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.

Canada

Regional Flood Frequency Analysis by Nonparametric Methods

By

Denis Gingras

A thesis presented to the University of Ottawa in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Civil Engineering

DEPARTMENT OF CIVIL ENGINEERING
UNIVERSITY OF OTTAWA
Ottawa, Ontario, Canada

ISBN 0-315-93580-4

Canada

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ABSTRACT

Parametric methods, currently used in regional flood frequency analysis, have numerous drawbacks and limitations, especially with regard to flood distribution selection and regional relationship form. Alternative approaches involving nonparametric methods are investigated in this thesis on a set of New Brunswick annual maximum floods. Nonparametric methods were employed at the three steps of regional analysis: at-site flood frequency analysis, homogeneous region delineation and regional relationship development.

Nonparametric flood frequency analysis indicated that an annual maximum flood data set from New Brunswick contained some unimodal distributions along with many mixed distributions of bimodal and heavy-tailed shapes. A simulation study showed that sampling variability from a unimodal distribution could not account for the bimodality in nonparametric frequency analysis, confirming the existence of mixed distributions. L-moment analysis, a parametric method, confirmed that the entire set of floods from New Brunswick could not be appropriately described by a unimodal distribution.

In this study, a new method is proposed for the purpose of homogeneous region delineation which effectively combines geographical considerations and flood data characteristics. The technique is based on the grouping of stations with similar density function shape, which reflect similar flood

generating mechanisms. In New Brunswick, flood densities of three different shapes were grouped on a geographical basis to delineate homogeneous regions. Statistical tests based on L-moment analysis confirmed that the stations within a homogeneous bimodal region came from the same distribution. But L-moment analysis would propose either the Generalized Logistic or the Generalized Extreme Value as the regional distribution. Nonparametric frequency analysis revealed, however, that the floods within that region actually came from a mixed distribution.

Nonparametric regression was employed for regional relationship development in New Brunswick; however, no significant improvement over the parametric approach of linear regression resulted. Using bootstrapping of pairs, a new method to compute the confidence interval at the center of a nonparametric regression was investigated. A comparison of linear and nonparametric regression confidence intervals can assist in evaluating the appropriateness of a linear model, and thus the need to employ nonparametric regression. Nonparametric regression was shown to be useful in screening irrational relationships that could be developed with the parametric approach. A new regional analysis methodology, involving nonparametric methods at the three steps of regional analysis, is proposed in this study, resulting in improved homogeneous region delineation, in more accurate at-site quantile estimates, and more realistic regional relationships.

ACKNOWLEDGEMENTS

The author wishes to gratefully acknowledge the support, advice and encouragement of Dr. Kaz Adamowski, of the Department of Civil Engineering of the University of Ottawa, throughout the study. I am also indebted to Dr. Mayer Alvo, of the Department of Mathematics of the University of Ottawa, for his assistance with regard to the confidence interval determination aspect of the research.

Financial assistance by the Ontario Graduate Scholarship and Natural Sciences and Engineering Research Council grants is acknowledged as well.

LIST OF SYMBOLS

a, b, c, d, e	regression coefficients
B	vector of regression coefficients
C	kernel variance
DA	drainage area
d_{ij}	analytical expression
EV1	Extreme Value Type I distribution
$F(x)$	cumulative distribution function
$f(x)$	probability density function
$\hat{f}(x)$	estimated probability density function
GEV	Generalized Extreme Value distribution
GLS	generalized least squares
H	heterogeneity ratio
h	smoothing factor
$IMSE$	integrated mean square error
ISE	integral square error
$J(x)$	integral of bivariate density
$K(\ )$	kernel function
LCV	L-coefficient of variation
LN3	Three-parameter lognormal distribution
$m(x)$	analytical expression
n	sample size
n_i	frequency in class interval i

O	single station flood estimates
OLS	ordinary least squares
P	predicted flood estimates from regression equations
Q	design flood
$R(h)$	risk function
$S.E.$	standard error
TCEV	Two-Component Extreme Value distribution
V_1, V_2, V_3	heterogeneity measures
X	matrix of independent variables
X_i^*	bootstrapped sample from X_i
$x_1, x_2, ...$	physiographic/climatic variables
$\bar{x}$	mean of x
Y	vector of dependent variables
Y_p	predicted value
α, β	EV1 distribution parameters
λ_r	rth L-moment
τ_r	rth L-moment ratio
μ, σ	mean and standard deviation
μ_v, σ_v	mean and standard deviation of V
θ	estimated statistic
$\varepsilon, \epsilon, \epsilon_1, \epsilon_2$	random numbers
$\xi_1, \xi_2, \theta_1, \theta_2$	TCEV distribution parameters

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Motivation

The estimation of a design flood is a common task for a water resources engineer. Estimates of design floods of given return periods are needed for the design of bridges, culverts, spillways, navigation works, water intakes and so on. More often than not, this estimation is required at an ungauged site because gauging stations exist at only a limited number of locations. As well, it may be required at a gauged site with too short a period of record to provide a reliable estimate.

It is possible to prorate flows from a nearby hydrometric station with a sufficiently long record, but if that gauged basin is not similar to the

watershed of interest, or very close in size, the resulting estimate may be highly erroneous. Another possibility is to develop a hydrologic model, such as HYMO, but this approach requires considerable site-specific data not usually available. In practice, a frequently used method is based on transferring the information from gauged sites to ungauged sites through a regionalization technique. There are two regionalization methods called the index flood method and the multiple regression method (Kite, 1977).

In the first approach, the index flood method, the ratios of the design floods to an index flood in a particular region are established. This index flood is then estimated by a multiple linear regression equation relating it to physiographic and climatic characteristics of the basin. In the second approach, a series of multiple linear regression equations are developed for a variety of design floods of different return periods. These will sometimes incorporate different physiographic characteristics as differing factors affect floods of small and large return periods.

Regional analysis involves three basic steps (Kite, 1977). The first is defining a homogeneous region, an area throughout which a developed relationship applies. The second is frequency analysis at the gauged sites, while the third is regional relationship development, which, in the past, always involved multiple linear regression. While the regionalization techniques provide relatively easy tools for design estimation, the standard errors of estimates for regional flood equations are rarely less than 20 percent and often exceed 50 percent.

Such large standard errors in flood estimation are undesirable. As well, a regional flood frequency study for the United Kingdom, which used an index flood approach (Natural Environment Research Council, 1975), has been criticized for yielding physically impossible estimates (Hosking et al, 1985). Therefore, the need exists for a methodology leading to regional relationships with lower standard errors and producing more physically meaningful estimates at ungauged sites.

1.2 Scope

All previous regionalization studies have employed parametric methods for single station flood frequency analysis and regional relationship development. Nonparametric techniques, which are described in detail later, provide a new frontier in hydrology (Adamowski, 1987), and are currently unexplored with regard to regional relationship development. It is thus proposed to investigate the use of nonparametric techniques in the three major steps of regional analysis, leading to the development of a new methodology for regional flood frequency analysis.

In terms of homogeneous region delineation, it is proposed to investigate how the shape of the probability density function, as obtained from nonparametric frequency analysis, can be compared on a geographical basis among gauged sites and become an additional tool in delineating homogeneous regions. Single site parametric frequency analysis has numerous

drawbacks, the most salient being that it is unsuitable for multimodal densities, which are known to occur in Canada. Nonparametric frequency analysis does not suffer from this drawback (Labatiuk and Adamowski, 1987). And finally, it is known that the basic assumptions of linear regression are not fulfilled in regional flood relationship development (Holder, 1985). Nonparametric methods are not constrained by these assumptions.

Before undertaking any kind of frequency analysis, it is very important that the streamflow data be screened for inconsistencies, errors, omissions, the impact of regulation and significant land use changes. For a previous regional flood frequency study in New Brunswick (Inland Waters/Lands and New Brunswick Department of Municipal Affairs and Environment, 1987), this screening was performed. As well, physiographic/climatic factors known to be correlated to flood peaks were derived. For these reasons, the already screened data of New Brunswick will be used in this study for numerical analysis.

A concern in using nonparametric frequency analysis to assess distribution shape is related to its ability to correctly indicate a unimodal distribution when the data are unimodal and correctly indicate a mixed distribution when the data come from a mixed distribution. Thus, a simulation study will investigate the impact of sampling variability on the density shape provided by nonparametric frequency analysis.

Additionally, it is important to determine whether the developed nonparametric regressions lead to regional relationships more accurate than

those of the multiple linear regression approach. Through Monte Carlo simulations and bootstrapping techniques, means of comparing parametric and nonparametric regressions are investigated. This aids to ascertain whether a given nonparametric regression is superior to the corresponding linear regression through the determination of confidence regions.

This thesis therefore provides a new methodology for regional analysis by nonparametric methods, resulting in improved homogeneous region delineation, in more accurate single station flood frequency analysis and more physically meaningful regional relationships.

1.3 Objectives

The main goal of the thesis is to develop a new regional flood frequency analysis approach employing nonparametric methods, and to compare it with the currently used parametric methods. A new methodology is investigated which considers the shape of the probability density function for homogeneous region delineation, and employs nonparametric frequency analysis and nonparametric regression analysis for regional relationship development. This leads to more accurate and reliable estimates of design floods.

The specific objectives are as follows:

1. Perform single station flood frequency analysis by nonparametric

methods using the entire period of record of annual maximum daily mean flows for New Brunswick hydrometric stations with at least 10 years of observations. Compare with the quantile estimates obtained by the parametric techniques.

2. Design a Monte Carlo simulation study to test whether the multimodal density estimation given by nonparametric frequency is due to sampling variability.

3. Investigate the delineation of homogeneous flood regions for New Brunswick using the shape of the probability density function.

4. Develop nonparametric regressions to determine the regional relationships for the 2, 10, 20, 50 and 100 year floods in New Brunswick. Compare with the regressions obtained by the parametric techniques.

5. Design a Monte Carlo simulation experiment to evaluate the accuracy criterion selected for comparison of nonparametric and parametric regression in regional analysis. Develop and apply a new bootstrapping technique of pairs in order to obtain confidence intervals for the regressions for further comparison.

6. Develop a new methodology for regional analysis employing nonparametric methods at all steps of the investigation.

1.4 Thesis Overview

Chapter I gives background detail on the need for regional analysis, presents how the limitations of the standard parametric approaches are proposed to be overcome, and summarizes the thesis.

Chapter II reviews the literature with regard to regional analysis, flood frequency analysis and regression analysis, pointing out unresolved problems and deficiencies with the parametric approaches. The advantages of nonparametric methods are highlighted.

Chapter III develops theoretical background on nonparametric frequency analysis and nonparametric regression analysis. As well, Monte Carlo simulation and bootstrapping are introduced. Use of these two approaches will be made later to compare parametric and nonparametric regressions. A recent parametric frequency method, L-moment analysis, is explained.

Chapter IV discusses the hydrometric network along with the climate of New Brunswick, and then presents and compares the single station flood frequencies, both parametric and nonparametric. Simulation studies and L-moment analysis results are given.

Chapter V presents and discusses the selected homogeneous regions and the regression equations. The simulation results are given so that the parametric and nonparametric regressions can be compared. A new method of estimating confidence regions at the center of a nonparametric regression

using bootstrapping is investigated. A new regional analysis methodology is proposed.

Chapter VI summarizes the study, provides major conclusions, and gives recommendations for further research.

Chapter 2

LITERATURE REVIEW

2.1 Regional Analysis

2.1.1 Methodology

Regional analysis consists of transferring streamflow information from a number of gauged sites within a homogeneous region to ungauged sites or gauged sites with a short record within that same region in order to arrive at an estimate of design flows. A given geographical area is considered homogeneous if it can be assumed that a given regional relationship applies to all streams in that area.

The delineation of homogeneous regions is a controversial issue in re-

gional studies. A common approach to delineate homogeneous regions is to examine the residual pattern from a multiple linear regression of a given design flood for the entire study area. Based on the geographical proximity of positive and negative residuals, homogeneous regions are then proposed. For a New Brunswick regional flood study (Inland Waters/Lands and New Brunswick Dept of Municipal Affairs and Environment, 1987), it was found that the residual pattern corresponded in some instances to the province's climatic zones (Inland Waters/Lands and New Brunswick Department of Municipal Affairs and Environment, 1988a).

Other definitions of homogeneous region are that all single station flood frequencies within the region follow the same probability distribution (Hosking and Wallis, 1991) or that standardized flow statistics are similar (Burn, 1989). Burn argues that similar extreme flood-return period relationships occur because of similar rainfall patterns combined with similar physical characteristics of the drainage basins. Thus, climate appears to play an important role in homogeneous region delineation.

Proper delineation of homogeneous regions is important in all regional transfer methods. There exist two basic regional analysis techniques, to which there are variations; namely, the multiple regression technique and the index relationship approach. The multiple regression technique involves the development of a linear relationship between the design floods and physiographic/climatic variables such as drainage area, area of lakes and swamps, mean annual precipitation and so on. For a number of hydrometric

stations within the study area, single station flood frequency analysis is performed to generate a set of design floods for all gauged sites. For the basins draining to the hydrometric stations, various physiographic/climatic parameters, assumed or known to be uncorrelated among themselves, are determined. These two types of variables are then related through multiple linear regression.

Because a logarithmic transformation is often applied, the final equation is generally of the following form:

$$logQ = a + blogx_1 + clogx_2 + ... + zlogx_n \qquad (2.1)$$

where Q is the design flood of a given return period, x_1 to x_n the physiographic/climatic variables and a to z the regression coefficients.

Usually, there are only one or two statistically significant physiographic/climatic variables; for example, drainage area was the only variable in Prince Edward Island (Inland Waters Directorate, 1989), while drainage area and mean annual precipitation were significant in New Brunswick (Inland Waters/Lands and New Brunswick Dept of Municipal Affairs and Environment, 1987). Rarely are three or more basin characteristics statistically significant due to degrees of freedom limitations caused by sample size availability.

The residuals, or the differences between the predicted design floods from the regression equations and the estimated design floods from single

station frequency analysis, along with geographical considerations, are then employed to define homogeneous regions. The geographical distribution of residuals in a homogeneous region should be random (Riggs, 1985).

In some of the variations of the multiple linear regression technique, one develops a relationship between the probability distribution parameters from parametric single station flood frequency analysis and physiographic data (Lettenmaier and Potter, 1985) or by yielding regional estimates of these parameters for the homogeneous region (Fiorentino et al, 1987). Some have proposed that higher-order parameters controlling the skewness be determined from larger regions than the parameters controlling the coefficient of variation (Gabriele and Arnell, 1991). These variations have been criticized for generating links between the standard deviation or the skew coefficient with the basin characteristics that are not always statistically significant in Canada (National Research Council of Canada - Associate Committee on Hydrology, 1989).

The multiple linear regression technique requires the appropriate physiographic/climatic variables to explain the naturally occuring changes in flood flows in the watershed, and these can sometimes be difficult to ascertain. Its most serious drawback, though, is that the model is assumed linear, and there is no physical justification for a linear relationship between floods and physiographic characteristics, even after some transformation.

The main advantage of the regression approach is that basin characteristics known to affect floods are part of the relationship. Another advantage

is that the application of linear regression yields standard errors of estimate and a coefficient of determination. These provide information as to the uncertainty of the estimates and permit comparison among equations using different physiographic/climatic factors or assuming different homogeneous regions.

In the index flood method, for the gauged basins of a given homogeneous region, the ratios of the floods of a given return period to an index flood, usually of a two year return period, are computed (Harvey et al, 1985). A plot of the median ratios versus the return period yields the index relationship, with the index flood estimated by a linear regression relationship. The main criticism of the index approach (Kite, 1977) is that the ratio to the index flood is very much basin size dependent and not a constant as assumed in this technique. Thus, there is a large margin of error in its application unless the relationship was developed over a narrow range of basin sizes and is applied within that range. Another major criticism is that other important physiographic/climatic factors affecting the ratio to the index flood, such as the area of lakes and swamps, are never considered.

It is interesting to note that in New Brunswick, the 2 year flood regression equation on a province-wide basis has a standard error of 18% while the 100 year flood has a standard error of 28%. Since variances are additive, an index relationship with a standard error of 21% or less would be required to obtain a more accurate estimate of the 100 year flood by combining the 2 year regression equation with an index relationship. The

ratios of the 100 year floods to the 2-year floods were computed and a coefficient of variation of 26.4% for that ratio was found. Thus, the nutliple linear regression equation for the 100 year flood would be less biased than a combination of the 2 year flood equation and a 100 year index ratio.

Caissie and El-Jabi (1991) reached a similar conclusion from a Canada-wide comparative study of the multiple regression and index-flood methods. They showed that the multiple regression approach yielded results that were similar or superior to those from the index-flood method.

While both regional analysis techniques have their limitations, the multiple linear regression approach has a sounder physical basis as it considers physiographic/climatic factors influencing floods. Since large return period floods are known to be caused by factors other than small return period floods, multiple linear regression may result in a series of equations for various return periods having differing independent variables. For these two reasons, it is favored by many and recommended by the author over the index method.

2.1.2 Recent Canadian Studies

There have been a number of regional flood frequency studies undertaken in Canada during the past few years. While they have provided the means to estimate more accurately design floods at ungauged sites, the large standard errors of the equations lead to great uncertainty.

A regional flood study for the Island of Newfoundland (Interagency Working Group, 1984), involving twenty-two hydrometric stations, developed multiple linear regression equations for the 2, 5, 10, 20, 25, 50 and 100 year instantaneous peak floods on a province-wide basis and for two homogeneous regions. Standard errors ranged from 13 to 26%.

A regional flood study for Ontario (Moin and Shaw, 1986) developed multiple linear regression equations and index flood frequency curves for the 2, 5, 10, 20, 50 and 100 year instantaneous peak floods. Up to 270 stations were employed in the analysis. For the index method, twelve homogeneous regions were defined, while three regions were defined for multiple regression. Even though a comparatively large number of stations were employed, the two-year flood standard errors ranged from 20 to 35%.

A regional flood frequency study for New Brunswick (Inland Waters/Lands and New Brunswick Dept of Municipal Affairs and Environment, 1987) involved multiple linear regressions based on 37 stations throughout the province of New Brunswick and surrounding areas for a concurrent period of record of sixteen years. Historical information was employed at four locations. Not only was there a province-wide equation developed for the 2, 10, 20, 50 and 100 year daily mean floods, but equations for four homogeneous regions were established as well. The main stem of the Saint John River was considered a fifth region because of its overwhelming size in comparison to the other streams.

Also developed in that study were regression equations for the ratio

of the annual maximum instantaneous peak flows to the annual maximum daily flows (QP/QD). Standard errors ranged from 7 to 37%; however, some of the low standard errors were for regions with ten or fewer hydrometric stations and with a large range of basin sizes. The present work will apply new regional techniques to the New Brunswick flood data base.

In Prince Edward Island (Inland Waters Directorate, 1989), a regional flood study developed linear regression equations for the 2, 10 and 20 year instantaneous peak floods. Longer return period flood equations were avoided because of the short record lengths of hydrometric stations in the study area. Because only eight stations with at least nine years of record were available, and these varied in size from 5.6 to 146 square kilometres, standard errors were from 45 to 56%.

As longer data bases are becoming available, it is expected that further regional studies will be undertaken in Canada. And previous studies are likely to be updated as hydrometric networks expand, as was done in New Brunswick. Because of the large standard errors of regional flood estimates, as mentioned earlier, it is important to develop a more accurate approach, such as the one proposed in this thesis investigating nonparametric methods.

2.1.3 Homogeneous Region Delineation

Homogeneous regions are usually defined as regions having similar hydrologic, climatic, physiographic characteristics and/or flood-related variables, such as coefficient of skew. The subdivision of a large area into smaller homogeneous regions for the purpose of flood frequency analysis, as well as the determination of a regionalized skew coefficient, is often quite controversial. For quantile estimation, regional equations are developed or a regional parameter evaluated which then apply to this homogeneous region.

Grouping basins of similar hydrologic response or following the same probability distribution or of similar statistical parameters is expected to yield regional relationships with lower standard errors than those for the entire study area. In some cases, the difference can be somewhat significant as in the New Brunswick study where for the 2 year flood, for example, the province-wide equation has a standard error of 18% and the equations for smaller homogeneous regions have standard errors of 14, 12, 11 and 15%. In the Newfoundland study, the province-wide 2 year flood equation has a standard error of 26%, but the two regional equations have standard errors of 20 and 21%. In Ontario, the province-wide 2 year flood equation has a standard error of 35% while the regional equations have standard errors of 38, 20 and 32%.

A major concern in delineating homogeneous regions is that the gauged basins within the homogeneous region also be representative of the un-

gauged sites within that region. The impact of an odd gauged basin on the developed regional flood relationship can be reduced by including as many gauged basins as possible in the analysis. A large sample also limits the presence of spurious correlations between the design flood and the physiographic/climatic variables.

Homogeneous region delineation through residual examination and geographical considerations, usually physiography and climate, has dominated in the past. However, homogeneous regions delineated on this basis were often found not to be hydrologically similar (Linsley, 1982; Cunnane, 1987). Consequently, some researchers defined homogeneity in terms of flood-related variables such as normalized annual floods and coefficient of variation, which led to non-geographical regions (Wiltshire, 1985; Burn, 1989). The argument of the proponents of non-geographical regions is that geographical proximity is no guarantee of homogeneity because neighbouring basins are often physically very different. Even tributaries need not have the same flood-producing properties as the main basin.

Burn (1990) developed the region of influence approach with Manitoba flood data. This technique involves each gauging site having a potentially different set of stations for at-site estimation. The problem with Burn's non-geographical regions is in the practical application to ungauged sites. One cannot easily determine which gauged watersheds a given ungauged site is most related to because streamflow data does not exist at that site for comparison.

Wiltshire (1985) partitioned some gauged basins on the basis of a single physiographic/climatic characteristic; for example, the data set could be divided in a set of "small" and "large" basins based on area, or a set of "wet" or "dry" basins based on mean annual precipitation. He found that, for flood data from the United Kingdom, partitioning the basins on the basis of a soil index and mean annual rainfall led to the best fit. Bhaskar and O'Connor (1988), using cluster analysis to determine non-geographical flood regions for the state of Kentucky in the United States, and Cavadias (1990), using canonical correlation on Newfoundland flood data, also found non-geographical regions differing from those of previous studies which used residuals from multiple linear regression and geographic considerations. If the concept of non-geographical regions is correct, then any new homogeneous region delineation approach must be flexible enough to provide a non-geographic solution.

Those homogeneous region delineation methods seeking an identical parametric distribution or regional values of parameters must contend with the limitations and subjectivity of parametric flood frequency analysis. An alternative to parametric flood frequency analysis which has been successfully employed is nonparametric frequency analysis (Adamowski, 1989). Nonparametric methods do not depend on an assumption regarding the type of distribution or on the estimation of many parameters.

In fact, a single parameter, the smoothing factor, is calculated based on available data in nonparametric frequency analysis. This smoothing

factor has no regional interpretation. Nonparametric methods, as shall be further discussed later, can handle multimodal densities, which have been found in the annual maximum flood distributions of Canadian rivers where different flood generation mechanisms are at work (Alila, 1988; Labatiuk and Adamowski, 1987).

Essentially all of the above discussed approaches for homogeneous region delineation are based either on some sort of geographical considerations (on the basis of general location, physiography, weather regimes) or flood data characteristics (probability distribution, regional statistical flood parameters). It is possible to combine geographical considerations and flood data characteristics for homogeneous region delineation through the use of the probability density function shape.

As it is known that maximum annual flood series typically contain floods arising from different mechanisms, probability density function shape will be dependent on the processes generating floods in the watershed. A basin with a single mechanism for flood generation will have a unimodal density while a basin with two or more mechanisms, for example fall floods and spring snowmelt floods, may have a bimodal probability density function. This shape, which can be determined from the nonparametric frequency density, is then compared on a geographical basis and used for delineation of homogeneous regions. Such an approach is essentially a reflection of climatic considerations.

It is proposed to develop regional flood relationships for the province of

New Brunswick, where floods occur from a variety of mechanisms, using a regression approach because of its lower standard errors over the index approach. For the developed regional flood equations to properly represent the existing natural links between flow and watershed characteristics, three criteria must be satisfied. First, the homogeneous area over which the regional equation applies must be correctly delineated, and the shape of the probability density function, along with geographical factors, will be considered in this delineation. Secondly, accurate regional flood equations depend on accurate estimates of the gauged site flood frequencies. And finally, accurate equations also depend on an accurate form of the relationship between the design floods and the influencing variables. At-site flood frequency analysis and regression relationship development literature are discussed in the next two sections.

2.2 Flood Frequency Analysis

2.2.1 Parametric Approach

The estimation of a flood of a given return period can be accomplished by the more conventional parametric methods or by nonparametric methods. The parametric approach involves fitting an assumed theoretical probability distribution to a series of observed annual maximum flows. This assumed distribution, along with its estimated parameters, are then employed to

estimate quantiles of the distribution. Numerous probability distributions have been proposed to fit annual maximum floods, but the most commonly used in Canada are the lognormal, the three-parameter lognormal, the Generalized Extreme Value, and the log-Pearson III (Pilon et al, 1985).

The normal distribution is the basis for the lognormal and the three-parameter lognormal. It has a bell-shaped curve with the following density function:

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

where μ is the population mean and σ^2 the population variance. Its skew coefficient equals 0.0, while its coefficient of kurtosis equals 3.0.

Because annual floods are typically skewed, a logarithmic transformation can be applied and then a lognormal distribution fitted to the data. The three-parameter lognormal (LN3) is a shifted lognormal distribution with the variate equal to $ln(x-c)$, where c is a lower boundary. Its probability density function is:

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

where μ_y and σ_y^2 are the mean and the variance of the $ln(x-c)$ values. The floods of many Canadian rivers were found to fit the LN3 well (Sangal

and Biswas, 1970). Regional studies for Ontario (Moin and Shaw, 1986), Newfoundland (Interagency Working Group, 1984) and New Brunswick (Inland Waters/Lands and the New Brunswick Department of Municipal Affairs and Environment, 1987) selected the LN3.

The General Extreme Value (GEV) distributions are also used to describe annual maximum floods. The Gumbel distribution, very popular in the United Kingdom, is also known as the Extreme Value Type I (EV1) or the double exponential distribution. It is a special case of the GEV, with a density function as follows:

$$f(x) = 1/\alpha \exp(-((x-\beta)/\alpha) - \exp(x-\beta)/\alpha) \qquad (2.4)$$

where α is a scale parameter and β a location parameter. With a skew coefficient of 1.14 and a coefficient of kurtosis of 5.4, the application of the Gumbel to Canadian rivers is limited.

The Pearson distributions are a set of distributions that solve a differential equation proposed by Karl Pearson. The log-Pearson type III is the recommended distribution in the United States and was selected for Prince Edward Island (Inland Waters Directorate, 1989). A very flexible distribution is the Wakeby (Houghton, 1978), which has five parameters. Not only are high order moments not required for parameter estimation, but the left-hand side observations are divorced from the right-hand side observations.

A more complex compounded distribution, although still unimodal like the others, is called the Two-Component Extreme Value (TCEV) (Rossi et al. 1984). This four-parameter distribution, which effectively combines two EV1's, has the following cumulative probability function:

$$F(x) = exp(-exp(-(x-\xi_1)/\theta_1)).exp(-exp(-(x-\xi_2/\theta_2) \qquad (2.5)$$

where ξ_1, ξ_2, θ_1 and θ_2 are the parameters. This distribution was found appropriate in Italy where flood distributions exhibit a large skew and have a heavy tail.

The hydrologist is thus left with a wide selection of distributions for fitting. There exist four standard fitting techniques (Kite, 1977) which, in order of increasing efficiency, are: graphical, least squares, moments and maximum likelihood. More recent and very promising approaches propose the use of probability weighted moments and L-moments (Hosking and Wallis, 1990).

It has been found that L-moments can characterize a wider range of distributions than conventional methods, are more robust in the presence of outliers and are less subject to bias. L-moment ratios are compared on a regional basis in order to determine the appropriateness of a single frequency distribution to describe all flood flows in a particular area (Hosking, 1990). L-moments have been applied to determine the regional distribution of precipitation in Washington State, U.S.A. (Schaefer, 1990) and Ontario (Pilon

et al. 1992). They have also been applied to annual maximum floods in Nova Scotia (Pilon and Adamowski, 1992) and Newfoundland (Pilon et al. 1992) where the GEV distribution was chosen as the regional distribution. More details on L-moments will be provided in section 3.4.

Distribution selection in L-moment analysis is performed by comparing L-moment ratios such as L-skewness and L-kurtosis to the theoretical values. This is possible for a variety of unimodal distributions including the compounded TCEV distribution (Gabriele and Arnell, 1991). For the latter, there exists a limited range of values of L-skewness and L-kurtosis which can yield a TCEV. The usual application of L-moments involves selecting the most appropriate unimodal distribution, as was done for Nova Scotia and Newfoundland. However, it is known that many rivers in Canada follow an annual flood distribution better described by a multimodal density; for example, it was shown that the probability density function of the Northeast Margaree river floods in Nova Scotia is bimodal (Labatiuk, 1985).

Fitting and selecting a probability distribution for a set of annual floods presents numerous difficulties. The available hydrometric records to perform frequency analysis are usually short in relation to the probabilities of events or design floods required. These design floods are typically for large return periods, for the right-hand tail of the distribution, where little information is available. In all parametric methods, the entire data set determines the shape of the tail and it has been shown that a small differ-

ence in the lower flood values can severely impact on the right-hand tail where the differences between the various distributions are very significant (Klemes, 1987).

Estimates of high order moments needed for parameter estimation and distribution selection, such as the coefficient of skew and kurtosis, are much more reliable the greater the record length. Outliers can greatly distort the analysis as well, although very large flood values provide important information about the right-hand tail. Furthermore, different distributions can fit the observed data equally well and yield very different large return period floods. In practice, an incorrect assessment of the design flood will lead to either an inadequate system that will fail more often than anticipated or to too large an expenditure on the construction of that system.

Still another problem with the conventional parametric flood frequency analysis approach is that floods in Canada are known to be the result of a variety of generating mechanisms such as snowmelt, frontal precipitation, local thunderstorms and hurricanes, and thus are not drawn from a single statistical population. Also, watersheds are not time invariant, as they produce different flood outputs for the same storm input at different times of the year. Obviously, in a cold climate, floods occurring in the spring, on frozen ground with snowmelt, are different from summer floods.

Throughout the North American continent, the large return period floods are very often caused by a different process than the smaller return period floods (Stoddart and Watt, 1970; Waylen and Woo, 1984; Diehl and Potter,

1987; Hirschboeck, 1987) and yet it is these smaller floods, along with the large ones, that are used to assess the larger design floods in parametric flood frequency analysis. The many drawbacks of parametric frequency analysis make distribution selection a difficult task, as exemplified by the complexity of an expert system for flood frequency analysis (Chow and Watt, 1990).

Because of the various mechanisms at work, a parametric unimodal distribution is inappropriate in many instances as the probability density function of floods is often bimodal (Labatiuk, 1985; Alila, 1988). Sometimes, there are two distinct peaks of different magnitude, as for the Nackawic River in New Brunswick, as shown in Figure 2.1. Some researchers have suggested the use of a combination of distributions, or the use of a heavy-tailed distribution to properly define the distribution of annual floods where they are multimodal (Waylen and Woo, 1982; Rossi et al, 1984; Singh, 1987; Ahmad et al, 1988). The problem with a mixed distribution approach again concerns distribution selection along with the increased number of parameters to be estimated.

The many difficulties associated with parametric flood frequency analysis, particularly related to tail behaviour and multimodality, have led to the investigation of other approaches. One such successful technique is nonparametric frequency analysis, which will be investigated in this study.

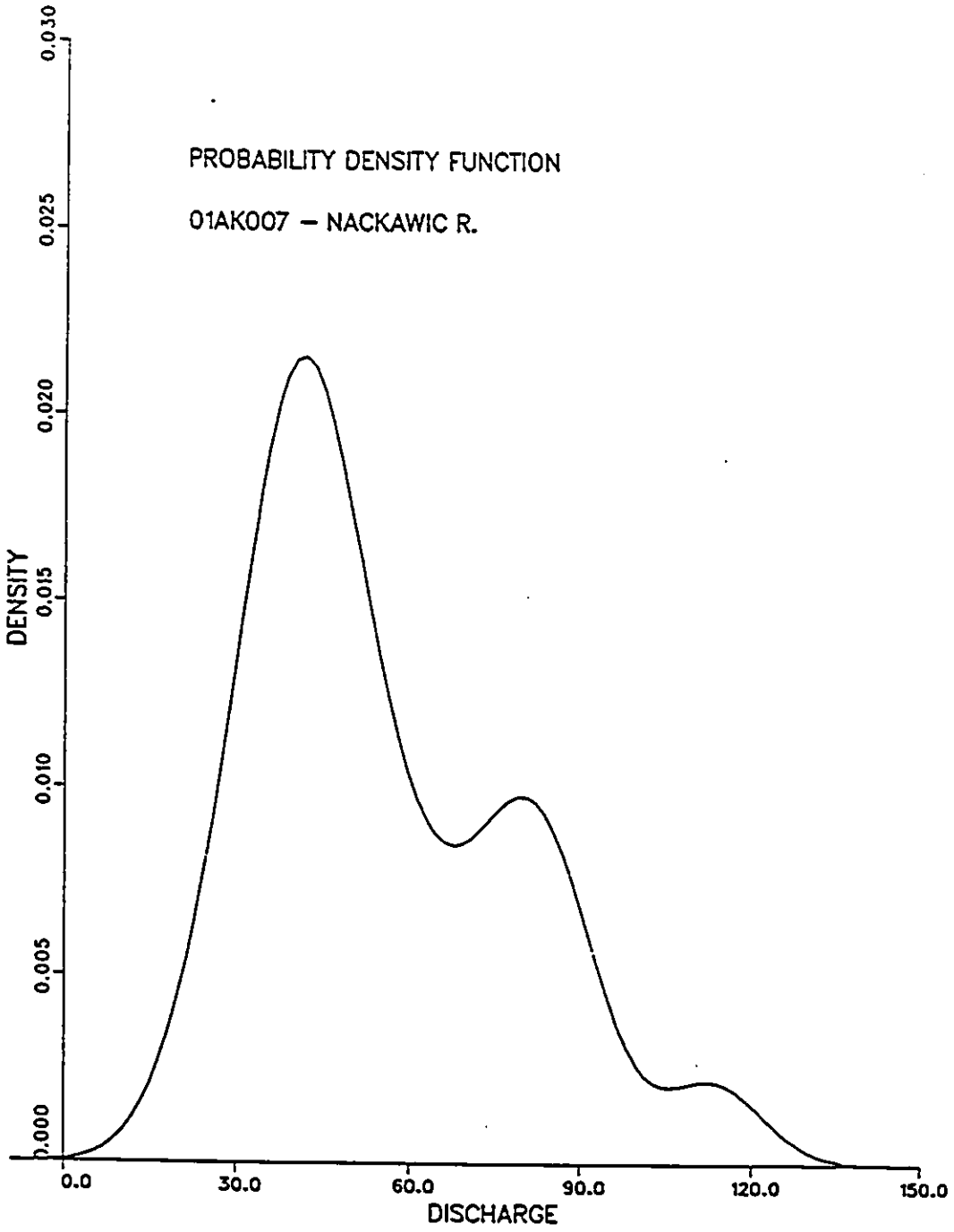


Figure 2.1: Probability Density Function of Nackawic River

2.2.2 Nonparametric Approach

The great advantage of nonparametric frequency techniques is the avoidance of distribution selection, which is always to some extent subjective (Adamowski, 1985; Bardsley, 1988). In nonparametric frequency, the probability density function is estimated from fitting Fourier series or kernel functions and does not require an assumption, that may turn out to be totally erroneous, with regard to distribution shape or type.

The most popular and commonly used nonparametric method in hydrology is the kernel estimator, but a Fourier series method has been successfully used by Wu and Woo (1989). In fitting the annual maximum floods for eight Canadian rivers, they found lower standard errors by the kernel and the Fourier nonparametric methods over four parametric distributions. The nonparametric methods were also capable of reproducing bimodality.

More details on nonparametric frequency analysis are given in section 3.1. At this point, however, it is important to explain that in the kernel technique, kernel functions, which are themselves probability density functions, are used. Their spread is defined by a smoothing parameter h which must be calibrated from all the data points. The resulting probability density function for the data is, in effect, the sum of the kernels, their centers being located at the observation points.

It has been shown (Labatiuk, 1985) that the selection of h is crucial because it defines the smoothness of the density function. The chosen criteria

for h selection has been minimizing of the integrated mean square error by a cross-validation technique and it is currently computed automatically by nonparametric flood frequency software. A much less critical factor is the choice of the kernel function.

Labatiuk (1985), using Monte Carlo simulation, showed that nonparametric flood estimates were of the same order as parametric flood estimates. The rectangular, quadratic and Gaussian kernels were investigated and they led to comparable results. Analysis of observed flood records indicated the presence of multimodal densities, which are ill-suited to parametric methods.

Alila (1988) pursued the research further by investigating the fixed, transformed, variable and adaptive kernels. It was concluded that the four kernels provided comparable results based on standard errors of estimate. He also studied the generating mechanisms behind the annual maximum floods when the density function was bimodal and found that the two modes corresponded to two different flood generation processes. His conclusion was that the nonparametric methods had successfully estimated a mixed distribution of flood data.

A Monte Carlo simulation (Adamowski, 1989) indicated that nonparametric methods provide more accurate flood estimates than parametric methods for large return periods because the right-hand tail is not dependent on lower flood values. For smaller return period floods, both parametric and nonparametic approaches are equivalent. Bardsley (1988) notes

that, as is the case in parametric frequency analysis, nonparametric methods are unreliable for design magnitudes beyond the largest data value. However, in extrapolation, the quantile estimates from a nonparametric approach are influenced by the larger data points and not the probability distribution selected from the entire data set. Thus, only the relevant information has an impact on the extrapolated values in nonparametric frequency analysis.

Bardsley (1989) performed a Monte Carlo simulation with a Gumbel kernel and found good agreement with the theoretical values but was concerned by underestimation beyond the 100 year return period. He selected the Gumbel kernel because a positively skewed form was deemed desirable if the density was not to decrease too rapidly beyond the largest data point.

The estimation of very large return period floods can be accomplished with the assistance of historical information, which is knowledge that a flood of a given magnitude was reached at a certain point in time. Its inclusion in single station flood frequency analyses, as well as in regional analyses, leads to a better definition of the right-hand tail and of larger return period floods according to recent Monte Carlo simulations (Jin and Stedinger, 1989; Adamowski and Feluch, 1990; Pilon and Adamowski, 1992).

Bardsley (1989) pointed out that the assumption of stationarity of the data series, of constant values in the statistical parameters of the generating process with time, a basic criterion for flood frequency analysis, may no longer apply with the record lengths involving historical information. Even

with relatively short records, the assumption of stationarity of flood data series has been questioned and led to the selection of a shorter concurrent period of record for regional flood frequency in New Brunswick (Gingras et al, 1988).

As well, the notion of estimating very large return period floods, even at only the 100 year level, has been criticized (Klemes, 1987). It is argued that the statistical analysis of short return period snowmelt floods is irrelevant to assigning a value to a large return period flood caused by a different mechanism such as a hurricane. This indicates that even more caution than usual must be exercised in using as well as interpreting any flood frequency analyses employing historical information.

For small return periods, it has been demonstrated that parametric and nonparametric methods provide comparable results, but for large return periods the nonparametric technique has been shown to provide less biased flood frequency estimates (Adamowski, 1989). For this reason, and because it handles multimodal densities and avoids the subjective process of distribution selection, nonparametric frequency is selected in this study for single station flood frequency analysis.

2.3 Regression Analysis

2.3.1 Linear Regression

Regression involves finding a relationship between variables for prediction purposes. The classical method is parametric regression, requiring the specification of a class of regression functions and the estimation of its parameters. Linear regression, with parameters estimated by ordinary least squares (OLS), is the common method of developing regional flood frequency relationships. It involves fitting the linear model $y = a_0 + a_1x_1 + a_2x_2 + ...$ by minimizing the sum of the squares of the deviations of the data points from the chosen regression line.

In matrix form, the model is Y = X B + ϵ, composed of a vector Y of n dependent variables (the design floods), a vector B of $p + 1$ regression coefficients and an n by p matrix X of p independent variables (the physiographic/climatic factors) for the n sites. The value of B resulting in the minimization of the sum of the squares of the deviations is:

$$B = (X^TX)^{-1}X^TY \tag{2.6}$$

where X^T is the transpose of X (Haan, 1977).

