANDOMNESS Runs above and below median = 62 Runs above median=70 Runs below median=70 For this test, Z=1.527	1.960	Not significant
MANN-WHITNEY HOMOGENEITY TEST Split by time span, Size of Sample 1 = 68 Size of Sample 2 = 72 For this test, Z= 1.527	-1.645	Not significant

N.B: At the 5 % level of significance, the correlation is significantly from zero, but it is not so at the 1 % level. That is, the dependence is significant, but not highly so.

Table A.6: Flood Data for Turtle River near Mine Center Station 05PB014

Year	Mth	Flow (cms)	One Day A.P.(mm)	Ant.Temp. (C)	Year	Mth	Flow	One Day A.P.(mm)	Ant.Temp (C)
1921	5	76	18.3	15	1954	5	171	17.5	
1922	4	100	43.2	3.3	1955	7	99	47.8	25
1923	4	134	11.2	4.4	1956	5	159	26.2	12.7
1924	5	67	7.1	6.1	1957	4	141	25.7	8.3
1925	5	81	9.1	16.7	1958	7	57	23.1	22.8
1926	6	84	45	10.6	1959	6	74	19.3	30
1927	7	252	34.3	25	1960	6	105	10.2	17.8
1928	10	94	37.1	17.2	1961	4	71	*	16.7
1929	4	48	17.3	3.9	1962	5	164	25.4	26
1930	5	56	16.3	7.2	1963	6	154	19.3	31.7
1931	6	43	20.1	21.7	1964	5	192	18.0	23
1932	4	104	27.9	1.1	1965	5	140	19.1	17.2
1933	5	124	1.5	20.6	1966	5	195	16.5	16.1
1934	5	158	14	17.8	1967	4	102	18.5	4.4
1935	5	134	11.7	16.7	1968	6	163	49.0	22.2
1936	5	72	21.8	22.8	1969	6	124	56.6	22.2
1937	5	166	28.4	21.1	1970	5	167	21.1	15.6
1938	5	184	17.8	15.6	1971	11	159	11.4	1.7
1939	6	40	14.2	12.2	1972	4	154	15.7	3.3
1940	5	51	4.8	14.4	1973	10	84	32.0	15
1941	10	305	32.5	12.2	1974	6	230	51.6	17.2
1942	9	115	29.2	11.1	1975	5	121	5.6	30.0
1943	5	208	20.8	11.1	1976	4	126	4.6	7.8
1944	8	127	74.9	21.1	1977	9	93	40.9	16.1
1945	4	93	25.4	*	1978	6	155	21.3	24
1946	4	105	15.2	1.1	1979	4	184	44.7	3.9
1947	6	163	39.6	15.6	1980	4	58	0.8	2.5
1948	5	206	31.8	13.9	1981	5	61	18.0	12
1949	5	84	22.9	11.7	1982	5	126	34.0	*
1950	5	261	38.1	*					
1951	5	188	5.9	19.7					
1952	7	155	41.7	25.6					
1953	6	69	17.0	10					

Note: Mth = Month, A.P = Antecedent precipitation and Ant. Temp. = One day antecedent temperature (Stars indicate data not available)

Table A.7: Statistical Testings of Turtle River near Mine Center Station 05PB014

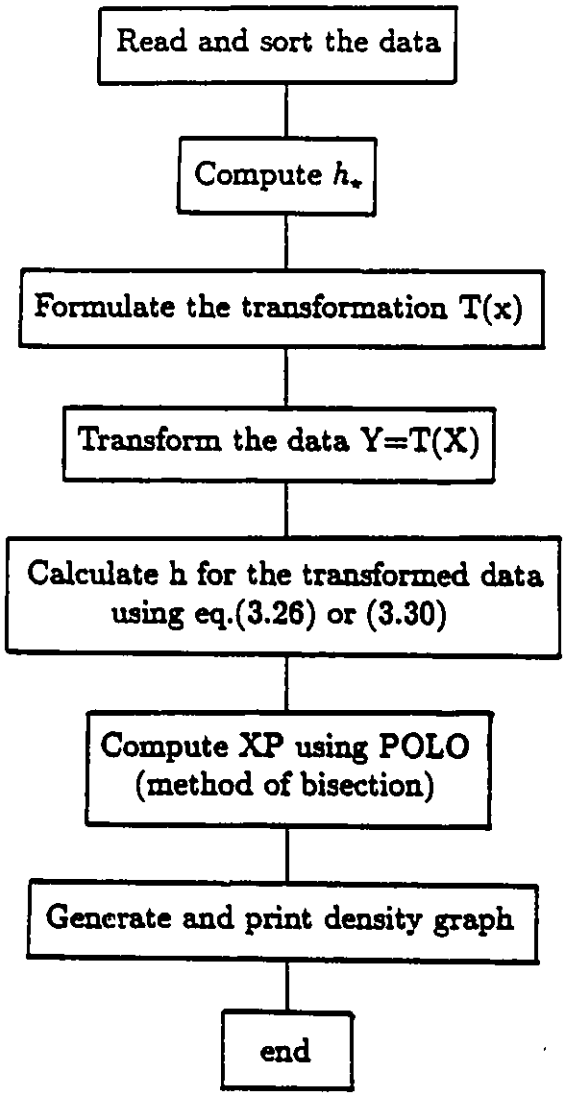
Test	Critical Value at 5 % level	Test Result
SPEARMAN TEST FOR INDEPENDENCE Spearman Rank Order Serial Coeff.=0.036 D.F = 58 Corresponds to students T = 0.273	1.672	Not significant
SPEARMAN TEST FOR TREND Spearman Rank Order Correlation Coef.=-0.209 D.F. = 59 Corresponds to students T = - 1.644	-2.001	Not significant
RUN TEST FOR GENERAL RANDOMNESS Runs above and below median = 31 Runs above median=30 Runs below median=30 For this test, Z=0.000	1.960	Not significant
MANN-WHITNEY HOMOGENEITY TEST Split by time span, Size of Sample 1 = 30 Size of Sample 2 = 31 For this test, Z=-1.183	-1.645	Not significant

Appendix B

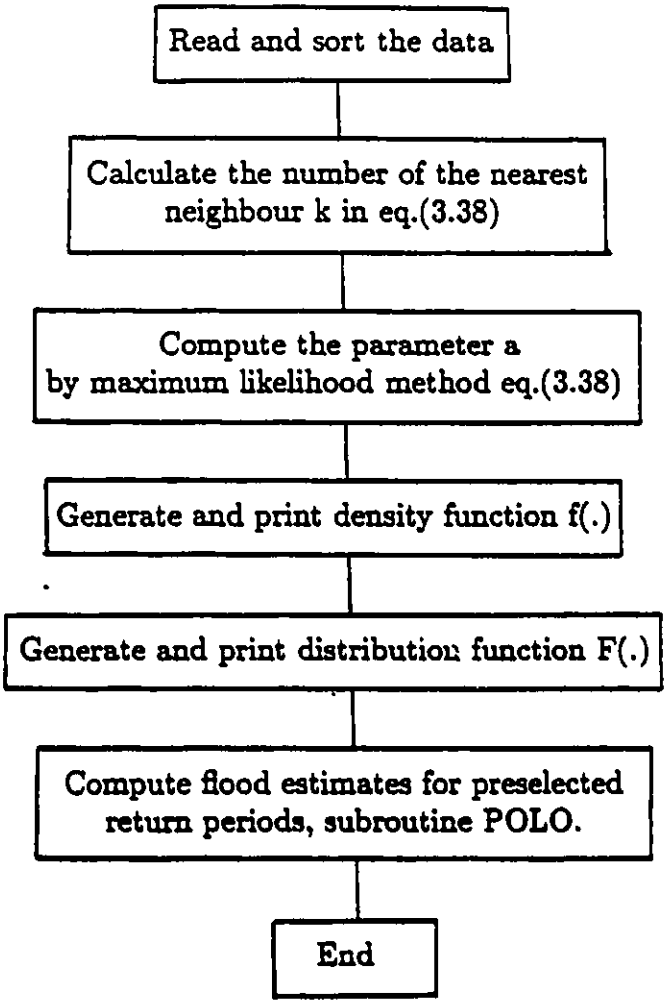
COMPUTER PROGRAMS

B.1 Flowcharts

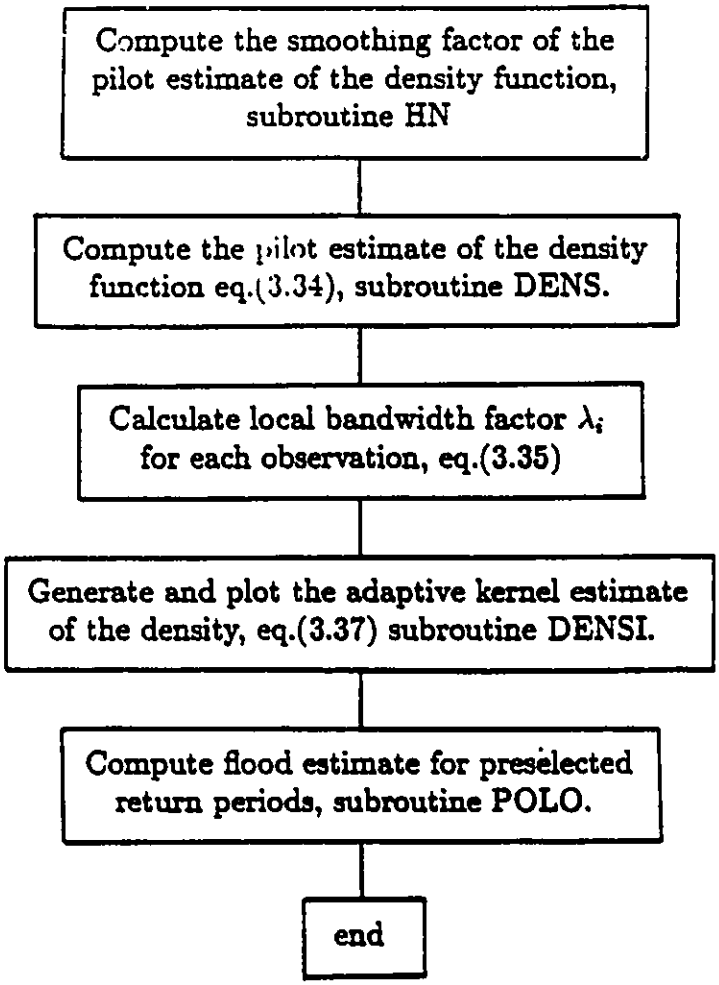
COMMON FLOWCHARTS OF THE TRANSFORMED KERNEL PROGRAMS
TRTKER and CITKER



FLOWCHART OF THE VARIABLE KERNEL PROGRAM
VARKER



FLOWCHART OF THE ADAPTIVE KERNEL PROGRAM ADAKER



B.2 Program Listings

```
C        TRIANGULAR TRANSFORMED KERNEL PROGRAM
C        PROGRAMMER: DR. M. FELLOH - 1990
C
C THIS PROGRAM EMPLOYS AN ANNUAL FLOOD SERIES AS AN INPUT TO CALCULATE
C FLOOD ESTIMATES FOR A PRE-SELECTED RETURN PERIOD USING THE TRIANGULAR
C TRANSFORMED KERNEL ESTIMATOR SUGGESTED BY DEVROYE ET AL. (1985).
C IN THE NUMERICAL PROCEDURE THE EPANECHNIKOV OPTIMAL KERNEL AND THE
C FLANAGAN OPTIMAL SMOOTHING FACTOR ARE APPLIED. THE PROGRAM PROVIDES
C ALSO THE FIXED KERNEL ESTIMATOR AS AN OPTION. DENSITY FUNCTION CAN BE
C PLOTTED UNDER REQUEST.
C
C INPUT FILE
C
C THE INPUT FILE COMPRISES :
C (1) NAME OF THE STATION IN CHARACTER VARIABLE ( FORMAT 18A4 )
C (2) N RECORD LENGTH OF FLOOD DATA (INTEGER)
C (3) ANNUAL FLOOD SERIES (REAL VARIABLE FREE FORMAT)
C (4) KL, KEY AND SCAL CONTROL PARAMETERS
C
C IMPORTANT VARIABLE
C
C XOBS : INPUT VECTOR OF FLOOD DATA
C N : NUMBER OF FLOOD OBSERVATION IN AN ANNUAL FLOOD SERIES.
C HH : SMOOTHING FACTOR
C T : RETURN PERIOD
C XP : FLOOD MAGNITUDE AS ESTIMATED BY THE TRANSFORMED KERNEL
C KL : INTEGER CONTROLLING THE ESTIMATION METHOD
C        KL= 0 FOR TRANSFORMED KERNEL ESTIMATOR
C        KL= 1 FOR THE FIXED KERNEL ESTIMATOR
C KEY: INTEGER CONTROLLING OUTPUT OPTIONS
C        KEY=0 PRINTING ONLY FLOOD ESTIMATES FOR SPECIFIED RETURN PERIOD
C        KEY=1 PRINTING EMPIRICAL PROBABILITIES
C        KEY=2 PRINTING THEORETICAL PROBABILITIES
C SCAL: VARIABLE CONTROLLING THE STARTING POINT ON THE Y Axis OF THE
C       DENSITY FUNCTION (SCAL=0.0 NO NEED TO SPECIFY THE STARTING POINT
C A,B : LIMITS FOR METHOD OF BISECTION
C EPS : ACCURACY LEVEL FOR THE METHOD OF BISECTION
C
C DESCRIPTION OF MAIN SUBROUTINES
C
C  SORT : RANK THE DATA FROM MAXIMUM TO MINIMUM.
C  HN   : CALCULATE THE SMOOTHING FACTOR FOR THE FIXED KERNEL.
C  TR   : TRANSFORM THE FLOOD DATA.
C  FCN  : FOR A GIVEN RETURN PERIOD THIS FUNCTION IS EQUATED TO ZERO
C         AND SOLVED FOR XP USING POLO.
C  POLO : COMPUTES ROOTS OF A FUNCTION BY METHOD OF BISECTION.
C  W    : CALCULATE THE INTEGRAL OF THE KERNEL FUNCTION K(.).
C  DENS : CALCULATE DENSITY FUNCTION AT EACH FLOOD OBSERVATION.
C  PLOT : GENERATE THE PLOT OF THE DENSITY FUNCTION.
C  OUTX : PRINTS FLOOD DATA IN COLUMNS
C  PEM  : COMPUTES EMPIRICAL PROBABILITIES.
C
C
      DIMENSION T(13),VALUE(2,333),ROB(3,3),XMEM(3)
     *,IY(3)
      CHARACTER*8 TT(8,8),IZ(8),SP,ST
      COMMON /OBS/M,XOBS(333),YT(333),SUM,FN,H,HH
     */MIN/XMAX,FIP,KL
     */KY/MP,TITL/DE/D
      DATA T/1.005,1.05,1.25,2.,5.,10.,20.,
     *50.,100.,200.,500.,1000.,2000./
      DATA (TT(I,1),I=1,8)/'THE TR',
     *'ANSFOR','MED KE','RNEL E','STIMAT','E     '/
      DATA (TT(I,2),I=1,8)/'TABULA','TED NO','NPARAM'
     *,'ETRIC ','DENSIT','Y FUNC','TION  ','      '/
      DATA SP,ST/'      ','THE KE'/
      PI=3.1415926536
      EPS=0.5
      MP=2
      TITL=0
      READ(5,*) KL,KEY,SCAL
      CALL INTIT(3)
      IF(KL.NE.1) WRITE(6,20)
      TT(1,1)=SP
      TT(2,1)=SP
      TT(3,1)=ST
C 14  WRITE(6,20)(TT(I,1),I=1,6)
      IF(KL.NE.1) WRITE(6,31)
      READ(5,*) N
      READ(5,*)(XOBS(I),I=1,N)
      IF(KEY.LE.0.OR.KEY.EQ.2) CALL OUTX(N,XOBS)
      IF(KEY.EQ.1) CALL SORTP(N,XOBS)
      IF(KEY.EQ.2) CALL SORT(N,XOBS)
      XMAX=-10.E33
      XMIN=-XMAX
      DO 8 I=1,N
      XI=XOBS(I)
      IF(XI.GT.XMAX) XMAX=XI
      IF(XI.LT.XMIN) XMIN=XI
   8  CONTINUE
      FN=FLOAT(N)
      HH=HN(N,XOBS)
      WRITE(6,29) HH
      IF(KL.EQ.1) GO TO 6
      DO 7 I=1,N
      XI=XOBS(I)
   7  YT(I)=TR(XI)
      H=5./182./PI/FN
      H=H**0.2
      WRITE(6,30) H
```

```
      SUM=0.