Unfortunately, least squares will provide the optimal value of y based on

values of x_1, x_2 and so on only if certain assumptions are met. The relationship between the parameters must be linear and there must be constant variance of the "errors" about the regression line, with these errors being uncorrelated and normally distributed with a zero mean. These assumptions are known to be rarely met in flow regionalization (Holder, 1985).

A recent alternative to the standard least squares estimation of multiple linear regression parameters is generalised least squares (GLS) (Tasker and Stedinger, 1989). This technique takes into consideration the fact that the standard error of estimate of each single station flood frequency is different, mainly due to differing record lengths, and that some correlation exists among the annual maximums at the various sites within a homogeneous region. Application of the GLS procedures to New Brunswick regional flood frequencies led to standard errors of estimate for the regional equations slightly smaller than those of the OLS approach along with more physically realistic regression coefficients (Inland Waters/Lands and New Brunswick Dept of Municipal Affairs and Environment, 1988).

Other regression approaches are ridge regression, principal components and weighted least squares. All these techniques, however, imply a linear relationship between parameters. Even after applying a logarithmic transformation, there is usually no justification for a design flood in a particular homogeneous region to be linearly related to physiographic data.

As well, the scatter around the regression line for a particular homogeneous region is often not uniform. Because of differences in basin slopes

and in the amount of wetlands, there is a large variation in flood magnitude for a given basin size, particularly for the smaller basins. If part of the "homogeneous region" is much steeper or wetter than the remainder of the region, the variations around the regression line are not expected to be normally distributed. Also, as mentioned earlier, large standard errors of estimate for regional equations are obtained with linear regression, sometimes exceeding 50%.

In order to alleviate some of these problems, it is possible to select a nonlinear regression technique or some approach not limited by various assumptions which do not apply in regional analysis. Nonlinear regression, however, would require the assumption of a given shape of the relationship, which may not be the correct one. Thus, nonparametric regression is proposed for this study.

2.3.2 Nonparametric regression

In nonparametric regression, there is no need to assume a relationship of prespecified form as a relationship of possibly complicated or unorthodox shape is generated by the data points themselves. The value of the predicted variable as a function of the explanatory variable or variables for any regression can be established as the ratio of two probability density functions, which are estimated with nonparametric frequency techniques (Muller, 1988; Prakasa Rao, 1983).

Further details on nonparametric regression are available in section 3.2. A very important consideration is that the nonparametric approach does not require the assumption of linearity and of constant variance of normally distributed variations about the regression line, which are known not to apply in regional relationship development.

A fundamental concern in multiple linear regression is for the number of degrees of freedom, which equals the number of data points minus the number of variables. While it is possible and statistically valid to obtain an equation with as little as a single degree of freedom, the resulting equation is of very limited practical value as insufficient data was used to develop it. Thus, the larger the data set used to develop linear regression relationships, the greater the confidence in that relationship.

In nonparametric regression, it is also advisable to employ a large data set. It has been shown (Silverman, 1986) that for density estimation in the unidimensional case, a sample size of seven is required to obtain a mean square error of 0.1. In two dimensions, the sample size required to achieve the same error increases to thirty-two and in three dimensions it becomes one hundred and fourteen. Luckily, most regional equations involve two dimensions, on a few occasions three, and very rarely more.

Another concern in regression is with multicollinearity, or near-linear dependencies among the explaining variables (Wetherill, 1986). The selection of physiographic/climatic variables uncorrelated among themselves, which is the usual practice, will eliminate the problem if the relationship among

the variables is linear.

A final concern is with the wide range of the data sets, often covering many orders of magnitude, encountered in some regressions. The difficulties these entail can be minimized by applying a transformation in order to reduce the range of the data, such as the logarithmic transformation usually taken in regional analysis and mentioned earlier. It is still possible, however, for a small sample covering a wide range in a given variable to have a zone lacking a nonparametric density because there are no data points nearby. Thus, an adequate spread of the physiographic/climatic variables within their range is important.

Therefore, nonparametric regression as applied to logarithmically transformed variables, as long as a good spread of these variables is present, fulfills all the requirements of regression analysis without the limitations of linear parametric regression. While nonparametric regression has already been applied in hydrology to the prediction of groundwater levels from runoff by Adamowski and Feluch (1991) as well as to a long-range streamflow forecasting model (Smith, 1991), its application to regional analysis has not been as yet investigated.

2.3.3 Confidence Intervals

The calculation of confidence intervals is one method of comparing the results of parametric and nonparametric regression. In the linear regression

case, the theory has been developed and the following equation is known to provide the standard error of an estimate from a linear regression relationship (Haan, 1977):

$$S.E. = \left(\left(\frac{\sum (Y_p(i) - y(i))^2}{n-2} \right) \left(\frac{1}{n} + \frac{(x(i) - \bar{x})^2}{\sum (x(i) - \bar{x})^2} \right) \right)^{0.5} \qquad (2.7)$$

where x and y are, respectively, the independent and dependent variables, $x(i)$ is the independent variable at a given data point, $y(i)$ the corresponding value of the dependent variable y at $x(i)$, $Y_p(i)$ the predicted value at $x(i)$ from the fitted regression equation, $\bar{x}$ is the mean of x, and n is the number of data points. Thus, the confidence intervals are not constant but depend on their proximity to the mean value of x.

Unfortunately for nonparametric regression, no theoretically exact solution exists for the construction of its confidence intervals. However, resampling theory permits the calculation of confidence intervals through the use of a technique called bootstraping (Efron, 1982) to be explained in section 3.3.2. In order to develop confidence intervals for a regression, bootstrapping of residuals has been proposed (Freedman, 1981). However, since residuals are not uniform over a regression relationship, other methods have been investigated.

To develop confidence intervals for a nonparametric regression, some multi-step bootstrapping techniques have been investigated (Hardle and Marron, 1991). Here, the residual at a given location is used to recon-

struct the distribution of residuals and a simultaneous confidence region for the entire nonparametric regression is obtained. This technique is a time-consuming exercise. It will be shown in section 5.4 that a simpler technique using the bootstrapping of pairs duplicates the theoretical confidence intervals at the center of a linear regression. By a similar process, confidence intervals will be investigated at the center of the nonparametric regression relationship so that both parametric and nonparametric approaches can be compared.

Chapter 3

THEORETICAL DEVELOPMENT

3.1 Nonparametric Frequency Analysis

The easiest way to present the concept of nonparametric frequency is to begin with the histogram. If a sample of n observations $x_1, x_2, ...x_n$ is available, by selecting a class width h, a histogram can be built like the one in Figure 3.1. If the vertical axis contains relative frequency, then each class interval has a relative frequency of n_i/n, where n_i is the number of data points within the class interval.

However, since the area under a probability density function must equal

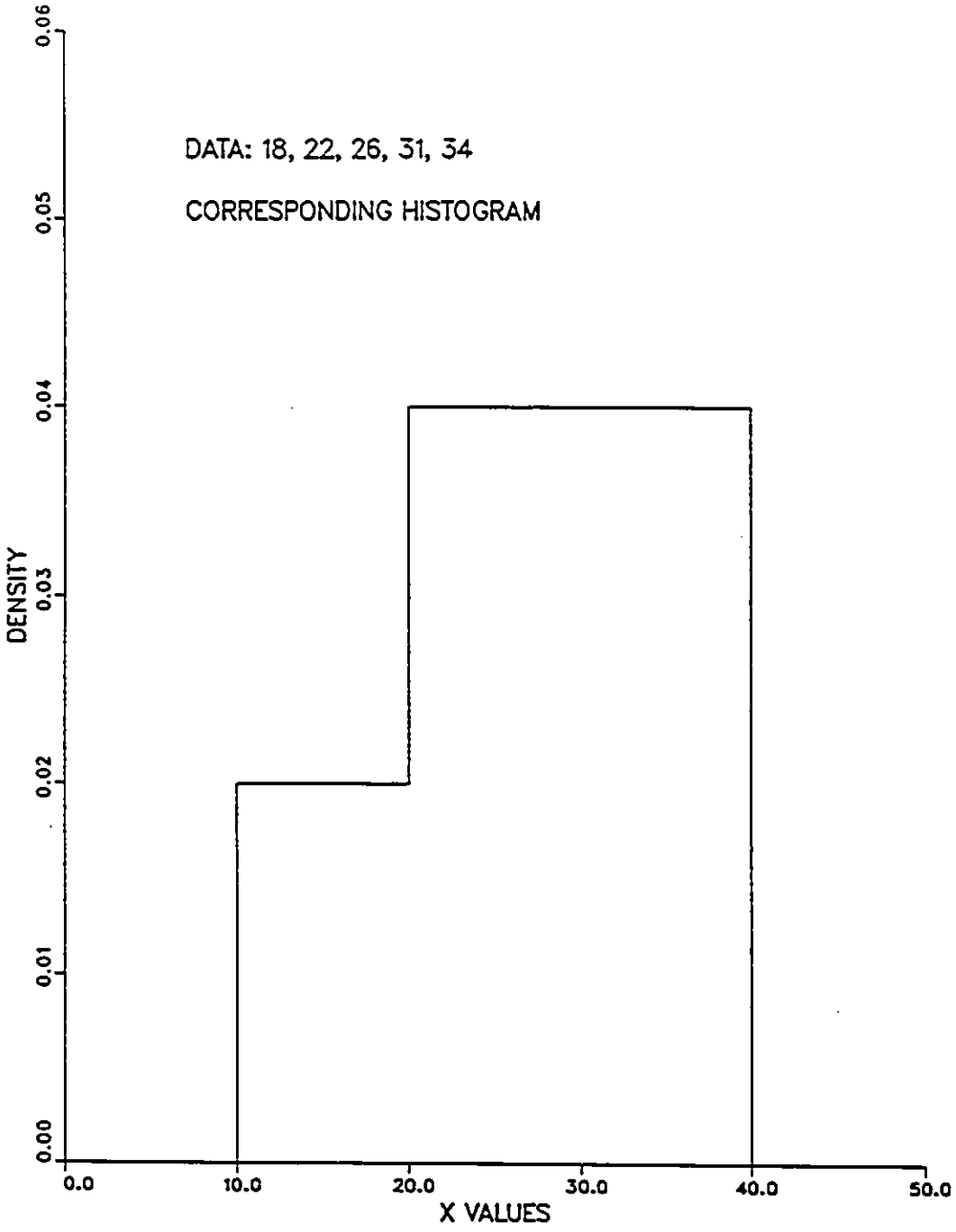


Figure 3.1: Histogram Construction

unity, by dividing each relative frequency by h, the corresponding density value is obtained. Assigning a probability of $1/n$ to each data point, the estimated density $\hat{f}(x)$ can be generalized as:

$$\hat{f}(x) = \frac{no.\ of\ observations\ within\ class}{n \quad h} \tag{3.1}$$

It is important to note that in building a histogram, a block of height $1/nh$ centered at the middle of the class interval is added for every data point within that interval.

In nonparametric frequency analysis by the kernel method, an approach in many ways similar to that of histogram building is employed. But instead of using a rectangular block, a kernel of a given shape, which is not necessarily rectangular, is chosen. And as well, instead of centering the kernel at the middle of an interval, it is located at the data point location itself. Once again, the probability density function is the sum of the kernels, as shown in Figure 3.2. Figure 3.3 shows the corresponding cumulative density.

The following equation is used to express mathematically the nonparametric density estimate $\hat{f}(x)$ (Silverman, 1986):

$$\hat{f}(x) = \frac{1}{n \quad h} \sum_{i=1}^{n} K((x - x_i)/h) \tag{3.2}$$

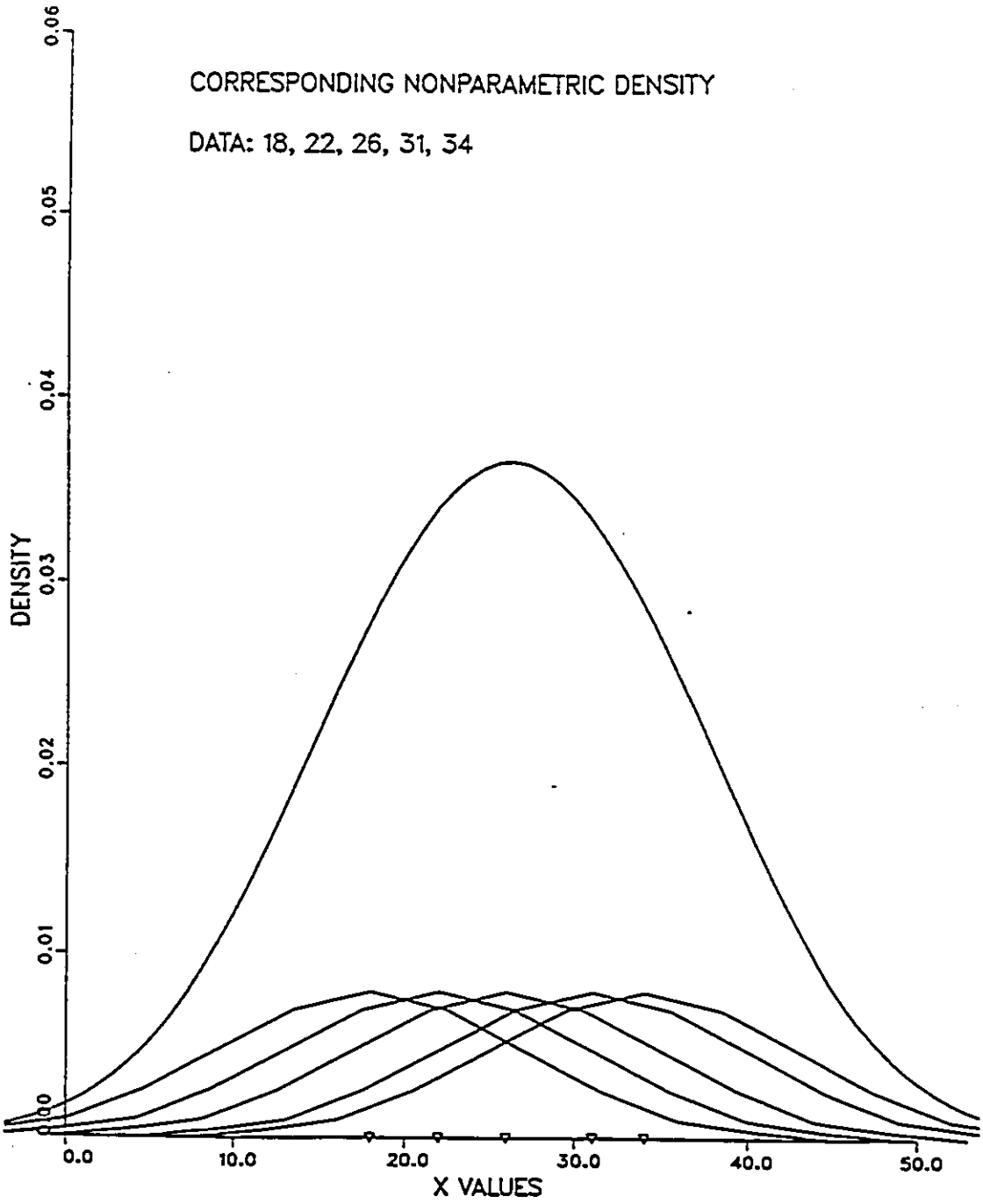


Figure 3.2: Nonparametric Density Construction by the Kernel Method

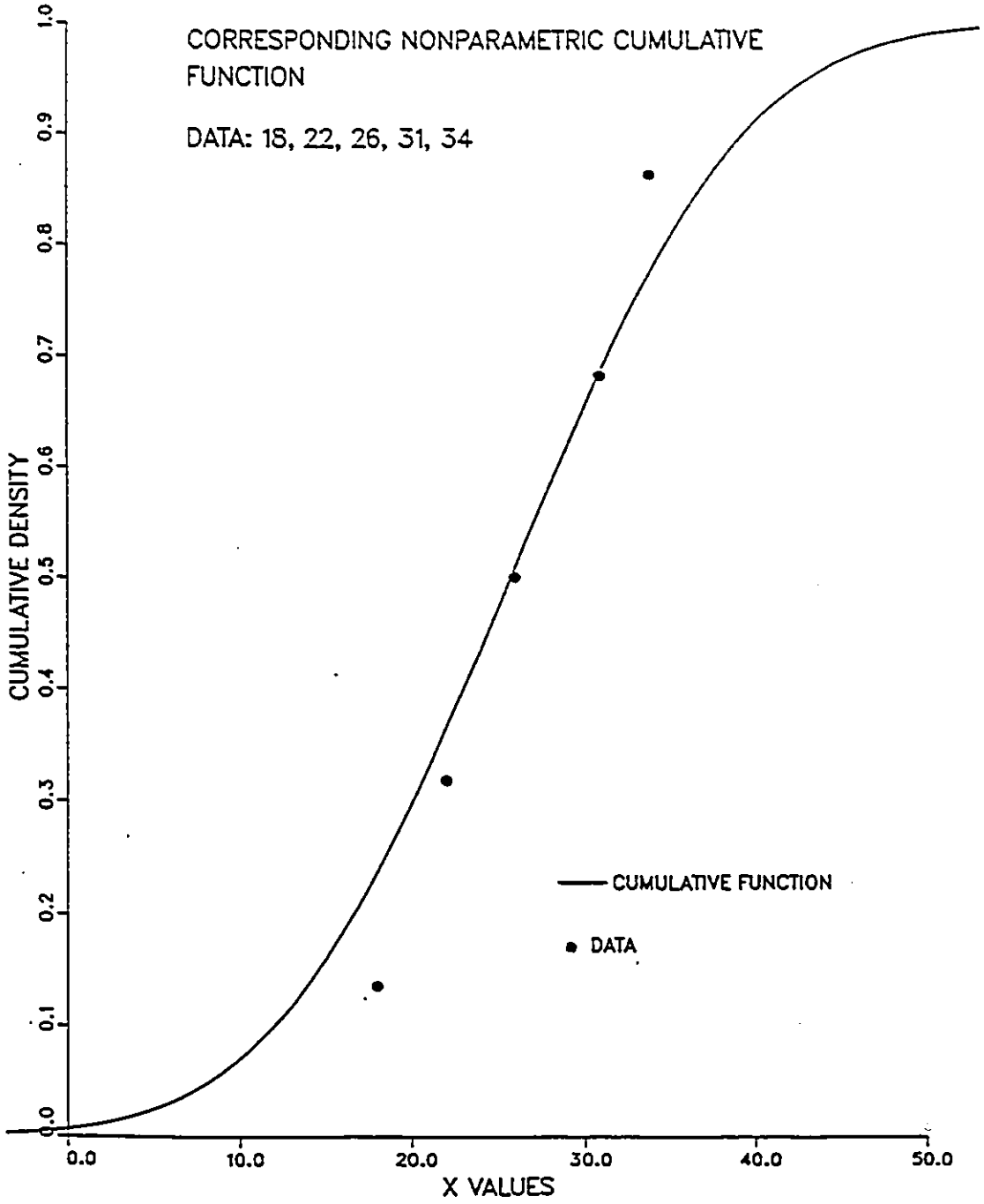


Figure 3.3: Cumulative Probability Function by the Kernel Method

where x_1 to x_n are the observations, $K()$ is the assumed kernel function which is itself a probability density function such as a normal or a rectangular distribution, and h is a smoothing factor defining the spread of the kernel. As was the case for the histogram, the total area of each individual kernel is once again $1/n$.

The kernel must satisfy the following conditions (Silverman, 1986):

$$\int K(x)dx = 1 \tag{3.3}$$

$$\int xK(x)dx = 0 \tag{3.4}$$

$$\int x^2K(x)dx = C \neq 0 \tag{3.5}$$

where C is the kernel variance. For a Gaussian kernel, a normal distribution with standard deviation equal to h is used to represent the density of every point. The normal kernel is given by:

$$K(x) = \frac{1}{\sqrt{2\pi h^2}} \quad exp(-(x - x_i)^2/2h^2) \tag{3.6}$$

One method of computing the smoothing factor h has been minimizing by a cross-validation technique the integrated mean square error (IMSE) which is expressed by (Silverman, 1986):

$$IMSE = E\ (\int (\hat{f}(x) - f(x))^2 \quad dx) \tag{3.7}$$

where $\hat{f}(x)$ is an estimate of the unknown density function $f(x)$. A few mathematical manipulations will follow in order to lead to an equation permitting the use of cross-validation. The previous equation can be expanded to:

$$IMSE = E\ (\int \hat{f(x)}^2 dx - 2\int \hat{f(x)} f(x) dx + \int f(x)^2 dx) \qquad (3.8)$$

Because IMSE must be minimized with regard to $\hat{f(x)}$, the last term, which does not involve $\hat{f(x)}$, can be discarded. Thus, the sum of the first two terms are to be minimized.

In cross-validation, a new estimate $\hat{f}_{-i}$ is constructed using all data points except x_i. Thus, $\hat{f}_{-i}(x)$ is the nonparametric kernel estimate ignoring a single data point:

$$\hat{f}_{-i}(x) = \frac{1}{(n-1)h} \sum_{j \neq i} K((x - x_j)/h) \qquad (3.9)$$

It has been shown (Silverman, 1986) that:

$$E\ \frac{1}{n} \sum_i \hat{f}_{-i}(x_i) = E\ \int \hat{f(x)} f(x) dx \qquad (3.10)$$

Replacing equation (3.10) into equation (3.8) and ignoring its last term, the risk function to be minimized, $R(h)$, which depends on the smoothing

factor h, is:

$$R(h) = E\ (\int \hat{f}(x,h)^2 dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(x_i,h)) \tag{3.11}$$

where $\hat{f}_{-i}$ is the density estimate based on the entire data set except for x_i, beginning with an assumed h. The basic principle of least squares cross-validation is to construct an estimate of $R(h)$ from the data themselves and then to minimize this estimate over h to give the smoothing factor.

After some computations, it has been shown (Rudemo, 1982) that equation (3.11), for a normal kernel as defined by equation (3.6), is equivalent to:

$$R(h) = \frac{1}{2\sqrt{\pi}nh}\ (\ 1 + \sum_{i=1, i\neq j}^{n}\sum_{j=1}^{n}\frac{2}{n}\ exp(d_{ij}/4) - \sum_{i=1, i\neq j}^{n}\sum_{j=1}^{n}\frac{2\sqrt{2}n}{n-1}\ exp(d_{ij}/2)\ ) \tag{3.12}$$

where $d_{ij} = -((x_i - x_j)/h)^2$.

By taking the derivative of the above $R(h)$ expression with respect to h and equating to 0, the following equation results:

$$\frac{1}{2\sqrt{\pi}nh} \quad (\sum_{i=1,i\neq j}^{n}\sum_{j=1}^{n} exp(d_{ij}/4) \; (\; (1 - \frac{4\sqrt{2}n}{n-1} exp(d_{ij}/4))(\frac{x_i - x_j}{h} - 1)$$
$$(\frac{x_i - x_j}{h} + 1) - 1 \;) \;) = 0. \qquad (3.13)$$

Therefore, the value of h can be determined by numerically solving equation (3.13). Scott and Terrell (1987) have shown that the cross-validation procedure leads to consistent and asymptotically optimal nonparametric density estimates.

The selection of the smoothing factor h is crucial as it defines the smoothness of the density function. Too large an h value leads to a unimodal nonparametric density regardless of the multimodality of the data, while too small an h leads to a distorted multimodal density regardless of the unimodality of the data. Figures 3.4 and 3.5 show the resulting densities when a smoothing factor too large or too small is chosen. The h values are, respectively, 3.0 and 15.0 instead of 8.964 as for Figure 3.2. The latter was obtained by the cross-validation procedure.

The kernel choice is not critical, as various kernels lead to comparable densities and result in very little change with respect to IMSE (Prakasa Rao, 1983). Some have favoured the Gumbel kernel (Bardsley, 1989) while others have favoured the Gaussian kernel (Adamowski, 1989). The latter was selected for this study.

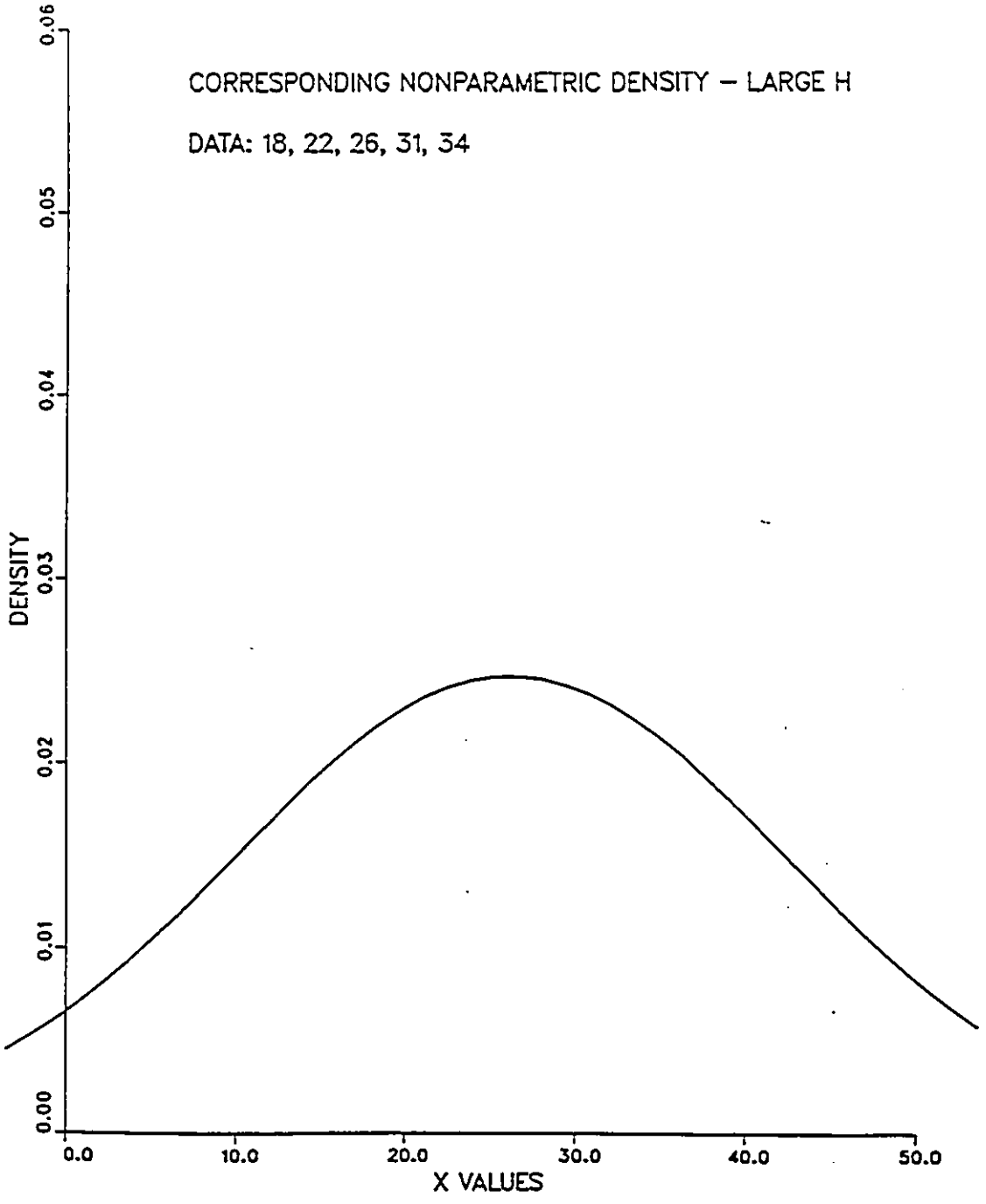


Figure 3.4: Nonparametric Density with Large H

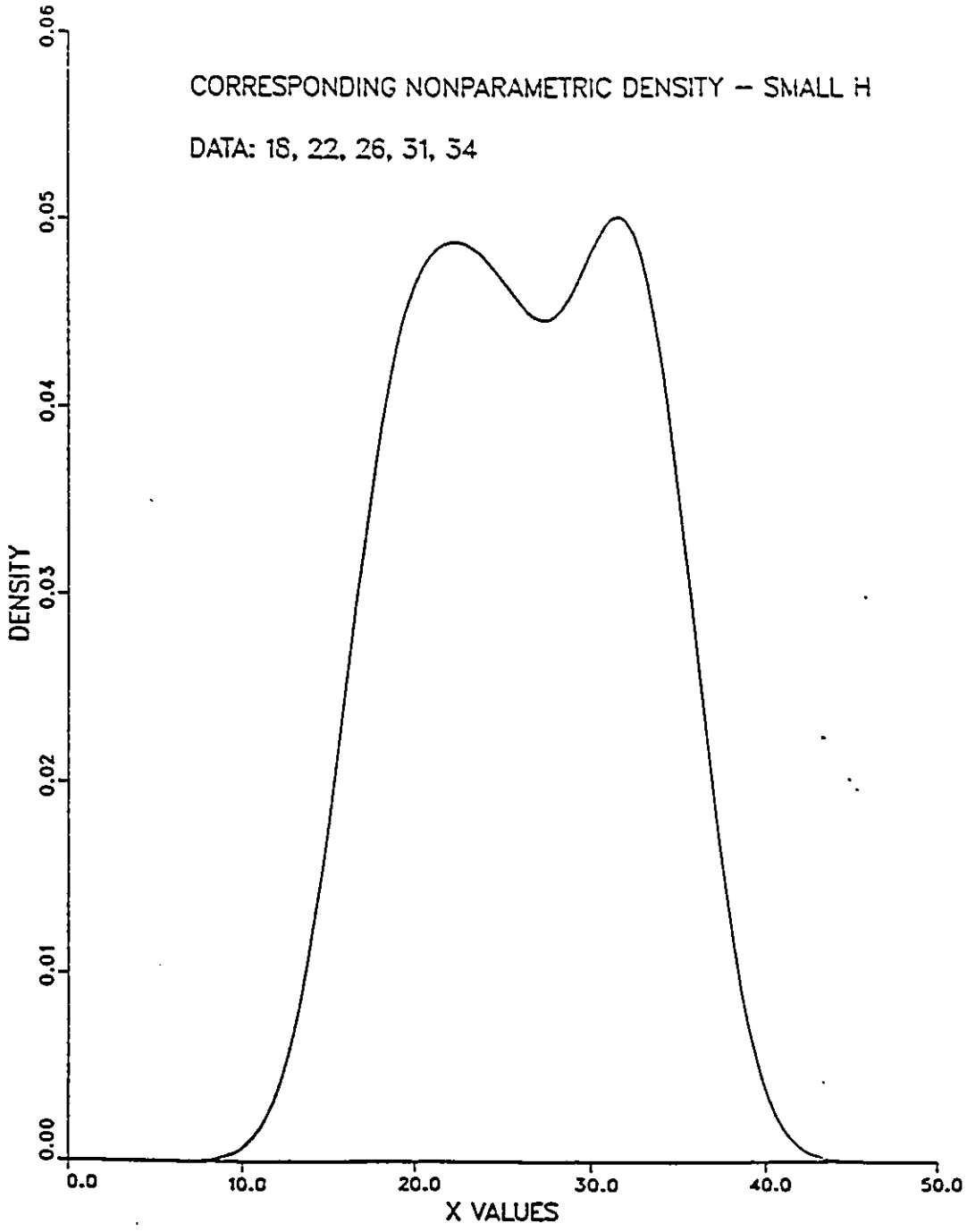


Figure 3.5: Nonparametric Density with Small H

3.2 Nonparametric Regression

In a regression of any kind, it is assumed that a random variable Y is related to a vector X of dimension p of other random variables. In order to estimate a regression relationship nonparametrically, the notion is used that the regression of Y on X is in effect a conditional mean. Therefore, the predicted value from a regression relationship can be defined as the ratio of two probability density functions, each of which is estimated nonparametrically.

The optimal estimate of a function Y given that variable X equals x can be expressed as follows (Prakasa Rao, 1983):

$$E(Y/X = x) = \frac{\int y \quad f(x,y)dy}{f_X(x)} \tag{3.14}$$

where $f(x,y)$ is the joint density function of X and Y, and $f_x()$ is the marginal density function of X. $E(Y/X = x)$ represents the optimal estimate from a regression of Y on X.

Both the numerator and denominator of the above equation can be estimated nonparametrically. Expanding the nonparametric kernel density function to the multivariate case considering p variables, then:

$$f(x) = \frac{1}{n} \sum_{i=1}^{n} \Pi_{l=1}^{p} \frac{K((x_l - x_{li})/h_l)}{h_l} \tag{3.15}$$

where K() is the kernel function, i the data counter going up to n sites, l the dimension counter going up to p variables, and h_l the smoothing factor for dimension l.

Similarly, the joint bivariate density function can be expressed nonparametrically as:

$$f(x,y) = \frac{1}{n\ h_l\ h} \sum_{i=1}^{n} K((x - x_i)/h_l, (y - y_i)/h) \tag{3.16}$$

or in multivariate form as:

$$f(x,y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^p} K((y - y_i)/h)\ \Pi_{l=1}^{p} \frac{1}{h_l} K((x_l - x_{li})/h_l) \tag{3.17}$$

In order to develop the nonparametric form of equation (3.14), a few terms must be defined and some mathematical manipulation must take place. If $K(x,y)$ is a bivariate density function, set

$$J(x) = \int_{-\infty}^{\infty} K(x,y)dy \tag{3.18}$$

Thus, the marginal density of X, $f_X(x)$, can be defined as:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \tag{3.19}$$

$$= \frac{1}{n\ h_l} \sum_{i=1}^{n} J((x - x_i)/h_l) \tag{3.20}$$

Thus, the regression estimate $E(Y/X = x)$, when substituting equations (3.16) and (3.20) into equation (3.14), becomes:

$$E(Y/X = x) = h_l \frac{\sum_{i=1}^{n} m((x - x_i)/h_l)}{\sum_{i=1}^{n} J((x - x_i)/h_l)} + \frac{\sum_{i=1}^{n} y_i\ J((x - x_i)/h_l)}{\sum_{i=1}^{n} J((x - x_i)/h_l)} \tag{3.21}$$

where

$$m(x) = \int_{-\infty}^{\infty} y\ K(x,y)dy \tag{3.22}$$

One of the basic kernel conditions, equation (3.4), is that equation (3.22) equals zero. As a consequence of $m(x) = 0$, equation (3.21) simplifies to:

$$E(Y/X = x) = \frac{\sum_{i=1}^{n} y_i\ J((x - x_i)/h_l)}{\sum_{i=1}^{n} J((x - x_i)/h_l)} \tag{3.23}$$

Replacing the $J(x)$ expression of equation (3.18) and incorporating the multivariate version of $f(x)$ from equation (3.15), then the equation for the optimal estimate from a regression of Y on X becomes (Muller, 1988):

$$E(Y/X) = \frac{\sum_{i=1}^{n} y_i \prod_{l=1}^{p} \frac{K((x_l - x_{il})/h_l)}{h_l}}{\sum_{i=1}^{n} \prod_{l=1}^{p} \frac{K((x_l - x_{il})/h_l)}{h_l}} \qquad (3.24)$$

with the variables as defined earlier.

To obtain the values of the smoothing factors h_l, least squares cross-validation is once again used. An equation defining a risk function to be minimized, similar to equation (3.12) but in a multivariate form, must be developed. Its derivative with respect to the smoothing factors must equal zero, resulting in an equation to be solved for all p values of h_l. This equation is given by (Adamowski and Feluch, 1991):

$$\frac{-n}{2} + \frac{1}{2}\sum_{i=1, i\neq j}^{n} exp(d_{ij}/4)\,\Big(\,(1 - \frac{4n\sqrt{2}}{n-1} exp(d_{ij}/4))((x_{il} - x_{jl})/h_l - 1)\,((x_{il} - x_{jl})/h_l + 1) - 1\,\Big) = 0 \qquad (3.25)$$

where $d_{ij} = -((x_{il} - x_{jl})/h_l)^2$.

A very important consideration is that the nonparametric approach does not require the assumption of linearity and of constant variance of normally distributed variations about the regression line, which are known not to apply in regional relationship development. As with nonparametric frequency, the choice of the kernel function is not crucial, but the selection of the smoothing factor h is. A Gaussian kernel was selected for this study.

3.3 Simulation Techniques

Simulation is the development and application of mathematical models in order to replicate a real-world data situation. Analysis performed on the generated data is expected to increase our knowledge of the real-world system (Fleming, 1975). Data generated from known probability functions, called Monte Carlo simulations, have been used extensively in hydrology to study the probabilistic behavior of water resources systems, particularly reservoirs (Fiering, 1967). In bootstrapping and other resampling techniques, data is simulated by selecting from an already available data base. Analysis of a large number of bootstrapped samples will yield statistical results comparable to the theoretical (Efron, 1982).

3.3.1 Monte Carlo Simulation

Monte Carlo simulation is employed to generate data from a known probability function. Simulated flow data from a given distribution can be any positive number but there exists a probability associated with each flow magnitude. Using a selected distribution and its parameters, a generating probability distribution function is first defined. Then, through a random number generating technique, a synthetic series of data is produced (Haan, 1977).

The simulated data, considered representative of a real-world situation,

then serve to enhance our understanding of various water resources systems. In this thesis, Monte Carlo simulation is used to acquire an understanding of how sampling variability will impact on density function shape if the data actually come from a unimodal distribution. By generating a number of samples from a unimodal Gumbel (or EV1) distribution and then from some bimodal distribution, a better understanding of density shape variability is achieved, as will be explained in section 4.4.

The simulation is achieved by transforming the cumulative probability function of the EV1, which is given by (Kite, 1977):

$$F(x) = exp(-exp(-x - \beta)/\alpha)) \tag{3.26}$$

where $F(x)$ is the cumulative probability with α and β as parameters, into the following:

$$x = \beta - \alpha \ ln(-ln(F(x))) \tag{3.27}$$

By generating a series of uniformly distributed random numbers between 0 and 1, which correspond to the cumulative function $F(x)$, a series of flow values are obtained from equation (3.27). The flow series is representative of streamflows following an EV1 distribution with parameters β and α of values as chosen in the generating process.

Simulation is also used in this thesis to compare the performance of parametric and nonparametric regression. As will be explained in more detail in section 5.3, data will be generated from known linear and slightly nonlinear regional relationships to which will be added a random component greater than that observed in New Brunswick. Following analysis of the generated data by both parametric and nonparametric regressions, the usefulness of a given statistical comparison criterion, the Integral Square Error (ISE), to be defined later, will be evaluated.

Without the simulation experiments, the conclusions reached about density shape variability and about the use of the ISE criterion would have been impossible.

3.3.2 Bootstrapping

Bootstrapping is one of a number of resampling techniques and their variations which can be used to construct confidence intervals (Efron, 1982). Given an observed data sample X containing observations $x_1, x_2, ...x_n$, a new sample is created, called X_i^*, of size n drawn with replacement from the n observed values. It has been shown that the X_i^*'s have the same mean, the same variance and the same density function as those of the original n observations.

Thus, by simulating a large number of samples of size n drawn with replacement from the original data, a process known as bootstrapping, var-

ious symmetrically defined statistics of an unknown population from which a sample is available can be estimated. A comparison of bootstrapped and theoretical estimates of various symmetrical population characteristics shows identical results as long as a sufficiently large number of bootstrapped samples are taken. Thus, computational power is in effect substituted for theoretical analysis. However, for complex data structures, such as nonparametric statistical problems, it may be the only avenue (Diaconis and Efron, 1983).

Bootstrapping of (x, y) pairs will be employed in section 5.4 in order to investigate a new approach in confidence interval determination for nonparametric regression. This will allow a comparison between parametric and nonparametric regressions more effective than visual comparison or the ISE criterion.

3.4 L-Moment Analysis

A recent and very promising technique in parametric flood frequency analysis is L-moment analysis. L-moments, analogous to conventional moments, are the expectations of certain linear combinations of order statistics. They have the theoretical advantages over conventional moments of being able to characterize a wider range of distributions, of being virtually bias free even for short samples and of being more robust to the presence of outliers in the data (Hosking, 1990). In mathematical terms, L-moments are defined

as:

$$\lambda_r = \int_0^1 x(F) \;\; P_{r-1}(F) \;\; dF \tag{3.28}$$

where

$$P_r(F) = \sum_{k=0}^{r} (-1)^{r-k} \left(\begin{array}{c} r \\ k \end{array} \right) \left(\begin{array}{c} r+k \\ k \end{array} \right) F^k \tag{3.29}$$

Here, λ_r is the rth L-moment, x is a random variable and $F(x)$ its cumulative distribution function. Specific relationships between λ_r and distribution parameters have been derived for a number of well-known distributions (Hosking, 1990).

It is often convenient to standardize L-moments of order 3 or greater so that they are independent of the units of measurement of x. The rth L-moment ratio is thus defined as:

$$\tau_r = \lambda_r / \lambda_2 \tag{3.30}$$

While λ_1 and $\lambda 2$ are considered measures of location and scale, τ_3 and τ_4 are regarded as measures of skewness and kurtosis. For this reason, they are also known as the coefficients of L-skewness and L-kurtosis. The ratio of λ_2 to λ_1 is known as the L-coefficient of variation, LCV.

For various distributions, the relationship between τ_3 and τ_4 have been determined and are given in Table 3.1 (Hosking, 1990). The corresponding graphical relationship, called the L-moment ratio diagram, is shown in Figure 3.6. For the compounded TCEV, only certain combinations of L-skewness and L-kurtosis will yield this distribution (Gabriele and Arnell, 1991). Figure 3.7 shows the feasible space of the TCEV.

In numerical calculations, U-statistics are used in order to estimate the L-moments from a random sample drawn from an unknown distribution. If $x_1, x_2 \ldots x_n$ is the sample and $x_{1:n} \leq x_{2:n} \leq \ldots \leq x_{n:n}$ the ordered sample, it can be shown that the rth sample L-moment l_r can be estimated from (Hosking, 1990):

$$l_r = \sum_{i=1}^{n} P_{r-1}(p_{i:n}) \quad x_{i:n} \tag{3.31}$$

where $p_{i:n}$ is the cumulative probability as expressed by an empirical plotting formula, such as this one deemed appropriate by Hosking (1990):

$$p_{i:n} = (i - 0.35)/n \tag{3.32}$$

where i is the rank and n the sample size.

In practice, a series of observed samples is available which is drawn from an unknown distribution. By computing τ_3 and τ_4 for each of the

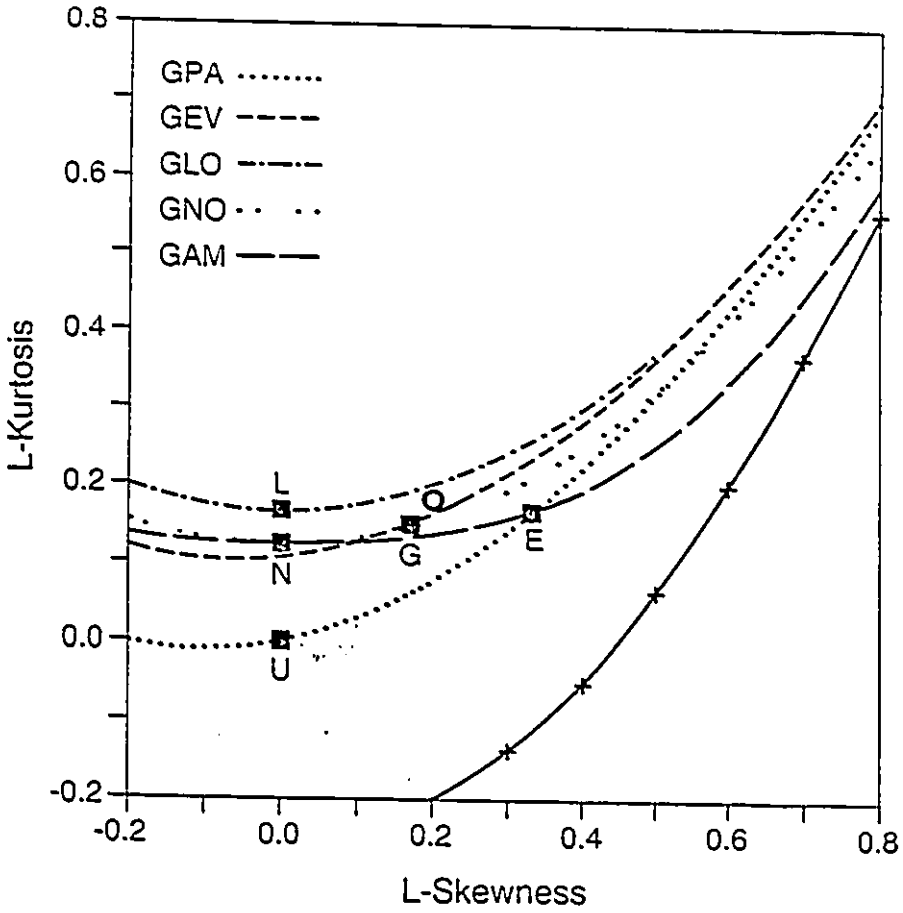
Table 3.1: L-Moment Relationships for Some Common Distributions

Distribution	τ_3	τ_4
Uniform	0	0
Normal	0	0.1226
Exponential	0.3333	0.1667
Gumbel	0.1699	0.1504
Generalized Logistic*	$-k$	$(1+5k^2)/6$
Generalized Extreme Value**	$\frac{2(1-3^{-k})}{1-2^{-k}} - 3$	$\frac{1-6.2^{-k}+10.3^{-k}-5.4^{-k}}{1-2^{-k}}$

$$* \quad x = c + \tfrac{a}{k}(1 - ((1-F)/F)^k)$$

$$** \quad x = c + \tfrac{a}{k}(1 - (-lnF)^k)$$

where a, c and k are parameters while F is the cumulative probability function.

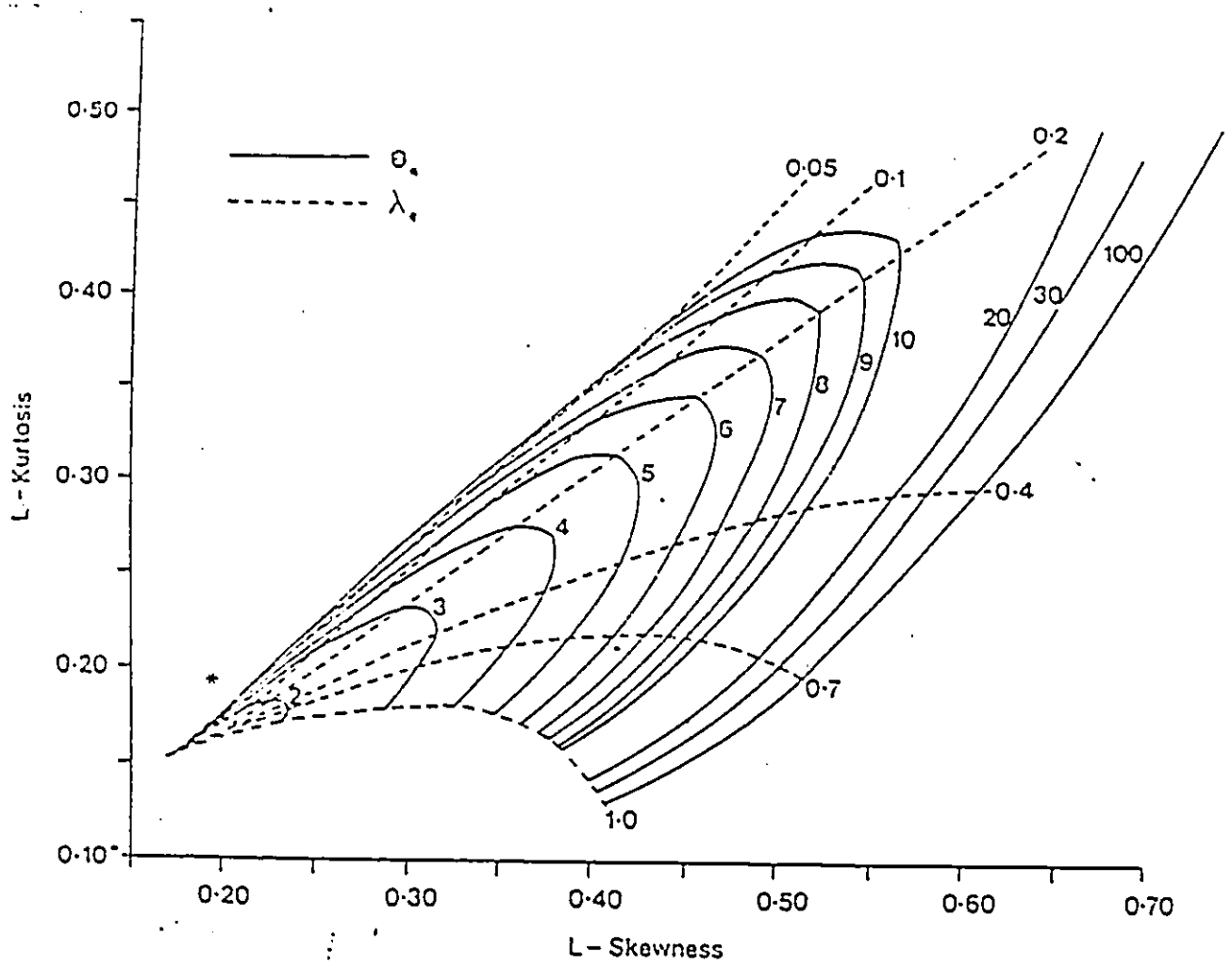


L-moment ratios of some common distributions for the values of usual interest E = exponential, G = Gumbel, L = logistic, U = uniform, GPA = generalized Pareto, GEV = generalized extreme value, GLO = generalized logistic, GNO = generalized normal, GAM = gamma, while vertical crosses are the lower bound for all other distributions.

O regional mean for New Brunswick floods

taken from Wallis (1989) and Hosking (1990)

Figure 3.6: Theoretical L-Moment Diagram



* weighted regional mean for New Brunswick floods

taken from Gabriele and Arnell (1991)

Figure 3.7: TCEV Feasible Space

samples and comparing these values with the theoretical relationships of the well-known distributions, the distribution from which the samples were presumably drawn is selected. L-moments have been shown to be superior to conventional moments in distribution identification (Hosking, 1990).

The L-moment ratio diagram of Figure 3.6 is used for this undertaking. The entire series of τ_3 and τ_4, along with the weighted values of L-skewness and L-kurtosis, are plotted on the diagram. The weighted regional statistic θ_k is computed from (Hosking and Wallis, 1990):

$$\theta_k = \sum_{i=1}^{N} n_i \theta_{ki} / \sum_{i=1}^{N} n_i \tag{3.33}$$

where N is the number of series of samples, n_i the sample length at site i, and θ_{ki} is the L-statistic estimated at site i based on n_i observations. From such a plot, a distribution is selected on the basis of proximity to a theoretical distribution curve and spread of the data around the weighted mean.

In order to test the homogeneity of a data set, in other words to test that the data come from the same probability distribution, three heterogeneity measures denoted by V_1, V_2 and V_3 have been developed by Hosking and Wallis (1991). They are computed as follows:

$$V_1 = \sum_{i=1}^{N} n_i (LCV_i - \bar{LCV})^2 / \sum_{i=1}^{N} n_i \tag{3.34}$$

$$V_2 = \sum_{i=1}^{N} n_i((LCV_i - \overline{LCV})^2 + (\tau_{3i} - \bar{\tau}_3)^2)^{0.5} / \sum_{i=1}^{N} n_i \qquad (3.35)$$

$$V_3 = \sum_{i=1}^{N} n_i((\tau_{3i} - \bar{\tau}_3)^2 + (\tau_{4i} - \bar{\tau}_4)^2)^{0.5} / \sum_{i=1}^{N} n_i \qquad (3.36)$$

Using weighted regional values of the L-moments computed from the observed data, a four-parameter kappa distribution is used to generate a large number of samples by Monte Carlo simulation. From these simulations, the mean μ_v and standard deviation σ_v of the three statistics V_1, V_2 and V_3 are computed. These represent the mean and standard deviation of the measures from a homogeneous population with the average characteristics of the observed data.

The actual values of V_1, V_2 and V_3 using the observed data are found and then the ratio H is computed as:

$$H = (V - \mu_v)/\sigma_v \qquad (3.37)$$

Hosking and Wallis (1991) suggest that a value of H less than 1 means a region is "acceptably homogeneous", a ratio between 1 and 2 means "possibly heterogeneous", while a ratio above 2 means "definitely heterogeneous". A computer package computing these ratios also uses the simulation information for goodness of fit purposes, but it will only test the adequacy of certain unimodal distributions.

L-moment analysis and the heterogeneity measures will be used in section 4.5 in order to compare parametric and nonparametric frequency analysis.

Chapter 4

NUMERICAL ANALYSIS

4.1 New Brunswick Hydrometric Network

As of 1988, the hydrometric network of the province of New Brunswick was composed of 85 active stations (Inland Waters Directorate, Environment Canada and New Brunswick Department of Municipal Affairs and Environment, 1989). Many of these stations are on heavily regulated streams so that the measured flows do not correspond in any way to natural flow conditions. Other stations were begun recently and have an insufficient quantity of data for a proper flood frequency analysis. Still other stations were on the Saint John River, a basin of overwhelming size in comparison to the remainder of the watersheds in the province.

As the result of preliminary data screening, 53 natural flow hydrometric stations with at least 10 years of record in New Brunswick, or in the surrounding regions of Quebec, Maine and Nova Scotia, were selected for regional analysis. These include a few stations with regulation having an insufficient impact on the accuracy of high flows. A list of these 53 stations along with other information is provided in Appendix A. Figure 4.1 shows the station locations.

A previous regional study for New Brunswick mentioned earlier (Inland Waters/Lands and New Brunswick Department of Municipal Affairs and Environment, 1987) divided the province into the four homogeneous regions shown in Figure 4.2 plus the Saint John River main stem. A list of the stations in each region is in Table 4.1. Their analysis was based on 37 stations with a concurrent period of record from 1970 to 1985, employing only parametric methods. The present analysis, also using New Brunswick data, will investigate the use of both parametric and nonparametric methods.

4.2 Climate of New Brunswick

The climate of New Brunswick is very variable throughout the province (Inland Waters/Lands and New Brunswick Department of Municipal Affairs and Environment, 1987). As shown by Figure 4.3, the province is divided into four climatic regions: the Bay of Fundy, Eastern New Brunswick, the New Brunswick Lowlands and the New Brunswick Highlands, correspond-

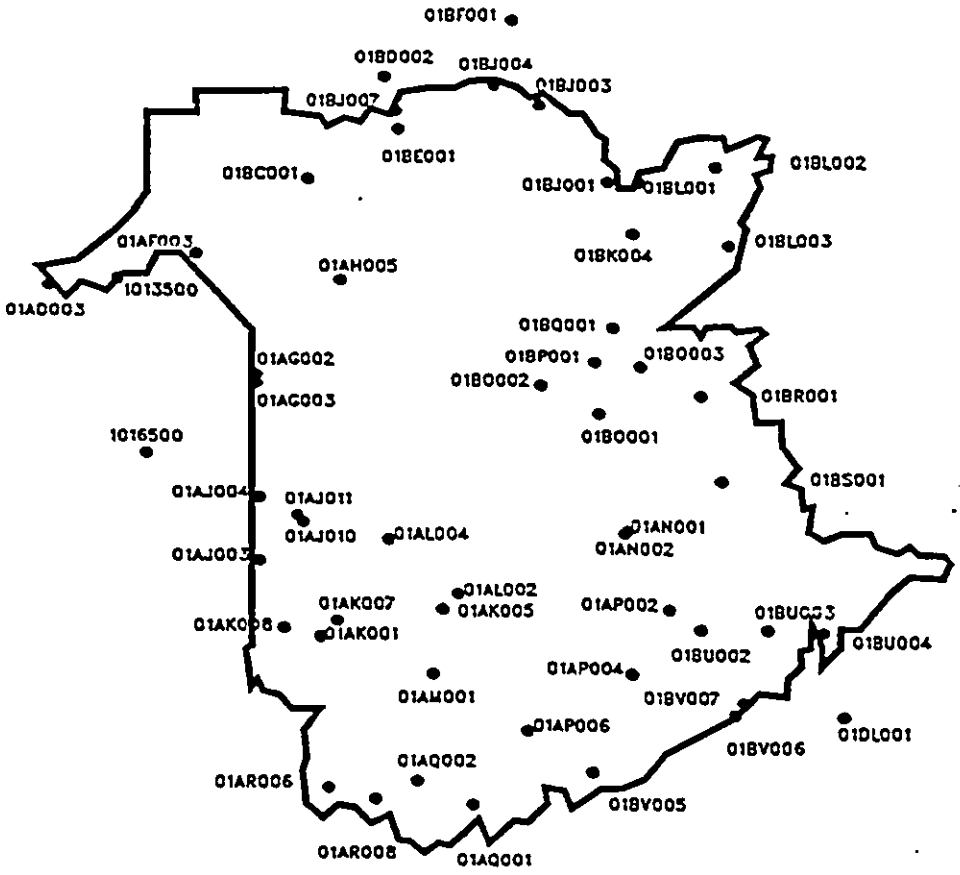
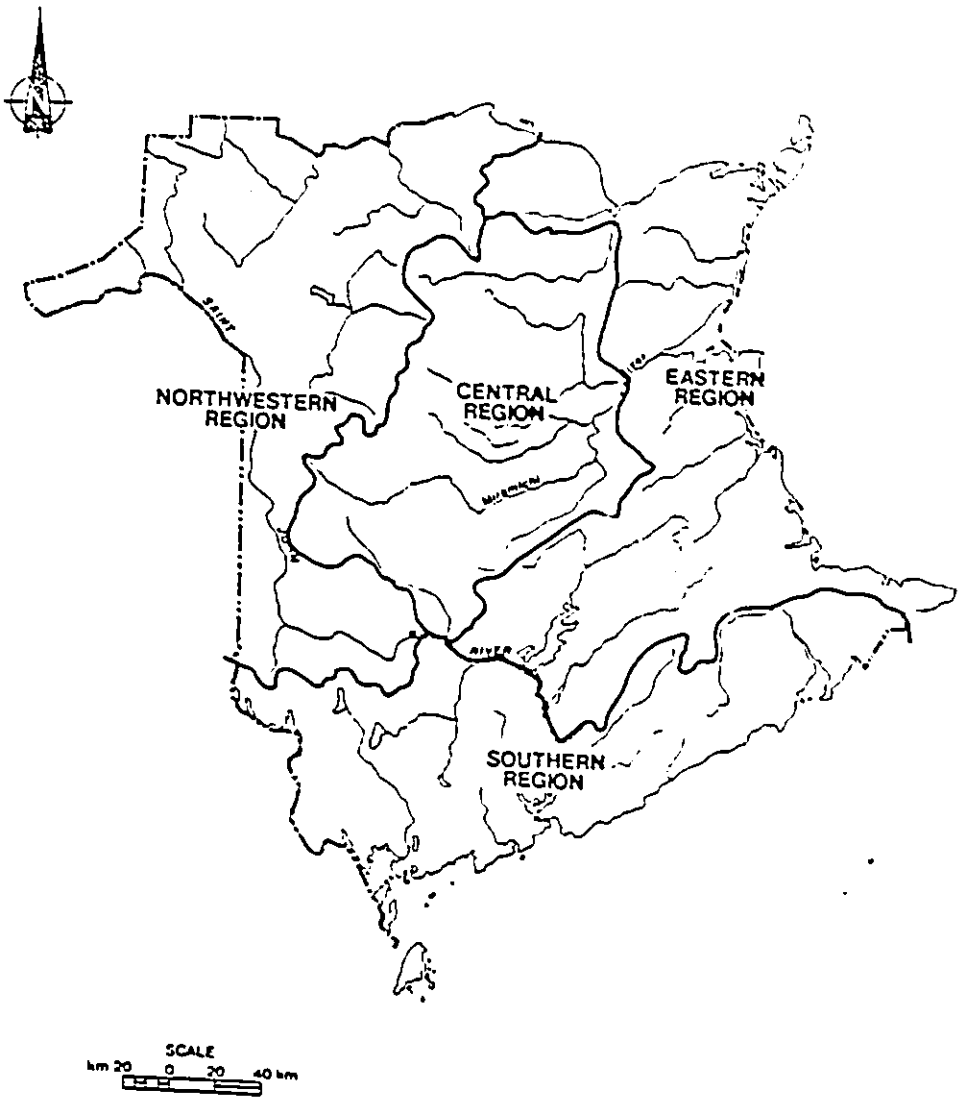


Figure 4.1: New Brunswick Hydrometric Stations



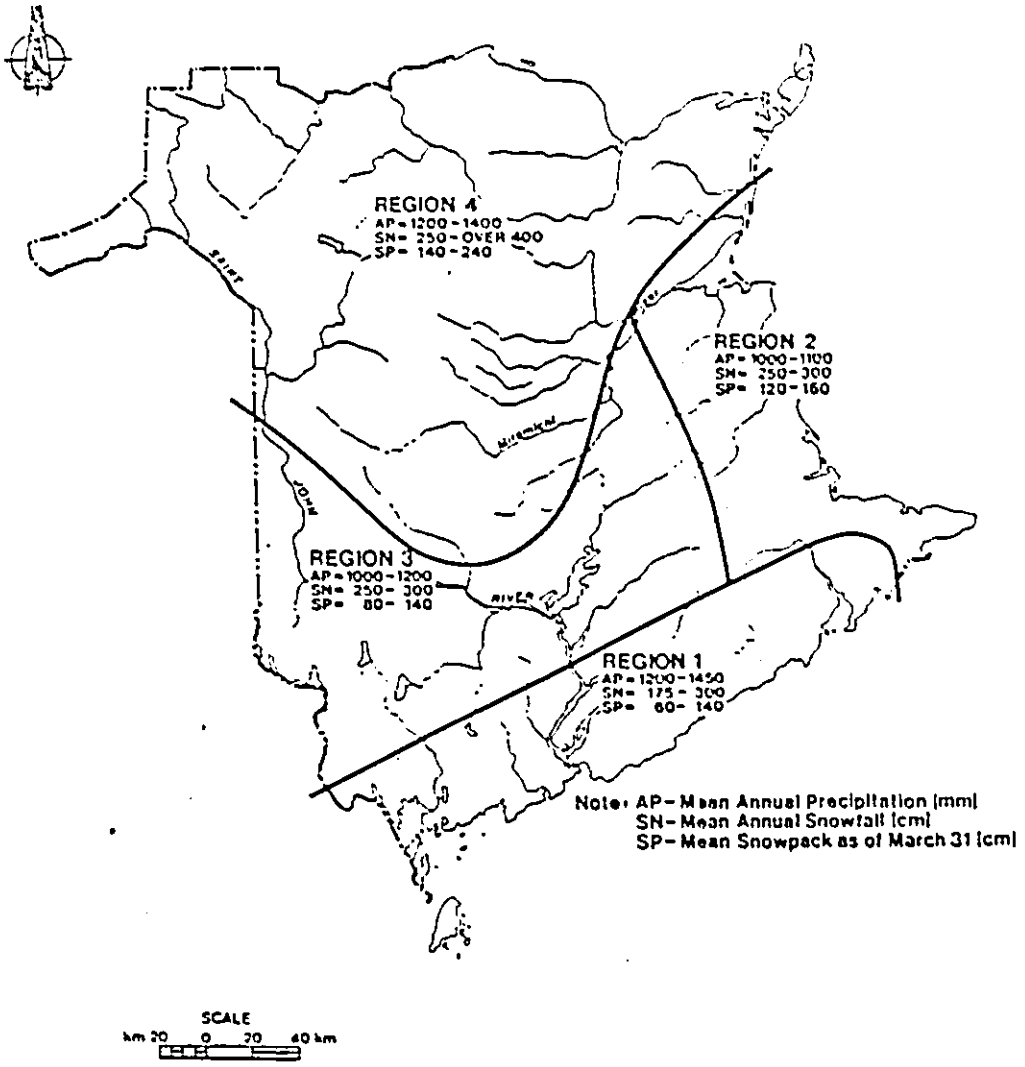
taken from Inland Waters/Lands and New Brunswick Department of Municipal Affairs and Environment, 1988b.

Figure 4.2: New Brunswick Homogeneous Flood Regions

Table 4.1: Stations Belonging to Homogeneous Regions

EASTERN	01AN001	01AN002	01AP002
.	01BJ001*	01BJ003*	01BL001
.	01BL002	01BL003	01BO003
.	01BP001	01BS001	01BU002
CENTRAL	01AJ010	01AJ011	01AK005
.	01AL002	01AL004	01BK004
.	01BO001	01BO002	01BP001
.	01BQ001		
SOUTHERN	01AM001*	01AP004	01AP006
.	01AQ001*	01AQ002*	01AR006*
.	01AR008*	01BU003	01BU004
.	01BV005	01BV006	01BV007
.	01DL001		
NORTHWESTERN	01AD003	01AF003	01AG002
.	01AG003	01AH005	01AJ003
.	01AJ004	01AK001	01AK007
.	01AK008	01BC001	01BE001*
.	01BJ004*	01BJ007	01BD002
.	01BF001	1013500	1016500

* - These stations do not belong to the modified regions, as explained in Section 5.2.



taken from Inland Waters/Lands and New Brunswick Department of Municipal Affairs and Environment, 1988a.

Figure 4.3: Climate Regions of New Brunswick

ing to numbers 1 to 4 on Figure 4.3. The Bay of Fundy region (Region 1) has high precipitation, which is in part due to the region's proximity to the major storm tracks on the east coast of North America, and low snow accumulation. Dying tropical storms will on occasion affect this area. Eastern New Brunswick (Region 2) and the New Brunswick Lowlands (Region 3) are areas of low precipitation and moderate snow accumulation.

The New Brunswick Highlands (Region 4), which cover roughly half of the province, form an area of moderate to high snow accumulation. The climate is actually very variable throughout Region 4 as the wide range of snow parameters on Figure 4.3 indicate. The most northern part of this area is subject to large quantities of snowfall and thus of very large snow accumulations in the spring.

Figure 4.4 is a map of average water content of snowpack on March 31 as simplified from a much more detailed map (Inland Waters/Lands and New Brunswick Department of Municipal Affairs and Environment, 1987). It can be observed that there is a climatologically different region in the north with average snowpack above 200 mm as well as another climatologically different area in the southwest with average snowpack below 120 mm.

Table 4.2 shows a partitioning of the New Brunswick annual maximum floods by seasons. The terms bimodal, unimodal and heavy-tailed in the table refer to the homogeneous regions to be delineated in the next chapter. It is observed that in the northern part of the province, the majority of floods occur in the spring, from snowmelt. As one goes southward, a

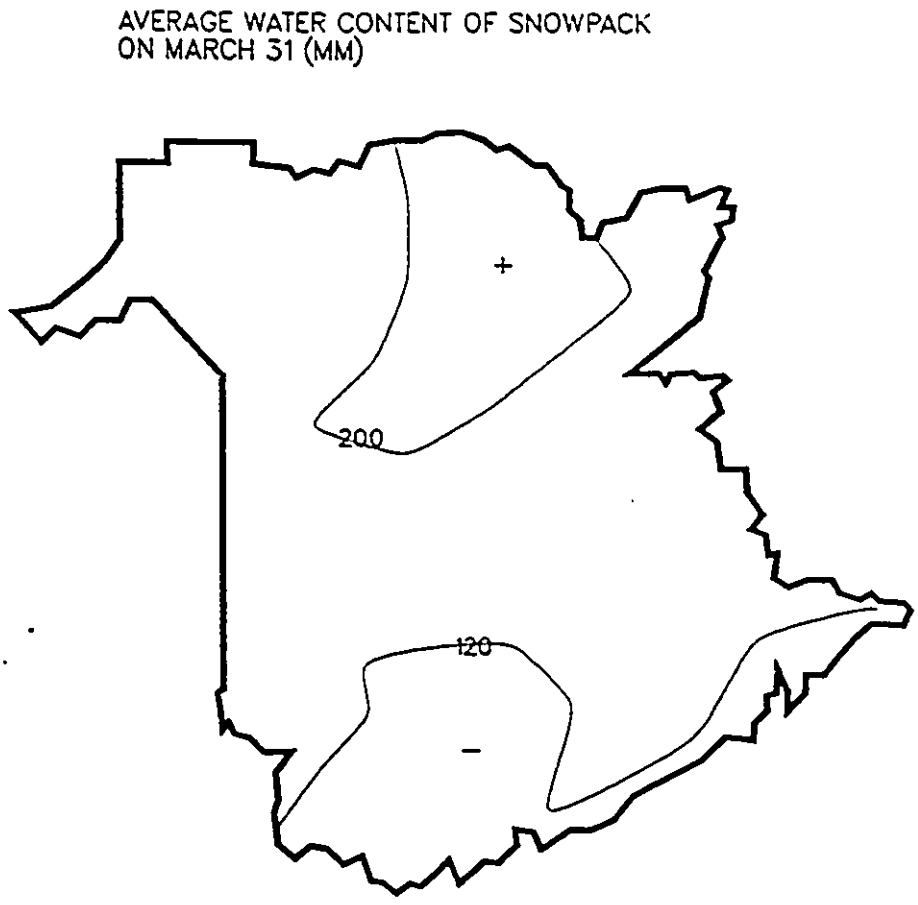


Figure 4.4: Simplified Snowpack Map of New Brunswick

Table 4.2: Seasonal Partitioning of Floods

Station Name	Partitioning Sp	Su	F	W	Station Name	Partitioning Sp	Su	F	W
Bimodal	.	.	.	.	01BO002	19	1	2	1
01AD003	36	1	0	0	01BO003	15	0	1	0
01AF003	25	0	0	0	01BP001	30	1	3	3
01AG002	17	1	2	1	01BQ001	24	0	3	0
01AG003	13	1	0	0	01BR001	19	0	1	1
01AH005	13	2	1	0	01BS001	19	2	2	1
01AJ003	18	0	2	1	01BU002	18	2	5	2
01AJ004	20	0	1	0	01BU003	15	0	7	4
01AJ010	13	0	1	1	01BU004	7	2	6	4
01AJ011	12	0	2	1	01BV005	3	0	3	5
01AK001	60	1	3	3	01BV006	9	1	10	4
01AK005	16	0	5	2	01BV007	5	0	3	3
01AK007	18	0	1	2	1013500	59	0	1	0
01AK008	14	0	1	0	1016500	30	1	1	0
01AL002	23	0	3	1	01BD002	15	0	0	0
01AL004	16	0	1	0	01BF001	25	0	0	0
01AN001	11	0	4	1	01DL001	9	2	5	3
01AN002	13	0	1	1	Unimodal				
01AP002	29	3	5	4	01BE001	53	3	3	0
01AP004	14	0	2	10	01BJ001	43	0	4	0
01AP006	9	0	3	1	01BJ003	23	0	1	0
01BC001	26	0	0	0	01BJ004	15	0	1	0
01BJ007	20	0	0	0	Heavy-Tailed				
01BK004	16	1	0	0	01AM001	18	0	6	2
01BL001	20	0	2	1	01AQ001	32	4	23	1
01BL002	16	0	2	1	01AQ002	44	0	10	5
01BL003	17	0	1	0	01AR006	14	0	2	6
01BO001	35	0	3	3	01AR008	4	0	4	5

Abbreviations used:
Sp: Spring - March 21 to June 21
Su: Summer - June 21 to September 21
F : Fall - September 21 to December 21
W : Winter - December 21 to March 21

greater proportion of floods come in the summer, fall or winter. In the southernmost parts of the province, spring floods may account for less than fifty percent of all annual maximum floods in some instances. Dying tropical storms also affect the southern coast on occasion. Therefore, different mechanisms are at play in different parts of the province; rain on snow is the dominant factor in some areas of the north while in the remainder of the province a combination of rain on snow and rainfall only floods occur.

An understanding of the climate of the various parts of New Brunswick is very important as this greatly influences the flood frequencies along with the probability density function shapes found throughout the province.

4.3 Single Station Frequencies

In a previous regional study of New Brunswick floods (Inland Waters/Lands and New Brunswick Department of Municipal Affairs and Environment, 1987), the coefficients of skew and kurtosis of the logarithmically transformed data as well as a visual fit of the cumulative frequency distributions from various distributions to the data were the determining factors in distribution selection. The three-parameter lognormal (LN3) distribution was the chosen at-site distribution, except for stations 01AR006 and 01AM001 where the Wakeby was used because the LN3 fit was considered extremely poor. In the present study, for comparative purposes, the LN3 was fit to all 53 stations with the results for the 2, 10, 20, 50 and 100 year return

period floods as tabulated in Table B.1 in the Appendix.

Nonparametric frequency using a fixed Gaussian kernel was also used for the same return periods, with the results of Table B.2 in the Appendix. Figure 4.5 shows the resulting parametric and nonparametric cumulative frequency functions for station 01AQ001 on the Lepreau River for comparison purposes. It is important to note how different the floods of return periods greater than ten years are in both methods.

It is observed that for low return periods such as the 2 year flood, the New Brunswick annual maximum flood estimates from both approaches are similar. However, for large return periods, especially at the 100 year flood level, the nonparametric frequency estimates tend to be less. Of the 53 nonparametric estimates, 33 were less than the parametric for the 100 year flood, 19 were greater and one was identical. As mentioned earlier, simulation studies (Adamowski, 1989) showed that nonparametric frequency provides less biased estimates for large return period flows; thus, it is concluded that parametric frequency is more likely to yield an overestimate of large return period floods in New Brunswick.

The probability density functions obtained by nonparametric frequency analysis were then plotted. All 53 plots are in Appendix A. Three of them, respectively for stations 01AP002, 01BE001 and 01AQ001, are reproduced in Figures 4.6 to 4.8, as examples of those encountered in New Brunswick. Many stations have a bimodal distribution, others a unimodal one and a few others a heavy-tailed distribution.