      DO 15 I=1,N
      Y=YT(I)
      Y1=Y-1
   15 SUM=SUM+W(Y/H)-W(Y1/H)
      SUM=SUM/FN
      WRITE(6,27)  SUM
    8 IF(KEY.LE.1) GO TO 10
      WRITE(6,24)
      N1=N-1
      N2=N-2
      DO 4 K=1,N,3
      DO 1 J=1,3
      I=K+J-1
      IF(I.GT.N) GO TO 4
      XI=XOBS(I)
      ROB(1,J)=DENS(XI)
      IF(KL.EQ.1) D=1.-D
      IF(KL.NE.1) D=ANAL(XI)
      ROB(2,J)=D
      ROB(3,J)=PEM(N,I)
      XMEM(J)=XI
    1 CONTINUE
      IF(K.GT.N2) GO TO 11
        WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
     *,XMEM(2),(ROB(L1,2),L1=1,3),XMEM(3),(ROB(L1,3),L1=1,3)
      GO TO 4
   11 IF(K.GT.N1) GO TO 12
      WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
     *,XMEM(2),(ROB(L1,2),L1=1,3)
      GO TO 4
   12 WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
    4 CONTINUE
   10 IF(KEY.LE.0) GO TO 13
C     PLOTTING OF DENSITY
      IY(2)=2
      IY(1)=1
      BOUND=XMAX
      DIF=AINT(XMAX/100.)
      IF(DIF.LE.1.5) DIF=1.
      IF(DIF.GT.1.5.AND.DIF.LE.2.5) DIF=2.
      IF(DIF.GT.2.5.AND.DIF.LE.7.5) DIF=5.
      IF(DIF.GT.7.5.AND.DIF.LE.12.5) DIF=10.
      IF(DIF.GT.12.5.AND.DIF.LE.27.5) DIF=20.
      IF(DIF.GT.27.5.AND.DIF.LE.75.0) DIF=50.
      IF(DIF.GT.75..AND.DIF.LE.125.) DIF=100.
      IF(DIF.GT.125..AND.DIF.LE.300.) DIF=200.
      IF(DIF.GT.300..AND.DIF.LE.700.) DIF=500.
      IF(DIF.GT.700..AND.DIF.LE.1300.) DIF=1000.
      IF(DIF.GT.1300.) DIF=2000.
      ISCAL=0
      IF(SCAL.LE.0.) ISCAL=1
      XI=0.
      XMID=XMIN+0.75
      I=0
    3 XI=XI+DIF
      IF(XI.LT.XMID) GO TO 3
      IF(XI.GT.BOUND) GO TO 2
      I=I+1
      VALUE(1,I)=XI
      VALUE(2,I)=DENS(XI)
      GO TO 3
    2 CALL PLOT(I,2,2,IY,VALUE,0.,SCAL,0,ISCAL,IPRIN)
C     COMPUTING OF FLOODS
   13 B=XMAX
      A=0.0
      BG=XMAX
      WRITE (6,22)
      DO 5 I=1,11
      FIP=1./T(I)
      XP=POLO(A,B,EPS)
      BG=BG*1.1
      B=BG
      A=XP
    5 WRITE(6,23) T(I),XP
      WRITE(6,28)
   22 FORMAT(/5X,15HRETURN    FLOOD/5X,
     *17HPERIOD   ESTIMATE/)
   23 FORMAT(2X,F9.3,3X,F7.0)
   24 FORMAT(/,'DENSITY FUNCTION FOR EMPIRICAL DATA :'.
     *//,3(4X,'XOBS',1X,'DENSITY',2X,2('PROBABILITY',1X))./.
     *3(17X,'THEORETICAL  EMPIRICAL  '),/)
   25 FORMAT(3(F7.0,3(1X,G11.5),1X))
   26 FORMAT(1X,' THE TRANSFORMED KERNEL ESTIMATE './)
   27 FORMAT('  SUM = ',G11.5/)
   28 FORMAT(////)
   29 FORMAT(' H FOR OBSERVATED DATA    = ',G11.5)
   30 FORMAT(  ' H FOR TRANSFORMATED DATA = ',G11.5/)
   31 FORMAT(' TRIANGULARY TRANSFORMATION'//)
      STOP
      END
      FUNCTION DENS(X)
C     DENSITY OF TRANSFORMED DATA - IF KL=0
C     DENSITY OF OBSERVATED DATA  - IF KL=1
      COMMON /OBS/N,XOBS(333),YT(333),SUM,FN,H,IPR
     */DE/D/MIN/XMAX,FIP,KL
      DENS=0.
      DY=0.
      I=IR(X)
      DO 2 I=1,N
```

```
      IF(KL.EQ.1) GO TO 6
      Y=(T-YT(I))/H
      AY=ABS(Y)
      Z=0.
      IF(AY.LT.1.) Z=0.75*(1.-Y*Y)
      DENS=DENS+Z
    6 Y=(X-XOBS(I))/HH
      AY=ABS(Y)
      Z=0.
      IF(AY.LT.1.) Z=0.75*(1.-Y*Y)
    2 DY=DY+Z
      DY=DY/FN/HH
C     DY - DENSITY OF OBSERVATED DATA
      IF(KL.NE.1) GO TO 5
      DENS=DY
      RETURN
    5 IF(D-0.5) 1,1,3
    1 DIFTR=DY/SQRT(6.*D)
      GO TO 4
    3 DIFTR=DY/SQRT(6.*(1.-D))
C     DIFTR - DIFFERENTIABLE OF TRANSFORMATION
    4 DENS=DENS*DIFTR/FN/H/SUM
      RETURN
      END
      FUNCTION TR(X)
C     TRANSFORMATION
      COMMON /DE/D
     */OBS/N,XOBS(333),YT(333),SUM,FN,H,HH
      D=0.
      DO 3 I=1,N
      Y=(X-XOBS(I))/HH
    3 D=D+W(Y)
      ON=D/FN
        P=0.5-ON
      D=0.5+ON
      IF(ON) 1,1,2
    1 TR=SQRT(D/2.)
      RETURN
    2 TR=1.-SQRT(P/2.)
      RETURN
      END
      FUNCTION FCN(N)
      COMMON /NIN/KMAX,FIP,KL/DE/D
      IF(KL.GT.0) GO TO 1
      FCN=FIP-ANAL(N)
      RETURN
    1 TN=TR(N)
      FCN=FIP+0-1.
      RETURN
      END
      FUNCTION ANAL(X)
      COMMON  /OBS/N,XOBS(333),YT(333),SUM,FN,H,HH
      S=0.
      TX=TR(X)
      DO 1 I=1,N
      YI=YT(I)
      YI=(1.-YI)/H
      Y=(TX-YI)/H
    1 S=S+W(YI)-W(Y)
      ANAL=S/FN/SUM
      RETURN
      END
      FUNCTION W(X)
      W=-0.5
      IF(X.GT.-1..AND.X.LT.1.) W=0.75*X*(1.-X*X/3.)
      IF(X.GE.1.) W=0.5
      RETURN
      END
      SUBROUTINE OUTX(N,Y)
      DIMENSION Y(N)
      FN=FLOAT(N)
      FL=10.
      R=FN/FL
      R=R-AINT(R)
      R=R*FL+0.5
      IR=INT(R)-1
      NIL=INT(FL)
      ML=NIL
      IL=1
      WRITE(6,8)
      DO 1 I=1,N,ML
      J=I+ML-1
      IF(NIL.LT.ML) J=I+IR
      WRITE(6,8)(Y(L),L=I,J)
      NIL=N-ML*IL
      IL=IL+1
    1 CONTINUE
      WRITE(6,7)
    9 FORMAT(7H XOBS :/)
    8 FORMAT(1X,20F7.0)
    7 FORMAT(1X//)
      RETURN
      END
      SUBROUTINE INTIT(NI)
      DIMENSION T(18)
      READ(5,10)(T(K),K=1,18)
      IF(NI.EQ.1) WRITE(6,21)(T(K),K=1,18)
      IF(NI.EQ.2) WRITE(6,22)(T(K),K=1,16)
      IF(NI.EQ.3) WRITE(6,23)(T(K),K=1,18)
   10 FORMAT(18A4)
```

```
 21   FORMAT(1H,///4X,80(1H*)/4X,2H* ,
      118A6,2H */4X,80(1H*)///)
 22   FORMAT(1H ///6X,68(1H-)/6X,2H* ,16A6,
      12H *,/6X,68(1H-)///)
 23   FORMAT(1H ,16A6/)
        RETURN
        END
        FUNCTION POLO(A,B,EPS)
        FA=FCN(A)
        FB=FCN(B)
        FAB=FA*FB
        IF(FAB.LE.0.) GO TO 10
C     WRITE(6,68) FA,FB,A,B
        WRITE(6,67)
        STOP
 68   FORMAT(///1X,10(1H*),3X,'TO CHANGE
      1BOUNDARY IN POLO',//1X,10(1H*),3X,
      2'STOP OF PROGRAM'/
      3/' FCN(A) =',E18.8,3X,' FCN(B) =',E18.8,
      4/' A= ',E18.8,3X,' B = ',E18.8)
 67   FORMAT(////)
 10   X=(A+B)/2
        FM=FCN(X)
        FAM=FA*FM
        IF(FAM) 2,3,4
  2     B=X
        GO TO 5
  4     FA=FM
        A=X
  5     IF(ABS(A-B).GT.EPS) GO TO 10
  3     POLO=(A+B)/2
        RETURN
        END
        SUBROUTINE SORT(N,A)
        DIMENSION A(N)
        N1=N-1
        DO 2 I=1,N1
        II=I+1
        AI=A(I)
        DO 2 J=II,N
        AJ=A(J)
        IF(AI.GT.AJ) GO TO 4
        A(J)=AI
        A(I)=AJ
        AI=AJ
  4     CONTINUE
  2     CONTINUE
        RETURN
        END
        FUNCTION HN(N,XOBS)
        DIMENSION XOBS(N),X(333)
        ALFA=0.5
        DO 1 I=1,N
  1     X(I)=XOBS(I)
        DO 2 I=1,N
        IHIGH=I
        ILOW=1
  3     IF(IHIGH.EQ.ILOW) GO TO 4
        L=(ILOW+IHIGH)/2
        IF(X(I)-X(L)) 5,7,6
  5     IHIGH=L
        GO TO 3
  6     ILOW=L+1
        GO TO 3
  7     K=L+1
        GO TO 8
  4     K=ILOW
  8     SAVE=X(I)
        ISORT=I-K
        DO 9 I2=1,ISORT
        L=I+1-I2
  9     X(L)=X(L-1)
  2     X(K)=SAVE
        S=0
        FN=FLOAT(N)
        KN=INT(FN**ALFA)
        DO 10 I=1,N
        XI=X(I)
        M=I-KN
        B=999999.
        IF(M.GT.0) B=XI-X(M)
        M=I+KN
        A=999999
        IF(M.LE.N) A=X(M)-XI
        D=B
        IF(A.LT.B) D=A
 10     S=S+D
        HN=S/FN
        RETURN
        END
        FUNCTION PEM(N,I)
C
C     THIS FUNCTION COMPUTES EMPIRICAL PROBABILITY PJ
C
        COMMON /KY/MP
        FI=FLOAT(I)
        FN=FLOAT(N)
        GO TO (1,2,3,4,5,6) ,MP
  1     PEM=FI/(FN+1 )
        RETURN
```

```
      PEM=(FI-0.25)/(FN+0.5)
        RETURN
    3 F=FN+1
        S=SQRT(F*FN)
        PEM=(2.*FI-F+S)/S/2.
        RETURN
    4 PEM=(FI-0.3)/(FN+0.4)
        RETURN
    5 PEM=(FI-0.4)/(FN+0.2)
        RETURN
    6 PEM=(FI-0.5)/FN
        RETURN
        END
        SUBROUTINE SORTP(N,A)
        DIMENSION A(N)
        COMMON /XY/NP,TITL
        NI=N-1
        DO 2 I=1,NI
        I1=I+1
        AI=A(I)
        ID=INT(TITL)
        DO 2 J=I1,N
        AJ=A(J)
        IF(AI.GT.AJ.AND.ID.EQ.0) GO TO 4
        IF(AI.LE.AJ.AND.ID.NE.0) GO TO 4
        A(J)=AI
        A(I)=AJ
        AI=AJ
    4 CONTINUE
    2 CONTINUE
        N3=N-3
        N2=N-2
        N1=N-1
        WRITE(6,23)
        DO 12 I=1,N,4
        I4=I+3
        I3=I+2
        I2=I+1
        IF(I.GT.N3) GO TO 33
        P1=PEM(N,I)
        P2=PEM(N,I2)
        P3=PEM(N,I3)
        P4=PEM(N,I4)
        WRITE(6,22) P1,A(I),P2,A(I2),P3,A(I3),P4,A(I4)
        GO TO 5
   33 IF(I.GT.N2) GO TO 24
        P1=PEM(N,I)
        P2=PEM(N,I2)
        P3=PEM(N,I3)
        WRITE(6,22) P1,A(I),P2,A(I2),P3,A(I3)
        GO TO 5
   24 IF(I.GT.N1) GO TO 13
        P1=PEM(N,I)
        P2=PEM(N,I2)
        WRITE(6,22) P1,A(I),P2,A(I2)
        GO TO 5
   13 P1=PEM(N,I)
        WRITE(6,22) P1,A(I)
    5 CONTINUE
   12 CONTINUE
        WRITE(6,21)
   22 FORMAT('1H',4(F7.4,1X,F8.1,3X))
   23 FORMAT(1H,/17H EMPIRICAL DATA : ,///4(4X,1HP,7X,
      *4HXOBS,3X)/)
   21 FORMAT(///)
        RETURN
        END
        SUBROUTINE PLOT(N,M,NY,IY,Y,YMN,YMX,IPO,ISCAL,IPR(N)
C
C *****  PLOTTING OF THE DENSITY FUNCTION
        DIMENSION IY(1),Y(M,1),SCL(13)
        CHARACTER*8 IZ(8),IH1,IH2,IH3,IH4,
       *LINE(121),IXIS(80)
        DATA IH1,IH2,IH3,IH4
       1/'+-','---','.',' '/
        DATA IZ/'*','$','X','0','S','O'/
        DO 6 I=1,12
        K=(I-1)*5+1
        IXIS(K)=IH1
        DO 6 J=1,4
        L=K+J
    6 IXIS(L)=IH2
        IF(ISCAL.EQ.0) GO TO 4
        YMN=1.E30
        YMX=-1.0E30
        DO 5 I=2,NY
        L=IY(I)
        DO 5 J=1,N
        YY=Y(L,J)
        IF(YY.LT.YMN) YMN=YY
        IF(YY.GT.YMX) YMX=YY
    5 CONTINUE
    4 I1=IY(1)
        DY=(YMX-YMN)*0.01
        DO 10 I=1,11
        AI=FLOAT(I)
   10 SCL(I)=((11.-AI)*YMN+(AI-1.)*YMX)*0.1
        IF(NY.GT.8) NY=8
        IF(IPO.EQ.0) GO TO 40
        IMX=M-1
```

```
      DO 30 I=1,IMX
      JMX=N-I
      DO 30 J=1,JMX
      IF( Y(II,J).LE. Y(II,J+1)) GO TO 30
      DO 20 K=1,M
      TMP=Y(K,J)
      Y(K,J)= Y(K,J+1)
      Y(K,J+1)= TMP
 20   CONTINUE
 30   CONTINUE
 40   IS=5
      IF(IPRIN.EQ.0) GO TO 1
      L=IY(1)
      WRITE(6,5000)  (Y(L,K),K=1,N)
      DO 2 I=2,MY
      L=IY(I)
      WRITE(6,5002)(Y(L,K),K=1,N)
 2    CONTINUE
 1    WRITE(6,5003)
      DO 3 I=2,MY
      II=I-1
      L=IY(I)
 3    WRITE(6,5004)  IZ(II)
      WRITE(6,5005)
      WRITE(6,1001)  (SCL(I),I=1,11),(IXIS(I),I=1,50)
      DO 200 J=1,N
      DO 50 I=1,101
 50   LINE(I)=IH4
      DO 100 IYY=2,MY
      II=IYY-1
      I=IY(IYY)
      YY=Y(I,J)
      IF(YY.GT.YMX.OR.YY.LT.YMN) GO TO 100
      BI=(YY-YMN)/DY
      IB=INT(BI)
      IF((BI-FLOAT(IB)).GE.0.5) IB=IB+1
      IB=IB+1
      IF(LINE(IB).NE.IH4) GO TO 60
      LINE(IB)=IZ(II)
      GO TO 100
 60   LINE(IB)=IZ(6)
100   CONTINUE
      DO 80 L=1,101,10
      IF(LINE(L).EQ.IH4) LINE(L)=IH3
 80   CONTINUE
      IF(J.NE.IS) GO TO 90
      IF(J.EQ.N) GO TO 90
      DO 70 L=1,101,2
      IF(LINE(L).EQ.IH4) LINE(L)=IH3
 70   CONTINUE
      IS=IS+5
 90   WRITE(6,2000)
     *Y(II,J),(LINE(L),L=1,101)
200   CONTINUE
      WRITE(6,2001) (IXIS(L),L=1,50),(SCL(L),L=1,11)
      RETURN
1001 FORMAT(5X,11E10.2,/12X,50A2,1H+)
2000 FORMAT(1X,F10.0,1X,101A1)
2001 FORMAT(12X,50A2,1H+/5X,11E10.2//////)
5000 FORMAT(/////
     *1X,'RANDOM VARIABLE :'//
     *8(4X,G12.6))
5002 FORMAT(/1X,'THE TABULATED NONPARAMETRIC DENSITY FUNCTION'
     *//8(4X,G12.6))
5003 FORMAT(//1X,'CHAIRS OF PLOT'//)
5004 FORMAT(1X,A1,' -  NONPARAMETRIC DENSITY FUNCTION')
5005 FORMAT(1X,'0 -  COMMON CHAIR FOR A FEW PLOTS'//)
      END
```

```
C        CIRCULAR TRANSFORMED KERNEL PROGRAM
C        PROGRAMMER: DR. W. FELUCH - 1988
C
C THIS PROGRAM EMPLOYS AN ANNUAL FLOOD SERIES AS AN INPUT TO CALCULATE
C FLOOD ESTIMATES FOR A PRE-SELECTED RETURN PERIOD USING THE CIRCULAR
C TRANSFORMED KERNEL ESTIMATOR SUGGESTED BY ADAMOWSKI AND FELUCH (1988).
C IN THE NUMERICAL PROCEDURE THE EPANECHNIKOV OPTIMAL KERNEL AND THE
C FLANAGAN OPTIMAL SMOOTHING FACTOR ARE APPLIED. THE PROGRAM PROVIDES
C ALSO THE FIXED KERNEL ESTIMATOR AS AN OPTION. DENSITY FUNCTION CAN BE
C PLOTTED UNDER REQUEST.
C
C INPUT FILE
C
C THE INPUT FILE COMPRISES :
C (1) NAME OF THE STATION IN CHARACTER VARIABLE ( FORMAT 18A4 )
C (2) N RECORD LENGTH OF FLOOD DATA (INTEGER)
C (3) ANNUAL FLOOD SERIES (REAL VARIABLE FREE FORMAT)
C (4) KL, KEY AND SCAL CONTROL PARAMETERS
C
C IMPORTANT VARIABLE
C
C XOBS : INPUT VECTOR OF FLOOD DATA
C N : NUMBER OF FLOOD OBSERVATION IN AN ANNUAL FLOOD SERIES.
C HH : SMOOTHING FACTOR
C T : RETURN PERIOD
C XP : FLOOD MAGNITUDE AS ESTIMATED BY THE TRANSFORMED KERNEL
C KL : INTEGER CONTROLLING THE ESTIMATION METHOD
C        KL= 0 FOR TRANSFORMED KERNEL ESTIMATOR
C        KL= 1 FOR THE FIXED KERNEL ESTIMATOR
C KEY: INTEGER CONTROLLING OUTPUT OPTIONS
C        KEY=0 PRINTING ONLY FLOOD ESTIMATES FOR SPECIFIED RETURN PERIOD
C        KEY=1 PRINTING EMPIRICAL PROBABILITIES
C        KEY=2 PRINTING THEORETICAL PROBABILITIES
C SCAL: VARIABLE CONTROLLING THE STARTING POINT ON THE Y Axis OF THE
C       DENSITY FUNCTION (SCAL=0.0 NO NEED TO SPECIFY THE STARTING POINT
C A,B : LIMITS FOR METHOD OF BISECTION
C EPS : ACCURACY LEVEL FOR THE METHOD OF BISECTION
C
C DESCRIPTION OF MAIN SUBROUTINES
C
C  SORT . RANK THE DATA FROM MAXIMUM TO MINIMUM.
C  HH   : CALCULATE THE SMOOTHING FACTOR FOR THE FIXED KERNEL.
C  TR   : TRANSFORM THE FLOOD DATA.
C  FCN  : FOR A GIVEN RETURN PERIOD THIS FUNCTION IS EQUATED TO ZERO
C         AND SOLVED FOR XP USING POLO.
C  POLO : COMPUTES ROOTS OF A FUNCTION BY METHOD OF BISECTION.
C  W    : CALCULATE THE INTEGRAL OF THE KERNEL FUNCTION K(.).
C  DENS : CALCULATE DENSITY FUNCTION AT EACH FLOOD OBSERVATION.
C  PLOT : GENERATE THE PLOT OF THE DENSITY FUNCTION.
C  OUTX : PRINTS FLOOD DATA IN COLUMNS
C  PEM  : COMPUTES EMPIRICAL PROBABILITIES.
C
C
      CHARACTER*6 TT(8,6),IZ(6),SP,ST
      DIMENSION VALUE(2,333),ROB(3,3),XMEM(3),IY(3),T(13)
      COMMON /OBS/N,XOBS(333),YY(333),SUM,FN,H,HH
     1/MIN/XMAX,FIP,KL
     2/KY/MP,TITL/DE/D
      DATA T/1.005,1.05,1.25,2.,5.,10.,20.,
     *50.,100.,200.,500.,1000.,2000./
      DATA (TT(I,1),I=1,6)/'THE TR',
     *'ANSFOR','MED KE','RNEL E','STIMAT','E     '/
      DATA (TT(I,2),I=1,8)/'TABULA','TED HO','MPARAM'
     *,'ETRIC ','DENSIT','Y FUNC','TION  ','      '/
      DATA SP,ST/'      ','THE KE'/
      PI=3.1415926536
      EPS=.005
      MP=2
      TITL=0
      READ(5,*) KL,KEY,SCAL
      CALL INFIT(3)
      IF(KL.NE.1) WRITE(6,28)
      TT(1,1)=SP
      TT(2,1)=SP
      TT(3,1)=ST
C14   WRITE(6,26)
      IF(KL.NE.1) WRITE(6,31)
      READ(5,*) N
      READ(5,*)(XOBS(I),I=1,N)
C     WRITE(7,*) N
      IF(KEY.LE.0.OR.KEY.EQ.2) CALL OUTX(N,XOBS)
      IF(KEY.EQ.1) CALL SORTP(N,XOBS)
      IF(KEY.EQ.2) CALL SORT(N,XOBS)
      XMAX=-10.E33
      XMIN=-XMAX
      DO 9 I=1,N
      XI=XOBS(I)
      IF(XI.GT.XMAX) XMAX=XI
      IF(XI.LT.XMIN) XMIN=XI
    9 CONTINUE
      FN=FLOAT(N)
      HH=HH(N,XOBS)
      WRITE(6,29) HH
      IF(KL.EQ.1) GO TO 6
      DO 7 I=1,N
      XI=XOBS(I)
    7 YY(I)=TR(XI)
      H=5.*PI/16./FN
      H=H**0.2
      WRITE(6,30) H
```

```
      SUM=0.
      DO 15 I=1,N
      Y=YT(I)
      Y1=1.+Y
      Y=1.-Y
15    SUM=SUM+W(Y/H)+W(Y1/H)
      SUM=SUM/FN
      WRITE(6,27) SUM
6     IF(KEY.LE.1) GO TO 10
      WRITE(6,24)
      M1=M-1
      M2=M-2
      DO 4 K=1,M,3
      DO 1 J=1,3
      I=K+J-1
      IF(I.GT.N) GO TO 1
      XI=XOBS(I)
      ROB(1,J)=DENS(XI)
      IF(KL.EQ.1) D=1.-D
      IF(KL.NE.1) D=AMAL(XI)
      ROB(2,J)=D
      ROB(3,J)=PEM(N,I)
      XMEM(J)=XI
1     CONTINUE
      IF(K.GT.M2) GO TO 11
      WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
     *,XMEM(2),(ROB(L1,2),L1=1,3),XMEM(3),(ROB(L1,3),L1=1,3)
      GO TO 4
11    IF(K.GT.M1) GO TO 12
      WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
     *,XMEM(2),(ROB(L1,2),L1=1,3)
      GO TO 4
12    WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
4     CONTINUE
10    IF(KEY.LE.0) GO TO 13
C     PLOTTING OF DENSITY
      IY(2)=2
      IY(1)=1
      BOUND=XMAX
      DIF=AINT(XMAX/100.)
      IF(DIF.LE.1.5) DIF=1.
      IF(DIF.GT.1.5.AND.DIF.LE.2.5) DIF=2.
      IF(DIF.GT.2.5.AND.DIF.LE.7.5) DIF=5.
      IF(DIF.GT.7.5.AND.DIF.LE.12.5) DIF=10.
      IF(DIF.GT.12.5.AND.DIF.LE.27.5) DIF=20.
      IF(DIF.GT.27.5.AND.DIF.LE.75.0) DIF=50.
      IF(DIF.GT.75..AND.DIF.LE.125.) DIF=100.
      IF(DIF.GT.125..AND.DIF.LE.300.) DIF=200.
      IF(DIF.GT.300..AND.DIF.LE.700.) DIF=500.
      IF(DIF.GT.700..AND.DIF.LE.1300.) DIF=1000.
      IF(DIF.GT.1300.) DIF=2000.
      ISCAL=0
      IF(SCAL.LE.0.) ISCAL=1
      XI=0.
      XMID=XMIN*0.75
      I=0
3     XI=XI+DIF
      IF(XI.LT.XMID) GO TO 3
      IF(XI.GT.BOUND) GO TO 2
      I=I+1
      VALUE(1,I)=XI
      VALUE(2,I)=DENS(XI)
      GO TO 3
2     CALL PLOT(I,2,2,IY,VALUE,0.,SCAL,0,ISCAL,IPRIN)
C     COMPUTING OF FLOODS
13    B=XMAX
      A=0
      BO=XMAX
      WRITE (6,22)
      DO 5 I=1,11
      FIP=1./T(I)
      XP=POLO(A,B,EPS)
      BO=BO*1.1
      B=BO
      A=XP
5     WRITE(6,23) T(I),XP
      WRITE(6,28)
22    FORMAT(/5X,15HRETURN    FLOOD/5X,
     *17HPERIOD   ESTIMATE/)
23    FORMAT(2X,F9.3,3X,F7.0)
24    FORMAT(/,'DENSITY FUNCTION FOR EMPIRICAL DATA  '.
     *//,3(4X,'XOBS',1X,'DENSITY',2X,2('PROBABILITY',1X)),/.
     *3(17X,'THEORETICAL  EMPIRICAL  '),/)
25    FORMAT(3(F7.0,3(1X,G11.5),1X))
26    FORMAT(1X,'THE TRANSFORMED KERNEL ESTIMATE'./)
27    FORMAT(' SUM = ',G11.5/)
28    FORMAT(////)
29    FORMAT(' H FOR OBSERVATED DATA    = ',G11.5)
30    FORMAT(' H FOR TRANSFORMATED DATA = ',G11.5/)
31    FORMAT(' CIRCLERY TRANSFORMATION'//)
      STOP
      END
      FUNCTION DENS(X)
C     DENSITY OF TRANSFORMATED DATA -IF KL=0
C     DENSITY OF OBSERVATED DATA   -IF KL=1
      COMMON /OBS/N,XOBS(333),YT(333),SUM,FN,H,HH
     */DE/D/XMIN/XMAX,FIP,KL
      PI=3.1415926536
      DENS=0.
      DY=0.
```

```
      T=TR(X)
      DO 2 I=1,M
      IF(KL.EQ.1) GO TO 6
      Y=(T-YT(I))/H
      AY=ABS(Y)
      Z=0.
      IF(AY.LT.1.) Z=0.75*(1.-Y*Y)
      DENS=DENS+Z
    6 Y=(X-XOBS(I))/HH
      AY=ABS(Y)
      Z=0.
      IF(AY.LT.1.) Z=0.75*(1.-Y*Y)
    2 DY=DY+Z
      DY=DY/FM/HH
C     DY - DENSITY OF OBSERVATION DATA
      IF(KL.NE.1) GO TO 5
      DENS=DY
      RETURN
    5 Z=1.-2.*D
      DIFTR=-4.*DY*SIN((4.*PI+ACOS(Z))/3.)/SQRT(1.-Z*Z)/3.
C     DIFTR - DIFFERENTIABLE OF TRANSFORMATION
    4 DENS=DENS*DIFTR/FM/H/SUM
      RETURN
      END
      FUNCTION TR(X)
C     TRANSFORMATION
      COMMON /DE/D
     */OBS/M,XOBS(333),YT(333),SUM,FM,H,HH
      PI=3.1415926536
      D=0.
      DO 3 I=1,M
      Y=(X-XOBS(I))/HH
    3 D=D+W(Y)
      D=0.5+D/FM
      TR=2.*COS((4.*PI+ACOS(1.-2.*D))/3.)
      RETURN
      END
      FUNCTION FCN(M)
      COMMON /MIN/XMAX,FIP,KL/DE/D
      IF(KL.GT.0) GO TO 1
      FCN=FIP-ANAL(M)
      RETURN
    1 TM=TR(M)
      FCN=FIP+D-1
      RETURN
      END
      FUNCTION ANAL(X)
      COMMON /OBS/M,XOBS(333),YT(333),SUM,FM,H,HH
      S=0.
      TX=TR(X)
      DO 1 I=1,M
      YI=YT(I)
      YI=(1.-YI)/M
      Y=(TX-YI)/M
    1 S=S+W(YI)-W(Y)
      ANAL=S/FM/SUM
      RETURN
      END
      FUNCTION W(X)
      W=-0.5
      IF(X.GT.-1..AND.X.LT.1.) W=0.75*X*(1.-X*X/3.)
      IF(X.GE.1.) W=0.5
      RETURN
      END
      FUNCTION ASIN(X)
      ASIN=1.5707963268-SQRT(1.-X)*(1.5707963050+
     1X*(-.2145988016+X*(.0889789874+X*(-.0501743046+
     2X*(.0308918810+X*(-.0170881256+X*(.0066700901
     3-X*.0012624911)))))))
      RETURN
      END
      FUNCTION ACOS(X)
      A=ASIN(SQRT(1.-X*X))
      IF(X.GE.-1..AND.X.LT.0.) ACOS=3.1415926536-A
      IF(X.GE.0..AND.X.LE.1.) ACOS=A
      RETURN
      END
      SUBROUTINE OUTX(M,Y)
      DIMENSION Y(M)
      FM=FLOAT(M)
      FL=10.