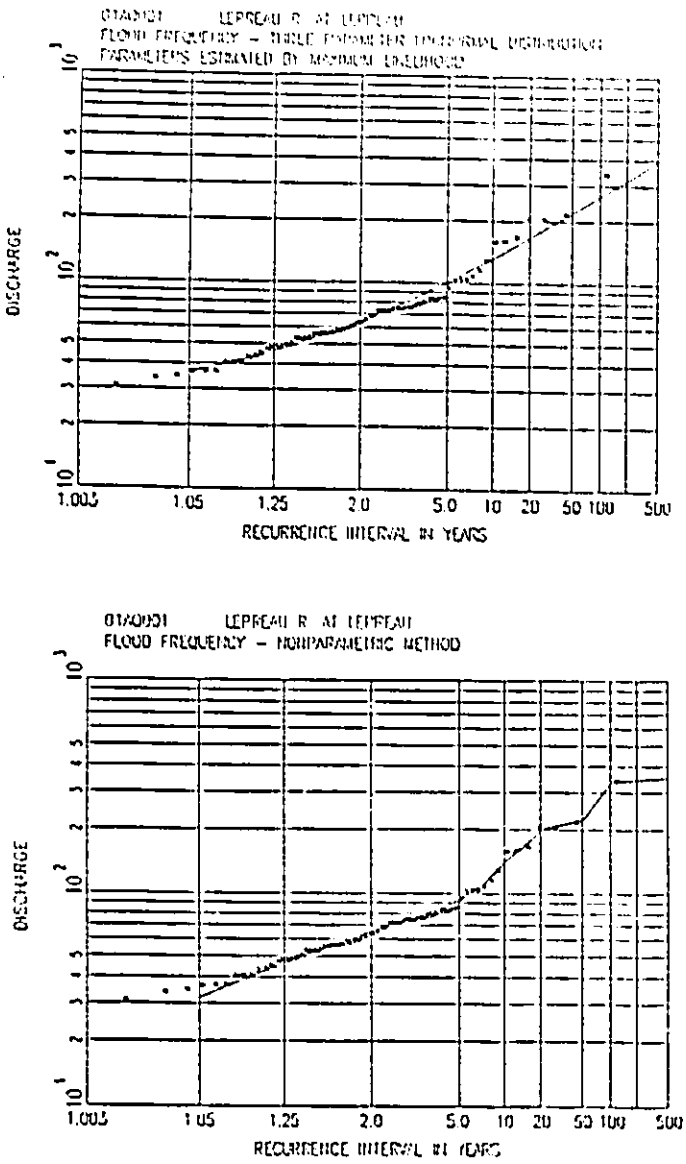


Figure 4.5: Parametric and Nonparametric Frequencies for Lepreau River

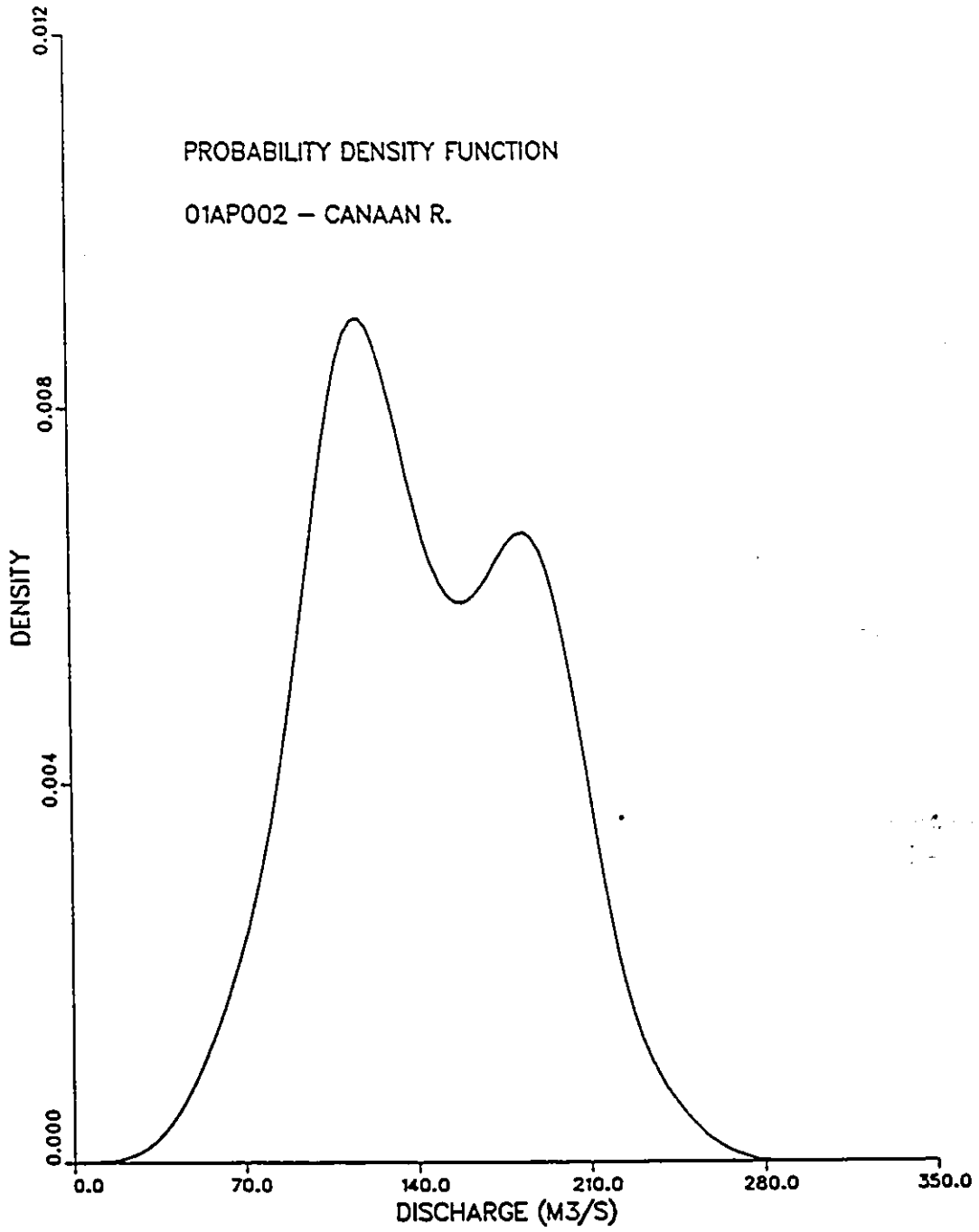


Figure 4.6: Probability Density Function for Canaan River

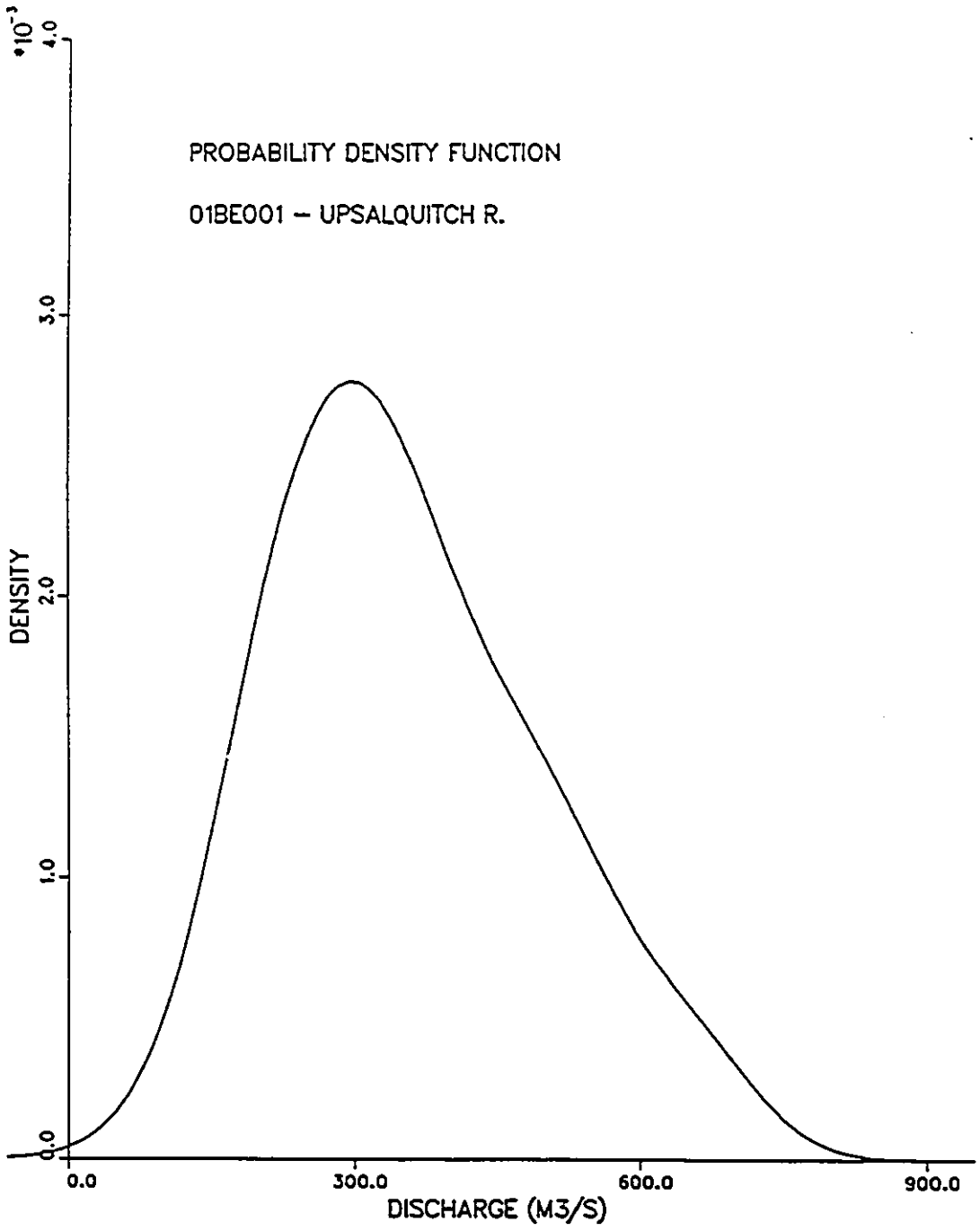


Figure 4.7: Probability Density Function for Upsalquitch River

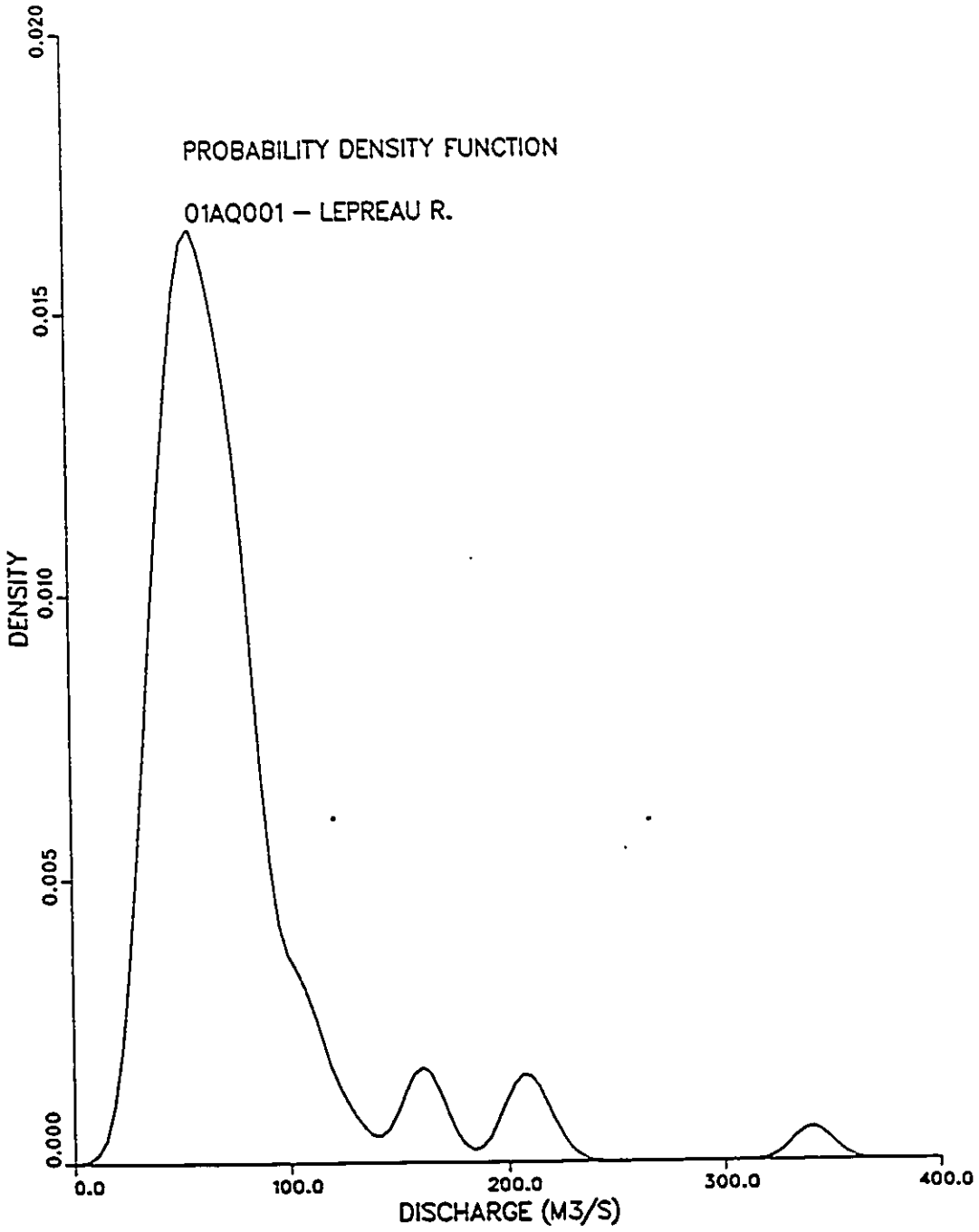


Figure 4.8: Probability Density Function for Lepreau River

As will be explained in the next chapter, basins with annual flood distributions of similar shapes are located near each other. This will be used to delineate homogeneous regions, even though some subjectivity is involved. These groupings of similarly shaped densities occur because the different shapes reflect different flood generation mechanisms, as confirmed by the seasonal partitioning of Table 4.2.

Because the heavy tail of Figure 4.8 contained an unrealistic gap, the question arose whether poor quality streamflow data, from the early part of the record, were the cause of the gap. A density function using only the last 30 years of record instead of the full 70 years has a smoother, but nonetheless heavy, tail as shown in Figure 4.9. With the availability of a much longer record, the unrealistic gap in the heavy tail, which causes no problem when exceedance probability is estimated, is expected to disappear.

4.4 Simulation Studies

Since the large grouping of bimodal densities found in New Brunswick had never been reported before, it was decided, through simulation studies, to investigate whether their cause might simply be due to sampling variability. Monte Carlo type simulations were therefore undertaken in order to show that nonparametric frequency would not indicate a bimodal distribution when the data actually came from a unimodal distribution, and

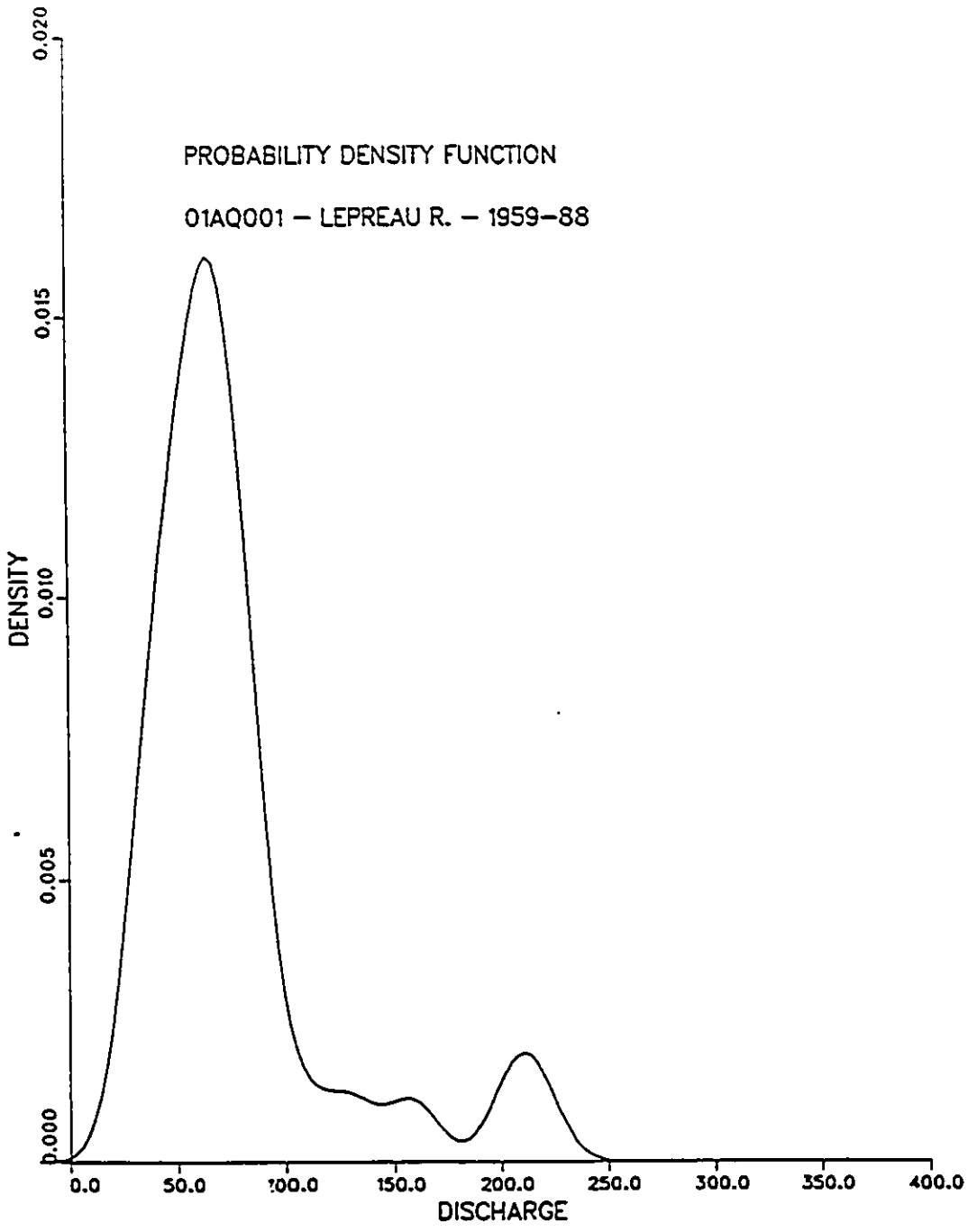


Figure 4.9: Probability Density Function for Lepreau R. - 1959-88

that nonparametric frequency could also accurately depict a bimodal distribution. It should be emphasized that the purpose of simulation was to verify whether nonparametric methods would properly identify unimodal or bimodal densities if present in the data. No attempt is made here to simulate the existing flood characteristics of the data being analyzed in New Brunswick.

One hundred sets of 24 floods were generated from a unimodal Extreme Value Type I (EV1) or Gumbel distribution with cumulative probability function as given by equation (3.26). The EV1 distribution was arbitrarily chosen because it is a well-known unimodal distribution and its generating function is fairly simple. Parameters α and β were arbitrarily chosen to be, respectively, equal to 400 and 1000, which could be realistic values for a given New Brunswick basin. A sample size of 24 was selected because this is roughly the median record length of available New Brunswick annual maximum floods.

All one hundred sets of 24 simulated floods were fit nonparametrically by kernel function and had their densities plotted. It was found that out of one hundred simulated sets, 49 were smoothly unimodal, 32 had some distortion in the tail of the type shown in Figure 4.10 for simulation 10, and 19 were multimodal as shown in Figure 4.10 by simulation 9.

A bimodal distribution with a very high second peak, as found for some New Brunswick floods, and as shown in Figure 4.6, was never encountered. It therefore appeared that these could not be due to sampling variability. It

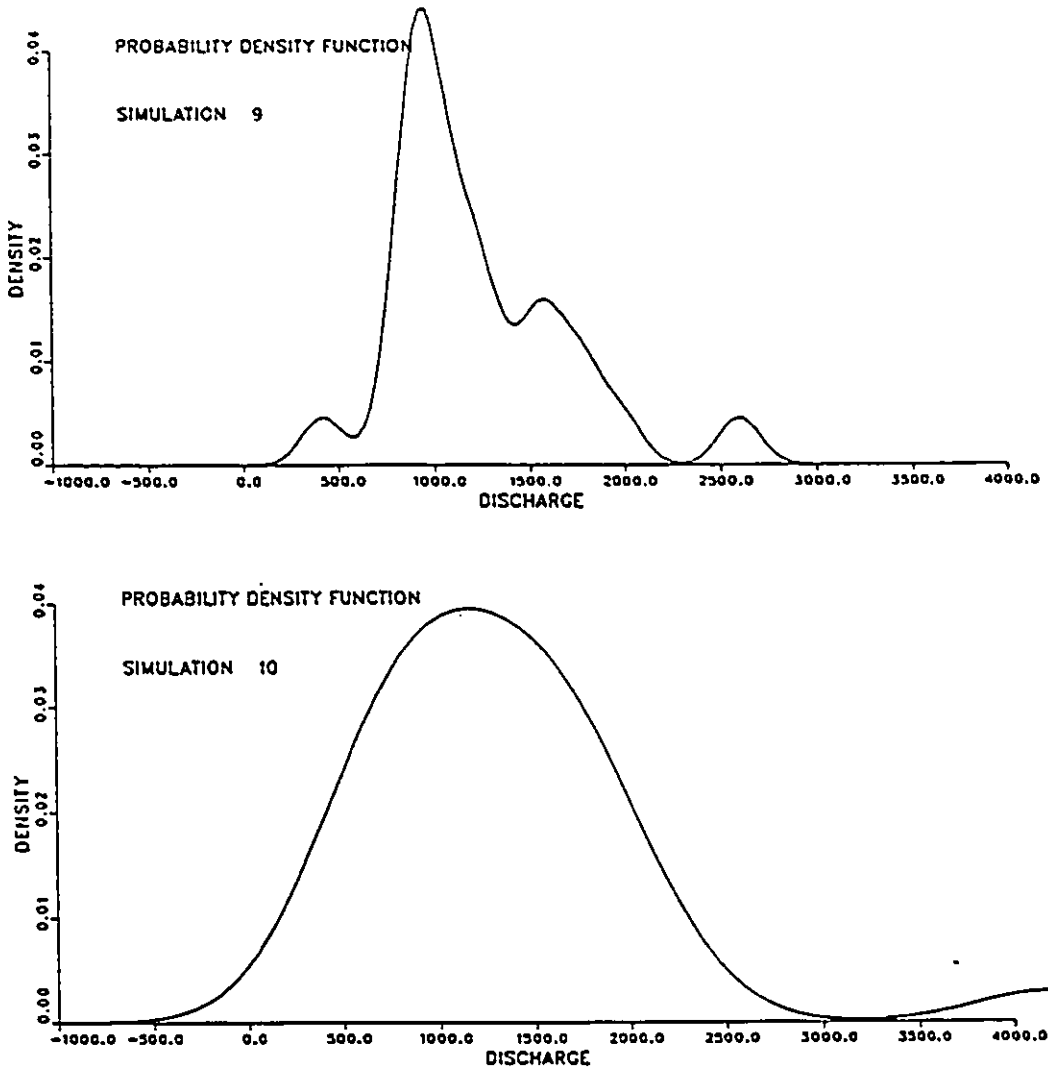


Figure 4.10: Examples of Densities from Unimodal Simulation

could be argued that some of the New Brunswick bimodal distributions with a low second peak were due to sampling variability; however, their repeated occurrences could not be accounted for solely by sampling variability. As well, out of the one hundred simulations, only six had a distortion in the tail in any way similar to the ones seen in Figures 4.8 and 4.9.

Since the question of the impact of record length on density shape variability arose, the previous simulation was repeated but using 70 years of data instead of 24. A period of 70 years was chosen because one of the stations with a heavy tail, 01AQ001, had 70 years of record. Out of the 100 resulting plots, 42 were smoothly unimodal, 32 had some distortion and 26 were multimodal. It thus appeared that sample length had no or a minimal influence on shape variability for the lengths of record available and employed in this study.

Then, one hundred sets of 24 floods were generated from a bimodal distribution consisting of 16 values from an EV1 with β equal to 1000 and α equal to 400, and 8 values from an EV1 with β equal to 2000 and α equal to 400. The plots revealed 48 unimodals, 15 bimodals with a clear second peak, 15 with some tail distortion and 22 multimodals. Because the two peaks were very close, almost half of the simulations were unimodal.

As a final simulation, one hundred sets of 24 floods were generated from a bimodal distribution consisting of 16 values from an EV1 with β equal to 1000 and α equal to 400, and 8 values from an EV1 with β equal to 3000 and α equal to 400. The plots of the density functions revealed 9 unimodal

distributions, 43 bimodal distributions with two high distinct peaks, and 48 multimodal distributions, many of which, like simulation 30 in Figure 4.11, were bimodal distributions with a distortion in the tail.

It was thus concluded that only data from a bimodal distribution would produce a density function with two distinct peaks by nonparametric frequency analysis. Thus, some of the New Brunswick annual maximum floods must be generated from a bimodal distribution and those two peaks are not due to sampling variability. As for the heavy-tailed densities, it was concluded that the probability was very remote that many distributions of such shape could randomly be found so close to each other given their low chance of occurrence due to sampling variability. It was therefore concluded that the previously selected and currently used parametric unimodal LN3 was clearly incorrect in describing New Brunswick's mixed distribution floods.

4.5 L-Moment Analysis

As explained earlier in section 3.4, L-moment analysis is a powerful parametric method for determining the distribution of a data set assumed to come from an identical distribution. For the 53 New Brunswick hydrometric stations mentioned earlier, L-moment analysis was performed with the resulting L-moment ratios of Table 4.3. The weighted regional means, computed from equation (3.33) and plotted on the theoretical L-moment diagram of Figure 3.6, are very close to the Generalized Logistic and the

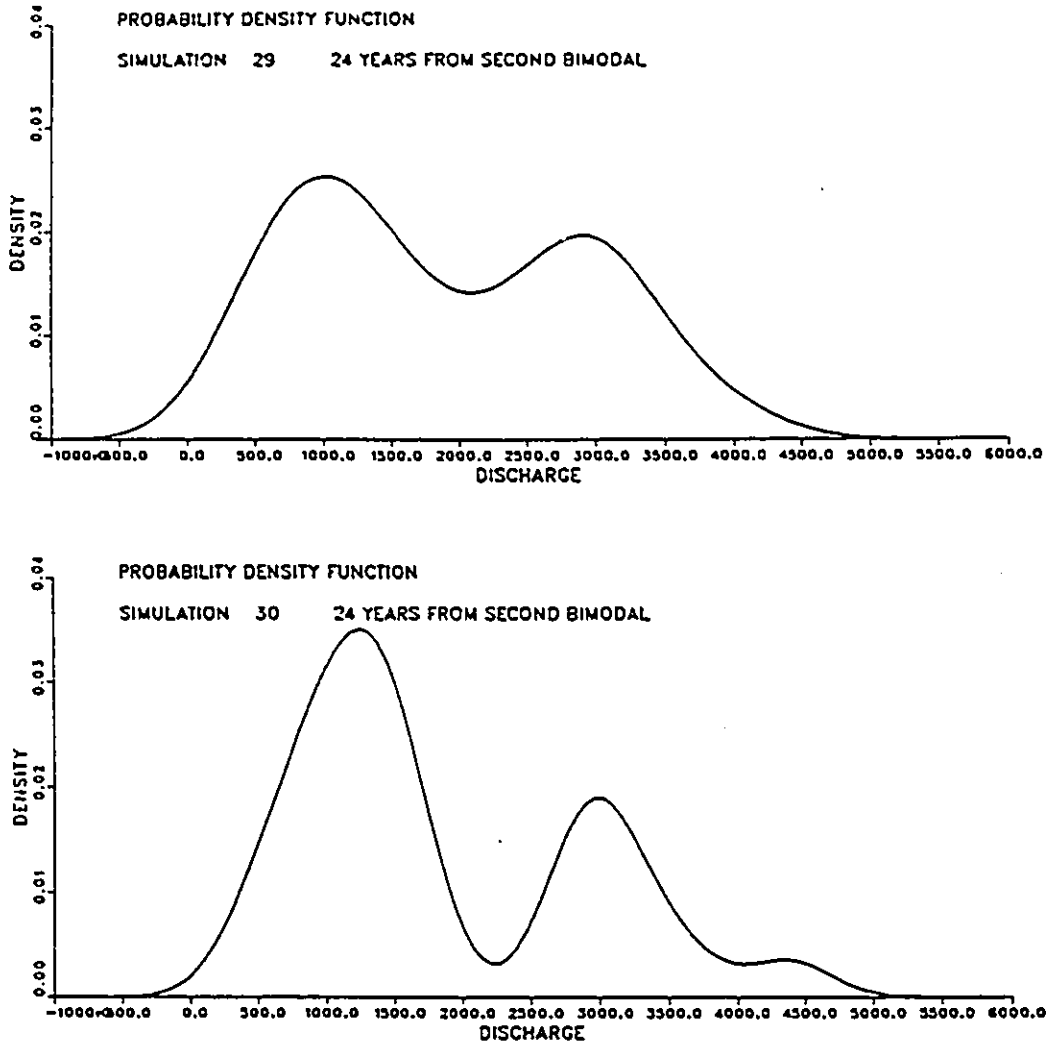


Figure 4.11: Examples of Densities from Bimodal Simulation

Table 4.3: L-Moment Ratios

Station Name	L-Moments τ_3	τ_4	Station Name	L-Moments τ_3	τ_4
01AD003	0.089	0.073	01BJ001	0.127	0.168
01AF003	0.128	0.217	01BJ003	0.058	0.194
01AG002	0.277	0.288	01BJ004	0.210	0.026
01AG003	0.139	-0.047	01BJ007	0.220	0.154
01AH005	0.216	0.195	01BK004	0.159	0.021
01AJ003	0.208	0.187	01BL001	0.304	0.260
01AJ004	0.321	0.237	01BL002	0.307	0.331
01AJ010	0.047	0.063	01BL003	0.172	0.326
01AJ011	0.310	0.373	01BO001	0.240	0.256
01AK001	0.208	0.229	01BO002	0.361	0.259
01AK005	0.303	0.174	01BO003	0.263	0.106
01AK007	0.236	0.091	01BP001	0.431	0.287
01AK008	0.071	0.228	01BQ001	0.278	0.236
01AL002	0.307	0.239	01BR001	0.203	0.020
01AL004	0.489	0.299	01BS001	-0.016	-0.002
01AM001	0.480	0.359	01BU002	0.119	0.261
01AN001	0.162	0.240	01BU003	0.115	0.194
01AN002	0.152	0.078	01BU004	0.213	0.123
01AP002	0.047	0.039	01BV005	0.236	0.094
01AP004	0.190	0.078	01BV006	0.205	0.140
01AP006	0.321	0.294	01BV007	0.246	0.107
01AQ001	0.433	0.325	1013500	0.042	0.142
01AQ002	0.334	0.289	1016500	0.146	0.124
01AR006	0.317	0.268	01BD002	0.163	0.340
01AR008	0.344	0.361	01BF001	0.108	0.070
01BC001	0.185	0.189	01DL001	0.249	0.078
01BE001	0.141	0.092			

The weighted means are 0.208 and 0.185 respectively.

Generalized Extreme Value curves. Since the Generalized Logistic is not usually employed in flood frequency analysis, a GEV distribution was selected. The L-moment ratio values, along with the regional weighted means, were plotted on an L-moment diagram which included the GEV, as shown in Figure 4.12.

Proximity of the weighted values to the theoretical GEV curve suggests that distribution as a likely generator. Because some of the nonparametric densities were bimodal, due to different flood generating mechanisms, the Two-Component Extreme Value (TCEV), a compounded distribution, was investigated as a possible data generator. However, the weighted means were outside the feasible space for the TCEV distribution, as shown in Figure 3.7 (Gabriele and Arnell, 1991). Therefore, the GEV distribution would still be suggested as the generating distribution of New Brunswick floods.

The three heterogeneity measures discussed in section 3.4 were computed and yielded H ratios between 1 and 2 on all three counts. This implied that the New Brunswick flood data set was possibly heterogeneous, or that the data did not likely come from a single distribution. If a single distribution were assumed as fitting the data, the Generalized Logistic would be selected by the heterogeneity tests. This distribution, of course, is not commonly used in flood frequency analysis, and thus its appropriateness can be questioned. Therefore, the parametric L-moment frequency analysis supports the findings from nonparametric frequency analysis which pointed out the

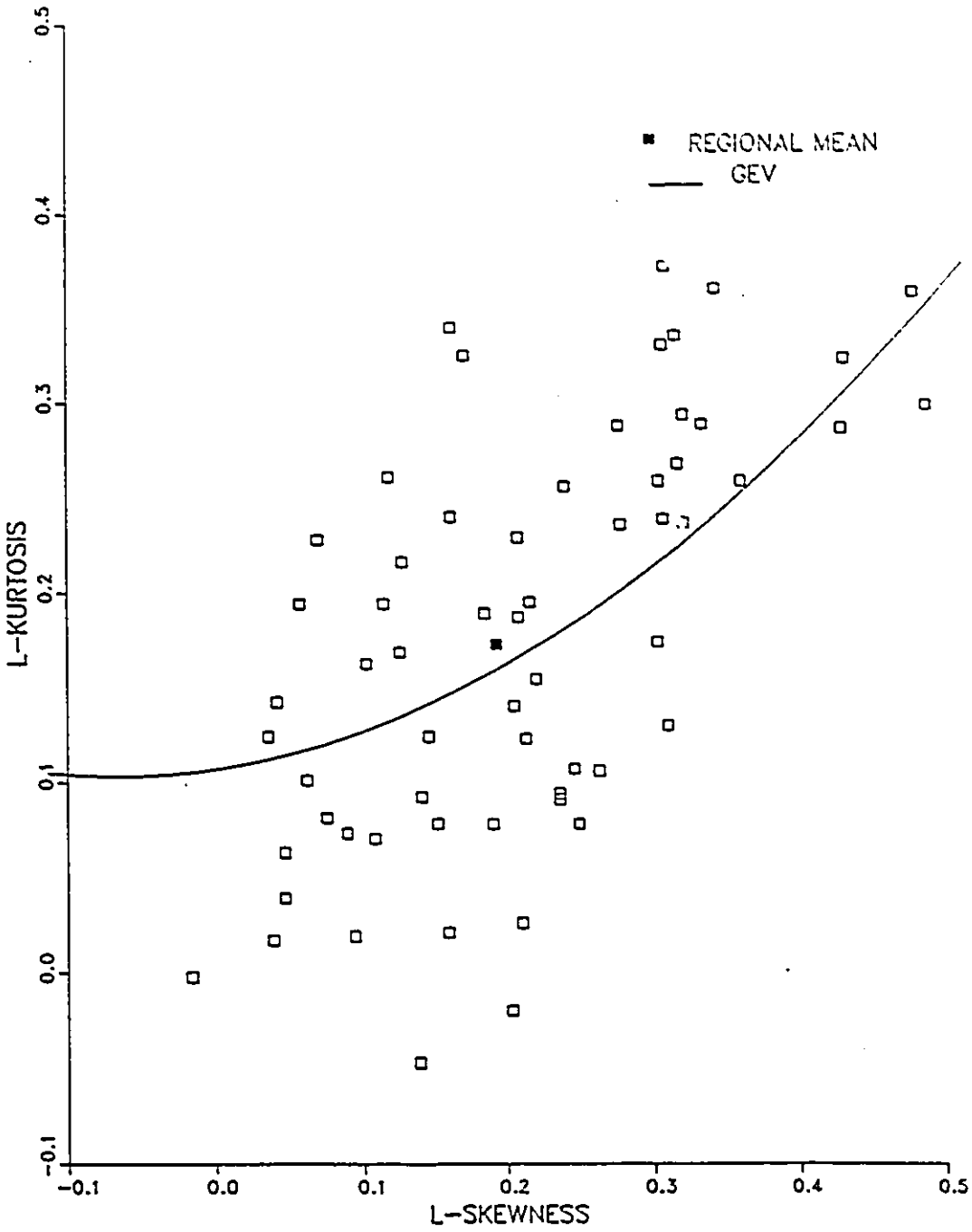


Figure 4.12: New Brunswick L-Moment Diagram

existence of a variety of distribution shapes suggesting a number of possibly different distributions for the data set.

As will be discussed in the next chapter, a portion of the province incorporating 44 of the 53 stations will be proposed as a homogeneous bimodal region. The details of this delineation are left to Section 5.1. L-moment analysis was performed on this set of 44 stations and the three heterogeneity measures computed, resulting in three H ratios less than 1. These 44 stations were therefore considered homogeneous by the tests, and two distributions were selected as appropriate to describe the data, namely the Generalized Extreme Value and the Generalized Logistic. Because nonparametric frequency analysis indicated a bimodal distribution for that region, the use of the TCEV was investigated. However, the location of the weighted means on the L-moment ratio diagram did not permit the selection of the TCEV as the data generator for the homogeneous region.

Relying solely on the parametric L-moment approach, the Generalized Extreme Value, a unimodal distribution, would have been probably mistakenly selected as the data generator of these 44 stations which are assumed to come from the same distribution. Using nonparametric frequency, however, it was concluded that the 44 stations are homogeneous but that the data are bimodal. This points out a weakness of parametric approaches such as L-moment analysis; namely, their difficulty, if not inability, to identify mixed distributions.

4.6 Comparison of Parametric and Nonparametric Frequency

The analysis presented in the previous three sections shows how nonparametric frequency analysis can be used to identify the distribution of floods and compute flood quantiles. A previous study (Inland Waters/Lands and New Brunswick Department of Municipal Affairs and Environment, 1987) had concluded that floods in New Brunswick were best described by the LN3 while an analysis performed in this study involving L-moments would select either the Generalized Logistic or the Generalized Extreme Value distributions. Such an improper unimodal distribution selection, especially when the data come from a mixed distribution, might lead to incorrect quantile estimates.

Nonparametric frequency analysis clearly showed that three distinct probability density function shapes existed for New Brunswick flood data. Some stations came from unimodal distributions, many others from bimodal distributions, and a few others from heavy-tailed distributions. These various density shapes were a reflection of different flood generating mechanisms. Parametric analysis would never have identified such mixed distributions as part of the New Brunswick flood sample.

Therefore, nonparametric frequency analysis can be considered as an important screening tool in distribution identification. A combination of nonparametric frequency analysis and L-moment analysis should confirm

the selection of a unimodal distribution or indicate the presence of a mixed distribution, as was the case with New Brunswick floods. Thus, the nonparametric approach helps to identify the underlying probability distribution, particularly when samples arise from a mixed distribution. Once a mixed distribution is detected, then a nonparametric method would be recommended for frequency analysis.

Chapter 5

RESULTS AND DISCUSSION

5.1 Homogeneous Region Delineation

The shape of the probability density function of the New Brunswick annual maximum flood series, as obtained from nonparametric frequency analysis, is employed to delineate homogeneous regions. Plots of density function for the 53 stations indicate that annual maximum floods basically follow unimodal, bimodal and heavy-tailed distributions. Table 4.2 shows a seasonal partitioning of floods for the 53 stations, as there is a strong, although imperfect, correlation between seasonal occurrence and flood mechanism. Sample sizes of less than 20 years for many stations made the analysis more difficult.

Spring floods are commonly due to rain on snow events, usually within a prolonged snowmelt period. However, on occasion, a rainfall only flood will occur in late spring after all the snow has melted. The limited availability of snow cover data makes the division between spring floods with snowmelt and spring floods without snowmelt uncertain for some stations. Snow courses are typically along the Saint John River or along the coast, as are most of the meteorological stations. Thus, the division of the spring flood series into generating mechanisms is imperfect.

The division by mechanisms for floods outside the spring season is more straightforward. Winter floods occur from rain on snow while fall floods are due to frontal precipitation events as are some of the few summer floods. Some hydrometric stations in the south of the province, however, are also affected on occasion by dying tropical storms during the summer.

The above indicates that different flood generating mechanisms are at work in New Brunswick, and the proportion of their occurrence by mechanism has a high correlation to density shape. Geographical grouping of basins of similar annual maximum flood density shape, and therefore subject to similar flood generation mechanisms, led to the delineation of three homogeneous regions.

The majority of New Brunswick rivers have annual maximum floods following a bimodal distribution such as Figure 4.6 for the Canaan River. The maximum density of the second mode is usually less than half that of the first mode. These rivers are subject to spring floods and the occasional

fall or winter floods. The fall and winter floods account for usually between ten and no more than thirty percent of all annual maximum floods.

A number of rivers have a smooth unimodal density as for the Upsalquitch River of Figure 4.7, where a majority of floods occur from the spring snowmelt. Summer and fall floods usually account for less than ten percent of annual maximum floods. Another group have a heavy-tailed density as for the Lepreau River of Figure 4.S. These rivers are usually mainly subject to spring floods but also on many occasions to fall and winter floods. In the case of one station (01AQ001), floods took place during the mid-August to mid-September period, an indication of tropical storms. Two of the five stations with heavy-tailed densities have over fifty percent of their annual maximum floods occurring at a time other than the spring.

As shown in Figure 5.1, most rivers with unimodal probability density functions come from the same geographical area in northern New Brunswick, while those with a heavy tail come from southwestern New Brunswick. These two areas correspond very closely to the climatologically different areas shown in Figure 4.4. Therefore, based on these findings, a homogeneous bimodal region excluding these two areas with different density shapes was established. As a consequence, grouped together are basins subject to the same flood generating mechanisms which appear to respond in a similar fashion to these mechanisms because of their similar density shape. While a small number of unimodal densities were found in the bimodal region, since these had floods resulting from different mechanisms, it

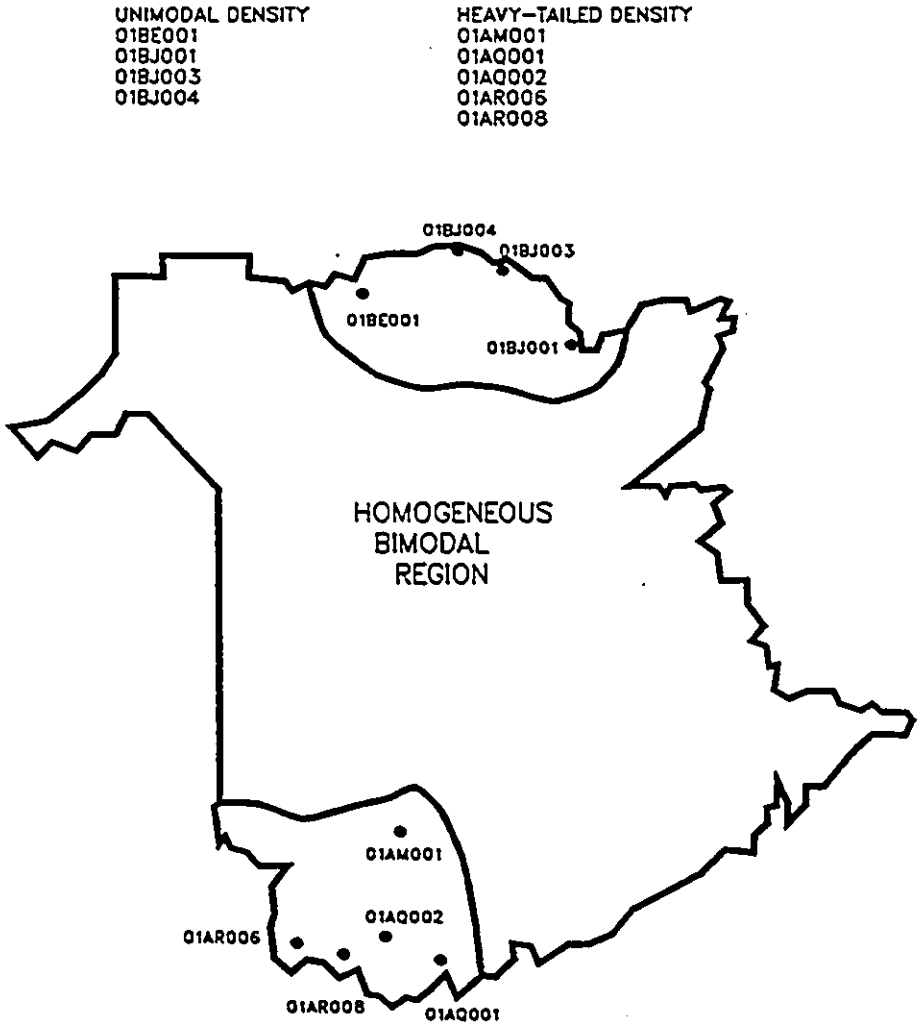


Figure 5.1: Homogeneous Region Delineation for New Brunswick

was assumed that the two modes had been very close to each other, leading to a unimodal density.

It should be noted that there exists some subjectivity in the delineation of those regions, as is the case in other homogeneous region delineation methods. For example, the density plots of stations 01BC001 and 01BJ007, while not smoothly unimodal, do suggest their generator could be unimodal. One could also, of course, argue that stations 01BV006 or 01AL002 belong to the heavy-tailed region based on density shape.

As the simulations of the previous chapter pointed out, data from a unimodal distribution may appear bimodal on occasion or vice-versa. Therefore, a global inspection of density shape is required in order to evaluate the existence of homogeneous regions. A test to investigate the number of modes exists (Silverman, 1981), but while it could provide a likely number of modes at a specific site, it would not provide a global outlook on the number of modes or differentiate between a bimodal distribution with two distinct peaks and a heavy-tailed distribution. As well, it could not differentiate between a truly heavy-tailed distribution or a heavy-tailed distribution due to sampling variation from a unimodal distribution.

Admittedly, due to the effect of sampling variability, some subjectivity is involved in delineating homogeneous regions, which is a weakness of the approach. However, the geographical proximity of similarly-shaped densities cannot be accounted for by random sampling variation. Because there will always exist zones of transition between climatic influences and between

physiographic features, it is expected that the delineation of homogeneous regions will never be exact, irrespective of methods employed.

5.2 Regression Analysis

The unimodal shape and heavy-tailed shape regions that were found contained, respectively, only four and five hydrometric stations, numbers too small to develop useful regression equations. Each of these groupings most likely belongs to a larger homogeneous region, only a small portion of which is located in New Brunswick. The set of 53 stations minus the 9 stations having a different density shape, for a total of 44 stations, will thus be called the homogeneous bimodal region. Regional analysis was performed and compared for both the homogeneous bimodal region and on a province-wide basis in order to determine whether the delineation of a homogeneous region would result in a more accurate regional equation.

Relationships for the previously delineated homogeneous regions shown in Figure 4.2 were also found. As well, equations for modified versions of these regions, not including the stations from the unimodal and heavy-tailed regions, were developed. Table 4.1 provides a list of the stations in all regions. Because of the great uncertainty involved in estimating floods for return periods much greater than the sample length, estimates of the 50 and 100 year floods for records less than 15 years in length were not included. Thus, the province-wide equation for the 50 and 100 year floods

included 45 stations, and for the homogeneous bimodal region included 37 stations. The smaller homogeneous regions were also similarly reduced.