      R=FM/FL
      R=R-AINT(R)
      R=R*FL+0.5
      IR=INT(R)-1
      NIL=INT(FL)
      ML=NIL
      IL=1
      WRITE(6,8)
      DO 1 I=1,M,ML
      J=I+ML-1
      IF(NIL.LT.ML) J=I+IR
      WRITE(6,8)(Y(L),L=I,J)
      NIL=M-ML*IL
      IL=IL+1
    1 CONTINUE
      WRITE(6,7)
    9 FORMAT(7H XOBS :/)
    8 FORMAT(1X,20F7.0)
    7 FORMAT(1X//)
```

```
      RETURN
      END
      SUBROUTINE INTIT(NI)
      DIMENSION T(18)
      READ(5,10)(T(K),K=1,18)
      IF(NI.EQ.1) WRITE(6,21)(T(K),K=1,18)
      IF(NI.EQ.2) WRITE(6,22)(T(K),K=1,18)
      IF(NI.EQ.3) WRITE(6,23)(T(K),K=1,18)
10    FORMAT(18A4)
21    FORMAT(1H,///4X,80(1H*)/4X,2H* ,
     118A4,2H */4X,80(1H*)///)
22    FORMAT(1H ///8X,68(1H-)/8X,2H* ,18A4,
     12H *,/8X,68(1H-)//)
23    FORMAT(1H ,18A4/)
      RETURN
      END
      FUNCTION POLO(A,B,EPS)
      FA=FCN(A)
      FB=FCN(B)
      FAB=FA*FB
      IF(FAB.LE.0.) GO TO 10
C     WRITE(6,66) FA,FB,A,B
      WRITE(6,67)
      STOP
66    FORMAT(///1X,10(1H*),3X,'TO CHANGE
     1BOUNDARY IN POLO',//1X,10(1H*),3X,
     2'STOP OF PROGRAM'/
     3/' FCN(A) =',E18.8,3X,' FCN(B) =',E18.8,
     4/' A= ',E18.8,3X,' B = ',E18.8)
67    FORMAT(////)
10    X=(A+B)/2
      FM=FCN(X)
      FAM=FA*FM
      IF(FAM) 2,3,4
2     B=X
      GO TO 5
4     FA=FM
      A=X
5     IF(ABS(A-B).GT.EPS) GO TO 10
3     POLO=(A+B)/2
      RETURN
      END
      SUBROUTINE SORT(N,A)
C
C
      DIMENSION A(N)
      N1=N-1
      DO 2 I=1,N1
      I1=I+1
      AI=A(I)
      DO 2 J=I1,N
      AJ=A(J)
      IF(AI.GT.AJ) GO TO 4
      A(J)=AI
      A(I)=AJ
      AI=AJ
4     CONTINUE
2     CONTINUE
      RETURN
      END
      FUNCTION HN(N,XOBS)
      DIMENSION XOBS(N),X(333)
      ALFA=0.5
      DO 1 I=1,N
1     X(I)=XOBS(I)
      DO 2 I=1,N
      IHIGH=N
      ILOW=1
3     IF(IHIGH.EQ.ILOW) GO TO 4
      L=(ILOW+IHIGH)/2
      IF(X(I)-X(L)) 5,7,6
5     IHIGH=L
      GO TO 3
6     ILOW=L+1
      GO TO 3
7     K=L+1
      GO TO 8
4     K=ILOW
8     SAVE=X(I)
      ISORT=I-K
      DO 9 I2=1,ISORT
      L=I+1-I2
9     X(L)=X(L-1)
2     X(K)=SAVE
      S=0
      FN=FLOAT(N)
      XN=INT(FN**ALFA)
      DO 10 I=1,N
      XI=X(I)
      M=I-XN
      B=999999.
      IF(M.GT.0) B=XI-X(M)
      M=I+XN
      A=999999
      IF(M.LE.N) A=X(M)-XI
      D=0
      IF(A.LT.B) D=A
10    S=S+D
      HN=S/FN
      RETURN
```

```
      END
      FUNCTION PEM(N,I)
C
C     THIS FUNCTION COMPUTES EMPIRICAL PROBABILITY PJ
C
      COMMON /XY/MP
      FI=FLOAT(I)
      FN=FLOAT(N)
      GO TO (1,2,3,4,5,6) ,MP
    1 PEM=FI/(FN+1.)
      RETURN
    2 PEM=(FI-0.25)/(FN+0.5)
      RETURN
    3 F=FN+1
      S=SQRT(F*FN)
      PEM=(2.*FI-F+S)/3/2.
      RETURN
    4 PEM=(FI-0.3)/(FN+0.4)
      RETURN
    5 PEM=(FI-0.4)/(FN+0.2)
      RETURN
    6 PEM=(FI-0.5)/FN
      RETURN
      END
      SUBROUTINE SORTP(N,A)
C
C
      DIMENSION A(N)
      COMMON /XY/MP,TITL
      N1=N-1
      DO 2 I=1,N1
      I1=I+1
      AI=A(I)
      ID=INT(TITL)
      DO 2 J=I1,N
      AJ=A(J)
      IF(AI.GT.AJ.AND.ID.EQ.0) GO TO 4
      IF(AI.LE.AJ.AND.ID.NE.0) GO TO 4
      A(J)=AI
      A(I)=AJ
      AI=AJ
    4 CONTINUE
    2 CONTINUE
      N3=N-3
      N2=N-2
      N1=N-1
      WRITE(6,23)
      DO 12 I=1,N,4
      I4=I+3
      I3=I+2
      I2=I+1
      IF(I.GT.N3) GO TO 33
      P1=PEM(N,I)
      P2=PEM(N,I2)
      P3=PEM(N,I3)
      P4=PEM(N,I4)
      WRITE(6,22) P1,A(I),P2,A(I2),P3,A(I3),P4,A(I4)
      GO TO 5
   33 IF(I.GT.N2) GO TO 24
      P1=PEM(N,I)
      P2=PEM(N,I2)
      P3=PEM(N,I3)
      WRITE(6,22) P1,A(I),P2,A(I2),P3,A(I3)
      GO TO 5
   24 IF(I.GT.N1) GO TO 13
      P1=PEM(N,I)
      P2=PEM(N,I2)
      WRITE(6,22) P1,A(I),P2,A(I2)
      GO TO 5
   13 P1=PEM(N,I)
      WRITE(6,22) P1,A(I)
    5 CONTINUE
   12 CONTINUE
      WRITE(6,21)
   22 FORMAT('1H',4(F7.4,1X,F8.1,3X))
   23 FORMAT(1H,/17H EMPIRICAL DATA : ,///4(4X,1HP,7X,
     *4HOBS,3X)/)
   21 FORMAT(///)
      RETURN
      END
      SUBROUTINE PLOT(N,M,NY,IY,Y,YMN,YMX,IPO,ISCAL,IPR(N)
C
C ***** PROCEDURE FOR PLOTTING THE DENSITY FUNCTION
      DIMENSION IY(1),Y(N,1),SCL(13)
      CHARACTER*8 IZ(8),IH1,IH2,IH3,IH4,
     *LINE(121),IXIS(80)
      DATA IH1,IH2,IH3,IH4
     1/'+-','---','  ','  '/
      DATA IZ/'*','$','X','0','5','D'/
      DO 6 I=1,12
      K=(I-1)*5+1
      IXIS(K)=IH1
      DO 6 J=1,4
      L=K+J
    6 IXIS(L)=IH2
      IF(ISCAL.EQ.0) GO TO 4
      YMN=1.E30
      YMX=-1.0E30
      DO 9 I=2,NY
      L=IY(I)
```

```
      DO 5 J=1,N
      YY=Y(L,J)
      IF(YY.LT.YMN) YMN=YY
      IF(YY.GT.YMX) YMX=YY
    5 CONTINUE
    4 II=IY(I)
      DY=(YMX-YMN)*0.01
      DO 10 I=1,15
      AI=FLOAT(I)
   10 SCL(I)=((11.-AI)*YMN+(AI-1.)*YMX)*0.1
      IF(NY.GT.6) NY=6
      IF(IPO.EQ.0) GO TO 40
      IMX=M-1
      DO 30 I=1,IMX
      JMX=N-I
      DO 30 J=1,JMX
      IF( Y(II,J).LE. Y(II,J+1)) GO TO 30
      DO 20 K=1,M
      TMP=Y(K,J)
      Y(K,J)= Y(K,J+1)
      Y(K,J+1)= TMP
   20 CONTINUE
   30 CONTINUE
   40 I5=5
      IF(IPRIN.EQ.0) GO TO 1
      L=IY(1)
      WRITE(6,5000)  (Y(L,K),K=1,M)
      DO 2 I=2,NY
      L=IY(I)
      WRITE(6,5002)(Y(L,K),K=1,M)
    2 CONTINUE
    1 WRITE(6,5003)
      DO 3 I=2,NY
      II=I-1
      L=IY(I)
    3 WRITE(6,5004)  IZ(II)
      WRITE(6,5005)
      WRITE(6,1001)  (SCL(I),I=1,11),(IXIS(I),I=1,50)
      DO 200 J=1,N
      DO 50 I=1,101
   50 LINE(I)=IH4
      DO 100 IYY=2,NY
      II=IYY-1
      I=IY(IYY)
      YY=Y(I,J)
      IF(YY.GT.YMX.OR.YY.LT.YMN) GO TO 100
      BI=(YY-YMN)/DY
      IB=INT(BI)
      IF((BI-FLOAT(IB)).GE.0.5) IB=IB+1
      IB=IB+1
      IF(LINE(IB).NE.IH4) GO TO 60
      LINE(IB)=IZ(II)
      GO TO 100
   60 LINE(IB)=IZ(6)
  100 CONTINUE
      DO 80 L=1,101,10
      IF(LINE(L).EQ.IH4) LINE(L)=IH3
   80 CONTINUE
      IF(J.NE.15) GO TO 90
      IF(J.EQ.N) GO TO 90
      DO 70 L=1,101,2
      IF(LINE(L).EQ.IH4) LINE(L)=IH3
   70 CONTINUE
      I5=I5+5
   90 WRITE(6,2000)
     *Y(II,J),(LINE(L),L=1,101)
  200 CONTINUE
      WRITE(6,2001) (IXIS(L),L=1,50),(SCL(L),L=1,11)
      RETURN
 1001 FORMAT(5X,11E10.2,/12X,50A2,1H+)
 2000 FORMAT(1X,F10.0,1X,101A1)
 2001 FORMAT(12X,50A2,1H+/5X,11E10.2//////)
 5000 FORMAT(/////
     *1X,'RANDOM VARIABLE :'//
     *8(4X,G12.6))
 5002 FORMAT(/1X,'THE TABULATED NONPARAMETRIC DENSITY FUNCTION'
     *//8(4X,G12.6))
 5003 FORMAT(//1X,'CHAIRS OF PLOT'//)
 5004 FORMAT(1X,A1,' -  NONPARAMETRIC DENSITY FUNCTION')
 5005 FORMAT(1X,'O -  COMMON CHAIR FOR A FEW PLOTS'//)
      END
```

```
C           VARIABLE KERNEL ESTIMATOR PROGRAM
C            PROGRAMMER DR. W. FELLOW - 1986
C
C THIS PROGRAM EMPLOYS AN ANNUAL FLOOD SERIES AS AN INPUT TO CALCULATE
C FLOOD ESTIMATES FOR PRESELECTED RETURN PERIODS USING THE VARIABLE
C KERNEL ESTIMATOR SUGGESTED BY BREIMAN ET AL 1977. IN THE NUMERICAL
C PROCEDURE THE EPANECHNIKOV OPTIMAL KERNEL IS APPLIED. THIS PROGRAM
C PROVIDE THE OPTION OF INCLUDING HISTORICAL ANALYSIS AND THE USE OF THE
C CIRCULAR TRANSFORMED AND FIXED KERNEL ESTIMATORS.
C THE PARAMETERS OF THE VARIABLE KERNEL ARE ESTIMATED USING THE
C MAXIMUM LIKELIHOOD METHOD. DENSITY FUNCTION CAN BE PLOTTED UNDER
C REQUEST OF THE USER.
C
C THE INPUT FILE COMPRISES :
C (1) NAME OF THE STATION IN CHARACTER VARIABLE ( FORMAT 18A4 )
C (2) N RECORD LENGTH OF FLOOD DATA (INTEGER)
C (3) ANNUAL FLOOD SERIES (REAL VARIABLE FREE FORMAT)
C (4) NUMBER OF CENSORED FLOODS LESS THAN XC
C (5) THE CENSORING THRESHOLD XC
C
C IMPORTANT VARIABLES
C
C XOBS : INPUT VECTOR OF FLOOD DATA
C N : NUMBER OF FLOOD OBSERVATION IN AN ANNUAL FLOOD SERIES.
C HH : SMOOTHING FACTOR
C T : RETURN PERIOD
C XP : FLOOD MAGNITUDE AS ESTIMATED BY THE TRANSFORMED KERNEL
C KL : INTEGER CONTROLLING THE ESTIMATION METHOD
C       KL= 0 FOR CIRCULAR TRANSFORMED KERNEL ESTIMATOR
C       KL= 1 FOR THE FIXED KERNEL ESTIMATOR
C KEY: INTEGER CONTROLLING OUTPUT OPTIONS
C       KEY=0 PRINTING ONLY FLOOD ESTIMATES FOR SPECIFIED RETURN PERIODS
C       KEY=1 CALCULATE FLOOD QUANTILES, PLOT DENSITY AND EXCEEDANCE
C             PROBABILITY.
C       KEY=2 AS KEY=1 AND PRINT MATRIX OF PROBABILITY.
C       KEY=3 AS KEY=2 AND MATRIX OF NEAREST NEIGHBOURS.
C SCAL: VARIABLE CONTROLLING THE STARTING POINT ON THE Y Axis OF THE
C       DENSITY FUNCTION (SCAL=0.0 NO NEED TO SPECIFY THE STARTING POINT
C IPAL: INTEGER VARIABLE CONTROLS THE OPTION OF USING HISTORICAL
C       ANALYSIS.
C       IPAL=0 NO HISTORICAL INFORMATION, NO VARIABLE KERNEL ESTIMATOR.
C       IPAL=1 USE HISTORICAL ANALYSIS WITH OR WITHOUT HISTORICAL DATA.
C NCD : NUMBER OF CENSORED FLOODS LESS THAN THE CENSORING THRESHOLD.
C XC  : CENSORING THRESHOLD
C A,B : LIMITS FOR METHOD OF BISECTION
C EPS : ACCURACY LEVEL FOR THE METHOD OF BISECTION
C
C DESCRIPTION OF MAIN SUBROUTINES
C
C  SORT : RANK THE DATA FROM MAXIMUM TO MINIMUM.
C  MR   : CALCULATE THE SMOOTHING FACTOR FOR THE FIXED KERNEL.
C  TR   : TRANSFORM THE FLOOD DATA.
C  FCN  : FOR A GIVEN RETURN PERIOD THIS FUNCTION IS EQUATED TO ZERO
C         AND SOLVED FOR XP USING POLO.
C  POLO : MAXIMIZING AND SOLVING FOR AK THE LOGARITHMIC OF THE
C         LIKELIHOOD FUNCTION.
C  W    : CALCULATE THE INTEGRAL OF THE KERNEL FUNCTION K(.).
C  DENS : CALCULATE DENSITY FUNCTION AT EACH FLOOD OBSERVATION.
C  PLOT : GENERATE THE PLOT OF THE DENSITY FUNCTION.
C  OUTX : PRINTS FLOOD DATA IN COLUMNS
C  PEM  : COMPUTES EMPIRICAL PROBABILITIES.
C  DIST : CALCULATES THE K-TH NEAREST NEIGHBOUR VALUES AND THE MEAN
C         OF THE K-TH NEAREST DISTANCE VALUES (COMPUTE K)
C  SPB  : GENERATE AND PLOT THE DISTRIBUTION FUNCTION F(.).
C
C ****************************************************************
      DIMENSION T(13),ROB(3,3),XMEM(3)
     *,IY(3),TT(8,8)
      COMMON /OBS/N,XOBS(222),YT(222),SUM,FN,H,HH
     */DIS/DST(222)/MIN/XMAX,FIP,KL,A
     */KY/HP,TITL/DE/D
     */XP/XX,XYZ/PAL/IPAL,LPAL(20),MPAL,MY,CD,XC,VALUE(3,222),KEY
     */KLU/KFIRST
      DATA T/1.005,1.05,1.25,2.,5.,10.,20.,
     *50.,100.,200.,500.,1000.,2000./
      PI=3.1415926536
      EPS=0.5
      NP=2
      TITL=0.
      READ(5,*) IPAL,KL,KEY,SCAL
      CALL INTIT(1)
      IF (KL.NE.1) WRITE(6,28)
      IF (KL.EQ.1) WRITE(6,35)
      IF(KL.NE.1) WRITE(6,31)
      READ(5,*) N,(XOBS(I),I=1,N)
      KFIRST=0
      IF(KEY.LE.0.OR.KEY.GE.2) CALL OUTX(N,XOBS)
      IF(IPAL.EQ.0) GO TO 18
      READ(5,*)NCD,XC
      WRITE(6,36)NCD,N,XC
      CD=FLOAT(NCD)
C     IF(KEY.LT.1) GO TO 18
C 19  READ(5,*)MPAL,MY
C     MAD=MPAL+1
C     LPAL(1)=1
C     READ(5,*) (LPAL(I),I=2,MAD)
C     WRITE(6,33)MPAL,MY,(LPAL(I),I=1,MAD)
   18 IA=2
      CALL SORT(0,N,XOBS)
      XMAX=-10.E33
```

```
      XMIN=-XMAX
      DO 9 I=1,N
      XI=XOBS(I)
      IF(XI.GT.XMAX) XMAX=XI
      IF(XI.LT.XMIN) XMIN=XI
9     CONTINUE
      FN=FLOAT(N)
      IF(IR.GT.0) GO TO 8
      R=FN-1.
      R=R+SQRT(R*R+8.*FN)
      R=R/4.
      IR=INT(R)
8     R=FLOAT(IR)
      IF(IPAL.GT.0) GO TO 18
      HH=HH(R,N,XOBS,ERR)
      WRITE(6,29) HH
      GO TO 17
18    KK=2
      CALL DIST
      A=0.0001
      B=2.
      EP=0.001
      STEP=(B-A)/1000.
      HH=POLO(A,B,EP,STEP)
C     HH=1.0
      WRITE(6,33) HH
17    CONTINUE
      IF(KL.EQ.1) GO TO 6
      DO 7 I=1,N
      XI=XOBS(I)
7     YY(I)=TR(XI)
      H=3.*PI/18./FN
      H=H**.2
      WRITE(6,30) H
      SUM=0.0
      DO 15 I=1,N
      Y=YY(I)
      Y1=1.+Y
      Y=1.-Y
15    SUM=SUM+W(Y/H)+W(Y1/H)
      SUM=SUM/FN
      WRITE(6,27) SUM
6     IF(KEY.LE.1) GO TO 10
      WRITE(6,24)
      N1=N-1
      N2=N-2
      DO 4 K=1,N,3
      DO 1 J=1,3
      I=K+J-1
      IF(I.GT.N) GO TO 1
      XI=XOBS(I)
      ROB(1,J)=DENS(XI)
      IF(KL.EQ.1) D=1.-D
      IF(KL.NE.1) D=ANAL(XI)
      ROB(2,J)=D
      ROB(3,J)=PEN(N,I)
      XMEM(J)=XI
1     CONTINUE
      IF(K.GT.N2) GO TO 11
      WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
     *,XMEM(2),(ROB(L1,2),L1=1,3),XMEM(3),(ROB(L1,3),L1=1,3)
      GO TO 4
11    IF(K.GT.N1) GO TO 12
      WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
     *,XMEM(2),(ROB(L1,2),L1=1,3)
      GO TO 4
12    WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
4     CONTINUE
10    IF(KEY.LE.0) GO TO 13
C     PLOTTING OF DENSITY
      IY(2)=2
      IY(1)=1
      BOUND=XMAX
      DIF=XMAX/100.
      IF(DIF.LE.1.5) DIF=1.
      IF(DIF.GT.1.5.AND.DIF.LE.2.5) DIF=2.
      IF(DIF.GT.2.5.AND.DIF.LE.7.5) DIF=5.0
      IF(DIF.GT.7.5.AND.DIF.LE.12.5) DIF=10.0
      IF(DIF.GT.12.5.AND.DIF.LE.27.5)DIF=20.
      IF(DIF.GT.27.5.AND.DIF.LE.75.) DIF=50.
      IF(DIF.GT.75..AND.DIF.LE.125.) DIF=100.
      IF(DIF.GT.125..AND.DIF.LE.300.) DIF=200.
      IF(DIF.GT.300..AND.DIF.LE.700.) DIF=500.
      IF(DIF.GT.700..AND.DIF.LE.1300.) DIF=1000.
      IF(DIF.GT.1300.) DIF=2000.
      ISCAL=0
      IF(SCAL.LE.0.) ISCAL=1
      XI=-DIF
      IF (KL.EQ.0) XMID=XMIN*0.75
      I=0
3     XI=XI+DIF
      IF(KL.EQ.0.AND.XI.LT.XMID) GO TO 3
      IF(XI.GT.BOUND) GO TO 2
      I=I+1
      VALUE(1,I)=XI
      VALUE(2,I)=DENS(XI)
      GO TO 3
2     CONTINUE
      IPRIM=KEY-1
      MM=2
```

```
      YMM=0.
      IPO=0
      CALL PLOT(1,MMY,MMY,IY,VALUE,YMM,SCAL,IPO,ISCAL,IPRIN)
C     PLOTTING OF RETURN PERIOD
      CALL SPB(N,XOBS,EPS,T,KEY)
C     COMPUTING OF FLOODS
13    B=XMAX
      A=0.
      BG=XMAX*2.
      STEP=BG/1000.
      WRITE (6,22)
      KK=0
      DO 5 I=2,11
      FIP=1./T(I)
      XP=POLO(A,B,EPS,STEP)
      BG=BG*1.1
      B=BG
      A=0.0
    5 WRITE(6,23) T(I),XP
      WRITE(6,26)
      IF(IPAL.EQ.0.OR.KEY.NE.3) STOP
      KFIRST=1
      DO 20 I=1,N
   20 DST(I)=OST(I)*HH
      CALL OUTX(N,DST)
   22 FORMAT(///5X,15HRETURN    FLOOD/5X,
     *17HPERIOD   ESTIMATE/)
   23 FORMAT(2X,F8.3,3X,F7.0)
   24 FORMAT(///' DENSITY FUNCTION FOR EMPIRICAL DATA :',
     *//,3(4X,'XOBS',1X,'DENSITY',3X,2('PROBABILITY',1X)),/,
     *3(18X,'THEORETICAL   EMPIRICAL   '),/)
   25 FORMAT(3(F7.0,3(1X,G11.5),1X))
   26 FORMAT(1X,'THE TRANSFORMED KERNEL ESTIMATE',/)
   27 FORMAT(' SUM = ',G11.5/)
   28 FORMAT(////)
   29 FORMAT(' N FOR OBSERVATED DATA    = ',G11.5)
   30 FORMAT( ' N FOR TRANSFORMATED DATA = ',G11.5/)
   31 FORMAT(' CIRCLERY TRANSFORMATION'//)
   32 FORMAT(' THE PARAMETER OF VARIABLE KERNEL  A =',G12.5//)
C33 FORMAT(' MPAL =',I3/
C    *' MY   =',I3/
C    *' LPAL :',20I4/)
   34 FORMAT(' NUMBER OF CENSORED FLOODS      NCD =',I4/
     *' NUMBER OF FULLY DEFINED FLOODS   N =',I4/
     *' THE CENSORING THRESHOLD          XC =',G11.5)
   35 FORMAT(' KERNEL ESTIMATE'/)
      STOP
      END
C ***************************************************************
      FUNCTION DENS(X)
C ***************************************************************
C     DENSITY OF TRANSFORMATED DATA -IF KL=0
C     DENSITY OF OBSERVATED DATA   -IF KL=1 OR KL=1 AND IPAL=1
      COMMON /OBS/N,XOBS(222),YY(222),SUM,FN,H,HH
     */DIS/DST(222)/MIN/XMAX,FIP,KL,R/KY/NPP,TITT/DE/D
     */KP/KK,XYZ/PAL/IPAL,LPAL(20),MPAL,NY,CO,XC
      PI=3.1415926536
      DENS=0.
      DY=0.
      IF(KL.EQ.0) T=TR(X)
      DO 2 I=1,N
      IF(KL.EQ.1) GO TO 6
      FI=FLOAT(I)
      Y=(T-YY(FI))/H
      AY=ABS(Y)
      Z=0.
      IF(AY.LT.1.) Z=0.75*(1.-Y*Y)
      DENS=DENS+Z
    6 Y=(X-XOBS(I))/HH
      IF(IPAL.LE.0) GO TO 8
      DSTI=OST(I)
      Y=Y/DSTI
    8 AY=ABS(Y)
      Z=0.
      IF(AY.GE.1.) GO TO 7
      ZZ=1.
      IR=INT(R)
      DO 3 K=1,IR
    3 ZZ=ZZ*AY
      Z=(1.-ZZ)*(R+1.)/2./R
    7 CONTINUE
      IF (IPAL.GT.0) Z=Z/DSTI
    2 DY=DY+Z
      DY=DY/FN/HH
C     WRITE(6,666) DY
C666  FORMAT(' DENSITY = ',E18.6/)
C     DY - DENSITY OF OBSERVATION DATA
      IF(KL.NE.1) GO TO 5
      DENS=DY
      RETURN
    5 Z=1.-2.*D
      DIFTR=-4.*DY*SIN((4.*PI+ACOS(Z))/3.)/SQRT(1.-Z**2)/3.
C     DIFTR - DIFFERENTIABLE OF TRANSFORMATION
    4 DENS=DENS*DIFTR/FN/H/SUM
      RETURN
      END
C ***************************************************************
      FUNCTION TR(X)
C ***************************************************************
C     TRANSFORMATION
```

```
      COMMON /OBS/N,XOBS(222),YT(222),SUM,FN,H,HH
     */DIS/DST(222)/MIN/XMAX,FIP,KL,R
     */XY/NPP,TITT/DE/D
     */XP/KK,XYZ/PAL/IPAL,LPAL(20),MPAL,NY,CO,XC
      PI=3.1415926536
      TR=0.
      D=0.
      DO 3 I=1,N
      XI=XOBS(I)
      IF(IPAL.GT.0) XI=YT(I)
      Y=(X-XI)/HH
      IF(IPAL.GT.0) Y=Y/DST(I)
3     D=D+W(Y)
      D=0.5+D/FN
      IF(KL.EQ.0) TR=2.*COS((4.*PI+ACOS(1.-2.*D))/3.)
      RETURN
      END
C     ******************************************************************
      FUNCTION FCN(NN)
C     ******************************************************************
      COMMON /OBS/N,XOBS(222),YT(222),SUM,FN,H,HH/DIS/DST(222)
     1/MIN/XMAX,FIP,KL,RRR
     */XY/NPP,TITT/DE/D
     2/XP/KK,PROB/PAL/IPAL,LPAL(20),MPAL,NY,CO,XC
      IF(KK.LE.0) GO TO 2
      IF(KK.EQ.2) GO TO 3
      FIP=PROB
      GO TO 1
2     CONTINUE
      IF(KL.GT.0) GO TO 1
      FCN=FIP-ANAL(NN)
      RETURN
1     TR=TR(NN)
      FCN=FIP+D-1.
      RETURN
3     D=0.
      IF(CO.LE.0.) GO TO 5
      S=0.
      R=0.
      DO 4 J=1,N
      DJ=DST(J)
      XJ=YT(J)
      XX=XC-XJ
      WY=XX/HH/DJ
      Y=ABS(WY)
      IF(Y.EQ.0.) GO TO 4
      Z=0.
      IF(Y.LT.1 ) Z=0.75*(1.-Y*Y)
      S=S+XX*Z/DJ
      R=R+W(WY)
4     CONTINUE
      R=0.5+R/FN
      D=CO*S/R/FN/FN
5     QZ=0.
      DO 6 I=1,N
      S=0.
      R=0.
      XI=YT(I)
      DO 7 J=1,N
      IF(J.EQ.I) GO TO 7
      DJ=DST(J)
      XJ=YT(J)
      XX=XI-XJ
      WY=XX/HH/DJ
      Y=ABS(WY)
      Z=0.
      IF(Y.LT.1.) Z=-1.5*Y
      S=S+(XX/DJ)*Z/DJ
      Z=0.
      IF(Y.LT.1.) Z=0.75*(1.-Y*Y)
      R=R+Z/DJ
7     CONTINUE
C     WRITE(8,*) QZ,S,R
6     IF(R.NE.0.)QZ=QZ+S/R
      Q=Q+QZ/FN+NN
      FCN=Q
C     WRITE(8,*) FCN
      RETURN
      END
C     ******************************************************************
      SUBROUTINE DIST
C     ******************************************************************
      LOGICAL LL
      DIMENSION IY(2)
      COMMON /OBS/N,XOBS(222),YT(222),SUM,FN,ZXYZ,YZX/DIS/DST(222)
     */MIN/XMAX,FIP,KL,R
     */XY/NPP,TITT/DE/DDD
     */XP/KK,XYZ/PAL/IPAL,LPAL(20),MPAL,NY,CO,XC,VALUE(2,222),KEY
      DO 1 I=1,N
1     YT(I)=XOBS(I)
      DO 2 I=1,N
      IHIGH=I
      ILOW=1
3     IF(IHIGH.EQ.ILOW) GO TO 4
      L=(ILOW+IHIGH)/2
      IF(YT(I)-YT(L)) 5,7,6
5     IHIGH=L
      GO TO 3
6     ILOW=L+1
      GO TO 3
```

```
      K=L+1
        GO TO 8
    4 K=ILOW
    8 SAVE=YT(I)
        ISORT=I-K
        DO 9 I2=1,ISORT
        L=I+1-I2
    9 YT(L)=YT(L-1)
    2 YT(K)=SAVE
        IF(IPAL.GT.0.AND.KEY.EQ.3) CALL OUTX(N,YT)
        N1=N
        DO 14 KN=1,N1
        S=0.
        DO 15 I=1,N
   15 S=S+RNN(I,KN)
        VALUE(1,KN)=FLOAT(KN)
   14 VALUE(2,KN)=S/FN
        IF(KEY.LE.2) GO TO 16
        NNY=2
        YMN=0.
        IPO=0
        IY(1)=1
        IY(2)=2
        ISCAL=1
        IPRIN=KEY
        SCAL=0.
        CALL PLOT(N1,NNY,NNY,IY,VALUE,YMN,SCAL,IPO,ISCAL,IPRIN)
   16 KONTR=-1
        FN2=FN/2.
        KN=INT(FN2)
   11 LL=.TRUE.
        DO 10 I=1,N
        D=RNN(I,KN)
        IF(D.LE.0.) LL=.FALSE.
   10 DST(I)=D
        IF(LL) WRITE(6,35) KN
        IF(LL) RETURN
        KONTR=KONTR+1
        WRITE(6,33)
        IF(KONTR.LE.1) GO TO 12
        WRITE(6,34)
        STOP
   12 KN=KN+1
        GO TO 11
   33 FORMAT(//' DISTANCE ZERO')
   34 FORMAT(//' YOU GIVE KEY=3 AND TO SEE PLOT OF THE K-TH',
     *' NEARST NEIGHBOR DISTANCE'///)
   35 FORMAT(/' NUMBER OF THE NEARST NEIGHBOR KN =',I4/)
        END
C ***************************************************************
```

```
      FUNCTION RNN(I,KN)
C ***************************************************************
C
C     THIS IS A PART OF NEAREST NEIGHBOR ALGORITM -
C     -  RELATED TO SUBROUTINE DIST
C
      COMMON /OBS/N,XOBS(222),YT(222),SUM,FN,ZXYZ,YZX
     */DIS/DST(222)/MIN/XMAX
      YI=YT(I)
      M=I-KN
      B=10.E33
      IF(M.GT.0) B=YI-YT(M)
      M=I+KN
      A=10.E33
      IF(M.LE.N)A=YT(M)-YI
      D=B
      IF(A.LT.B) D=A
      RNN=D
      RETURN
      END
C ***************************************************************
      FUNCTION ANAL(X)
C ***************************************************************
      COMMON /OBS/N,XOBS(222),YT(222),SUM,FN,N,HH
      S=0.