The 2, 10, 20, 50 and 100 year flood estimates were subjected to a linear regression using physiographic and climatic parameters such as drainage area, mean annual precipitation, percentage of lakes and swamps, and average water content of snow on March 31 as computed for the basins draining to the hydrometric stations in a previous study (Inland Waters/Lands and New Brunswick Department of Municipal Affairs and Environment, 1987). The following linear parametric regression models were assumed:

$$logQ = a + blogx_1 \tag{5.1}$$

or

$$logQ = c + dlogx_1 + elogx_2 \tag{5.2}$$

where Q is the design flood of a given return period, x_1 and x_2 are physioclimatic variables, while a, b , c, d and e are regression coefficients for the linear regression. For nonparametric regression, a relationship between the logarithms of the variables was found.

The most significant factor to enter the equation was always drainage area, with mean annual precipitation always coming second. The addition

of further variables was not statistically significant. For comparison of goodness of fit between the equations, the integral square error (ISE), in percentage, was calculated. It is defined by (Adamowski and Middleton, 1977):

$$ISE = \frac{\sum_{i=1}^{n}((O_i - P_i)^2)^{0.5} \quad X \quad 100}{\sum_{i=1}^{n} O_i} \tag{5.3}$$

where O are the single station flood estimates, P the predicted flood estimates from the regression equation and n the sample size. This is a commonly used statistical criterion for comparison of linear and nonparametric regression (Adamowski and Feluch, 1991).

Tables B.3 to B.7 in the Appendix provide the regression results from both the linear and the nonparametric regressions for, respectively, the 2, 10, 20, 50 and 100 year floods. Included are the linear regression equation parameters, the ISE's for both regressions and the number of data points. Figures 5.2 and 5.3 are examples of the resulting curves.

The entire set of two-dimensional relationships are in Appendix B. For the province-wide or the homogeneous bimodal region equations, where the number of data points varied from 37 to 53, nonparametric and linear regression provide relationships yielding flood estimates not very different from each other. Residuals of the nonparametric regression, shown in Figures 5.4 and 5.5 for the two-dimensional and the three-dimensional nonparametric regressions, have no pattern suggesting an incorrect rela-

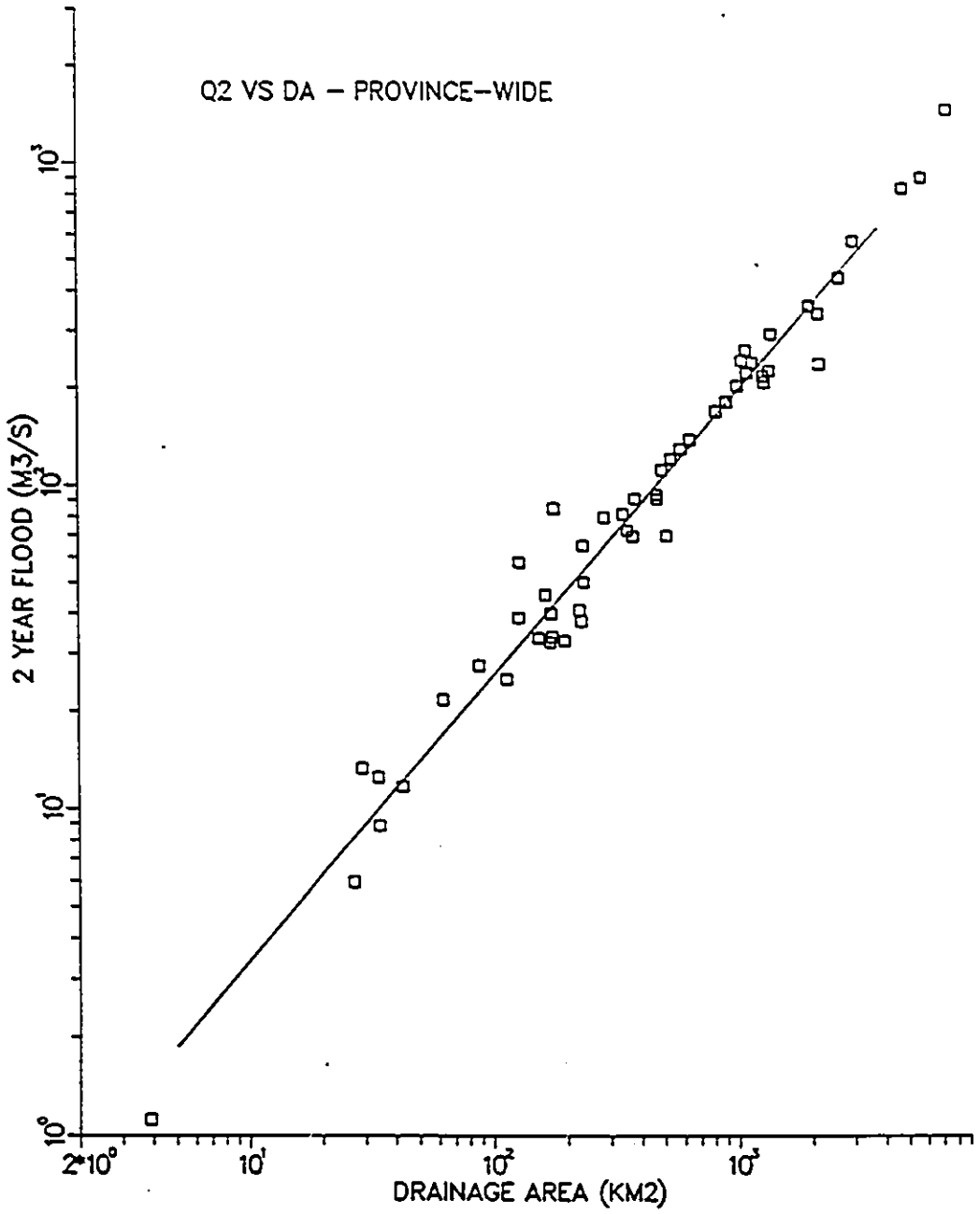


Figure 5.2: Linear Regression for Two-Year Flood Province-wide

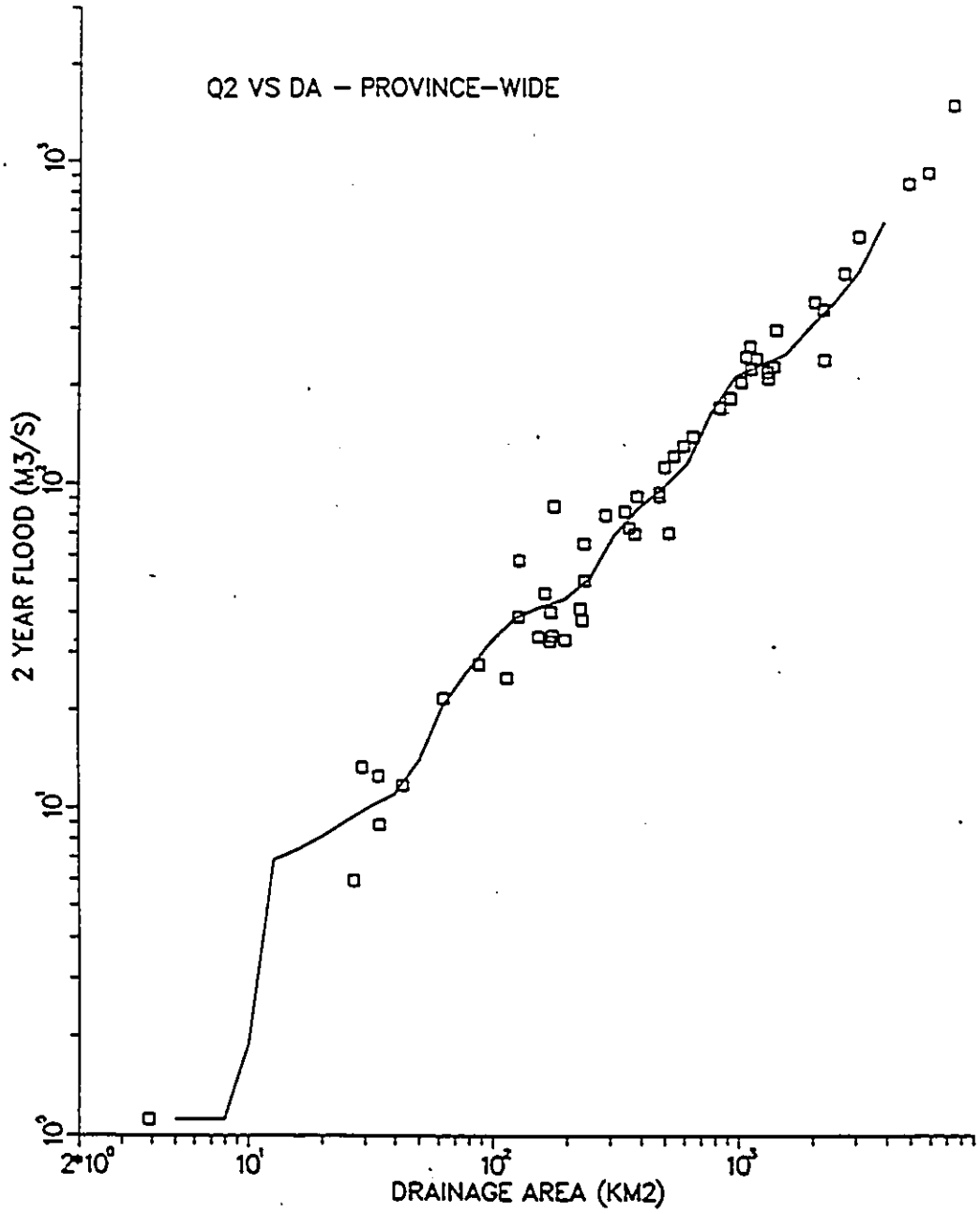


Figure 5.3: Nonparametric Regression for Two-Year Flood Province-wide

tionship.

For the smaller homogeneous regions, where the number of data points varied from 8 to 18, there is a greater difference between the linear and nonparametric regressions as the nonparametric regressions tend to define relationships joining together the smaller quantity of data points. However, the flood estimates from both approaches tend not to be very different again. On occasion, one could argue that the nonparametric regressions follow the data perhaps too closely, especially for the Northwestern region.

A comparison of the parametric and nonparametric ISE values indicates that, in general, the nonparametric regression has a lower ISE. For the Central region, however, the parametric ISE is consistently less, which could be due to the greater range of drainage areas. Equations for the modified homogeneous regions were only developed for the two-dimensional and not for the three-dimensional cases as the number of degrees of freedom was getting very low in some cases. The flood versus drainage area relationships for the modified regions did not provide lower ISE's on a consistent basis.

An interesting observation is that for the smaller return periods (2 and 10 year floods for the two-dimensional case and 2 year flood for the three-dimensional case), the ISE's are lower for the province-wide equations as opposed to the bimodal region equations. This is as expected because the province-wide equations were developed with a larger sample of stations. However, for the other return periods, the bimodal region equations have lower ISE's. This means that, once the stations with non-bimodal densities

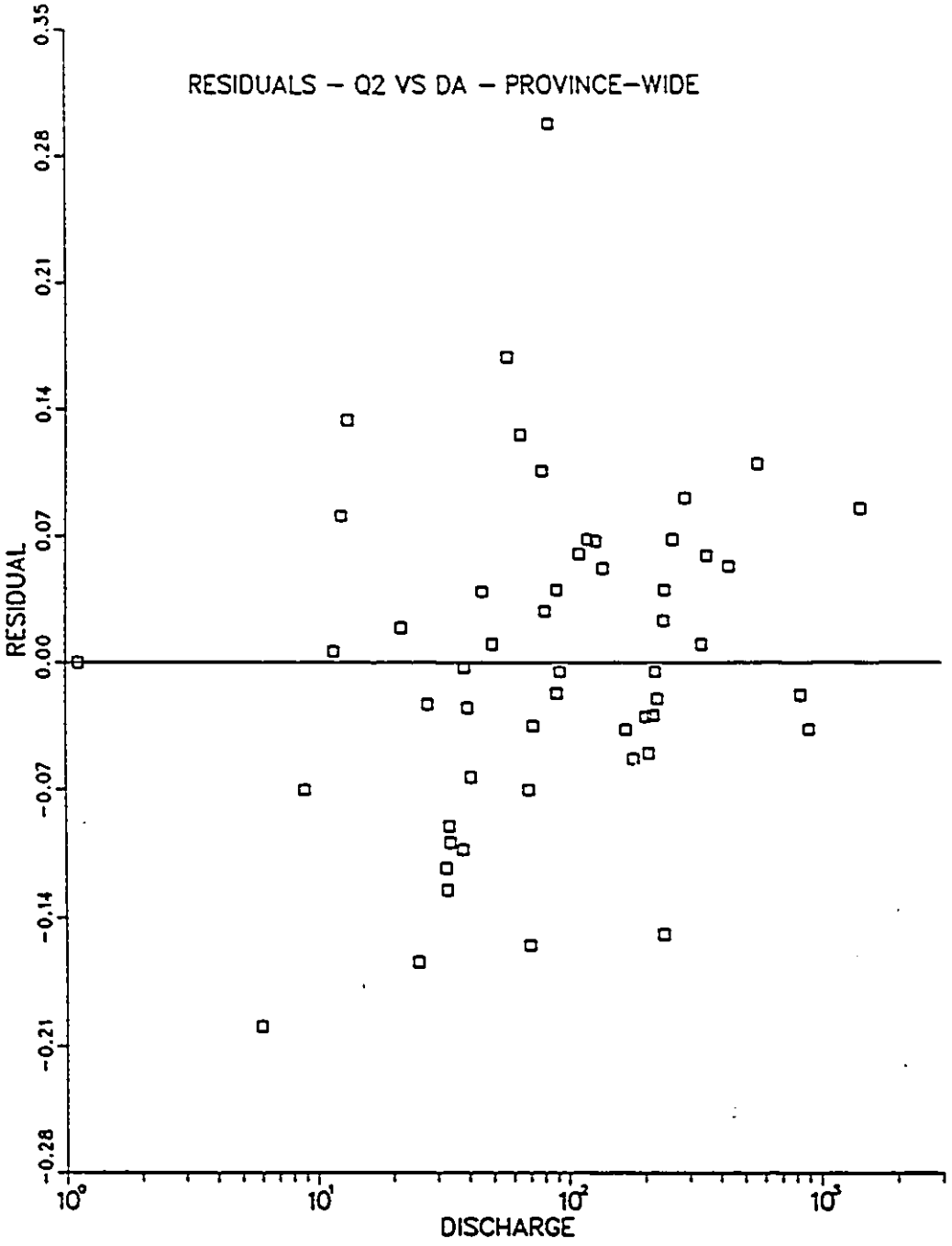


Figure 5.4: Residuals for Two-Dimensional Nonparametric Regression

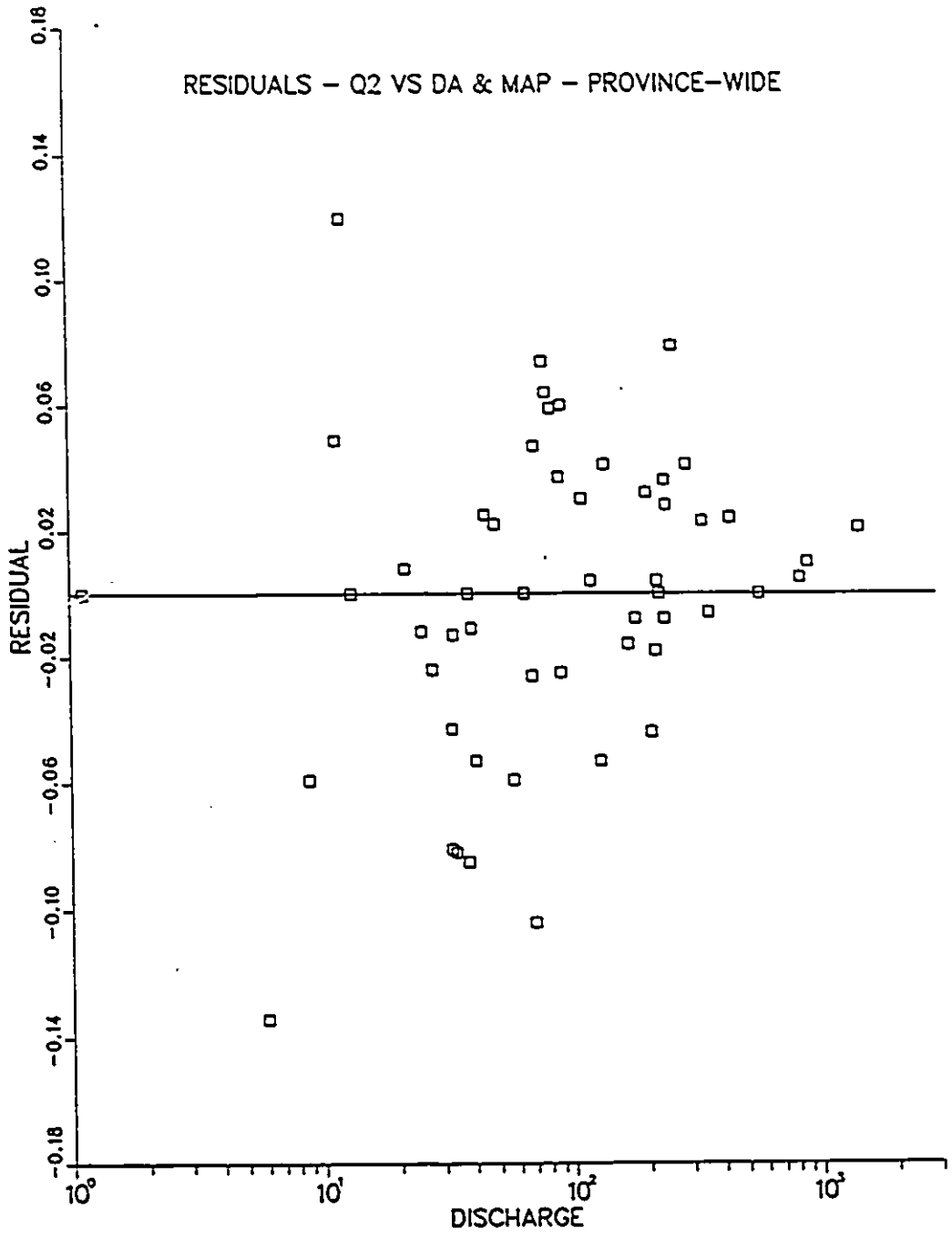


Figure 5.5: Residuals for Three-Dimensional Nonparametric Regression

are removed, the resulting regression equations have lower ISE's.

The reason for this is that the large return period floods of the non-bimodal stations were caused by different processes than for the remainder of the province. By defining a homogeneous bimodal region, in other words by removing some stations from where they did not belong, their adverse impact on the regressions was eliminated. This also suggests that different return periods may have different homogeneous regions.

In estimating a design flood at an ungauged site from a regional relationship, it is always recommended that a number of valid approaches be compared (Inland Waters/Lands and New Brunswick Department of Municipal Affairs and Environment, 1987). The results of Tables B.3 to B.7 thus provide the user with potentially four equations: province-wide, homogeneous bimodal, regional (such as Central, Northwestern, Eastern or Southern), and modified regional.

The equations with a lower ISE are expected to be more accurate. Thus, there are many instances where the user will favor a modified regional equation over a regional equation. As well, for return periods of 20 years or greater, the homogeneous bimodal equation will be favored over the province-wide equation where it is applicable.

A study of the parametric and nonparametric regression plots of Appendix B reveals that nonparametric regression can serve as a useful screening tool for parametric regression. As shown by Figure 5.6, nonparametric

regression will not yield a relationship between data that are too far apart; linear regression will yield an equation, but there are no data to support it. Similarly, the linear equations can be used to extrapolate while nonparametric regression cannot be used beyond the last data point; but one, of course, should always exercise great caution when extrapolating a model beyond the data limits.

When dealing with three dimensions, as shown by Figures 5.7 and 5.8, nonparametric regression is also a useful screening tool. The linear regression lines are, of course, parallel while the nonparametric relationships meet and then cross due to lack of data. As shown by Figure 5.9, there are no stations with both low drainage area and low mean annual precipitation, or both very high area and precipitation. Thus, linear regression provides an unrealistic assurance in predicting for concurrently low and concurrently high values of area and precipitation.

While it could be argued that both of the above examples are relatively simple to detect without using nonparametric regression as a screening agent, when the number of dimensions in a regression increases, such detection becomes very difficult. This could lead to the use of a linear relationship where it should not be employed and to providing the user with an unwarranted assurance in its application. By pointing out deficiencies in the application of regression analysis, nonparametric regression is therefore a useful screening tool for any regression.

An interesting observation is that the ISE statistic of Tables B.3 to

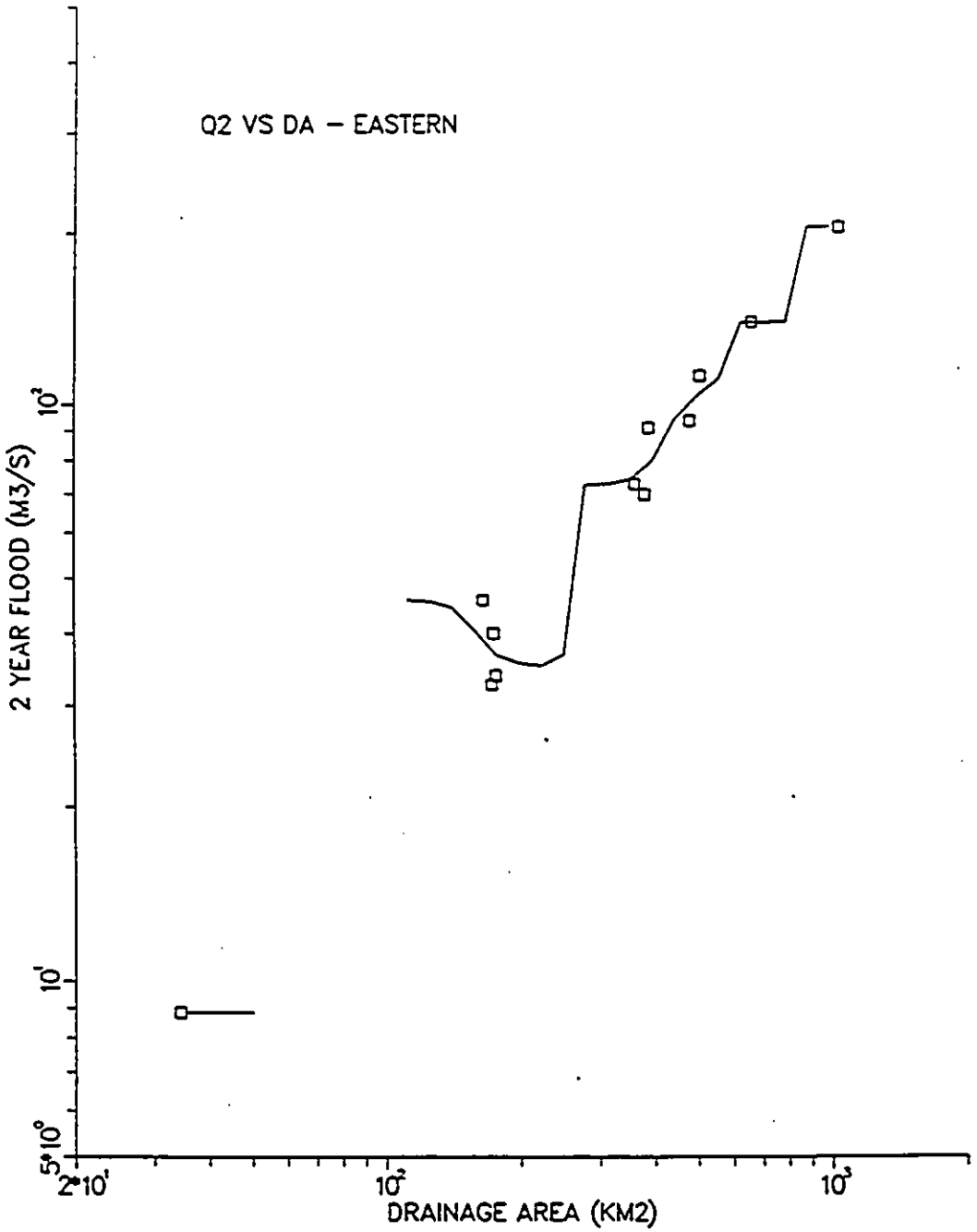


Figure 5.6: Nonparametric Regression for Two-Year Flood Eastern Region

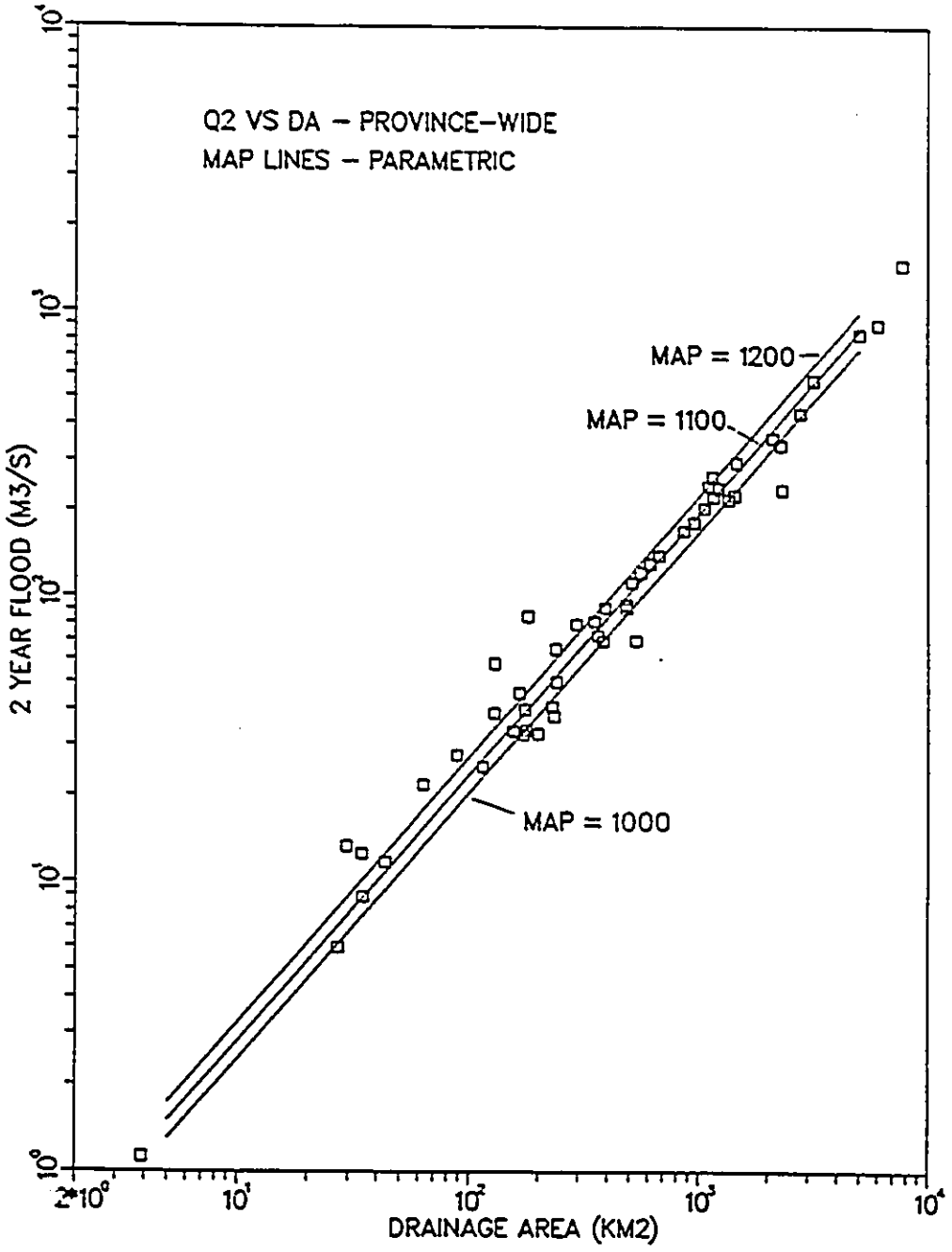


Figure 5.7: Three-Dimensional Linear Regression for Two-Year Flood Province-wide

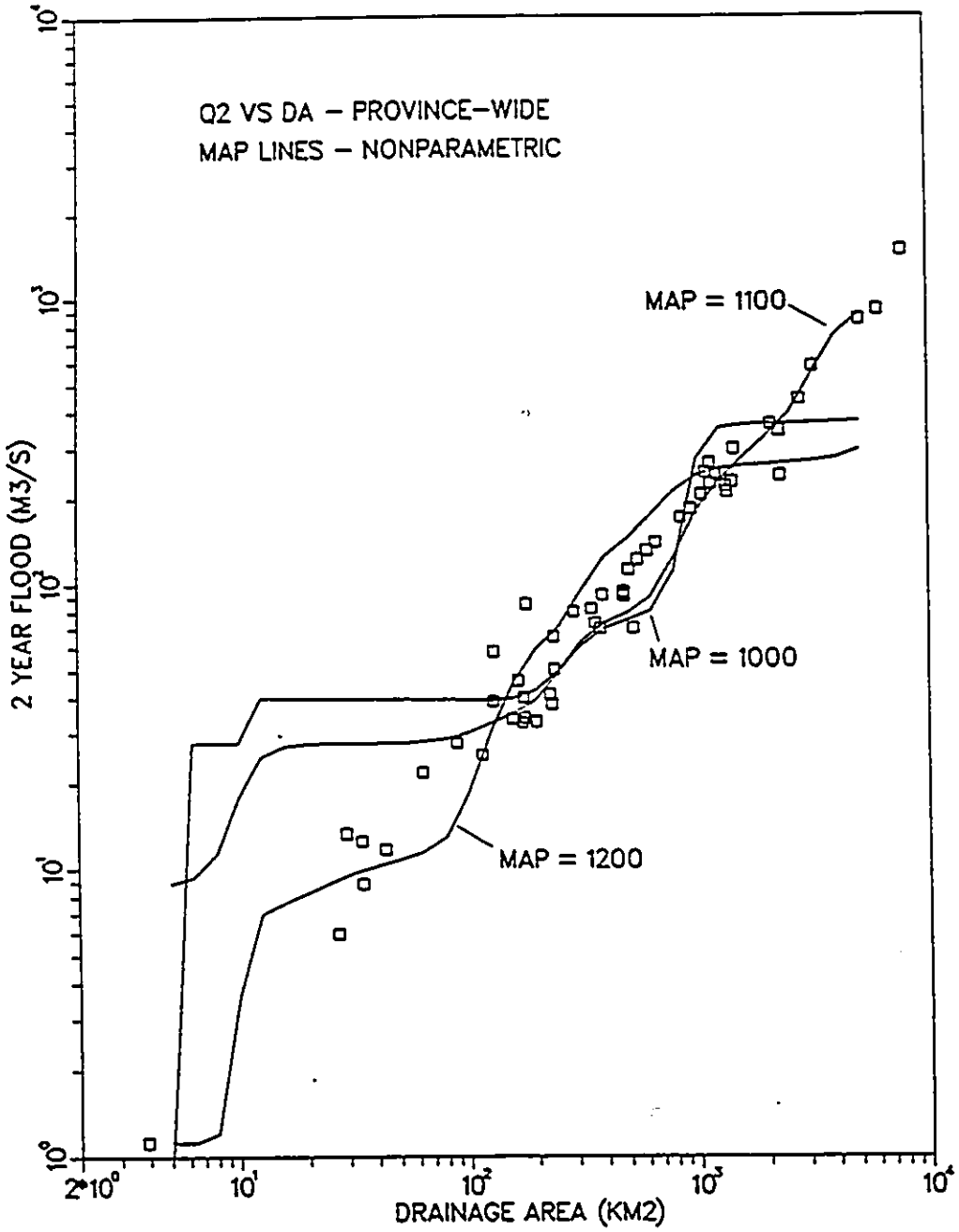


Figure 5.8: Three-Dimensional Nonparametric Regression for Two-Year Flood Province-wide

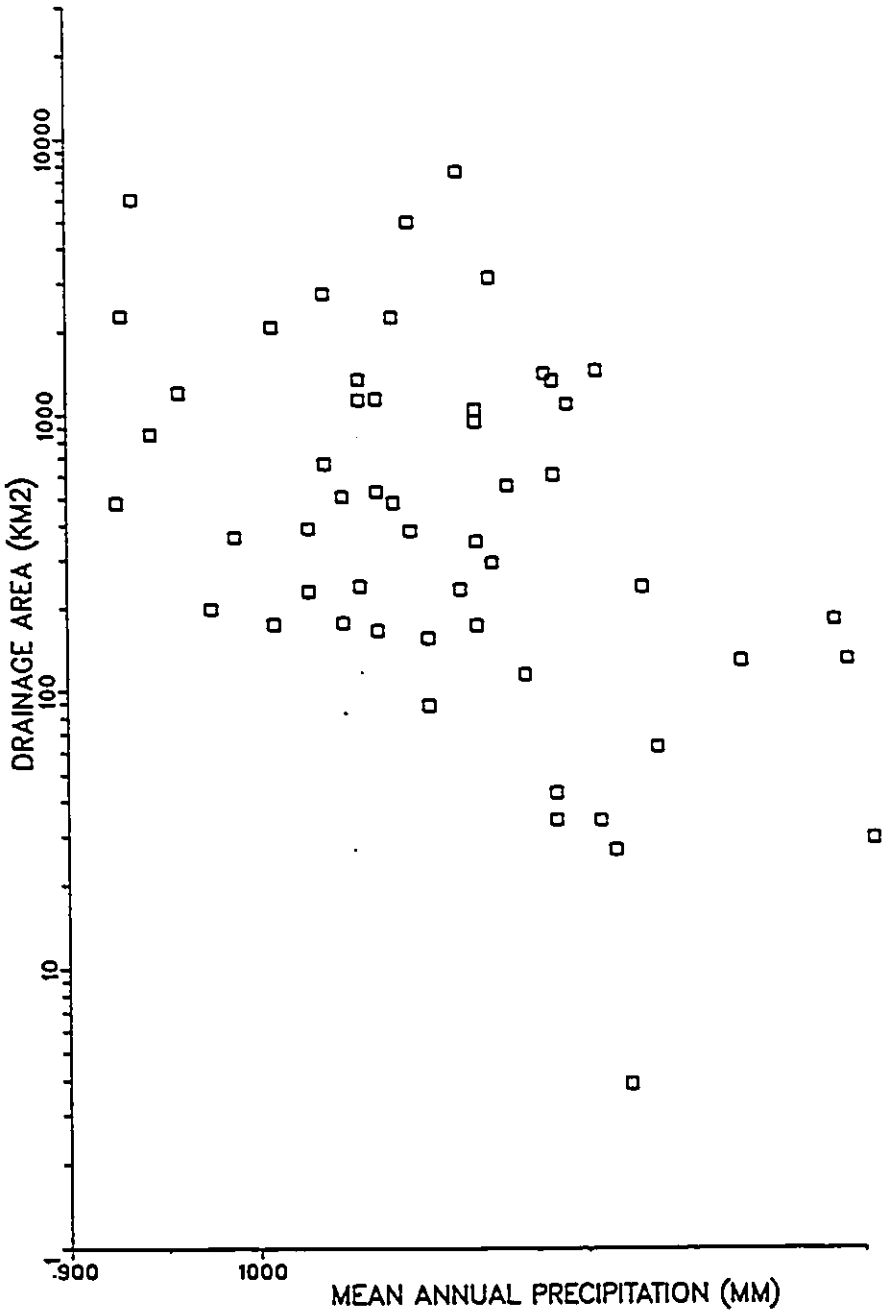


Figure 5.9: Drainage Area and Mean Annual Precipitation Variations

B.7 is almost always less for the nonparametric regression even though visually both parametric and nonparametric regression appear to provide comparably suitable fits. As well, the value of this statistic decreases when the nonparametric regression is unable to find a relationship at some points, as for example the 2 year flood equation for the Eastern region shown in Figure 5.6.

Such occurrences put into question the validity of the ISE statistic. This led to the further investigation of means of comparing parametric and nonparametric regressions. As explained in the next section, simulations were undertaken in order to better understand the ISE statistic and led to the investigation of a new approach for confidence interval computation.

5.3 Monte Carlo Simulations

In order to evaluate the impact of sample size on comparing parametric and nonparametric regressions, and to study what differences, if any, occur in relationship development when data come from linear and slightly nonlinear models, a Monte Carlo simulation was developed. Hypothetical at-site flood frequencies were generated for drainage areas between ten and a thousand square kilometers from linear and slightly nonlinear models representative of New Brunswick data. To samples of sizes 10, 20, 30, 40 and 50, both linear and nonparametric regressions were applied and the ISE statistic computed.

For generation from a linear relationship, the following model was used:

$$logQ = -0.5 + 0.9logDA + \varepsilon \tag{5.4}$$

where DA is a random number uniformly distributed between 10 and 1000, ε is a normally distributed random number of mean zero and standard deviation 0.2, and Q is the predicted design flood from the above relationship.

For generation from a slightly nonlinear relationship, the following model was used:

$$logQ = -0.45 + 0.9logDA - 0.05(logDA)^2 + \varepsilon \tag{5.5}$$

with variables as defined earlier.

It is important to note that $logQ$ and $logDA$ are the variables regressed both parametrically and nonparametrically. Because the observed random component about the regression lines observed for New Brunswick was very small, the resulting parametric and nonparametric fits appeared visually equivalent. Thus, a larger random component, of standard deviation 0.2, was selected to investigate what impact this would have on nonparametric regression. While this large value is clearly not representative of New Brunswick, it would be representative of a larger area with relatively few

stations.

Tables B.8 and B.9 in the Appendix show the ISE's from both parametric and nonparametric regressions for, respectively, the simulations from the linear and the slightly nonlinear models. In both instances, the ISE statistic is usually less for the nonparametric regression. Out of 50 simulations from a linear model, only four had a larger ISE for the nonparametric regression. Out of 50 simulations from a slightly nonlinear model, only nine had a larger ISE for the nonparametric regression. For both models, a sample size of 10, 20 or 30 was more likely to yield a larger ISE for nonparametric regression; three out of the four times for the linear model, and eight out of the nine for the slightly nonlinear model.

Figures 5.10 and 5.11 show the linear and nonparametric regressions for simulation 8 of sample size 10. Even though the ISE was larger for the nonparametric regression, it is visually not clear why one regression is supposed to be superior to the other.

As well, the ISE statistic resulting from a parametric regression applied to slightly nonlinear data was usually close to that of the nonparametric regression, even for a large sample size. Thus, based on ISE, there appeared to be little, if any, improvement in employing nonparametric regression rather than linear regression when the model did not depart significantly from a linear relationship.