      TX=IR(X)
      DO 1 I=1,N
      FI=FLOAT(I)
      YI=YT(FI)
      YI=(1.-YI)/H
      Y=(TX-YI)/H
    1 S=S+W(YI)-W(Y)
      ANAL=S/FN/SUM
      RETURN
      END
C ***************************************************************
      FUNCTION W(X)
C ***************************************************************
      A=1.7320508
      W=-0.5
      IF(X.GT.-1..AND.X.LT.1.) W=0.75*X*(1.-X/A)*(1.+X/A)
      IF(X.GE.1.) W=0.5
C     WRITE(6,6)W
C 6   FORMAT(' W =',E16.6)
C     STOP
      RETURN
      END
C ***************************************************************
      FUNCTION ASIN(X)
C ***************************************************************
```

```
      AS(N=1.5707963268-SQRT(1.-X)*(1.5707963050+
     *X*(-.2145988016+X*(.0889789874+X*(-.0501743046+
     *X*(.0308918810+X*(-.0170881256+X*(.0066700901
     *-X*.0012624911)))))))
      RETURN
      END
C ***************************************************************
      FUNCTION ACOS(X)
C ***************************************************************
      A=ASIN(SQRT(1.-X*X))
      IF(X.GE.-1..AND.X.LT.0.) ACOS=3.1415926536-A
      IF(X.GE.0..AND.X.LE.1.) ACOS=A
      RETURN
      END
C ***************************************************************
      FUNCTION HR(R,N,XOBS,ERR)
C ***************************************************************
      DIMENSION XOBS(N),X(222)
      R=2.
      DO 2 I=1,N
 2    X(I)=XOBS(I)
      CALL SORT(1,N,X)
      HR=0.
      NI=N-1
      DO 3 I=2,NI
 3    HR=FLOAT(2*I-N-1)*X(I)+HR
      HR=(X(N)-X(1))*FLOAT(N-1)+HR
      FN=FLOAT(N)
      R1=R+1
      R2=R*(2.*R+1.)
      HR=HR/FN/(FN-R2/R1)
      ERR=2.*R*FN*HR
      ERR=R1/ERR
      RETURN
      END
C ***************************************************************
      SUBROUTINE OUTX(N,Y)
C ***************************************************************
      COMMON /XLU/NFIRST
      DIMENSION Y(N)
      FN=FLOAT(N)
      FL=10.
      R=FN/FL
      R=R-AINT(R)
      R=R*FL+0.5
      IR=INT(R)-1
      NIL=INT(FL)
      ML=NIL
      IL=1
      IF(NFIRST.EQ.0) WRITE(6,9)
      IF(NFIRST.EQ.1) WRITE(6,6)
      DO 1 I=1,N,ML
      J=I+ML-1
      IF(NIL.LT.ML) J=I+IR
      WRITE(6,8)(Y(L),L=I,J)
      NIL=N-ML*IL
      IL=IL+1
 1    CONTINUE
      WRITE(6,7)
 9    FORMAT(7H XOBS :/)
 8    FORMAT(1X,20F7.0)
 7    FORMAT(1X//)
 6    FORMAT(' THE VALUES OF H(J) :'/)
      RETURN
      END
C ***************************************************************
      SUBROUTINE INTIT(NI)
C ***************************************************************
      DIMENSION T(18)
      READ(5,10)(T(K),K=1,18)
      IF(NI.EQ.1) WRITE(6,21)(T(K),K=1,18)
      IF(NI.EQ.2) WRITE(6,22)(T(K),K=1,18)
      IF(NI.EQ.3) WRITE(6,23)(T(K),K=1,18)
 10   FORMAT(18A4)
 21   FORMAT(1H ///4X,80(1H*)/4X,2H* ,4X,
     *18A4,2H */4X,80(1H*)///)
 22   FORMAT(1H ///8X,68(1H-)/8X,2H* ,18A4,
     *2H *,/8X,68(1H-)///)
 23   FORMAT(1H ,18A4/)
      RETURN
      END
C ***************************************************************
      FUNCTION POLO(AA,BB,EPS,STEP)
C ***************************************************************
      COMMON /XP/KX
      POLO=0.
      IF(KX.EQ.2) GO TO 11
      A=AA
      FA=FCN(A)
 12   B=A+STEP
      FB=FCN(B)
      IF(B.GT.BB) RETURN
      FAB=FA*FB
      IF(FAB.LE.0.) GO TO 10
      A=B
      FA=FB
      GO TO 12
 11   B=BB
      FB=FCN(B)
 10   A=B-STEP
```

```
      FA=FCN(A)
      IF(A.LT.AA) RETURN
      FAB=FA*FB
      IF(FAB.LE.0.) GO TO 10
      B=A
      FB=FA
      GO TO 6
    7 WRITE(6,66) FA,FB,A,B
      WRITE(6,67)
      STOP
   66 FORMAT(///1X,10(1H*),3X,
     *'TO CHANGE BOUNDERY IN POLO CHOOCING'/
     *1X,10(1H*),3X,'PARAMETER  "A" IN VARIABLE KERNEL'//1X,
     *10(1H*),3X,'STOP OF PROGRAM ( IN POLO )'/
     *3/' FCN(A) =',E18.8,3X,' FCN(B) =',E18.8,
     *4/' A= ',E18.8,3X,' B = ',E18.8)
   67 FORMAT(////)
   10 X=(A+B)/2.
      FX=FCN(X)
      FAX=FA*FX
      IF(FAX) 2,3,4
    2 B=X
      GO TO 5
    4 FA=FX
      A=X
    5 ABSA=ABS(A-B)
      IF(ABSA.GT.EPS) GO TO 10
    3 POLO=(A+B)/2.
      RETURN
      END
C ***************************************************************
      SUBROUTINE SORT(ID,N,A)
C ***************************************************************
C
C     PROCEDURE FOR RANKING FLOOD DATA
C     FROM MAX. TO MIN. - ID=0
C     FROM MIN. TO MAX. - ID=1
C
      DIMENSION A(N)
      N1=N-1
      DO 2 I=1,N1
      II=I+1
      AI=A(I)
      DO 2 J=II,N
      AJ=A(J)
      IF(AI.GT.AJ.AND.ID.EQ.0) GO TO 4
      IF(AI.LE.AJ.AND.ID.NE.0) GO TO 4
      A(J)=AI
      A(I)=AJ
      AI=AJ
    4 CONTINUE
    2 CONTINUE
      RETURN
      END
C ***************************************************************
      FUNCTION PEM(N,I)
C ***************************************************************
C
C     THIS FUNCTION COMPUTES EMPIRICAL PROBABILITY PJ
C
      COMMON /KY/NP
      FI=FLOAT(I)
      FN=FLOAT(N)
      GO TO (1,2,3,4,5,6) ,NP
    1 PEM=FI/(FN+1.)
      RETURN
    2 PEM=(FI-0.25)/(FN+0.5)
      RETURN
    3 F=FN+1.
      S=SQRT(F*FN)
      PEM=(2.*FI-F+S)/S/2.
      RETURN
    4 PEM=(FI-0.3)/(FN+0.4)
      RETURN
    5 PEM=(FI-0.4)/(FN+0.2)
      RETURN
    6 PEM=(FI-0.5)/FN
      RETURN
      END
C ***************************************************************
      SUBROUTINE SORTP(N,A)
C ***************************************************************
C
C
      DIMENSION A(N)
      COMMON /KY/NP,TITL
      N1=N-1
      DO 2 I=1,N1
      II=I+1
      AI=A(I)
      ID=INT(TITL)
      DO 2 J=II,N
      AJ=A(J)
      IF(AI.GT.AJ.AND.ID.EQ.0) GO TO 4
      IF(AI.LE.AJ.AND.ID.NE.0) GO TO 4
      A(J)=AI
      A(I)=AJ
      AI=AJ
    4 CONTINUE
    2 CONTINUE
```

```
      N3=N-3
      N2=N-2
      N1=N-1
      WRITE(6,23)
      DO 12 I=1,N,4
      I4=I+3
      I3=I+2
      I2=I+1
      IF(I.GT.N3) GO TO 33
      P1=PEM(N,I)
      P2=PEM(N,I2)
      P3=PEM(N,I3)
      P4=PEM(N,I4)
C     WRITE(6,22) P1,A(I),P2,A(I2),P3,A(I3),P4,A(I4)
      GO TO 9
   33 IF(I.GT.N2) GO TO 26
      P1=PEM(N,I)
      P2=PEM(N,I2)
      P3=PEM(N,I3)
      WRITE(6,22) P1,A(I),P2,A(I2),P3,A(I3)
      GO TO 9
   26 IF(I.GT.N1) GO TO 13
      P1=PEM(N,I)
      P2=PEM(N,I2)
      WRITE(6,22) P1,A(I),P2,A(I2)
      GO TO 9
   13 P1=PEM(N,I)
      WRITE(6,22) P1,A(I)
    9 CONTINUE
   12 CONTINUE
      WRITE(6,21)
   22 FORMAT('1H',4(F7.4,1X,F8.1,3X))
   23 FORMAT(1H,/17H EMPIRICAL DATA : ,///4(4X,1HP,7X,
     *4HCOBS,3X)//)
   21 FORMAT(///)
      RETURN
      END
C ***************************************************************
      SUBROUTINE PLOT(N,M,NY,IY,Y,YMN,YMX,IPO,ISCAL,IPRIN)
C ***************************************************************
C
      DIMENSION IY(1),Y(M,1),SCL(13)
      CHARACTER*8 IZ(6),IH1,IH2,IH3,IH4,
     *LINE(121),IXIS(60)
      DATA IH1,IH2,IH3,IH4
     1/'+-','-—','.','.' /
      DATA IZ/'*','$','X','0','S','O'/
      DO 6 I=1,12
      K=(I-1)*5+1
      IXIS(K)=IH1
      DO 6 J=1,4
      L=K+J
    6 IXIS(L)=IH2
      IF(ISCAL.EQ.0) GO TO 4
      YMN=1.E30
      YMX=-1.0E30
      DO 5 I=2,NY
      L=IY(I)
      DO 5 J=1,N
      YY=Y(L,J)
      IF(YY.LT.YMN) YMN=YY
      IF(YY.GT.YMX) YMX=YY
    5 CONTINUE
    4 I1=IY(1)
      DY=(YMX-YMN)*0.01
      DO 10 I=1,11
      AI=FLOAT(I)
   10 SCL(I)=((11.-AI)*YMN+(AI-1.)*YMX)*0.1
      IF(NY.GT.6) NY=6
      IF(IPO.EQ.0) GO TO 40
      IMX=N-1
      DO 30 I=1,IMX
      JMX=N-I
      DO 30 J=1,JMX
      IF( Y(I1,J).LE. Y(I1,J+1)) GO TO 30
      DO 20 K=1,M
      TMP=Y(K,J)
      Y(K,J)= Y(K,J+1)
      Y(K,J+1)= TMP
   20 CONTINUE
   30 CONTINUE
   40 I5=5
      IF(IPRIN.EQ.0) GO TO 1
      IF(IPRIN.EQ.3) GO TO 7
      L=IY(1)
      WRITE(6,5000)  (Y(L,K),K=1,N)
      DO 2 I=2,NY
      L=IY(I)
    2 WRITE(6,5002)(Y(L,K),K=1,N)
      GO TO 1
    7 L=IY(1)
      WRITE(6,5006)(Y(L,K),K=1,N)
      L=IY(NY)
      WRITE(6,5007)(Y(L,K),K=1,N)
    1 WRITE(6,5003)
      DO 3 I=2,NY
      II=I-1
      L=IY(I)
      IF(IPRIN.NE.3) GO TO 8
      WRITE(6,5008) IZ(II)
```

```
      GO TO 3
 8    WRITE(6,5004)  IZ(II)
 3    CONTINUE
      WRITE(6,5005)
      WRITE(6,1001)  (SCL(I),I=1,11),(IXIS(I),I=1,50)
      DO 200 J=1,M
      DO 50 I=1,101
 50   LINE(I)=IH4
      DO 100 IYY=2,MY
      II=IYY-1
      I=IY(IYY)
      YY=Y(I,J)
      IF(YY.GT.YMX.OR.YY.LT.YMN) GO TO 100
      BI=(YY-YMN)/DY
      IB=INT(BI)
      IF((BI-FLOAT(IB)).GE.0.5) IB=IB+1
      IB=IB+1
      IF(LINE(IB).NE.IH4) GO TO 60
      LINE(IB)=IZ(II)
      GO TO 100
 60   LINE(IB)=IZ(8)
 100  CONTINUE
      DO 80 L=1,101,10
      IF(LINE(L).EQ.IH4) LINE(L)=IH5
 80   CONTINUE
      IF(J.NE.15) GO TO 90
      IF(J.EQ.M) GO TO 90
      DO 70 L=1,101,2
      IF(LINE(L).EQ.IH4) LINE(L)=IH3
 70   CONTINUE
      I5=I5+5
 90   WRITE(6,2000)
     *Y(II,J),(LINE(L),L=1,101)
 200  CONTINUE
      WRITE(6,2001) (IXIS(L),L=1,50),(SCL(L),L=1,11)
      RETURN
1001 FORMAT(5X,11E10.2,/12X,50A2,1H+)
2000 FORMAT(1X,F10.0,1X,101A1)
2001 FORMAT(12X,50A2,1H+/5X,11E10.2//////)
5000 FORMAT(/////
     *1X,'RANDOM VARIABLE :'//
     *8(4X,G12.6))
5002 FORMAT(//1X,'THE TABULATED NONPARAMETRIC DENSITY FUNCTION'
     *//8(4X,G12.6))
5003 FORMAT(///1X,'CHAIRS OF PLOT'//)
5004 FORMAT(1X,A1,' -  NONPARAMETRIC DENSITY FUNCTION')
5005 FORMAT(1X,'O -  COMMON CHAIR FOR A FEW PLOTS'//)
5006 FORMAT(///' K-TH NEARST NEIGHBOR VALUES :'//8(4X,G12.6))
5007 FORMAT(/' THE MEAN OF THE K-TH NEARST'
     *,' NEIGHBOR DISTANCE VALUES :'//8(4X,G12.6))
5008 FORMAT(1X,A1,' -  THE MEAN OF THE K-TH NEAREST'
     *,' NEIGHBOR DISTANCE')
      END
C     ******************************************************************
      FUNCTION SLL(AT,T,D,N)
C     ******************************************************************
      DIMENSION T(13),D(10)
      X=AT
      R= 8.*FLOAT(N)/25.43101134
      SLL=0.
      DO 2 I=2,11
      A=T(I)
      B=T(I-1)
      IF(X.GT.A) GO TO 2
      SLL=D(I-1)*R*(ALOG(X)-ALOG(B))+SLL
      GO TO 3
2     SLL=D(I-1)*R*(ALOG(A)-ALOG(B))+SLL
3     CONTINUE
      RETURN
      END
C     ******************************************************************
      FUNCTION RTT(AS,T,D,N)
C     ******************************************************************
      DIMENSION T(13),D(10)
      R= 8.*FLOAT(N)/25.43101134
      SLP=0.
      RTT=-1.
      DO 2 I=2,11
      A=T(I)
      B=T(I-1)
      C=D(I-1)
      OLD=SLP
      SLP=C*R*(ALOG(A)-ALOG(B))+SLP
      IF(AS.GT.SLP) GO TO 2
      RTT=B*EXP((AS-OLD)/C/R)
      RETURN
2     CONTINUE
      IF(RTT.LE.0.) RTT=A
      RETURN
      END
C     ******************************************************************
      SUBROUTINE SPB(M,XOBS,EPS,T,KEY)
C     ******************************************************************
      DIMENSION IY(3),Y(3),XOBS(N),T(13),D(10),ISCL(16)
     *,XPM(333),IRET(333)
      CHARACTER*3 IZ(3),IH1,IH2,IH3,IH4
     *,LINE(121),IXIS(75)
      COMMON /XP/XX,PROB/PAL/IPAL,LPAL(20),MPAL,MYY,CO
      DATA IH1,IH2,IH3,IH4
     */'+','-','.','.' */
```

```
      DATA (IZ(I),I=1,3)/'*','X','O'/
      DATA D/8.,7.,5.,3.,1.2,1.2,.8,.5,.4,.3/
      DO 4 I=1,3
4     IY(I)=I
      NY=5
      NSTEP=10
      NSCAL=12
C     NSTEP - NUMBER OF POINTS IN INTERVAL
C     NSCAL - NUMBER OF INTERVALS
C     NSTEP*NSCAL - NUMBER OF POINTS IN LINE
      FNL=FLOAT(NSCAL)
      N1=NSCAL+1
      N2=NSCAL*5
      N3=NSCAL*NSTEP+1
      DO 6 I=1,NSCAL
      K=(I-1)*5+1
      IXIS(K)=IN1
      DO 6 J=1,4
      L=K+J
6     IXIS(L)=IN2
      YMX=1.E20
      YMN=-1.0E20
      DO 5 J=1,N
      YY=XOBS(J)
      IF(YY.LT.YMN) YMN=YY
      IF(YY.GT.YMX) YMX=YY
5     CONTINUE
      YMX=YMX*1.1
      II=IY(1)
      DY=(YMX-YMN)/FLOAT(NSTEP)/FNL
      DO 10 I=1,N1
      AI=FLOAT(I)
      SCLI=((FLOAT(N1)-AI)*YMN+(AI-1.)*YMX)/FNL
10    ISCL(I)=SCLI
      WRITE(6,5003)
      WRITE(6,5004) IZ(1)
      WRITE(6,5006) IZ(2)
      WRITE(6,5005) IZ(NY)
      WRITE(6,1001) (ISCL(I),I=1,N1),(IXIS(I),I=1,N2)
      KK=1
      B=YMX
      A=0.
      STEP=0/500.
      BG=YMX
      NZ=N
C     IF(IPAL.GT.0) NZ=NYY
      T2=T(2)
      SLT=SLL(T2,T,D,NZ)
      LTT=INT(SLT)
      RET=1./PEM(NZ,NZ)
      SLT=SLL(RET,T,D,NZ)
      J=INT(SLT)-1
      K=1
      JM=0
C     IBB=MPAL+1
16    J=J+1
      SL=FLOAT(J)
      RTTH=RTT(SL,T,D,NZ)
      PROB=1./RTTH
      XP=0.
      IF(J.GE.LTT) XP=POLO(A,B,EPS,STEP)
      B=BG
      A=XP
      XOBSK=10.E22
      IF(K.GT.N) GO TO 17
      LPALK=K
      IF(IPAL.LE.0) GO TO 18
C     KK=N-K+1
C     IF(KK.LE.MPAL) LPALK=LPAL(KK)
C     IF(KK.GT.MPAL) LPALK=LPAL(IBB)+KK-IBB
C     NKA=LPALK
C     GO TO 19
18    NKA=NZ-LPALK+1
19    PK=PEM(NZ,NKA)
      SLT=SLL(1./PK,T,D,NZ)
      IF(SLT.LT.(SL-0.5).OR.SLT.GE.(SL+0.5))
     *GO TO 17
      NK1=N-K+1
      XOBSK=XOBS(NK1)
      IF(CO.GT.0.) XOBSK=10.E22
      K=K+1
17    CONTINUE
      Y(1)=SL
      Y(2)=XP
      Y(3)=XOBSK
      IF(KEY.LT.2) GO TO 49
      JM=JM+1
      XPM(JM)=XP
      TRET(JM)=RTTH
49    DO 50 I=1,N3
50    LINE(I)=IH4
      DO 100 IYY=2,NY
      II=IYY-1
      I=IY(IYY)
      YY=Y(I)
      IF(YY.GT.YMX.OR.YY.LT.YMN) GO TO 100
      BI=(YY-YMN)/DY
      IB=INT(BI)
      IF((BI-FLOAT(IB)).GE.0.5) IB=IB+1
      IB=IB+1
```

```
      IF(LINE(18).NE.IH4) GO TO 60
      LINE(18)=I2(11)
      GO TO 100
   60 LINE(18)=I2(3)
  100 CONTINUE
      DO 80 L=1,N3,NSTEP
      IF(LINE(L).EQ.IH4) LINE(L)=IH3
   80 CONTINUE
      DO 85 I=1,11
      TI=T(I)
      SLT=SLL(TI,T,D,NZ)
      IS=INT(SLT)
      IF(J.EQ.15) GO TO 66
   65 CONTINUE
      GO TO 90
   66 DO 70 L=1,N3,2
      IF(LINE(L).EQ.IH4) LINE(L)=IH3
   70 CONTINUE
      IF(I.EQ.1) WRITE(6,1980) TI,(LINE(L),L=1,N3)
      IF(I.GE.2.AND.I.LE.3)
     *WRITE(6,1990) TI,(LINE(L),L=1,N3)
      IF(I.GT.3) WRITE(6,2000) TI,(LINE(L),L=1,N3)
      TI=T(11)
      SLT=SLL(TI,T,D,NZ)
      IS=INT(SLT)
      IF(J.EQ.15) GO TO 210
      GO TO 16
   90 WRITE(6,2002) (LINE(L),L=1,N3)
      IF(XP.LT.YMX) GO TO 16
  210 WRITE(6,2001) (IXIS(L),L=1,N2),(ISCL(L),L=1,N1)
      IF(KEY.LT.2) RETURN
      WRITE(6,5007)
      WRITE(6,5008)(XMET(I),I=1,JM)
      WRITE(6,5009)
      WRITE(6,5008)(XPM(I),I=1,JM)
      RETURN
 1001 FORMAT(I8,12I10/6X,60A2,1H+)
 1980 FORMAT(1X,F5.3,121A1)
 1990 FORMAT(1X,F4.2,1X,121A1)
 2000 FORMAT(1X,F4.0,1X,121A1)
 2001 FORMAT(6X,60A2,1H+/I8,12I10/////)
 2002 FORMAT(6X,121A1)
 5003 FORMAT(///1X,'CHAIRS OF PLOT'/)
 5004 FORMAT(1X,A1,' - COMPUTED VALUES')
 5005 FORMAT(1X,A1,' - COMMON CHAIR FOR A TWO PLOTS'//)
 5008 FORMAT(1X,A1,' - OBSERVED VALUES')
 5007 FORMAT(///' RETURN PERIODS VALUES :'/)
 5008 FORMAT(8(4X,G12.6))
 5009 FORMAT(//' FLOODS VALUES :'/)
      END
```

```
C        ADAPTIVE KERNEL ESTIMATOR PROGRAM
C        PROGRAMMER: DR. W. FELLUCH - 1988
C        MODIFIED BY Y. ALILA - 1987
C THIS PROGRAM EMPLOYS AN ANNUAL FLOOD SERIES AS AN INPUT TO CALCULATE
C FLOOD ESTIMATES FOR A PRE-SELECTED RETURN PERIODS USING THE ADAPTIVE
C KERNEL ESTIMATOR SUGGESTED BY SILVERMAN (1986).
C IN THE NUMERICAL PROCEDURE THE EPANECHNIKOV OPTIMAL KERNEL AND THE
C FLANAGAN OPTIMAL SMOOTHING FACTOR ARE APPLIED. DENSITY FUNCTION CAN
C BE PLOTTED UNDER REQUEST OF THE USER
C
C INPUT FILE
C
C THE INPUT FILE COMPRISES :
C (1) NAME OF THE STATION IN CHARACTER VARIABLE ( FORMAT 16A4 )
C (2) N RECORD LENGTH OF FLOOD DATA (INTEGER)
C (3) ANNUAL FLOOD SERIES (REAL VARIABLE FREE FORMAT)
C (4) KL, KEY AND SCAL CONTROL PARAMETERS
C
C IMPORTANT VARIABLE
C
C XOBS : INPUT VECTOR OF FLOOD DATA
C N : NUMBER OF FLOOD OBSERVATION IN AN ANNUAL FLOOD SERIES.
C HH : SMOOTHING FACTOR
C T : RETURN PERIOD
C XP : FLOOD MAGNITUDE AS ESTIMATED BY THE TRANSFORMED KERNEL
C ALP: SENSITIVITY PARAMETER OF THE ADAPTIVE KERNEL.
C PAR: VECTOR COMPRISING THE LOCAL BANDWIDTH PARAMETER VALUES.
C KL : INTEGER CONTROLLING THE ESTIMATION METHOD
C        KL= 0 FOR TRANSFORMED KERNEL ESTIMATOR
C        KL= 1 FOR THE FIXED KERNEL ESTIMATOR
C KEY: INTEGER CONTROLLING OUTPUT OPTIONS
C        KEY=0 PRINTING ONLY FLOOD ESTIMATES FOR SPECIFIED RETURN PERIOD
C        KEY=1 PRINTING EMPIRICAL PROBABILITIES
C        KEY=2 PRINTING THEORETICAL PROBABILITIES
C SCAL: VARIABLE CONTROLLING THE STARTING POINT ON THE Y Axis OF THE
C       DENSITY FUNCTION (SCAL=0.0 NO NEED TO SPECIFY THE STARTING POINT
C A,B : LIMITS FOR METHOD OF BISECTION
C EPS : ACCURACY LEVEL FOR THE METHOD OF BISECTION
C
C DESCRIPTION OF MAIN SUBROUTINES
C
C  SORT . RANK THE DATA FROM MAXIMUM TO MINIMUM.
C  HH   : CALCULATE THE SMOOTHING FACTOR FOR THE FIXED KERNEL.
C  TR   : TRANSFORM THE FLOOD DATA.
C  FCN  : FOR A GIVEN RETURN PERIOD THIS FUNCTION IS EQUATED TO ZERO
C         AND SOLVED FOR XP USING POLO.
C  POLO . COMPUTES ROOTS OF A FUNCTION BY METHOD OF BISECTION.
C  H    : CALCULATE THE INTEGRAL OF THE KERNEL FUNCTION K(.).
C  DENS : CALCULATE DENSITY FUNCTION AT EACH FLOOD OBSERVATION
C         BY THE FIXED KERNEL METHOD.
C  DENSI: CALCULATE DENSITY FUNCTION AT EACH FLOOD OBSERVATION
C         BY THE ADAPTIVE KERNEL METHOD.
C  PLOT : GENERATE THE PLOT OF THE DENSITY FUNCTION.
C  OUTX : PRINTS FLOOD DATA IN COLUMNS
C  PEM  : COMPUTES EMPIRICAL PROBABILITIES.
C
C
C*****************************************************************
      DIMENSION T(13),VALUE(2,333),ROB(3,3),XMEM(3)
     *,IY(3)
      CHARACTER*6 TT(8,6),IZ(6),SP,ST
      COMMON /OBS/N,XOBS(333),YT(333),SUM,FN,H,HH,PAR(333)
     */MIN/XMAX,FIP,KL
     */KY/NP,TITL/DE/D
      DATA T/1.005,1.05,1.25,2.,5.,10.,20.,
     *50.,100.,200.,500.,1000.,2000./
      DATA (TT(1,I),I=1,8)/'THE TR',
     *'ANSFOR','MED KE','RNEL E','STIMAT','E     '/
      DATA (TT(1,2),I=1,8)/'TABULA','TED NO','NPARAM'
     *,'ETRIC ','DENSIT','Y FUNC','TION  ','      '/
      DATA SP,ST/'      ','THE KE'/
      PI=3.1415926536
      EPS=0.5
      ALP=.0
      NP=2
      TITL=0
      READ(5,*) KL,KEY,SCAL
      CALL INTIT(3)
      WRITE(6,444) ALP
444   FORMAT('THE PARAMETER OF THE ADAPTIVE KERNEL ALPHA IS  ',F5.3)
      IF(KL.NE.1) WRITE(6,26)
      TT(1,1)=SP
      TT(2,1)=SP
      TT(3,1)=ST
C 14  WRITE(6,26)(TT(I,1),I=1,6)
      IF(KL.NE.1) WRITE(6,21)
      READ(5,*) N
      READ(5,*)(XOBS(I),I=1,N)
      IF(KEY.LE.0.OR.KEY.EQ.2) CALL OUTX(N,XOBS)
      IF(KEY.EQ.1) CALL SORIP(N,XOBS)
      IF(KEY.EQ.2) CALL SORT(N,XOBS)
      XMAX=-10.E33
      XMIN=-XMAX
      DO 9 I=1,N
      XI=XOBS(I)
      IF(XI.GT.XMAX) XMAX=XI
      IF(XI.LT.XMIN) XMIN=XI
   9  CONTINUE
      FN=FLOAT(N)
      HH=HH(N,XOBS)
```

```
      WRITE(6,28) HH
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
      SUML=0.
      DO 99 I=1,N
      XI=XOBS(I)
      SUML=SUML+ALOG10(DENS(XI))
 99   CONTINUE
      SUML=SUML/N
      GEE=10**(SUML)
      DO 999 I=1,N
      XI=XOBS(I)
      PAR(I)=(DENS(XI)/GEE)**(-(ALP))
 999  CONTINUE
C     WRITE(6,*)(XOBS(I),I=1,N)
C     WRITE(6,*)(PAR(I),I=1,N)
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
      IF(KL.EQ.1) GO TO 6
      DO 7 I=1,N
      XI=XOBS(I)
  7   YT(I)=TR(XI)
      H=5./182./PI/FN
      H=H**0.2
      WRITE(6,30) H
      SUM=0.
      DO 15 I=1,N
      Y=YT(I)
      Y1=Y-1
 15   SUM=SUM+W(Y/H)-W(Y1/H)
      SUM=SUM/FN
      WRITE(6,27)  SUM
  6   IF(KEY.LE.1) GO TO 10
      WRITE(6,24)
      N1=N-1
      N2=N-2
      DO 4 K=1,N,3
      DO 1 J=1,3
      I=K+J-1
      IF(I.GT.N) GO TO 1
      XI=XOBS(I)
      ROB(1,J)=DENSI(XI)
      IF(KL.EQ.1) D=1.-D
      IF(KL.NE.1) D=ANAL(XI)
      ROB(2,J)=D
      ROB(3,J)=PEM(N,I)
      XMEM(J)=XI
  1   CONTINUE
      IF(K.GT.N2) GO TO 11
        WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
     *,XMEM(2),(ROB(L1,2),L1=1,3),XMEM(3),(ROB(L1,3),L1=1,3)
      GO TO 4
 11   IF(K.GT.N1) GO TO 12
      WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
     *,XMEM(2),(ROB(L1,2),L1=1,3)
      GO TO 4
 12   WRITE(6,25) XMEM(1),(ROB(L1,1),L1=1,3)
  4   CONTINUE
 10   IF(KEY.LE.0) GO TO 13
C     PLOTTING OF DENSITY
      IY(2)=2
      IY(1)=1
      BOUND=XMAX
      DIF=AINT(XMAX/100.)
      IF(DIF.LE.1.5) DIF=1.
      IF(DIF.GT.1.5.AND.DIF.LE.2.5) DIF=2.
      IF(DIF.GT.2.5.AND.DIF.LE.7.5) DIF=5.
      IF(DIF.GT.7.5.AND.DIF.LE.12.5) DIF=10.
      IF(DIF.GT.12.5.AND.DIF.LE.27.5) DIF=20.
      IF(DIF.GT.27.5.AND.DIF.LE.75.0) DIF=50.
      IF(DIF.GT.75..AND.DIF.LE.125.) DIF=100.
      IF(DIF.GT.125..AND.DIF.LE.300.) DIF=200.
      IF(DIF.GT.300..AND.DIF.LE.700.) DIF=500.
      IF(DIF.GT.700..AND.DIF.LE.1300.) DIF=1000.
      IF(DIF.GT.1300.) DIF=2000.
      ISCAL=0
      IF(SCAL.LE.0.) ISCAL=1
      XI=0.
      XMID=XMIN*0.75
      I=0
  3   XI=XI+DIF
      IF(XI.LT.XMID) GO TO 3
      IF(XI.GT.BOUND) GO TO 2
      I=I+1
      VALUE(1,I)=XI
      VALUE(2,I)=DENSI(XI)
      GO TO 3
  2   CALL PLOT(I,2,2,IY,VALUE,0.,SCAL,0,ISCAL,1)
C     COMPUTING OF FLOODS
 13   B=XMAX
      A=0.0
      BG=XMAX
      WRITE (6,22)
      DO 5 I=1,11
      FIP=1./T(I)
      XP=POLO(A,B,EPS)
      BG=BG*1.1
      B=BG
      A=XP
  5   WRITE(6,23) T(I),XP
      WRITE(6,28)
 22   FORMAT(/5X,15HRETURN    FLOOD/5X,
```

```
     *17HPERIOD   ESTIMATE/)
  23 FORMAT(2X,F8.3,3X,F7.0)
  24 FORMAT(/,'DENSITY FUNCTION FOR EMPIRICAL DATA :',
     *//,3(4X,'XOBS',1X,'DENSITY',2X,2('PROBABILITY',1X)),/,
     *3(17X,'THEORETICAL  EMPIRICAL  '),/)
  25 FORMAT(3(F7.0,3(1X,G11.5),1X))
  26 FORMAT(1X,' THE TRANSFORMED KERNEL ESTIMATE ',/)
  27 FORMAT('  SUM = ',G11.5/)
  28 FORMAT(////)
  29 FORMAT(' H FOR OBSERVATED DATA     = ',G11.5)
  30 FORMAT( ' H FOR TRANSFORMATED DATA = ',G11.5/)
  31 FORMAT(' TRIANGULARY TRANSFORMATION'//)
      STOP
      END
      FUNCTION DENS(X)
C     DENSITY OF TRANSFORMED DATA - IF KL=0
C     DENSITY OF OBSERVATED DATA  - IF KL=1
      COMMON /OBS/N,XOBS(333),YT(333),SUM,FN,H,HH,PAR(333)
     */DE/D/MIN/XMAX,FIP,KL
      DENS=0.