The general conclusion from the New Brunswick regressions and the

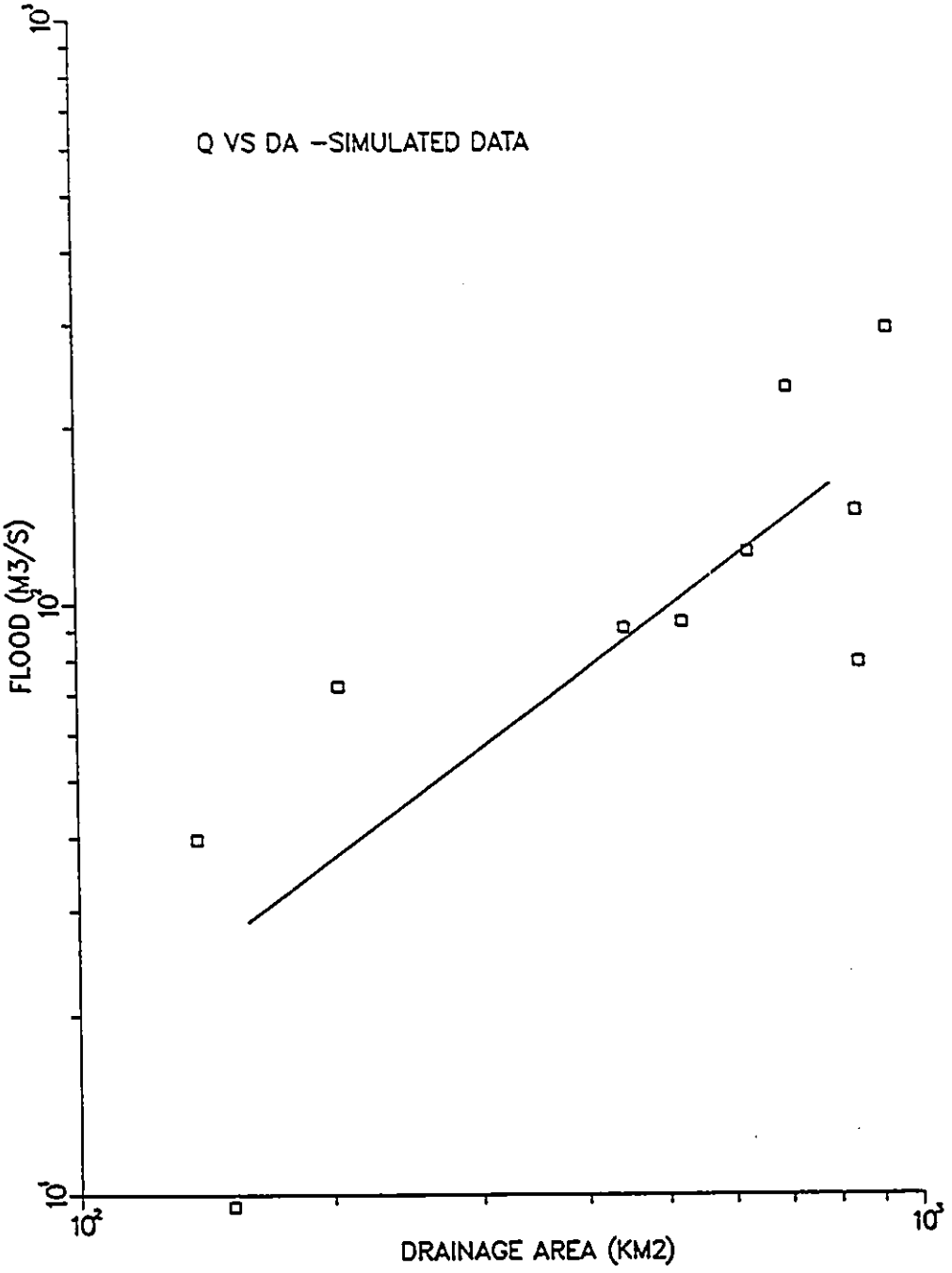


Figure 5.10: Linear Regression for Simulation 8 of Sample Size 10

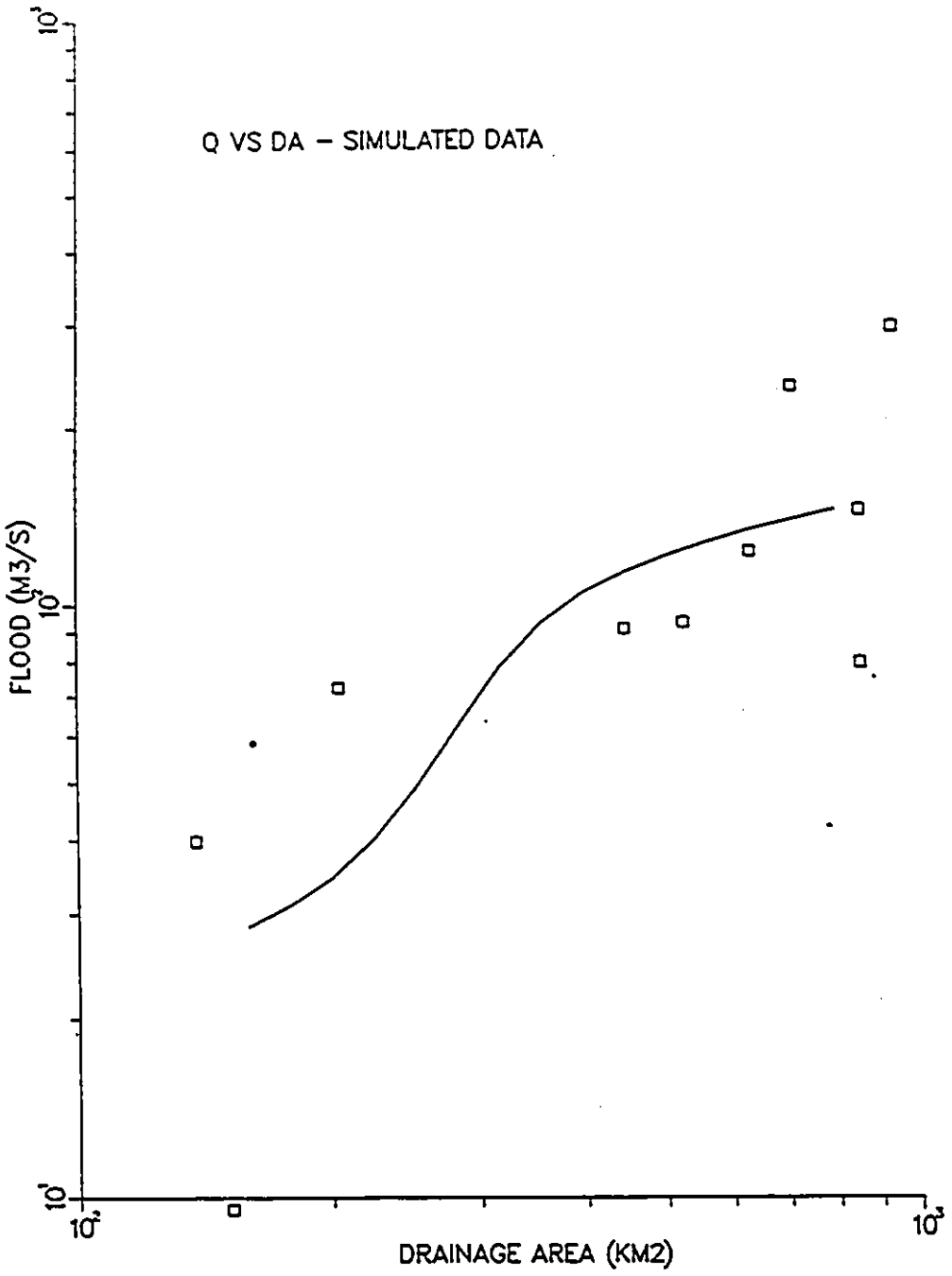


Figure 5.11: Nonparametric Regression for Simulation 8 of Sample Size 10

simulations is that the ISE statistic appears to be almost always less for the nonparametric regression whether the actual model is linear or slightly nonlinear, whether the random component is large or not, and whether the sample size is of ten values or fifty. Therefore, another criterion for comparison of parametric and nonparametric regressions had to be found. Confidence intervals, as explained in the next section, were selected as that criterion.

5.4 Confidence Interval Determination

As discussed earlier in sections 2.3.3 and 3.3.2, the theory of confidence intervals is well-known for parametric regression but no theoretically exact solution has been developed for nonparametric regression. In this thesis, confidence regions are investigated using bootstrapping of pairs, which will permit comparison of linear and nonparametric regressions.

Using bootstrapping of pairs, confidence regions can be obtained for a linear regression model, but only at the center of the line since, as mentioned earlier, bootstrapping is appropriate only in symmetrical situations. If a series of dependent and independent variables Y and X, each of sample size n, are believed to be related by $Y = a + bX$, then one can obtain confidence intervals for that relationship at the central value of the X's by bootstrapping sample pairs of size n of X and its corresponding Y. A linear regression is fit to each bootstrapped sample and a value of Y_p is

obtained at the chosen X. After a large number of repetitions, say in the thousands, the distribution of Y_p will be found to be normally distributed with its standard deviation equal to the value of equation (2.7).

To compute the confidence interval at the center of the X's for the nonparametric regression, a similar approach is undertaken. To a large number of bootstrapped samples of size n, a nonparametric regression is fit and Y_p is obtained. From the distribution of these numerous Y_p's, the confidence interval from the nonparametric regression can then be derived.

Simulations were undertaken to illustrate the above development. Data were generated from the linear model $y = 3x + \epsilon$ and the quadratic model $y = x^2 + \epsilon$. Two error terms ϵ were selected; one was uniformly distributed from -1 to 1, while the other was uniformly distributed from -0.1 to 0.1. These will be referred to as the large and small error scenarios. As well, the model $y = x^2$, without an error term, was employed. The X values varied from 0.1 to 3.0 in increments of 0.1, for a total of 30 pairs. Figure 5.12 is a graphical representation of the regression simulations. One hundred sets of 30 values were generated for both the linear and quadratic cases with small and large errors. For the quadratic model without error, only the original data could be used.

To determine the confidence intervals for each of the 100 trials in the cases with error as well as for the data fitting $y = x^2$, 1000 bootstrapped (x, y) pairs of size 30 were resampled from the generated data and then a linear parametric model fit. Using this model, the predicted value of Y at

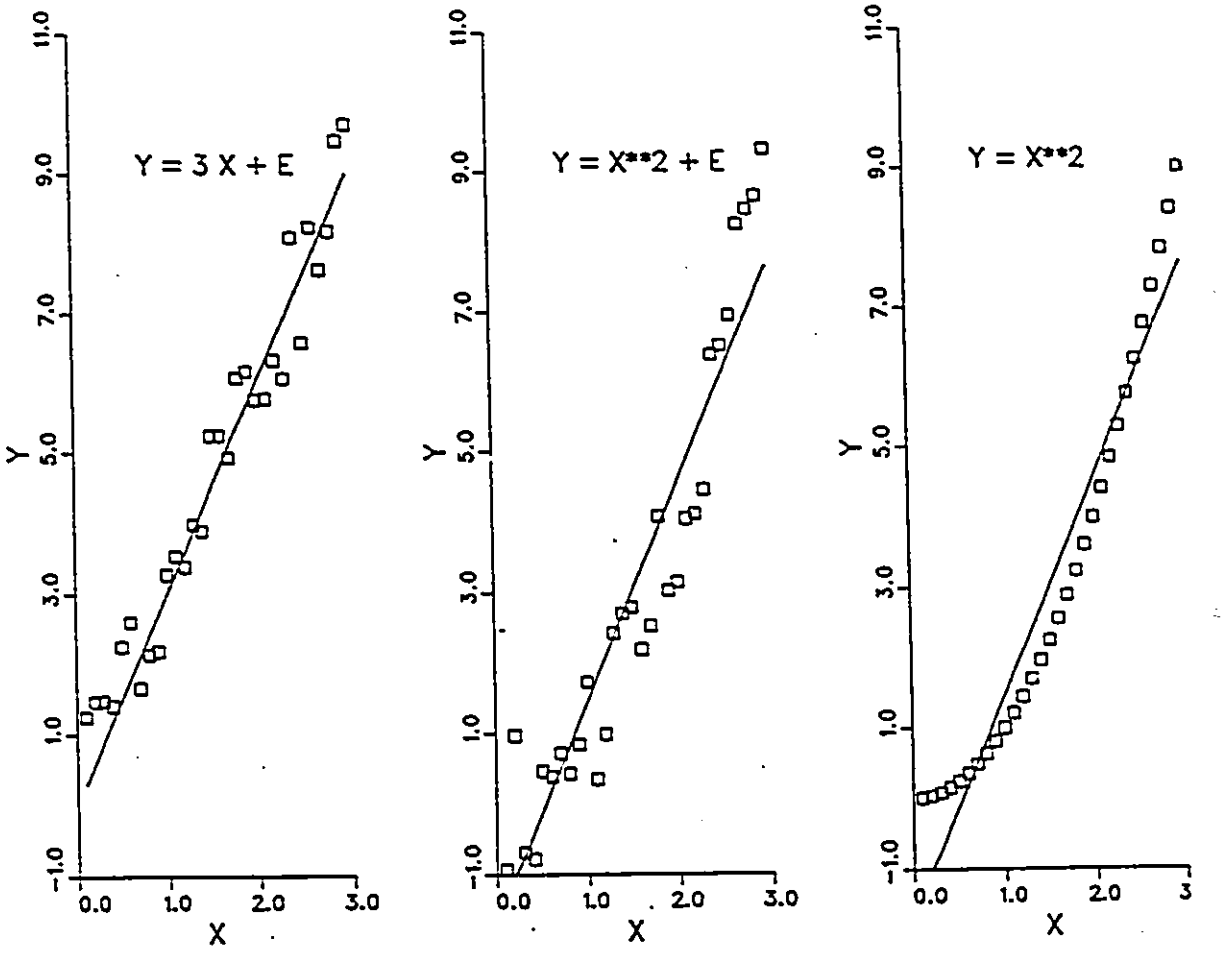


Figure 5.12: Regression Simulations

X equal to 1.55, at the center of the regression, was obtained. The mean and the standard deviation of those 1000 predicted values were found and it was determined that those bootstrapped estimates followed a normal distribution. The standard deviation of those 1000 predicted values was chosen as the standard error of estimate of the regression line.

For both the linear and the quadratic cases with each of the small and the large error scenarios, the 100 theoretical standard errors given by equation (2.7) were very close to the 100 standard errors estimated by bootstrapping. For the quadratic case without error, it was also very similar. For the linear case, the mean theoretical to bootstrapped standard error ratio was 1.020 with the standard deviation of the ratio equal to 0.025 for the large error scenario, while the mean theoretical to bootstrapped standard error ratio was 1.015 with a standard deviation of 0.025 for the small error scenario.

For the quadratic case with error, the mean theoretical to bootstrapped standard error ratio was 0.997 with the standard deviation of the ratio equal to 0.019 for the large error scenario, while the mean theoretical to bootstrapped standard error ratio was 0.987 with a standard deviation of 0.003 for the small error scenario. For the quadratic case without error, the theoretical to bootstrapped ratio was 0.987. It was thus concluded that bootstrapping at the center of a regression line can successfully reproduce the theoretical confidence intervals of that relationship.

The percentage of the 68 percent confidence regions which actually contained the theoretically correct value at X equal to 1.55 from equations

$y = 3x + \epsilon$ or $y = x^2 + \epsilon$ were also computed. The coverage from the linear case was 71 out of 100 sets for the large error scenario and 63 out of 100 sets for the small error scenario, which are very close to the theoretical 68 percent. For the quadratic case, it was zero, however, which is not unexpected since the linear model was incorrect.

Applying the same approach at a location other than the center of the regression line yielded confidence regions different from the theoretical, since bootstrapping is a correct procedure only for symmetrical situations. These confidence regions, which were much larger than the theoretical the farther away from the center, were, however, very close to the theoretical within the central one-fifth of the data range. The further away from the center, the more skewed were the distribution of the bootstrapped estimates.

For the nonparametric regression, 30 pairs of values for X ranging from 0.1 to 3.0 from equations $y = 3x + \epsilon$, $y = x^2 + \epsilon$ and $y = x^2$ were also used, considering large and small errors. To bootstrapped samples of size 30, nonparametric regression was applied and a distribution for the estimates at the center of the regression obtained. It is important to note that the distribution of the estimates was not a normal distribution but rather a peaky distribution with long tails. From this distribution, confidence regions were found, which were compared to the parametric confidence regions found earlier. The results are in Table 5.1.

Because the confidence region for the nonparametric regression is not a smooth curve as for parametric regression, a study of the predicted values

Table 5.1: 68 Percent Confidence Intervals at Center of Regressions

	Linear Model at X=1.55	Nonparametric Model at X=1.55	at X=1.5
$y = 3x + \epsilon_1$	(4.545,4.752) l=0.21	(4.132,4.915) l=0.78	(3.827,4.213) l=0.38
$y = 3x + \epsilon_2$	(4.639,4.659) l=0.02	(4.556,4.680) l=0.12	(4.402,4.551) l=0.15
$y = x^2 + \epsilon_1$	(2.957,3.284) l=0.33	(1.936,2.776) l=0.84	(1.614,1.983) l=0.37
$y = x^2 + \epsilon_2$	(2.998,3.255) l=0.26	(2.330,2.415) l=0.09	(2.185,2.241) l=0.06
$y = x^2$	(2.999,3.255) l=0.26	(2.405,2.472) l=0.07	(2.250,2.318) l=0.07

Note: The confidence intervals for the linear model at X=1.5 are almost identical to the values at X=1.55
l is the interval length ϵ_1 - U(-1,1) ϵ_2 - U(-0.1,0.1)

only for X equal to 1.55 can be misleading. Rather, the confidence region depends on the proximity of data points to the location of interest because the nonparametric regression estimate always tries to follow the data points closely. Figure 5.13, which shows both parametric and nonparametric confidence regions for the same data, illustrates the concept.

It must be pointed out that when the error is small or nonexistent, then the confidence interval in Table 5.1 may be very small. In such circumstances, probably more than 1000 bootstrappings are required in order to achieve the accuracy reported in Table 5.1. For example, comparing the $y = x^2 + \epsilon$ and the $y = x^2$ confidence regions at X equal to 1.5, one would expect a larger interval when there is an error.

Nonetheless, comparing the confidence regions at X equal to 1.5, which is near the center and for which there is a data point, and at X equal to 1.55 where there is no data point, it is seen that the confidence region of the parametric model for the data from $y = 3x + \epsilon$ for both error scenarios is entirely within the confidence region for the nonparametric model. Thus, for data from a linear model, the nonparametric regression confidence regions will be wider than the linear regression confidence regions.

For the case of data from $y = x^2$ and $y = x^2 + \epsilon$ with a small error, the confidence regions from the nonparametric regression are smaller than those from the parametric model. In these cases, the nonparametric model is clearly the model to select. As well, the predicted values from the nonparametric models are very close to the correct ones while the predicted

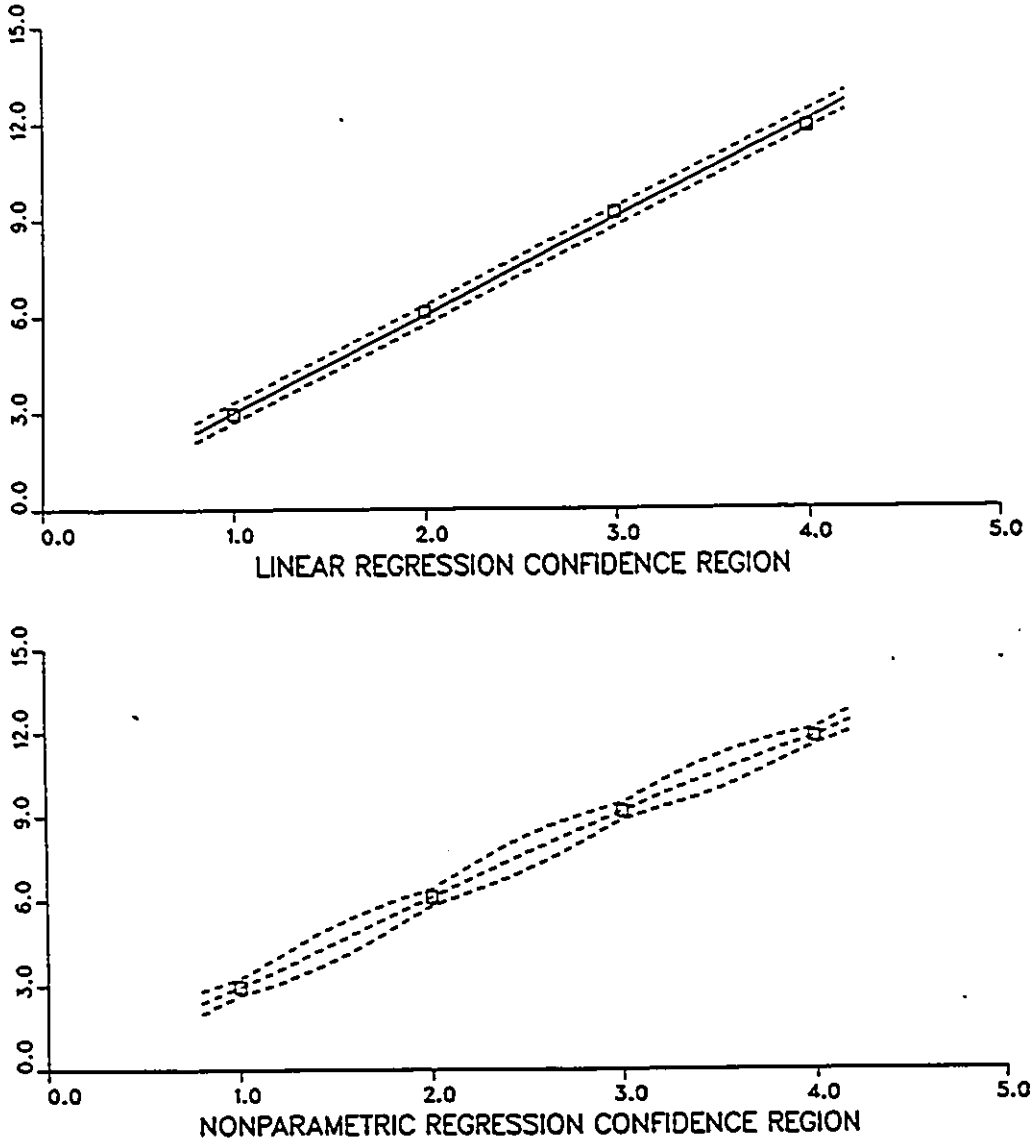


Figure 5.13: Linear and Nonparametric Regression Confidence Regions

values from the linear models are not.

For the case of data from $y = x^2 + \epsilon$ with a large error, the confidence region from the parametric model was found to be smaller than that of the nonparametric model. Thus, because of error contamination, it cannot be concluded that nonparametric regression provides the superior solution. In this instance, a linear parametric model appears correct, although the data came from a quadratic model with an error term.

Therefore, if data come from a clearly nonlinear model, the nonparametric regression confidence regions will be smaller than the parametric. If the data come from a linear model, or a nonlinear model with error contamination, the parametric regression confidence regions will be smaller. Consequently, a comparison of linear and nonparametric regression confidence intervals can assist in evaluating the appropriateness of a linear model.

The calculation of confidence intervals for a nonparametric regression thus provides a way of supplementing the analysis of residuals as a check on the appropriateness of the model. Often in parametric regression, residuals are visually inspected to reveal inadequacies in model selection or to support the choice of a linear or a nonlinear model. Unfortunately, those residuals cannot be checked against many variables simultaneously when a model with a large number of dimensions is considered. Nonparametric regression is a more quantitative objective technique, particularly useful when dealing with higher dimension data.

A Monte Carlo simulation was undertaken to check the confidence regions. A thousand sets of 30 values from $y = 3x + \epsilon$ with a small error were simulated and then both parametric and nonparametric regression applied. The value of Y at X equal to 1.5 was found for every simulation and confidence regions were built. The obtained confidence intervals for the parametric regression corresponded exactly to the value in Table 5.1, while for the nonparametric regression the length of the interval was 0.045 and not 0.15. Thus, further investigations and refinements are required to this technique.

Hence, a similar analysis was performed on the New Brunswick flood data. For both the linear and the nonparametric regressions of the 2-year flood against drainage area, confidence regions were computed by bootstrapping at the center of the regressions. Similarly, confidence regions were found for the relationship of the 2-year flood as a function of both drainage area and mean annual precipitation. As shown in Table 5.2, the linear regression confidence intervals are smaller, leading to the conclusion that a linear relationship is adequate in these instances. This should, however, be checked with another approach for confidence interval determination.

Table 5.2: 68 Percent Confidence Intervals at Center of Regression for 2 Year Flood Equations

	Linear Model at X=2.593	Nonparametric Model at X=2.593	at X=2.625
$logQ$ vs $logDA$	(1.911,1.938)	(1.881,1.959)	(1.908,1.959)
	l=0.03	l=0.08	l=0.05
$logQ$ vs $logDA$ & $logMAP$	(1.918,1.928)	(1.74,2.15)	(1.95,1.78)
	l=0.01	l=0.41	l=0.17

Note: The confidence intervals for the linear model at X=2.625 are almost identical to the values at X=2.593

l is the interval length

Q = 2 year flood; DA = area; MAP = precipitation

5.5 Comparison of Parametric and Nonparametric Regression

Based on the results of the previous two sections, it was concluded that a linear regression of logarithmically transformed variables adequately describes the regional relationships in New Brunswick. The computation of confidence intervals at the center of linear and nonparametric regressions assisted in evaluating the appropriateness of the linear model and is considered a more quantitative objective technique than visual inspection of the residuals. The use of the more sophisticated nonparametric regression does not lead to a significant improvement over linear regression and is thus not adding to the confidence of the latter. It is therefore not required for the development of regional flood equations in New Brunswick.

However, nonparametric regression can serve as a useful screening tool. In situations where there exists a lack of data, which are situations where a regression of any kind should not be used, nonparametric regression will point out deficiencies. This makes it a valuable screening tool in eliminating the formulation of irrational relationships and of relationships where data are lacking or insufficient.

5.6 Proposed New Methodology

The results of this chapter and the preceding one have shown how nonparametric methods can point out and improve over the limitations of the various parametric methods currently employed in regional flood frequency analysis. If regional analysis is performed with flood data coming from a well-known unimodal distribution and whose regional relationship is clearly linear and for which there is a sufficient spread of data, both parametric and nonparametric methods will yield almost identical regional estimates.

But if the data come from a bimodal or a heavy-tailed distribution, as is often the case in Canada, or if the regional relationship is actually nonlinear, or if the spread of data is uneven, then only through the use of nonparametric methods will the correct at-site flood frequencies and the correct regional relationships be found.

Therefore, based on its successful application to New Brunswick annual maximum floods in this thesis, the following step-by-step methodology for developing regional relationships by the multiple linear regression approach is proposed:

1. Following preliminary data screening, perform nonparametric frequency analysis using the fixed kernel method. Plot all probability density functions and attempt classification based on shape.

2. If the density shapes are all or mostly unimodal, perform L-moment

analysis in order to test for distribution homogeneity and to find a corresponding theoretical unimodal distribution describing the data. Division into smaller homogeneous regions may then proceed based on climatic or physiographic factors.

If the density shapes indicate mixed distributions, investigate the date of occurrence of the floods in order to determine whether a number of flood generating mechanisms are at work. With the assistance of the density function shapes and hydroclimatic information, delineate homogeneous regions. Perform the L-moment heterogeneity tests to determine whether the selected regions actually contain stations following the same distribution. Further division may or may not then be considered.

3. Find regional relationships for the selected regions using both parametric and nonparametric regressions. Anomalies in the nonparametric regressions will point out deficiencies in the relationship development procedure. If a visual inspection of the nonparametric regression suggests it is more appropriate than a linear regression, compute the confidence regions at the center of both regressions in order to evaluate the suitability of the linear model.

The above proposed methodology will ensure that an inappropriate flood frequency distribution is not selected; that homogeneous regions containing floods dependent on different processes, and thus having different distribution shapes, are not placed together; and that a linear relationship is not chosen when a nonlinear model is clearly appropriate.

Chapter 6

CONCLUSIONS

6.1 Summary of the Study

Nonparametric methods were employed at the three steps of regional analysis on a set of New Brunswick annual maximum floods. Nonparametric frequency analysis indicated the presence of some unimodal, some heavy-tailed and many bimodal distributions. Simulation studies showed that the bimodality could not be accounted for by sampling variability from a unimodal distribution, and thus is a result of naturally occuring generating mechanisms. Therefore, some of the New Brunswick floods follow mixed distributions.

Homogeneous regions were delineated based on flood density shape, and seasonal partitioning of the annual maximum floods confirmed this delineation. Heterogeneity tests from L-moment analysis also indicated that the entire flood data set from the province of New Brunswick did not come from a single distribution, but that the flood data from a delineated bimodal homogeneous region followed the same distribution. L-moment analysis proposed, however, the Generalized Extreme Value or the Generalized Logistic as the distribution describing the data from the homogeneous bimodal region. Nonparametric frequency analysis showed this to be improper, as mixed distributions were found.

Regression analysis was performed on a province-wide basis and for various regions, including the homogeneous bimodal region. Lower standard errors were obtained for the homogeneous bimodal region as opposed to the province-wide equation for large return periods even though the former contained a smaller quantity of data. This added further support to the delineation of homogeneous regions.

Both linear and nonparametric regressions were employed for regional relationship development. Usually lower integral square error (ISE) resulted for the nonparametric regression even though both regressions usually provided relationships not very different from each other. Nonparametric regression was found to be a useful screening tool, being able to detect instances where relationship development should not be attempted.

Simulation studies showed that the ISE criterion was inadequate. Using

bootstrapping pairs, a new method of computing confidence regions at the center of a nonparametric regression was investigated. The method showed that a linear regional relationship was adequate for New Brunswick floods. Based on its successful application in New Brunswick, a new methodology incorporating nonparametric methods at all three steps of regional flood frequency analysis was proposed. This approach lacks the many limitations of the parametric methods.

6.2 Major Conclusions

Based on the results of this thesis, many conclusions were reached with regard to the New Brunswick annual maximum floods. The flood data from that province come from three kinds of distributions of unimodal, bimodal and heavy-tailed shapes. The unimodal LN3 distribution used in a previous study is clearly improper in describing all of the New Brunswick data. Basins with floods of similar distribution shape, which reflect similar flood generating mechanisms, were successfully grouped on a geographical basis to form homogeneous regions.

Lower ISE's were found for the regional equations of a homogeneous bimodal region as opposed to province-wide equations for large return periods. This lends support to the homogeneous region delineation. After logarithmic transformation, a linear relationship was found to be acceptable between New Brunswick floods and physiographic/climatic variables.

A number of general conclusions about nonparametric methods having wide implications for hydrology were obtained:

1. Nonparametric frequency analysis can be used to determine the shape of the generating distribution in a given region. It will confirm the unimodal distribution selection of a parametric method such as the L-moment technique when the data are unimodal. More importantly, it will also point out a mixed distribution while the L-moment technique cannot. Thus, it is proposed that any parametric frequency analysis always be combined with nonparametric frequency analysis for confirmation.

2. Flood distribution shape, as represented by the nonparametric frequency density plot, is related to flood generation mechanisms. Floods of similar density shape can be grouped on a geographical basis to delineate homogeneous regions.

3. Using bootstrapping pairs, it is possible to obtain confidence regions at the center of a nonparametric regression. A comparison of linear and nonparametric regression confidence regions can then be used to test the appropriateness of a linear model and to assess the need for a nonparametric regression.

4. A new methodology for regional analysis involving nonparametric methods at all stages was formulated and successfully tested.

6.3 Recommendations for Further Research

The following areas are recommended for further research:

1. The new formulated methodology should be applied to flood regions in different parts of the country and in different parts of the globe where the flood generating mechanisms are not the same as in New Brunswick. This should confirm its general applicability and most likely discover further mixed flood distributions.

2. The new methodology may also be applied to a regional low flow study in order to evaluate its suitability to low flows, for which seasonal occurrences also take place.

3. Means of obtaining confidence regions for nonparametric flood estimates using resampling schemes such as the bootstrap should be investigated. Further investigations on confidence regions for nonparametric regression should be pursued.

4. It was shown that the TCEV could not describe New Brunswick floods parametrically. Other mixed or compounded parametric distributions could be investigated to describe New Brunswick floods.

5. Means of transferring historical information on a regional basis should be explored, as at-site historical information has been shown to greatly improve the accuracy of large return period flood estimates.

6. A robust mathematical test should be formulated to classify the various density shapes. It must be able to differentiate between a bimodal distribution with two high distinct peaks and a unimodal distribution with a heavy tail. It must also be able to take into account the impact of sampling variability on distribution shape.

REFERENCES

Adamowski, K., "Nonparametric Kernel Estimation of Flood Frequencies", Water Resources Research, 21(11), 1585-1590, 1985.

Adamowski, K., "Nonparametric Techniques for Analysis of Hydrological Events". Water for the Future: Hydrology in Perspective (Proceedings of the Rome Symposium, April 1987), IAHS Publication no. 164, 1987.

Adamowski, K., "A Monte Carlo Comparison of Parametric and Nonparametric Estimation of Flood Frequencies", Journal of Hydrology, 108, 295-308, 1989.

Adamowski, K. and A.C. Middleton, "Steady-State Dissolved Oxygen Model for the Rideau River", Canadian Journal of Civil Engineering, 4(4), 471-481, 1977.

Adamowski, K. and W. Feluch, "Nonparametric Flood Frequency Analysis with Historical Information", ASCE Journal of Hydraulic Engineering, vol. 116, no.8, 1035-1047, 1990.

Adamowski, K. and W. Feluch, "Application of Nonparametric Regression to Groundwater Level Prediction", Canadian Journal of Civil Engineering, 18, 600-606, 1991.

Ahmad, M.I., C.D. Sinclair and A. Werritty, "Log-logistic Flood Frequency Analysis", Journal of Hydrology, 98, 205-224, 1988.

Alila, Y., Nonparametric Flood Frequency Analysis with Historic Information and Hydroclimatically Defined Mixed Distributions, M.A.Sc. Thesis, University of Ottawa, Ottawa, 1988.

Bardsley, W.E., "Toward a General Procedure for Analysis of Extreme Random Events in the Earth Sciences", Mathematical Geology, 20(5), 513-528, 1988.

Bardsley, W.E., "Comment on 'Nonparametric Kernel Estimation of Flood Frequencies' by Kaz Adamowski", Water Resources Research, 24(6), 890, 1988.

Bardsley, W.E., "A Simple Parameter-Free Flood Magnitude Estimator", Hydrological Sciences Journal, 34(2), 129-137, 1989.

Bardsley, W.E., "Using Historical Data in Nonparametric Flood Estimation", Journal of Hydrology, 108, 249-255, 1989.

Bhaskar, N.R. and C.A. O'Connor, "Comparison of Method of Residuals and Cluster Analysis for Flood Regionalization", ASCE Journal of Water

Resources Planning and Management, 115(6), 793-808, 1989.

Burn, D.H., "Delineation of Groups for Regional Flood Frequency Analysis", Journal of Hydrology, 104, 345-361, 1989.

Burn, D.H., "An Appraisal of the Region of Influence Approach to Flood Frequency Analysis," Hydrological Sciences Journal, 35(2), 149-165, 1990.

Caissie, D. and N. El-Jabi, "A Stochastic Study of Floods in Canada: Frequency Analysis and Regionalization", Canadian Journal of Civil Engineering, 18, 225-236, 1991.

Cavadias, G.S., "The Canonical Correlation Approach to Regional Flood Estimation", in Regionalization in Hydrology, IAHS Publication no. 191, April 1990.

Chow, K.C.A. and W.E. Watt, "A Knowledge-Based Expert System for Flood Frequency Analysis", Canadian Journal of Civil Engineering, 17, 597-609, 1990.

Cunnane, C., "Review of Statistical Models for Flood Frequency Estimation", in Hydrologic Frequency Modeling, edited by V.P. Singh, 49-96, 1987.

Diaconis, P. and B. Efron, "Computer-intensive Methods in Statistics", Scientific American, 248(5), 116-130, 1983.

Diehl, T. and K.W. Potter, "Mixed Flood Distributions in Wisconsin", in Hydrologic Frequency Modeling edited by V.P. Singh, 213-226, 1987.

Draper, N.R. and H. Smith, Applied Regression Analysis, John Wiley and Sons, 1966.

Efron, B., The Jackknife, the Bootstrap and Other Resampling Plans, Society for Industrial and Applied Mathematics, 1982.

Fiering, M.B., Streamflow Synthesis, Harvard University Press, 1967.

Fiorentino, M., S. Gabriele, F. Rossi and P. Versace, "Hierarchical Approach for Regional Flood Frequency Analysis," in Regional Flood Frequency Analysis edited by V.P. Singh, 35-49, 1987.

Fleming, G., Computer Simulation Techniques in Hydrology, Elsevier Environmental Sciences Series, 1975.

Freedman, D.A., "Bootstrapping Regression Models", The Annals of Statistics, 9, 1218-1228, 1981.

Gabriele, S. and N. Arnell, "A Hierarchical Approach to Regional Flood Frequency Analysis", Water Resources Research, 27(6), 1281-1289, 1991.

Gingras, D. and K. Adamowski, "Coupling of Nonparametric Frequency and L-Moment Analysis for Mixed Distribution Identification", Water Resources Bulletin, 28(2), 263-272, 1992.

Gingras, D. and K. Adamowski, "Homogeneous Region Delineation Based on Annual Flood Generation Mechanisms", accepted by the Hydrological Sciences Journal, 1992.

Gingras, D., K. Adamowski and M. Alvo, "Regional Flood Relationships by Nonparametric Regression", submitted to Nordic Hydrology, 1992.

Gingras, D.G., J.D. Keefe and B.C. Burrell, "Regional Flood Frequency Analysis Using a Concurrent Period of Record", Canadian Water Resources Journal, 13(1), 20-25, 1988.

Guo, S.L., "Nonparametric Variable Kernel Estimation with Historical Floods and Paleoflood Information", Water Resources Research, 27(1), 91-98, 1991.

Haan, C.T., Statistical Methods in Hydrology, Iowa State University Press, 1977.

Hardle, W. and J.S. Marron, "Bootstrap Simultaneous Error Bars for Nonparametric Regression", The Annals of Statistics, 19, 778-796, 1991.

Harvey, K.D., R. Condie and P.J. Pilon, "Regional Flood Frequency Analysis with the Three-Parameter Lognormal Distribution", Proceedings CSCE 7th Canadian Hydrotechnical Conference, Saskatoon, 1985.

Hirschboeck, K.K., "Hydroclimatically-Defined Mixed Distributions in Partial Duration Flood Series", in Hydrologic Frequency Modeling edited

by V.P. Singh, 199-212, 1987.

Holder, R.L., Multiple Regression in Hydrology, Institute of Hydrology, Wallingford, United Kingdom, 1985.

Hosking, J.R.M., "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics", Journal of Royal Statistical Society, B, 52(1), 105-124, 1990.

Hosking, J.R.M. and J.R. Wallis, Regional Flood Frequency Analysis Using L-Moments, IBM Research Report RC 15658, Yorktown Heights, N.Y., 1990.

Hosking, J.R.M. and J.R. Wallis, Some Statistics Useful in Regional Frequency Analysis, IBM Research Report RC 17096, Yorktown Heights, N.Y., 1991.

Hosking, J.R.M., J.R. Wallis and E.F. Wood, "An Appraisal of the Regional Flood Frequency Procedure in the UK Flood Studies Report", Hydrological Sciences Journal, 30(1), 85-109, 1985.

Houghton, J.C., "Birth of a Parent: The Wakeby Distribution for Modeling Flood Flows", Water Resources Research, 14(6), 1105-1109, 1978.

Inland Waters/Lands Directorate, Environment Canada and the New Brunswick Department of Municipal Affairs and Environment, Flood Frequency Analyses New Brunswick, April 1987.

Inland Waters/Lands, Environment Canada and the New Brunswick Department of Municipal Affairs and Environment, New Brunswick Hydrometric Network Evaluation, Background Report no.2, Delineation of Physiographic-Climatic Zones, April 1988a.

Inland Waters/Lands, Environment Canada and the New Brunswick Department of Municipal Affairs and Environment, New Brunswick Hydrometric Network Evaluation, Background Report No. 3, Regional Hydrology and Data Transfer Methodologies, April 1988b.

Inland Waters Directorate, Environment Canada and New Brunswick Department of Municipal Affairs and Environment, New Brunswick Hydrometric Network Evaluation: Summary Report, January 1989.

Inland Waters Directorate - Atlantic Region, Streamflow Estimation Guidelines for Prince Edward Island, January 1989.

Interagency Working Group, Regional Flood Frequency Analysis for the Island of Newfoundland, a joint report of the Newfoundland Department of Environment and Environment Canada - Atlantic Region, 1984.

Jin, M. and J.R. Stedinger, "Flood Frequency Analysis with Regional and Historical Information", Water Resources Research, 25(5), 925-936, 1989.

Kite, G.W., Frequency and Risk Analysis in Hydrology. Water Resources Publications, Colorado, U.S.A., 1977.

Klemes, V., "Hydrological and Engineering Relevance of Flood Frequency Analysis", in Hydrologic Frequency Modeling edited by V.P. Singh. 1-18, 1987.

Labatiuk, C.W., A Nonparametric Approach to Flood Frequency Analysis, M.A.Sc. Thesis, University of Ottawa, Ottawa, 1985.

Labatiuk, C. and K. Adamowski, "Application of Nonparametric Density Estimation to Computation of Flood Magnitude/Frequency", in Stochastic Hydrology edited by I.B. MacNeill and G.J. Umphrey, 161-180, 1987.

Lettenmaier, D.P. and K.W. Potter, "Testing Flood Frequency Estimation Methods Using a Regional Flood Generation Model", Water Resources Research, 21(12), 1903-1914, 1985.

Linsley, R.K., "Flood Estimates. How Good Are They?" Water Resources Research, 22(9), 159-164, 1982.

Moin, S.M.A. and M.A. Shaw, Regional Flood Frequency Analysis for Ontario Streams, three volumes, Canada/Ontario Flood Damage Reduction Program, Environment Canada, November 1986.

Muller, H.G., Lecture Notes in Statistics Series: Nonparametric Regression Analysis of Longitudinal Data, Springer-Verlag, 1988.

National Research Council of Canada - Associate Committee on Hydrology, Hydrology of Floods in Canada: A Guide to Planning and Design,

Ottawa, 1989.

Natural Environment Research Council, Flood Studies Report, Institute of Hydrology, United Kingdom, 1975.

Pilon, P.J. and K. Adamowski, "The Value of Regional Information to Flood Frequency Analysis Using the Method of L-Moments", Canadian Journal of Civil Engineering, 19, 137-147, 1992.

Pilon, P.J. and K. Adamowski, "Estimation of the Asymptotic Variance of the T-Year Event With Historical Information in the Log Pearson Type III Distribution", accepted for publication by the Journal of Hydrology, 1992.

Pilon, P.J., K. Adamowski and Y. Alila, "Regional Analysis of Annual Maxima Precipitation Using L-Moments", Atmospheric Research, 27, S1-92, 1991.

Pilon, P.J., Y. Alila and K. Adamowski, "Assessment of Risk of Flooding Based on Regional Information", NATO/ASI series by Springer Verlag, 1992.

Pilon, P.J., R. Condie and K.D. Harvey, Consolidated Frequency Analysis Package - CFA - User Manual for Version 1 - DEC Pro Series, Water Resources Branch, Inland Waters Directorate, Environment Canada, Ottawa, 1985.

Prakasa Rao, B.L.S., Nonparametric Functional Estimation, Academic Press, 1983.

Riggs, H.C., Streamflow Characteristics, Development in Water Science 22, Elsevier, 1985.

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

Rudemo, M., "Empirical Choice of Histogram and Kernel Density Estimation", Scandinavian Journal of Statistics, 9, 65-78, 1982.

Sangal, B. and A.K. Biswas, "The Three-Parameter Lognormal Distribution and its Application in Hydrology", Water Resources Research, 6(2), 505-515,1970.

Schaefer, M.G., "Regional Analyses of Precipitation Annual Maxima in Washington State", Water Resources Research, 26(1), 119-131, 1990.

Scott, D.W. and G.R. Terrell, "Biased and Unbiased Cross-validation in Density Estimation", Journal of the American Statistical Association, 82, 1131-1146, 1987.

Silverman, B.W., "Using Kernel Density Estimates to Investigate Multimodality", Journal of Royal Statistical Society, B, 43(1), 97-99, 1981.

Silverman, B.W., Density Estimation for Statistics and Data Analysis, Chapman and Hall, 1986.

Singh, K.P., "Development of a Versatile Flood Frequency Methodology and its Application to Flood Series from Different Countries", in Hydrologic Frequency Modeling by V.P. Singh, 183-197, 1987.

Smith, J.A., "Long-Range Streamflow Forecasting Using Nonparametric Regression", Water Resources Bulletin, 27(1), 39-46, 1991.

Stoddart, R.B.L., and W.E. Watt, Flood Frequency Prediction for Intermediate Drainage Basins in Southern Ontario, Civil Engineering Research Report No. 66, Queen's University, Kingston, 1970.

Tapia, R.A., and J.R. Thompson, Nonparametric Probability Density Estimation, The John Hopkins University Press, 1978.

Tasker, G.D., and J.R. Stedinger, "An Operational GLS Model for Hydrologic Regression", Journal of Hydrology, 111, 361-375, 1989.

U.S. Water Resources Council, Guidelines for Determining Flood Flow Frequency, Bulletin 17B, Washington D.C., 1982.

Viessman, W., G.L. Lewis and J.W. Knapp, Introduction to Hydrology, Harper and Row, New York, 1989.

Wallis, J.R., Regional Frequency Studies Using L-Moments, IBM Re-

search Report R.C. 14597, 1989.

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

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

Wetherill, G.B., Regression Analysis with Applications, Chapman and Hall, 1986.

Wiltshire, S.E., "Grouping Basins for Regional Flood Frequency Analysis", Hydrological Sciences Journal, 30(1), 151-159, 1985.

Wu, K. and M.K. Woo, "Estimating Annual Flood Probabilities Using Fourier Series Method," Water Resources Bulletin, vol.25, no.4, 743-750, 1989.

Appendix A

NEW BRUNSWICK HYDROMETRIC STATION INFORMATION AND DENSITY PLOTS

Table A.1: New Brunswick Hydrometric Stations

Station Number	Name	Record (years)	Area (km^2)
01AD003	St Francis R. at Outlet of	1952-88	1350
.	Glasier Lake	.	.
01AF003	Green R. at Riviere-Verte	1963-88	1150
01AG002	Limestone R. at Four Falls	1968-88	199
01AG003	Aroostook R. near Tinker	1975-88	6060
01AH005	Mamozekel R. near	1973-88	230
.	Campbell River	.	.
01AJ003	Meduxnekeag R. near	1968-88	1210
.	Belleville	.	.
01AJ004	Big Presque Isle Stream	1968-88	484
.	at Tracey Mills (H)	.	.
01AJ010	Becaquimec Stream	1974-88	350
.	at Coldstream	.	.
01AJ011	Cold Stream at Coldstream	1974-88	156
01AK001	Shogomoc Stream near	1919-40	234
.	Trans Canada Highway	1944-88	.
01AK005	North Nashwaaksis Stream	1966-88	26.9
.	near Royal Road	.	.
01AK007	Nackawic R. near	1968-88	240
.	Temperance Vale	.	.

Table A.1: New Brunswick Hydrometric Stations (continued)

Station Number	Name	Record (years)	Area (km^2)
01AK00S	Eel R. near Scott Siding	1974-SS	531
01AL002	Nashwaak R. at Durham	1962-SS	1450
.	Bridge	.	.
01AL004	Narrows Mountain Brook	1972-SS	3.S9
.	near Narrows Mountain	.	.
01AM001	Northwest Oromocto R.	1963-SS	557
.	at Tracy	.	.
01AN001	Castaway Brook near	1972-SS	34.4
.	Castaway	.	.
01AN002	Salmon R. near Castaway	1974-SS	1050
01AP002	Canaan R. at East Canaan	1926-40	66S
.	.	1963-SS	.
01AP004	Kennebecasis R. at	1962-SS	1100
.	Apohaqui	.	.
01AP006	Nerepis R. near Fowlers	1976-88	293
.	Corner	.	.
01AQ001	Lepreau R. at Lepreau	1919-88	239
01AQ002	Magaguadavic R. at	1919-22	1420
.	Elmcroft	1924-32	.
.	.	1944-88	.

Table A.1: New Brunswick Hydrometric Stations (continued)

Station Number	Name	Record (years)	Area (km^2)
01AR006	Dennis Stream near	1967-88	115
.	St. Stephen (H)	.	.
01AR008	Bocabec R. above Tide	1967-79	43.0
01BC001	Restigouche R. below	1963-88	3160
.	Kedgwick River	.	.
01BE001	Upsalquitch R. at	1919-32	2270
.	Upsalquitch	1944-88	.
01BJ001	Tetagouche R. near West	1923-32	363
.	Bathurst	1952-88	.
01BJ003	Jacquet R. near Durham	1965-88	510
.	Centre	.	.
01BJ004	Eel R. near Eel River	1968-83	88.6
.	Crossing	.	.
01BJ007	Restigouche R. above	1969-88	7740
.	Rafting Ground Brook	.	.
01BK004	Nepisiquit R. near	1958-74	2090
.	Pabineau Falls	.	.
01BL001	Bass R. at Bass River	1966-88	175
01BL002	Southwest Caraquet R. at	1970-88	173
.	Burnsville	.	.

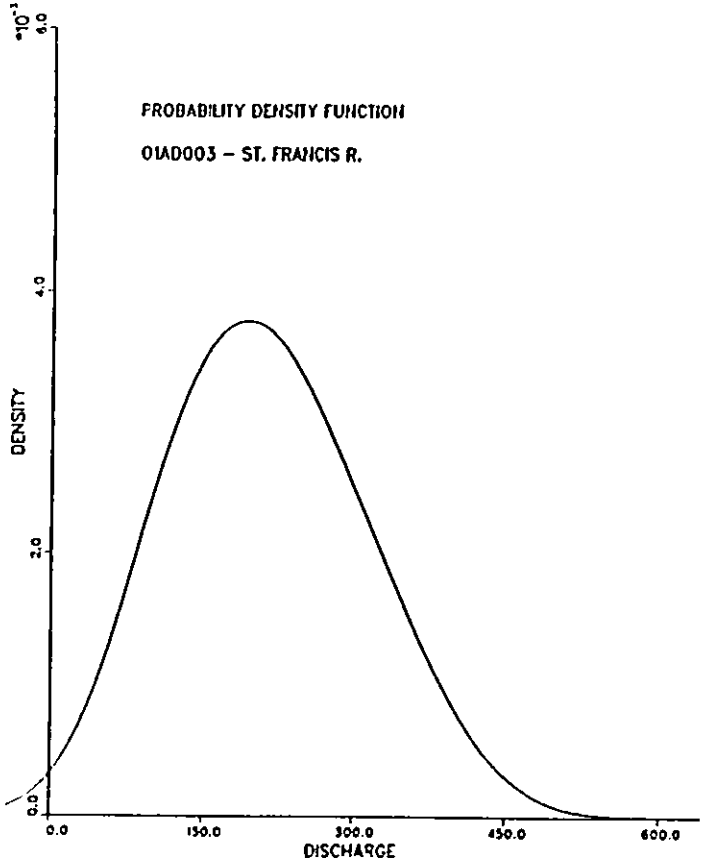
Table A.1: New Brunswick Hydrometric Stations (continued)

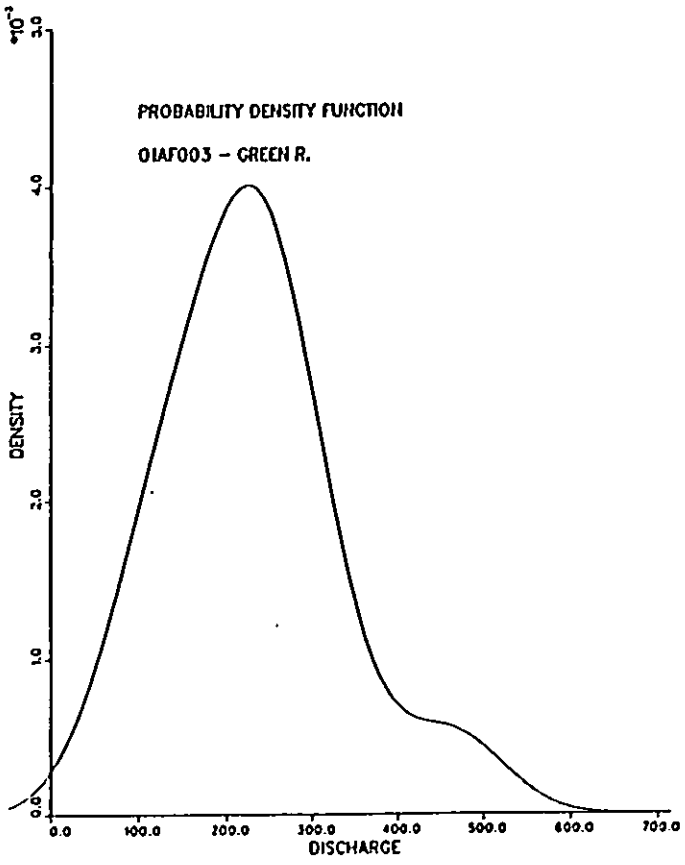
Station Number	Name	Record (years)	Area (km^2)
01BL003	Tracadie R. at Murphy	1971-88	383
.	Bridge Crossing	.	.
01BO001	Southwest Miramichi R.	1919-32	5050
.	at Blackville	1962-88	.
01BO002	Renous R. at McGraw	1966-88	611
.	Brook	.	.
01BO003	Barnaby R. below	1973-88	484
.	Semiwagan River	.	.
01BP001	Little Southwest Miramichi	1952-88	1340
.	R. at Lyttleton	.	.
01BQ001	Northwest Miramichi R. at	1962-88	948
.	Trout Brook (H)	.	.
01BR001	Kouchibouguac R. near	1931-32	177
.	Vautour	1970-88	.
01BS001	Coal Branch R. at	1965-88	166
.	Beersville	.	.
01BU002	Petitcodiac R. near	1962-88	391
.	Petitcodiac	.	.
01BU003	Turtle Creek near Turtle	1963-88	129
.	Creek	.	.

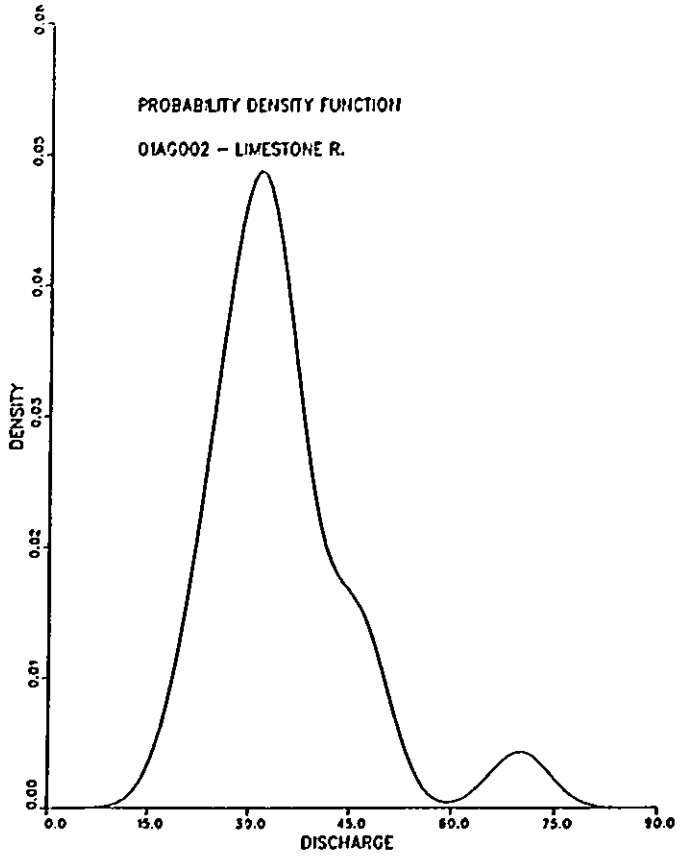
Table A.1: New Brunswick Hydrometric Stations (continued)

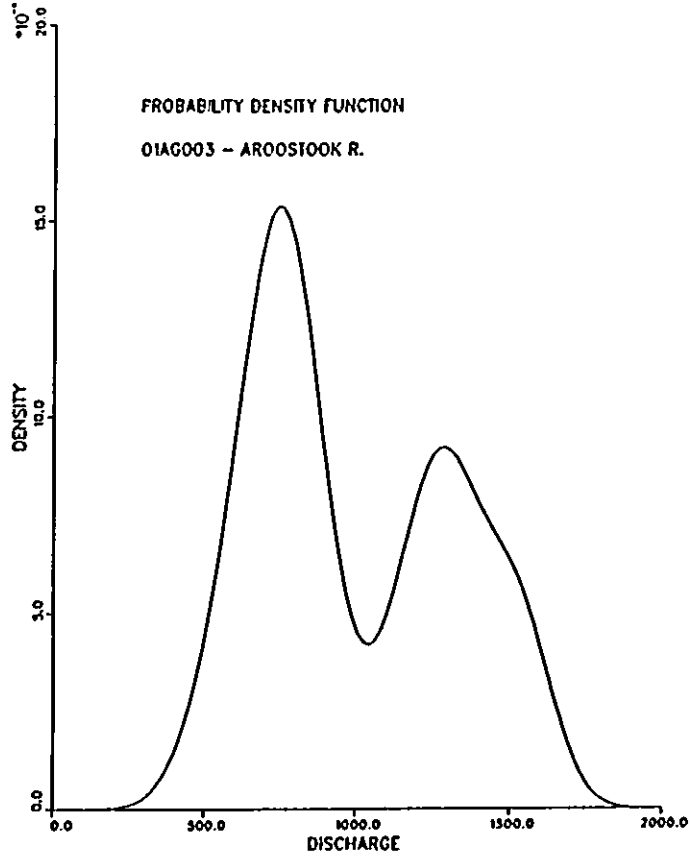
Station Number	Name	Record (years)	Area (km^2)
01BU004	Palmer's Creek near	1967-85	34.2
.	Dorchester	.	.
01BV005	Ratcliffe Brook below	1961-71	29.3
.	Otter Lake	.	.
01BV006	Point Wolfe R. at Fundy	1965-88	130
.	National Park	.	.
01BV007	Upper Salmon R. at Alma	1968-78	181
1013500	Fish R. near Fort Kent,	1904-08	2290
.	Maine, U.S.A.	1930-89	.
1016500	Machias R. near Ashland,	1952-83	855
.	Maine, U.S.A.	.	.
01BD002	R. Matapedia en amont	1970-88	2770
.	de la riviere Assemetquagan,	.	.
.	Quebec	.	.
01BF001	R. Nouvelle au pont,	1964-88	1140
.	Quebec	.	.
01DL001	Kelley R. at Eight Mile	1970-88	63.2
.	Ford, Nova Scotia	.	.
.	(H) - Historical information available		.

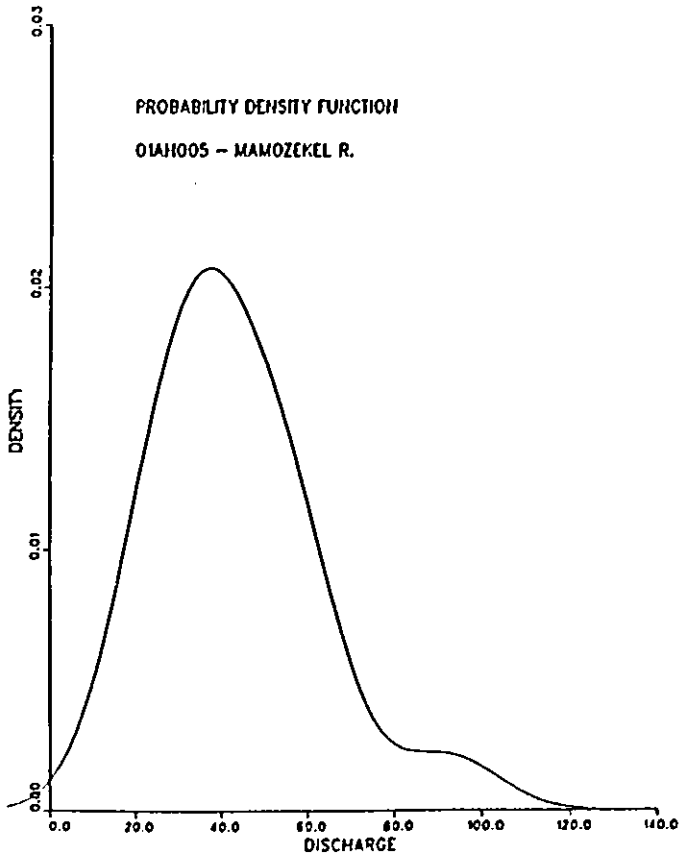
NEW BRUNSWICK PROBABILITY DENSITY PLOTS

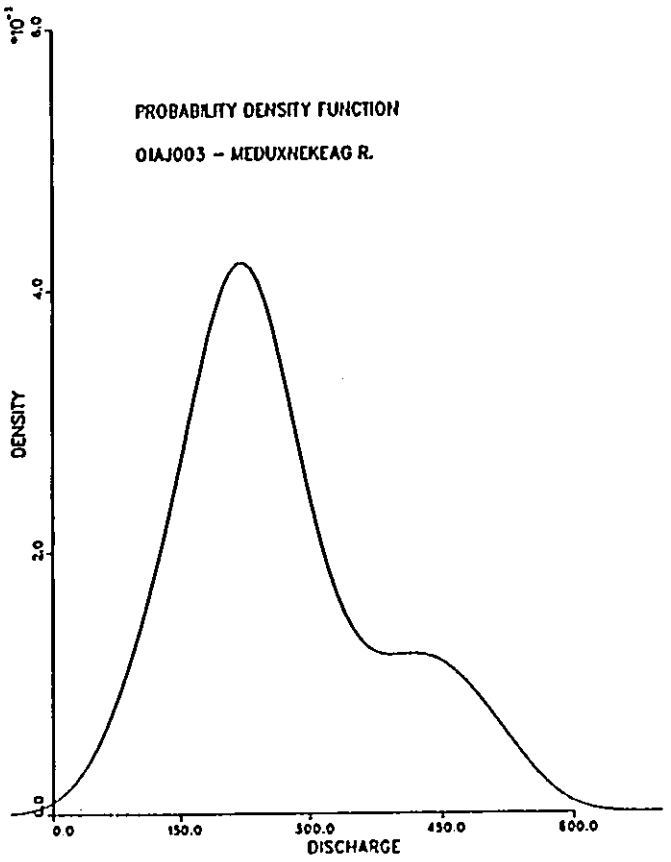

PROBABILITY DENSITY FUNCTION
01AD003 – ST. FRANCIS R.
*10⁻³
6.0
4.0
2.0
0.0
DENSITY
0.0
150.0
300.0
450.0
600.0
DISCHARGE

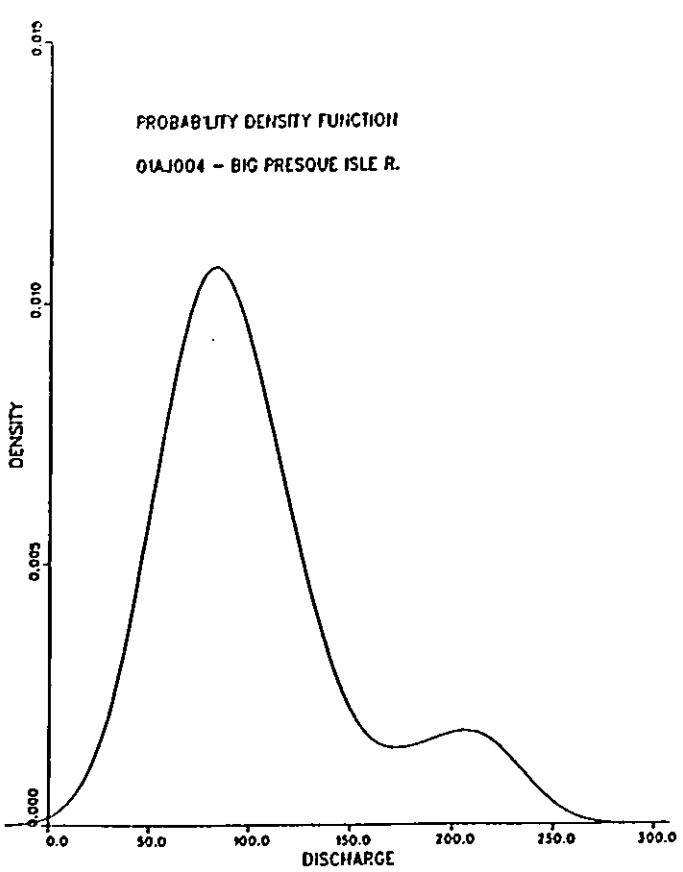

PROBABILITY DENSITY FUNCTION
01AF003 – GREEN R.
*10⁻³
4.0
3.0
2.0
1.0
0.0
DENSITY
0.0
100.0
200.0
300.0
400.0
500.0
600.0
700.0
DISCHARGE

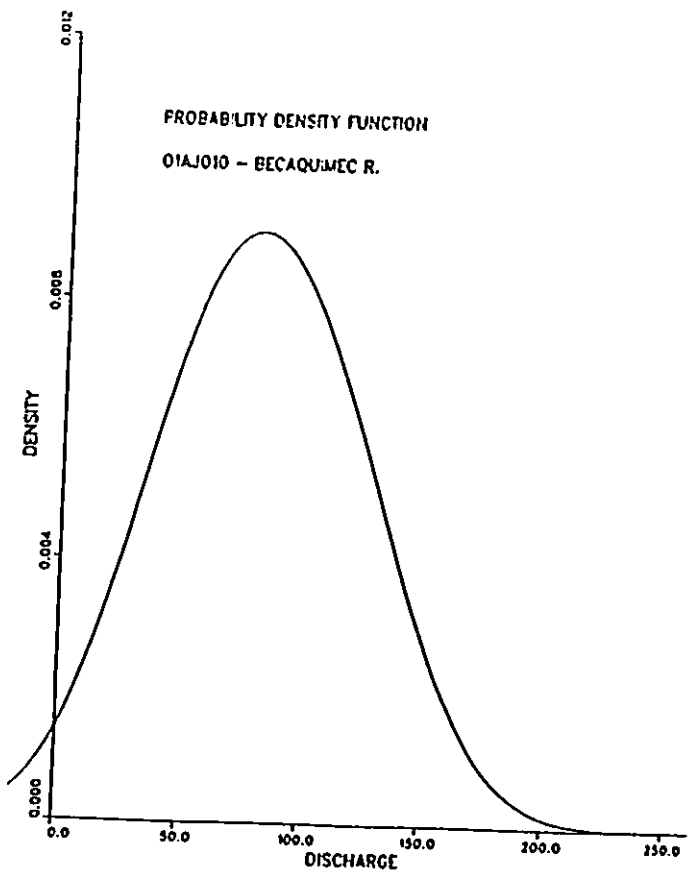

PROBABILITY DENSITY FUNCTION
01AG002 – LIMESTONE R.
DENSITY
0.00
0.01
0.02
0.03
0.04
0.05
0.0
15.0
30.0
45.0
60.0
75.0
90.0
DISCHARGE

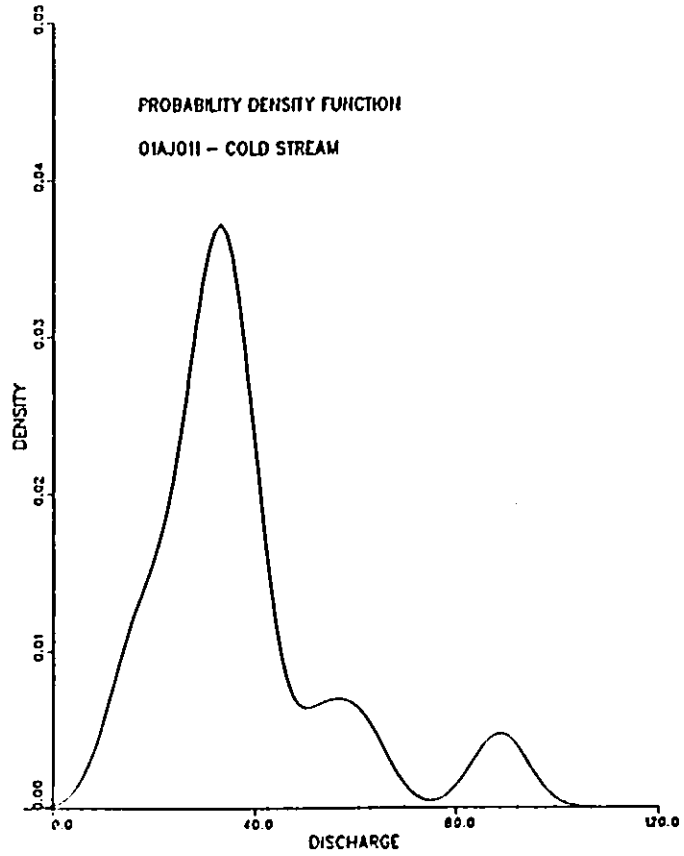

PROBABILITY DENSITY FUNCTION
01AG003 – AROOSTOOK R.
DENSITY
0.0
5.0
10.0
15.0
20.0
0.0
500.0
1000.0
1500.0
2000.0
DISCHARGE

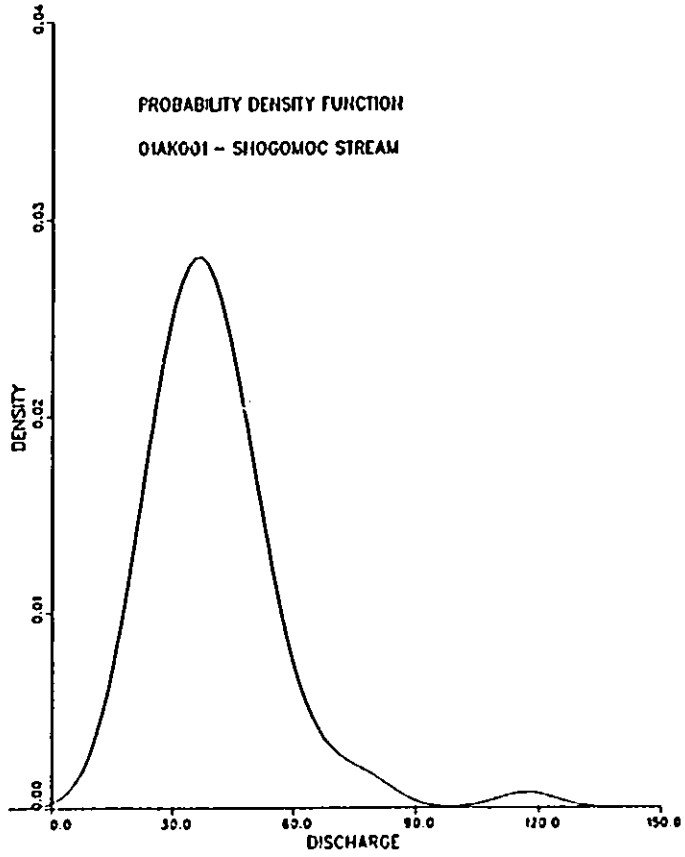

PROBABILITY DENSITY FUNCTION
01AH005 – MAMOZEKEL R.
DENSITY
0.03
0.02
0.01
0.00
0.0
20.0
40.0
60.0
80.0
100.0
120.0
140.0
DISCHARGE

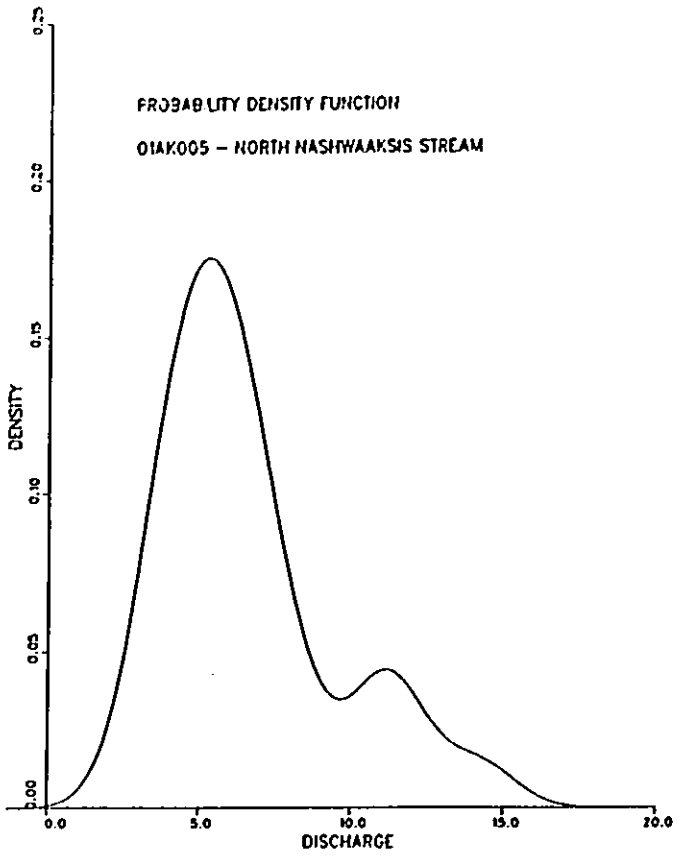

PROBABILITY DENSITY FUNCTION
01AJ003 – MEDUXNEKEAG R.
*10⁻³
6.0
4.0
2.0
0.0
DENSITY
0.0
150.0
300.0
450.0
600.0
DISCHARGE

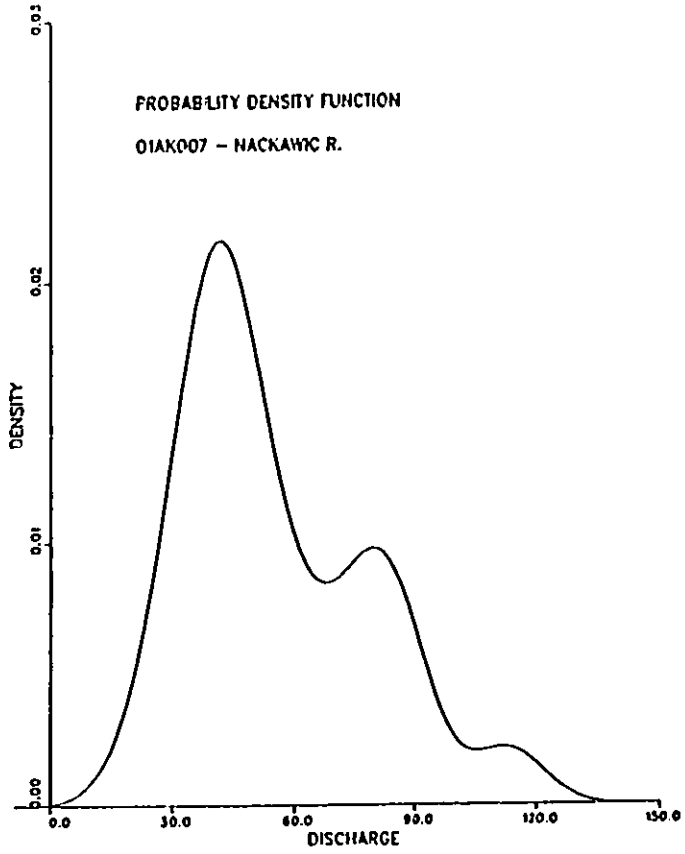

PROBABILITY DENSITY FUNCTION
01AJ004 – BIG PRESQUE ISLE R.
DENSITY
0.015
0.010
0.005
0.000
0.0
50.0
100.0
150.0
200.0
250.0
300.0
DISCHARGE

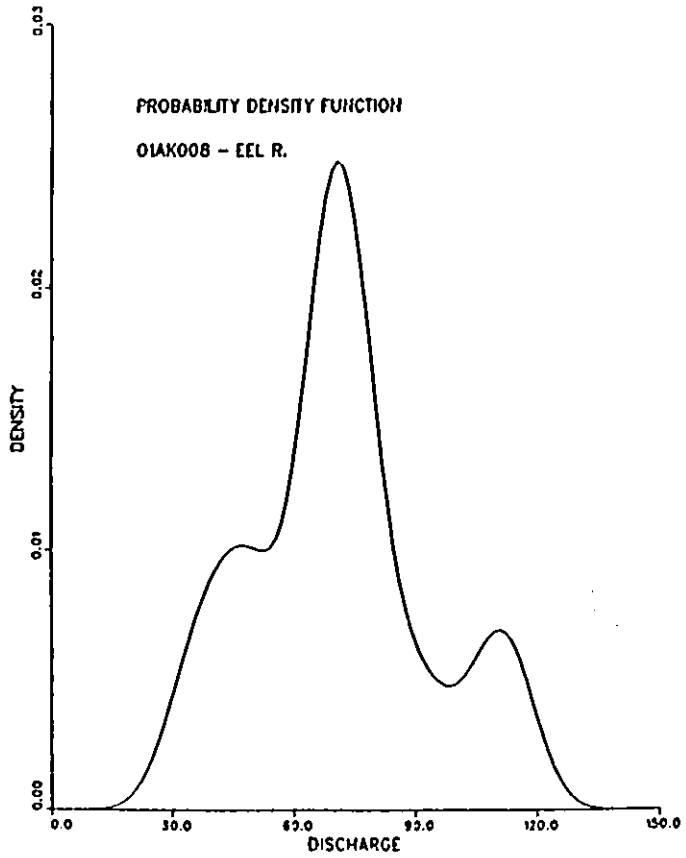

PROBABILITY DENSITY FUNCTION
01AJ010 – BECAQUIMEC R.
DENSITY
0.012
0.008
0.004
0.000
0.0
50.0
100.0
150.0
200.0
250.0
DISCHARGE

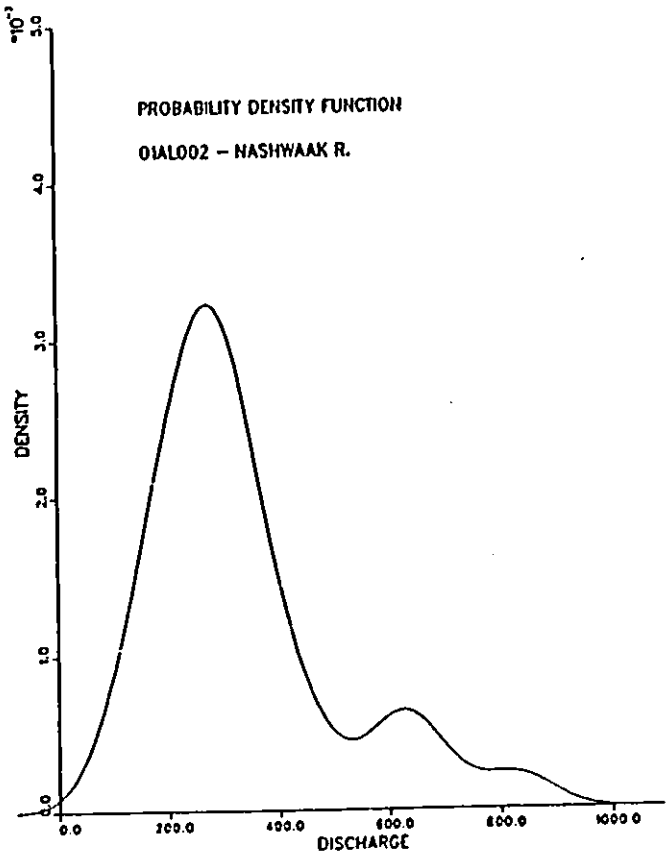

PROBABILITY DENSITY FUNCTION
01AJ011 – COLD STREAM
DENSITY
0.00
0.01
0.02
0.03
0.04
0.05
0.0
40.0
80.0
120.0
DISCHARGE

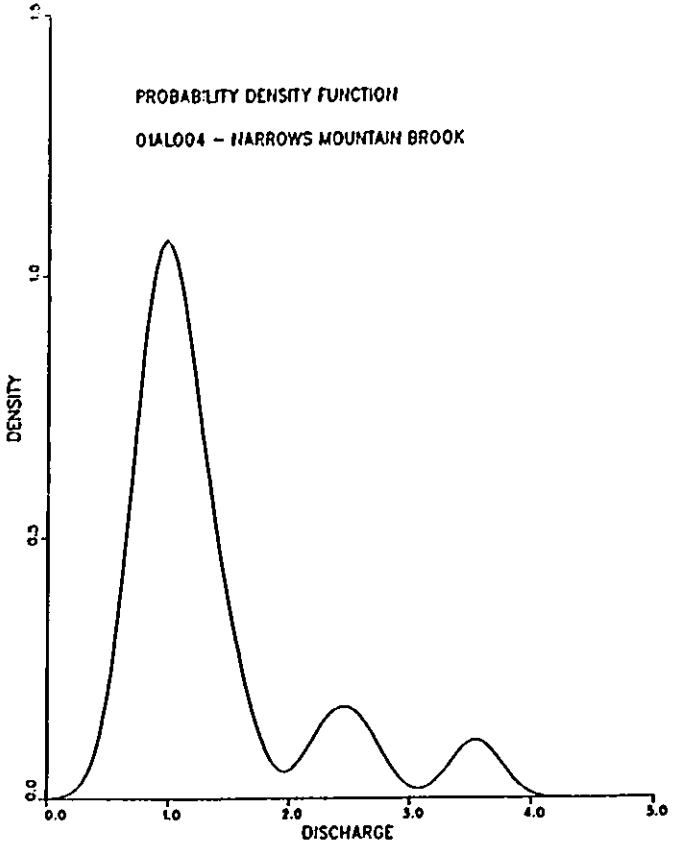

PROBABILITY DENSITY FUNCTION
01AK001 – SHOGOMOC STREAM
DENSITY
0.00
0.01
0.02
0.03
0.04
0.0
30.0
60.0
90.0
120.0
150.0
DISCHARGE

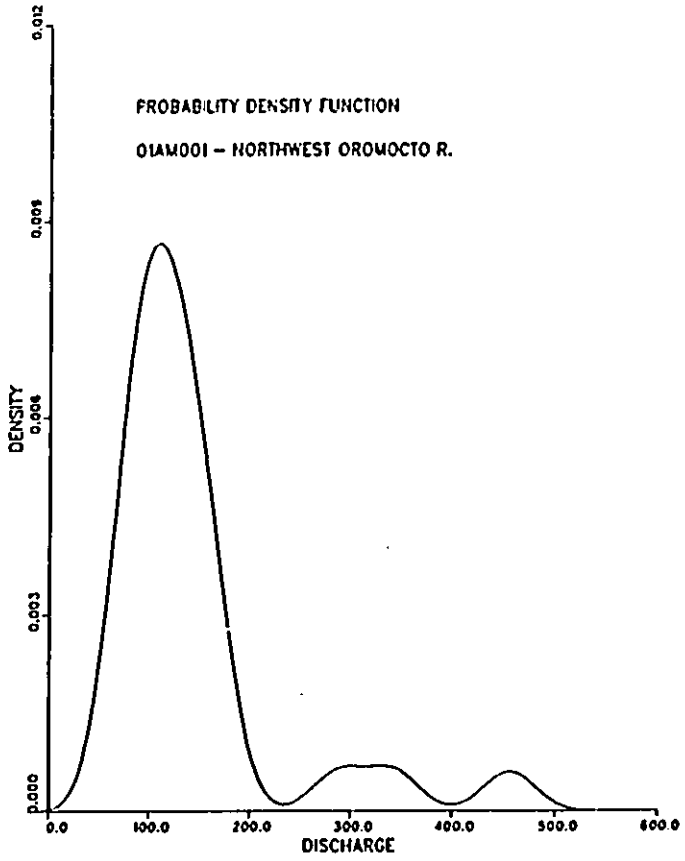

PROBABILITY DENSITY FUNCTION
01AK005 – NORTH NASHWAAKSIS STREAM
DENSITY
0.00
0.05
0.10
0.15
0.20
0.25
0.0
5.0
10.0
15.0
20.0
DISCHARGE

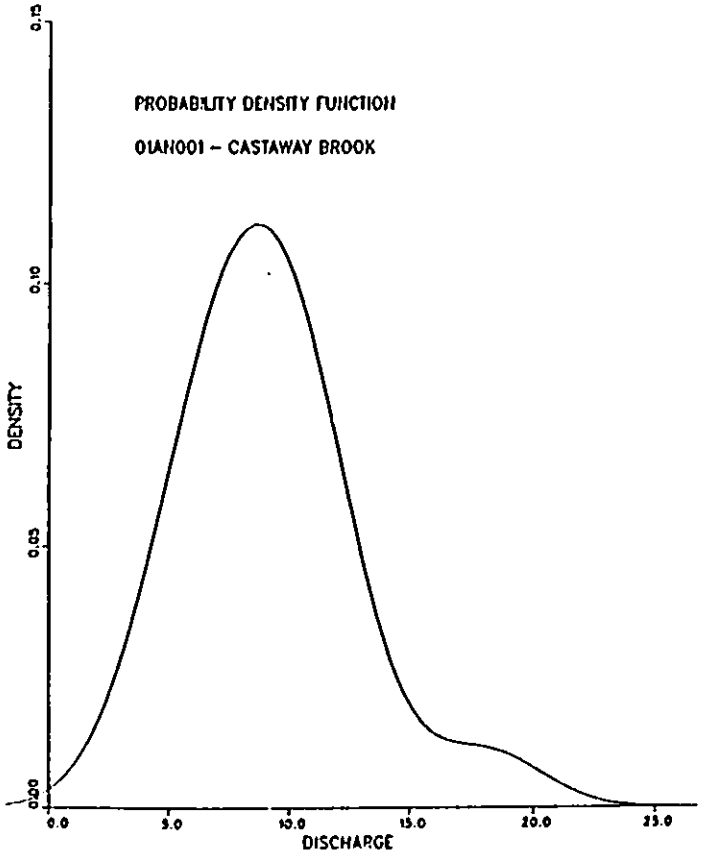

PROBABILITY DENSITY FUNCTION
01AK007 – NACKAWIC R.
DENSITY
0.00
0.01
0.02
0.03
0.0
30.0
60.0
90.0
120.0
150.0
DISCHARGE

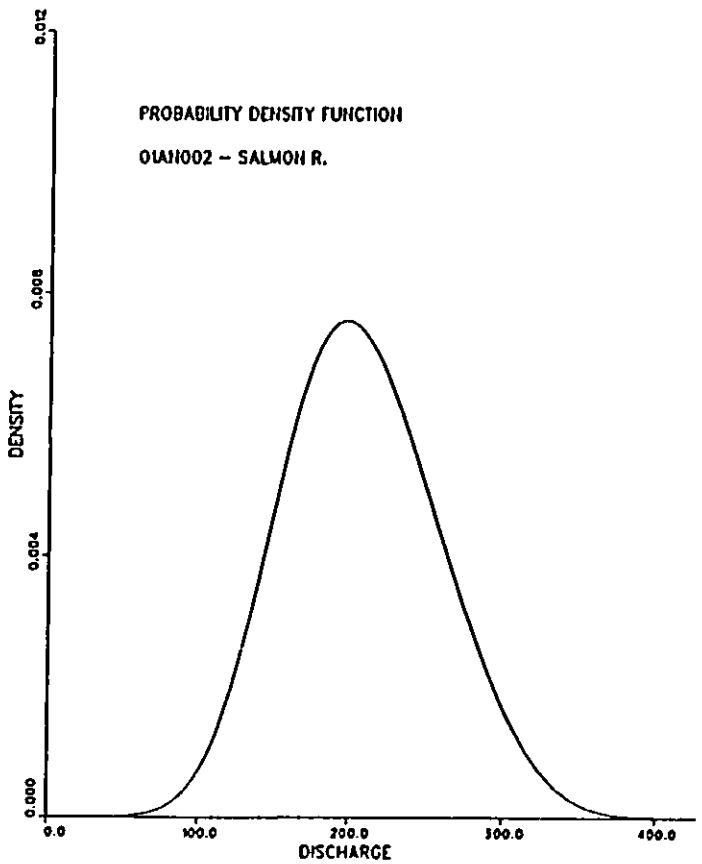

PROBABILITY DENSITY FUNCTION
01AK008 – EEL R.
DENSITY
0.03
0.02
0.01
0.00
0.0
30.0
60.0
90.0
120.0
150.0
DISCHARGE

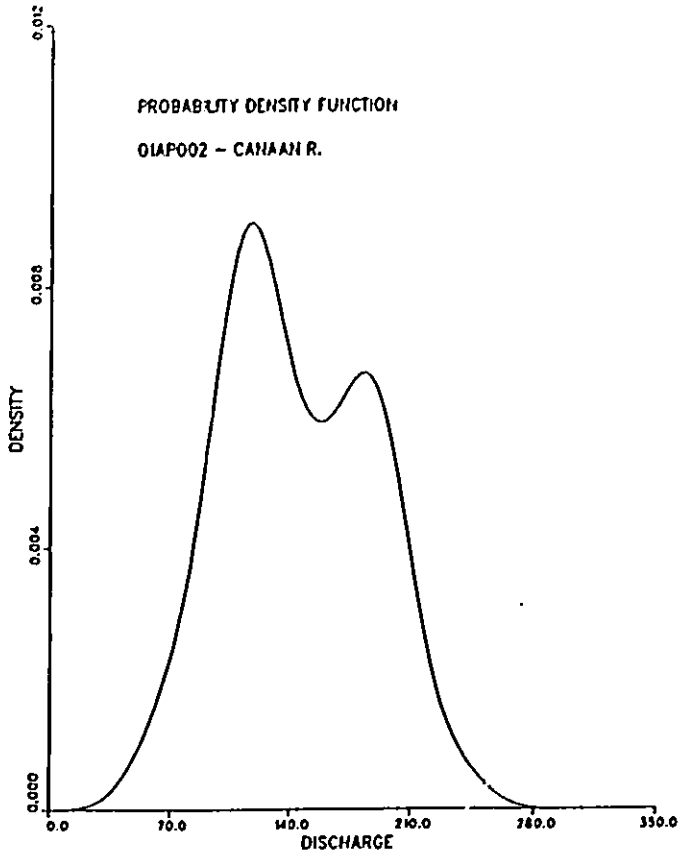

PROBABILITY DENSITY FUNCTION
01AL002 – NASHWAAK R.
DENSITY
0.0
200.0
400.0
600.0
800.0
1000.0
DISCHARGE

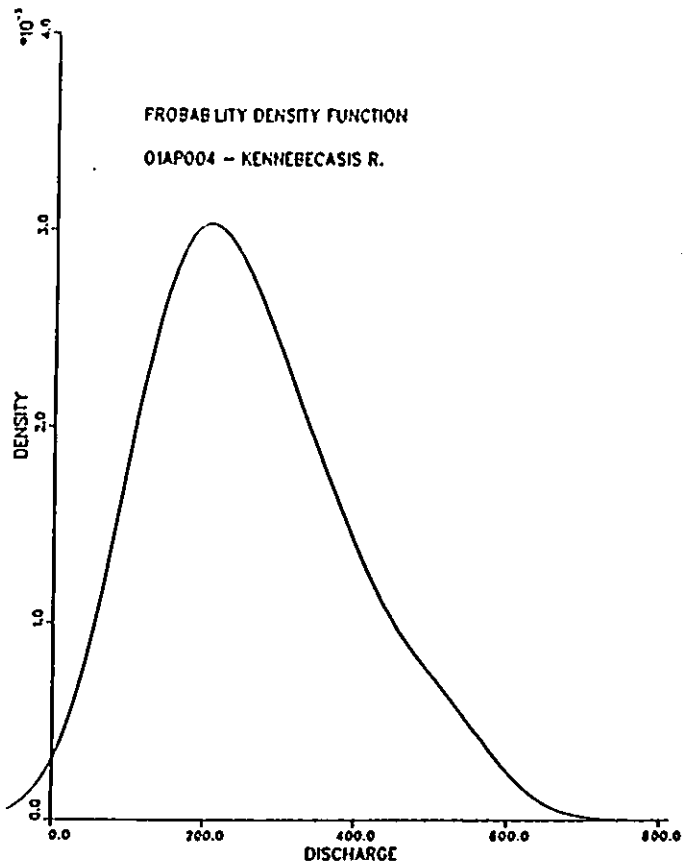

PROBABILITY DENSITY FUNCTION
01AL004 – NARROWS MOUNTAIN BROOK
DENSITY
1.5
1.0
0.5
0.0
0.0
1.0
2.0
3.0
4.0
5.0
DISCHARGE

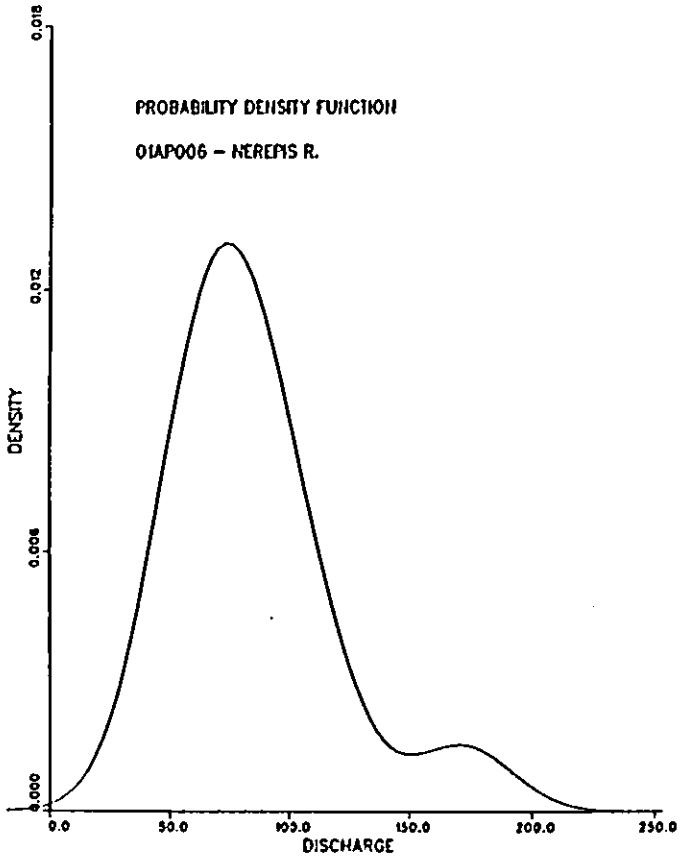

PROBABILITY DENSITY FUNCTION
01AM001 – NORTHWEST OROMOCTO R.
DENSITY
0.012
0.009
0.006
0.003
0.000
0.0
100.0
200.0
300.0
400.0
500.0
600.0
DISCHARGE

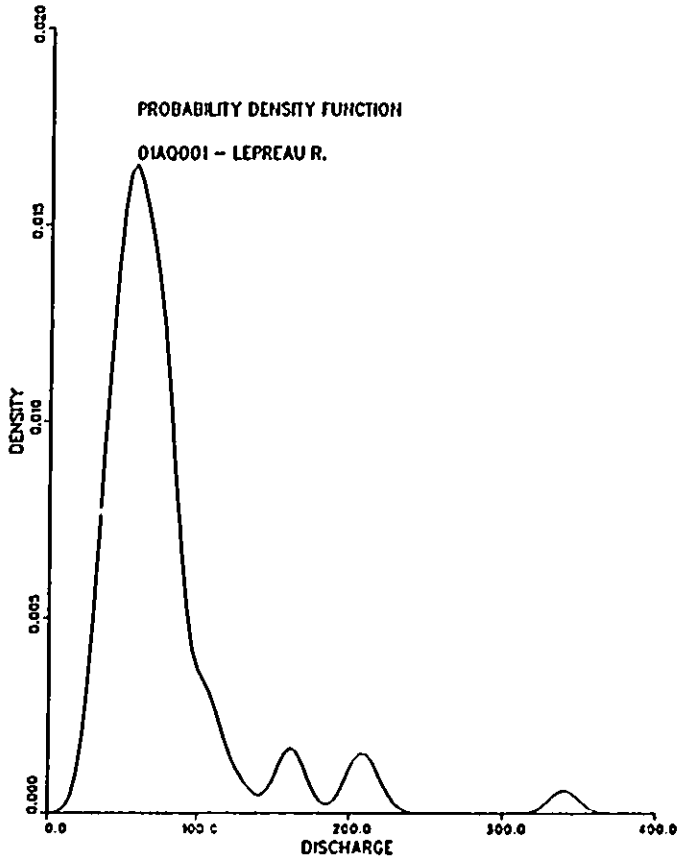

PROBABILITY DENSITY FUNCTION
01AN001 – CASTAWAY BROOK
DENSITY
0.15
0.10
0.05
0.00
0.0
5.0
10.0
15.0
20.0
25.0
DISCHARGE

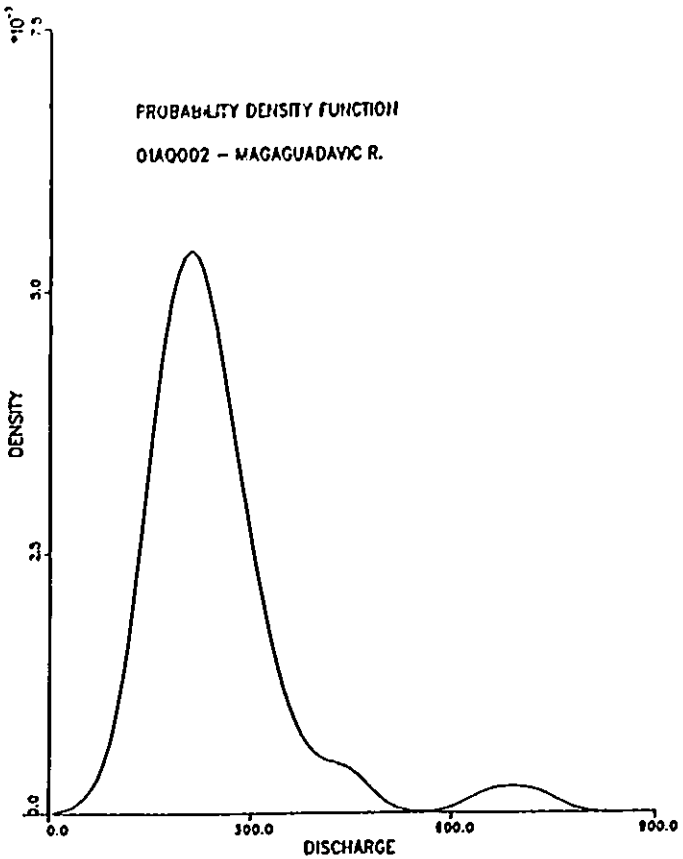

PROBABILITY DENSITY FUNCTION
01AN002 – SALMON R.
DENSITY
0.012
0.008
0.004
0.000
0.0
100.0
200.0
300.0
400.0
DISCHARGE

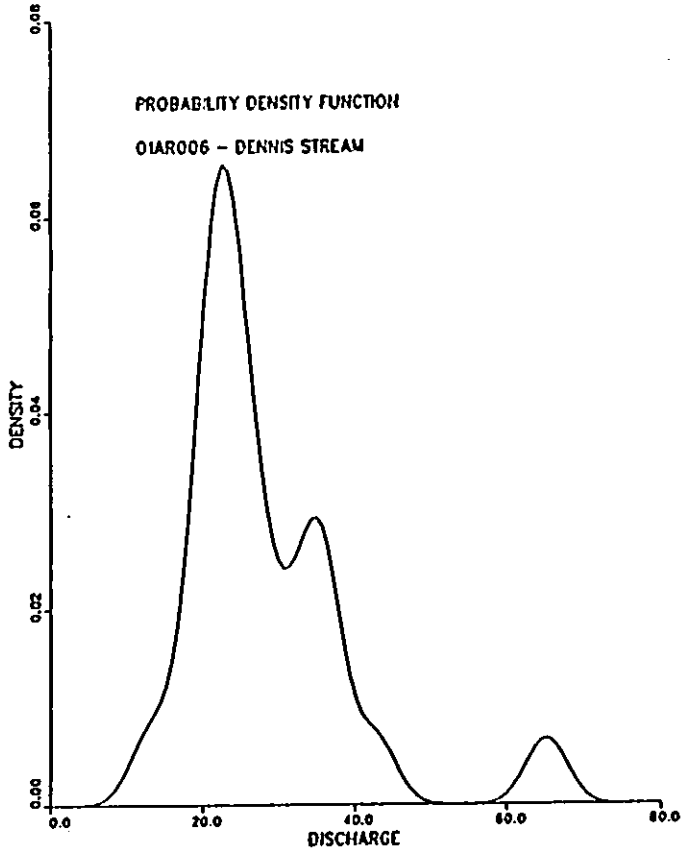

PROBABILITY DENSITY FUNCTION
01AP002 – CANAAN R.
DENSITY
0.012
0.008
0.004
0.000
0.0
70.0
140.0
210.0
280.0
350.0
DISCHARGE

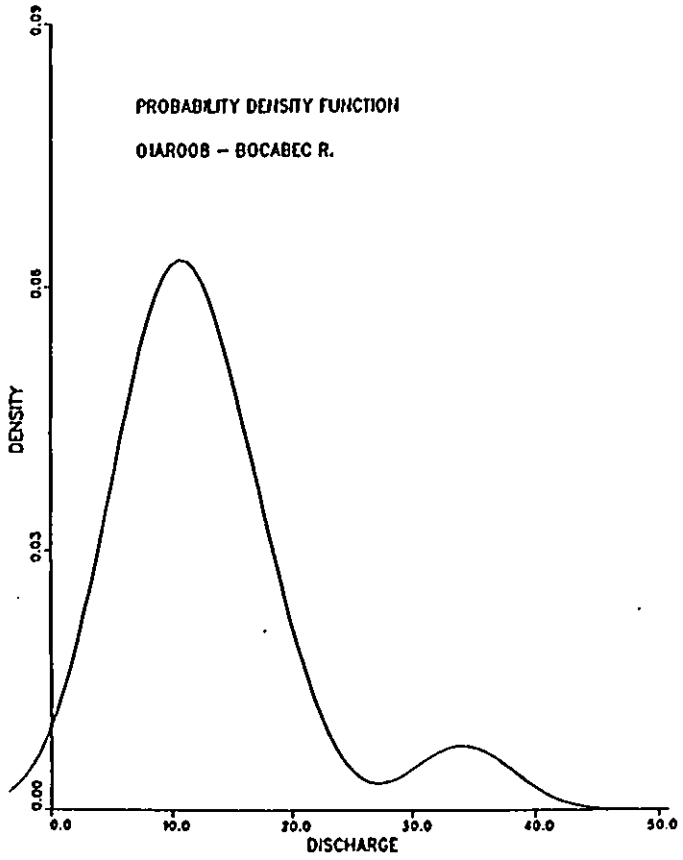

PROBABILITY DENSITY FUNCTION
01AP004 – KENNEBECASIS R.
*10⁻³
4.0
3.0
2.0
1.0
0.0
DENSITY
0.0
200.0
400.0
600.0
800.0
DISCHARGE

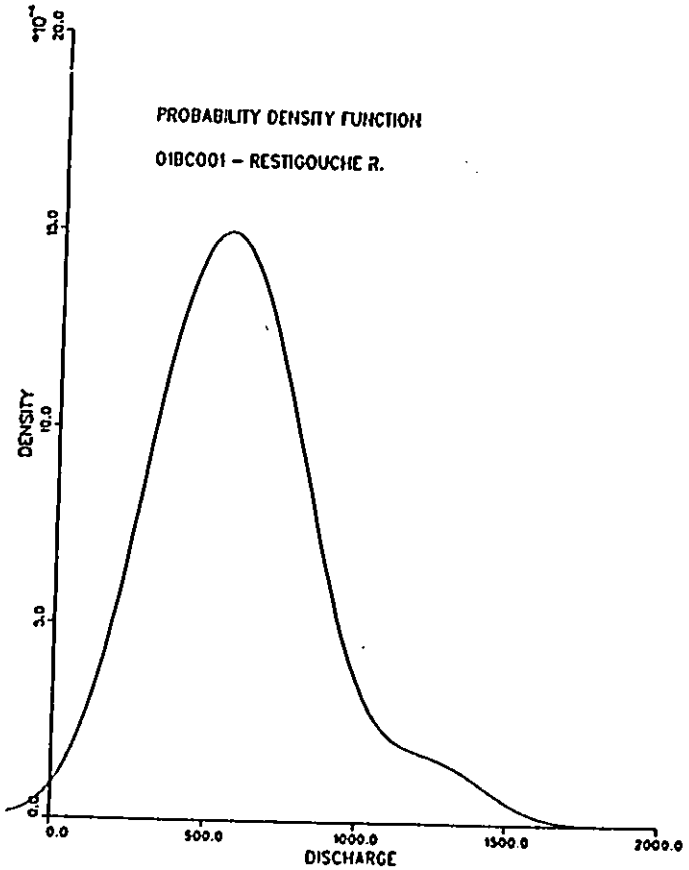

PROBABILITY DENSITY FUNCTION
01AP006 – NEREPIS R.
DENSITY
0.018
0.012
0.006
0.000
0.0
50.0
100.0
150.0
200.0
250.0
DISCHARGE

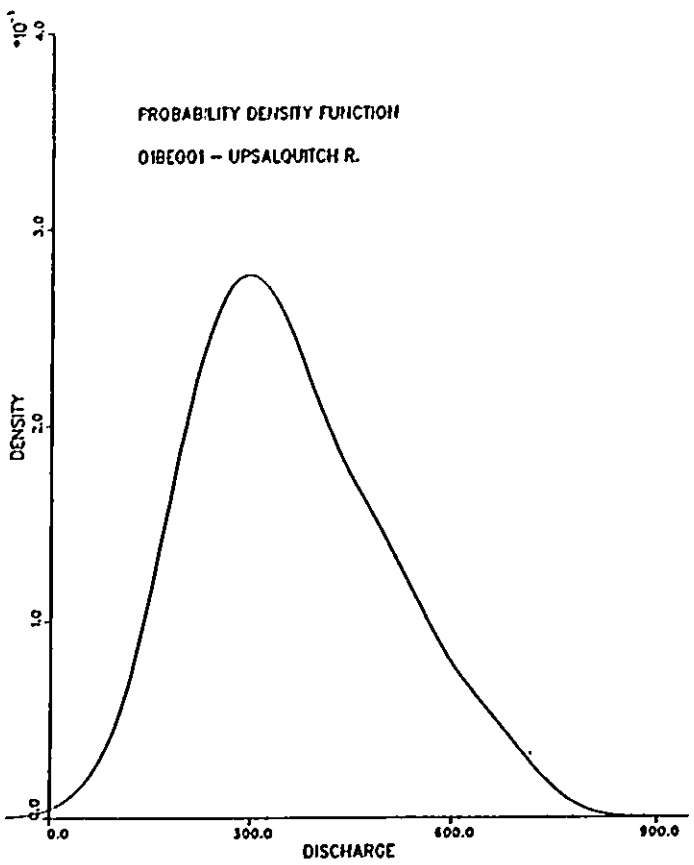

PROBABILITY DENSITY FUNCTION
01AQ001 – LEPREAU R.
DENSITY
0.020
0.015
0.010
0.005
0.000
0.0
100.0
200.0
300.0
400.0
DISCHARGE

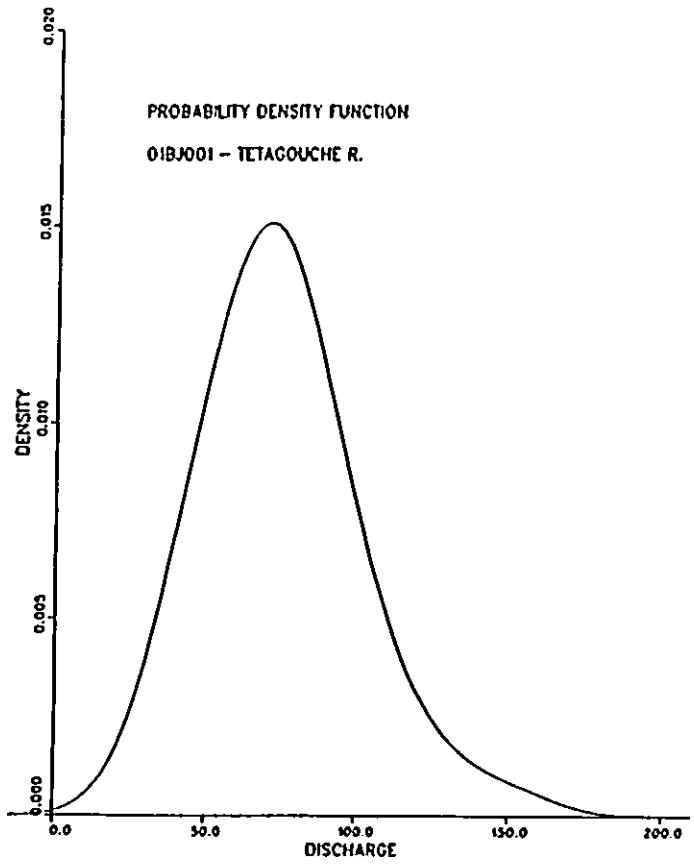

PROBABILITY DENSITY FUNCTION
01AQ002 – MAGAGUADAVIC R.
DENSITY
DISCHARGE
0.0
300.0
600.0
900.0

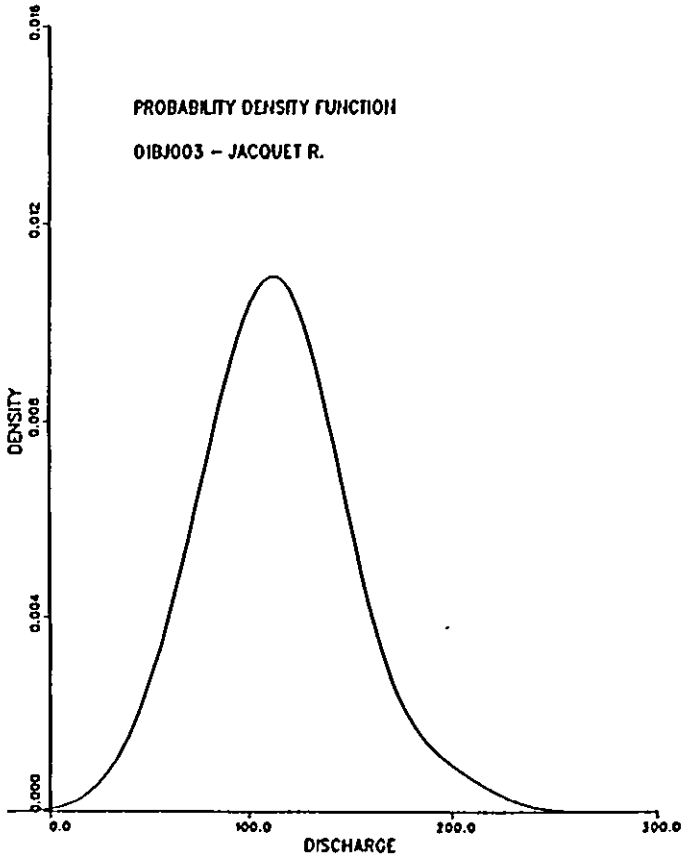

PROBABILITY DENSITY FUNCTION
01AR006 – DENNIS STREAM
DENSITY
DISCHARGE
0.0
20.0
40.0
60.0
80.0

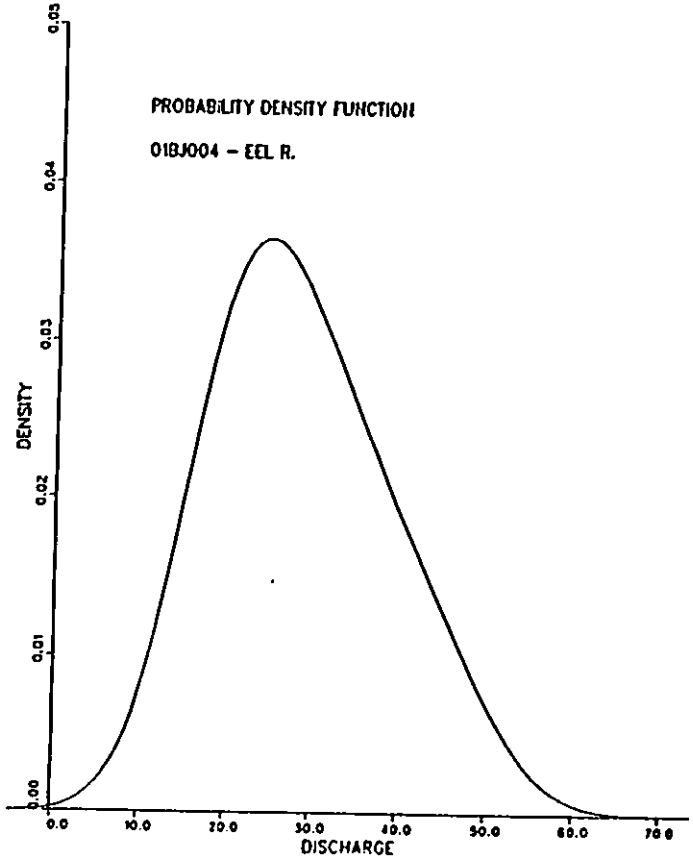

PROBABILITY DENSITY FUNCTION
01AR008 – BOCABEC R.
DENSITY
0.09
0.06
0.03
0.00
0.0
10.0
20.0
30.0
40.0
50.0
DISCHARGE

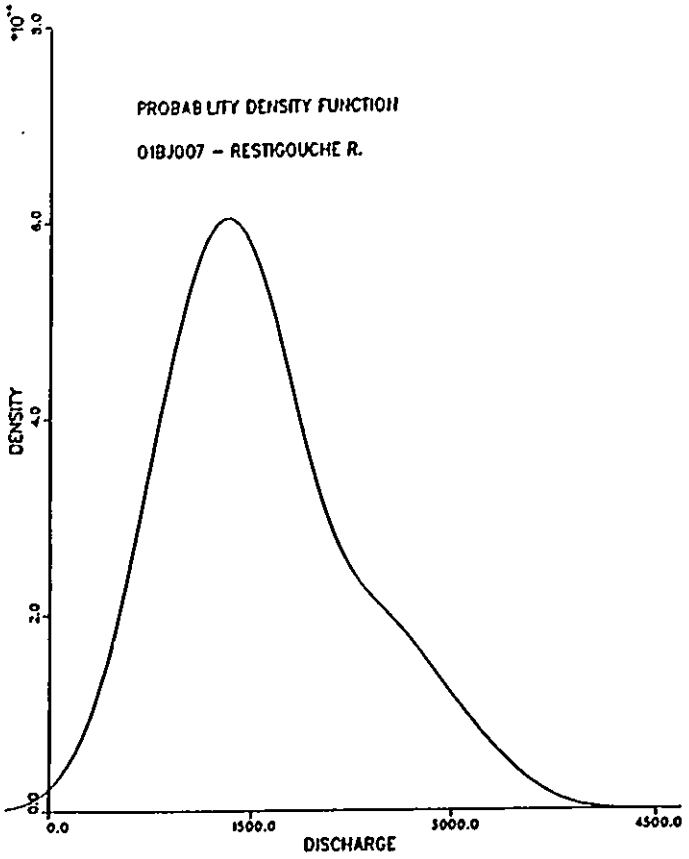

PROBABILITY DENSITY FUNCTION
01BC001 – RESTIGOUCHE R.
*10⁻⁴
20.0
15.0
10.0
5.0
0.0
DENSITY
0.0
500.0
1000.0
1500.0
2000.0
DISCHARGE

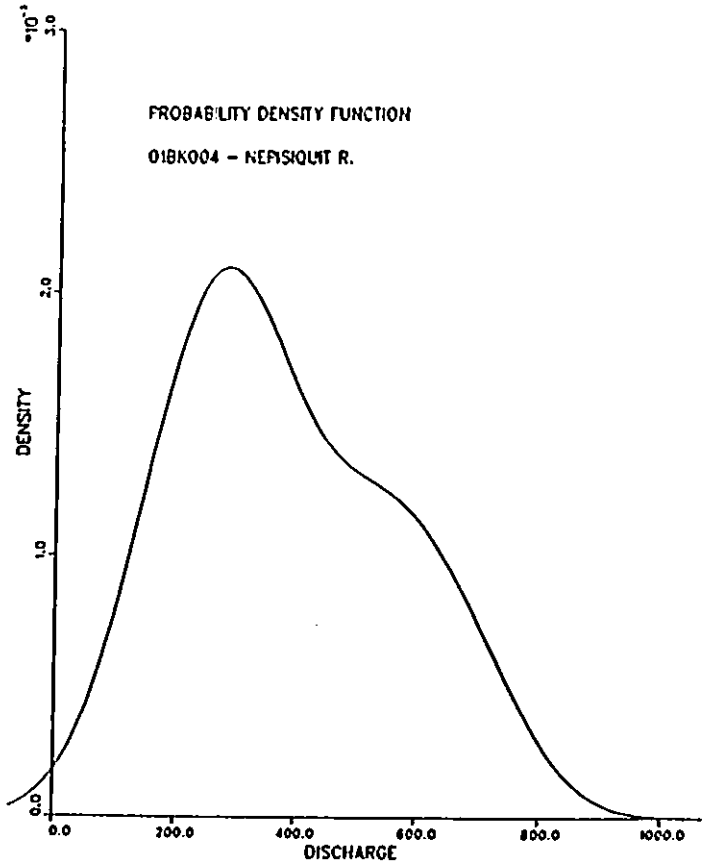

PROBABILITY DENSITY FUNCTION
01BE001 – UPSALQUITCH R.
DENSITY
*10⁻³
4.0
3.0
2.0
1.0
0.0
0.0
300.0
600.0
900.0
DISCHARGE

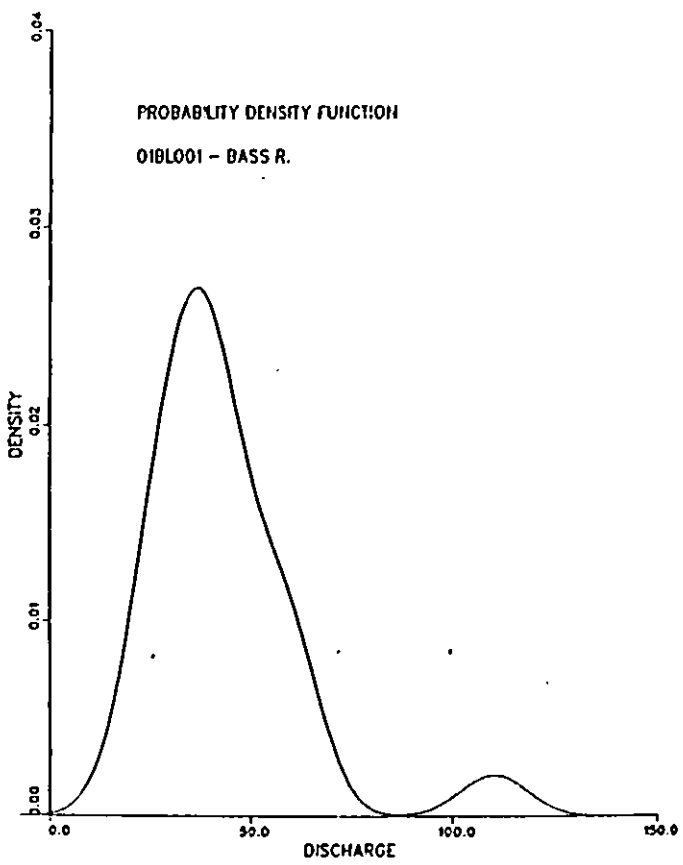

PROBABILITY DENSITY FUNCTION
01BJ001 – TETAGOUCHE R.
DENSITY
0.020
0.015
0.010
0.005
0.000
0.0
50.0
100.0
150.0
200.0
DISCHARGE

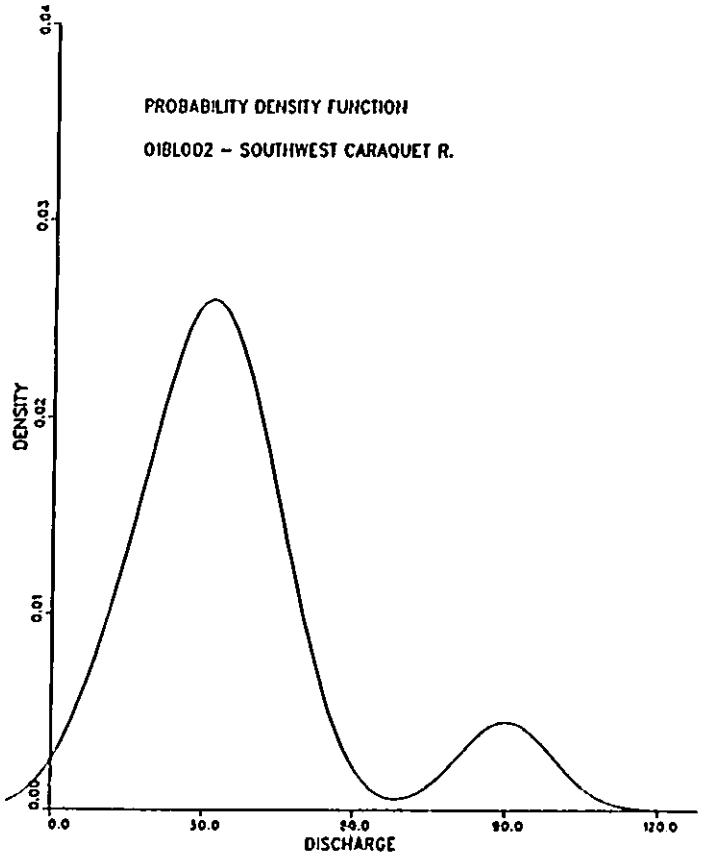

PROBABILITY DENSITY FUNCTION
01BJ003 – JACQUET R.
DENSITY
0.016
0.012
0.008
0.004
0.000
0.0
100.0
200.0
300.0
DISCHARGE