      DY=0.
      T=TR(X)
      DO 2 I=1,N
      IF(KL.EQ.1) GO TO 6
      Y=(T-YT(I))/H
      AY=ABS(Y)
      Z=0.
      IF(AY.LT.1.) Z=0.75*(1.-Y*Y)
      DENS=DENS+Z
   6  Y=(X-XOBS(I))/HH
      AY=ABS(Y)
      Z=0.
      IF(AY.LT.1.) Z=0.75*(1.-Y*Y)
   2  DY=DY+Z
      DY=DY/FN/HH
C     DY - DENSITY OF OBSERVATED DATA
      IF(KL.NE.1) GO TO 5
      DENS=DY
      RETURN
   5  IF(D-0.5) 1,1,3
   1  DIFTR=DY/SQRT(8.*D)
      GO TO 4
   3  DIFTR=DY/SQRT(8.*(1.-D))
C     DIFTR - DIFFERENTIABLE OF TRANSFORMATION
   4  DENS=DENS*DIFTR/FN/H/SUM
      RETURN
      END
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
      FUNCTION DENS1(X)
C     DENSITY OF TRANSFORMED DATA - IF KL=0
C     DENSITY OF OBSERVATED DATA  - IF KL=1
      COMMON /OBS/N,XOBS(333),YT(333),SUM,FN,H,HH,PAR(333)
     */DE/D/MIN/XMAX,FIP,KL
      DENS1=0.
      DY=0.
      T=TR1(X)
      DO 2 I=1,N
      IF(KL.EQ.1) GO TO 6
      Y=(T-YT(I))/H
      AY=ABS(Y)
      Z=0.
      IF(AY.LT.1.) Z=0.75*(1.-Y*Y)
      DENS1=DENS1+Z
   6  Y=(X-XOBS(I))/(HH*PAR(I))
      AY=ABS(Y)
      Z=0.
      IF(AY.LT.1.) Z=0.75*(1.-Y*Y)/PAR(I)
   2  DY=DY+Z
      DY=DY/FN/HH
C     DY - DENSITY OF OBSERVATED DATA
      IF(KL.NE.1) GO TO 5
      DENS1=DY
      RETURN
   5  IF(D-0.5) 1,1,3
   1  DIFTR=DY/SQRT(8.*D)
      GO TO 4
   3  DIFTR=DY/SQRT(8.*(1.-D))
C     DIFTR - DIFFERENTIABLE OF TRANSFORMATION
   4  DENS1=DENS1*DIFTR/FN/H/SUM
      RETURN
      END
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
      FUNCTION TR(X)
C     TRANSFORMATION
      COMMON /DE/D
     */OBS/N,XOBS(333),YT(333),SUM,FN,H,HH,PAR(333)
      D=0.
      DO 3 I=1,N
      Y=(X-XOBS(I))/HH
   3  D=D+W(Y)
      DN=D/FN
       P=0.5-DN
      D=0.5+DN
      IF(DN) 1,1,2
   1  TR=SQRT(D/2.)
      RETURN
   2  TR=1.-SQRT(P/2.)
      RETURN
      END
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
```

```
      FUNCTION TRI(X)
C     TRANSFORMATION
      COMMON /DE/D
     */OBS/N,XOBS(333),YT(333),SUM,FN,H,HH,PAR(333)
      D=0.
      DO 3 I=1,N
      Y=(X-XOBS(I))/(HH*PAR(I))
    3 D=D+W(Y)
      DN=D/FN
       P=0.5-DN
      D=0.5+DN
      IF(DN) 1,1,2
    1 TRI=SQRT(D/2.)
      RETURN
    2 TRI=1.-SQRT(P/2.)
      RETURN
      END
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
      FUNCTION FCN(W)
      COMMON /MIN/XMAX,FIP,KL/DE/D
      IF(KL.GT.0) GO TO 1
      FCN=FIP-ANAL(W)
      RETURN
    1 TW=TRI(W)
      FCN=FIP+D-1.
      RETURN
      END
      FUNCTION ANAL(X)
      COMMON  /OBS/N,XOBS(333),YT(333),SUM,FN,H,HH,PAR(333)
      S=0.
      TX=TR(X)
      DO 1 I=1,N
      YI=YT(I)
      YI=(1.-YI)/H
      Y=(TX-YI)/H
    1 S=S+W(YI)-W(Y)
      ANAL=S/FN/SUM
      RETURN
      END
      FUNCTION W(X)
      W=-0.5
      IF(X.GT.-1..AND.X.LT.1.) W=0.75*X*(1.-X*X/3.)
      IF(X.GE.1.) W=0.5
      RETURN
      END
      SUBROUTINE OUTX(N,Y)
      DIMENSION Y(N)
      FN=FLOAT(N)
      FL=10.
      R=FN/FL
      R=R-AINT(R)
      R=R*FL+0.5
      IR=INT(R)-1
      NIL=INT(FL)
      NL=NIL
      IL=1
      WRITE(6,9)
      DO 1 I=1,N,NL
      J=I+NL-1
      IF(NIL.LT.NL) J=I+IR
      WRITE(6,8)(Y(L),L=I,J)
      NIL=N-NL*IL
      IL=IL+1
    1 CONTINUE
      WRITE(6,7)
    9 FORMAT(7H XOBS :/)
    8 FORMAT(1X,20F7.0)
    7 FORMAT(1X//)
      RETURN
      END
      SUBROUTINE INTIT(NI)
      DIMENSION T(18)
      READ(5,10)(T(K),K=1,18)
      IF(NI.EQ.1) WRITE(6,21)(T(K),K=1,18)
      IF(NI.EQ.2) WRITE(6,22)(T(K),K=1,18)
      IF(NI.EQ.3) WRITE(6,23)(T(K),K=1,18)
   10 FORMAT(18A4)
   21 FORMAT(1H,///4X,80(1H*)/4X,2H* ,
     118A4,2H */4X,80(1H*)///)
   22 FORMAT(1H ///6X,68(1H-)/6X,2H* ,18A4,
     12H *,/6X,68(1H-)//)
   23 FORMAT(1H ,18A4/)
      RETURN
      END
      FUNCTION POLO(A,B,EPS)
      FA=FCN(A)
      FB=FCN(B)
      FAB=FA*FB
      IF(FAB.LE.0.) GO TO 10
C     WRITE(6,66) FA,FB,A,B
      WRITE(6,67)
      STOP
   66 FORMAT(///1X,10(1H*),3X,'TO CHANGE
     1BOUNDARY IN POLO',//1X,10(1H*),3X,
     2'STOP OF PROGRAM'/
     3/' FCN(A) =',E16.8,3X,' FCN(B) =',E16.8,
     4/' A= ',E16.8,3X,' B = ',E16.8)
   67 FORMAT(////)
   10 X=(A+B)/2
      FM=FCN(X)
```

```
      FAMA=FA*FN
      IF(FAMA) 2,3,4
2     B=X
      GO TO 5
4     FA=FN
      A=X
5     IF(ABS(A-B).GT.EPS) GO TO 10
3     POLO=(A+B)/2
      RETURN
      END
      SUBROUTINE SORT(N,A)
      DIMENSION A(N)
      N1=N-1
      DO 2 I=1,N1
      I1=I+1
      AI=A(I)
      DO 2 J=I1,N
      AJ=A(J)
      IF(AI.GT.AJ) GO TO 4
      A(J)=AI
      A(I)=AJ
      AI=AJ
4     CONTINUE
2     CONTINUE
      RETURN
      END
      FUNCTION HN(N,XOBS)
      DIMENSION XOBS(N),X(333)
      ALFA=0.5
      DO 1 I=1,N
1     X(I)=XOBS(I)
      DO 2 I=1,N
      IHIGH=I
      ILOW=1
3     IF(IHIGH.EQ.ILOW) GO TO 4
      L=(ILOW+IHIGH)/2
      IF(X(I)-X(L)) 5,7,6
5     IHIGH=L
      GO TO 3
6     ILOW=L+1
      GO TO 3
7     K=L+1
      GO TO 8
4     K=ILOW
8     SAVE=X(I)
      ISORT=I-K
      DO 9 I2=1,ISORT
      L=I+1-I2
9     X(L)=X(L-1)
2     X(K)=SAVE
      S=0
      FN=FLOAT(N)
      IOK=INT(FN**ALFA)
      DO 10 I=1,N
      XI=X(I)
      M=I-KN
      B=999999.
      IF(M.GT.0) B=XI-X(M)
      M=I+KN
      A=999999
      IF(M.LE.N) A=X(M)-XI
      D=B
      IF(A.LT.B) D=A
10    S=S+D
      HN=S/FN
      RETURN
      END
      FUNCTION PEM(N,I)
C
C     THIS FUNCTION COMPUTES EMPIRICAL PROBABILITY PJ
C
      COMMON /KY/NP
      FI=FLOAT(I)
      FN=FLOAT(N)
      GO TO (1,2,3,4,5,6) ,NP
1     PEM=FI/(FN+1.)
      RETURN
2     PEM=(FI-0.25)/(FN+0.5)
      RETURN
3     F=FN+1
      S=SQRT(F*FN)
      PEM=(2.*FI-F+S)/S/2.
      RETURN
4     PEM=(FI-0.3)/(FN+0.4)
      RETURN
5     PEM=(FI-0.4)/(FN+0.2)
      RETURN
6     PEM=(FI-0.5)/FN
      RETURN
      END
      SUBROUTINE SORTP(N,A)
C
C
      DIMENSION A(N)
      COMMON /KY/NP,TITL
      N1=N-1
      DO 2 I=1,N1
      I1=I+1
      AI=A(I)
      ID=INT(TITL)
```

```
      DO 2 J=II,N
      AJ=A(J)
      IF(AI.GT.AJ.AND.ID.EQ.0) GO TO 4
      IF(AI.LE.AJ.AND.ID.NE.0) GO TO 4
      A(J)=AI
      A(I)=AJ
      AI=AJ
    4 CONTINUE
    2 CONTINUE
      N3=N-3
      N2=N-2
      N1=N-1
      WRITE(6,23)
      DO 12 I=1,N,4
      I4=I+3
      I3=I+2
      I2=I+1
      IF(I.GT.N3) GO TO 33
      P1=PEM(N,I)
      P2=PEM(N,I2)
      P3=PEM(N,I3)
      P4=PEM(N,I4)
      WRITE(6,22) P1,A(I),P2,A(I2),P3,A(I3),P4,A(I4)
      GO TO 9
   33 IF(I.GT.N2) GO TO 24
      P1=PEM(N,I)
      P2=PEM(N,I2)
      P3=PEM(N,I3)
      WRITE(6,22) P1,A(I),P2,A(I2),P3,A(I3)
      GO TO 9
   24 IF(I.GT.N1) GO TO 13
      P1=PEM(N,I)
      P2=PEM(N,I2)
      WRITE(6,22) P1,A(I),P2,A(I2)
      GO TO 9
   13 P1=PEM(N,I)
      WRITE(6,22) P1,A(I)
    9 CONTINUE
   12 CONTINUE
      WRITE(6,21)
   22 FORMAT(4(F7.4,1X,F8.1,3X))
   23 FORMAT(1H,/17H EMPIRICAL DATA : ,///4(4X,1HP,7X,
     *6HRDBS,3X)/)
   21 FORMAT(///)
      RETURN
      END
      SUBROUTINE PLOT(N,M,NY,IY,Y,YMN,YMX,IPO,ISCAL,IPRIN)
C
      DIMENSION IY(1),Y(M,1),SCL(13)
      CHARACTER*8 IZ(8),IH1,IH2,IH3,IH4,
     *LINE(121),IXIS(60)
      DATA IH1,IH2,IH3,IH4
     1/'+-','-','-','.' /
      DATA IZ/'*','@','%','=','S','0'/
      DO 6 I=1,12
      K=(I-1)*5+1
      IXIS(K)=IH1
      DO 6 J=1,4
      L=K+J
    6 IXIS(L)=IH2
      IF(ISCAL.EQ.0) GO TO 4
      YMN=1.E30
      YMX=-1.0E30
      DO 5 I=2,NY
      L=IY(I)
      DO 5 J=1,N
      YY=Y(L,J)
      IF(YY.LT.YMN) YMN=YY
      IF(YY.GT.YMX) YMX=YY
    5 CONTINUE
    4 II=IY(1)
      DY=(YMX-YMN)*0.01
      DO 10 I=1,11
      AI=FLOAT(I)
   10 SCL(I)=((11.-AI)*YMN+(AI-1.)*YMX)*0.1
      IF(NY.GT.6) NY=6
      IF(IPO.EQ.0) GO TO 40
      IMX=M-1
      DO 30 I=1,IMX
      JMX=M-I
      DO 30 J=1,JMX
      IF( Y(II,J).LE. Y(II,J+1)) GO TO 30
      DO 20 K=1,N
      TMP=Y(K,J)
      Y(K,J)= Y(K,J+1)
      Y(K,J+1)= TMP
   20 CONTINUE
   30 CONTINUE
   40 IS=5
      IF(IPRIN.EQ.0) GO TO 1
      L=IY(1)
      WRITE(6,5000)  (Y(L,K),K=1,N)
      DO 2 I=2,NY
      L=IY(I)
      WRITE(6,5002)(Y(L,K),K=1,N)
    2 CONTINUE
    1 WRITE(6,5003)
      DO 3 I=2,NY
      II=I-1
      L=IY(I)
```

```
3   WRITE(6,5004)  IZ(II)
      WRITE(6,5005)
      WRITE(6,1001)  (SCL(I),I=1,11),(IXIS(I),I=1,50)
      DO 200 J=1,N
      DO 50 I=1,101
50  LINE(I)=IH4
      DO 100 IYY=2,NY
      II=IYY-1
      I=IY(IYY)
      YY=Y(I,J)
      IF(YY.GT.YMX.OR.YY.LT.YMN) GO TO 100
      BI=(YY-YMN)/DY
      IB=INT(BI)
      IF((BI-FLOAT(IB)).GE.0.5) IB=IB+1
      IB=IB+1
      IF(LINE(IB).NE.IH4) GO TO 60
      LINE(IB)=IZ(II)
      GO TO 100
60   LINE(IB)=IZ(6)
100  CONTINUE
       DO 80 L=1,101,10
       IF(LINE(L).EQ.IH4) LINE(L)=IH3
 80  CONTINUE
       IF(J.NE.15) GO TO 90
       IF(J.EQ.N) GO TO 90
       DO 70 L=1,101,2
       IF(LINE(L).EQ.IH4) LINE(L)=IH3
 70  CONTINUE
       IS=15+5
 90  WRITE(6,2000)
     *Y(II,J),(LINE(L),L=1,101)
200  CONTINUE
       WRITE(6,2001) (IXIS(L),L=1,50),(SCL(L),L=1,11)
       RETURN
1001 FORMAT(5X,11E10.2,/12X,50A2,1H+)
2000 FORMAT(1X,F10.0,1X,101A1)
2001 FORMAT(12X,50A2,1H+/5X,11E10.2//////)
5000 FORMAT(/////
     *1X,'RANDOM VARIABLE :'//
     *8(4X,G12.6))
5002 FORMAT(/1X,'THE TABULATED NONPARAMETRIC DENSITY FUNCTION'
     *//8(4X,G12.6))
5003 FORMAT(//1X,'CHAIRS OF PLOT'//)
5004 FORMAT(1X,A1,' -  NONPARAMETRIC DENSITY FUNCTION')
5005 FORMAT(1X,'0 -  COMMON CHAIR FOR A FEW PLOTS'//)
       END
```

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA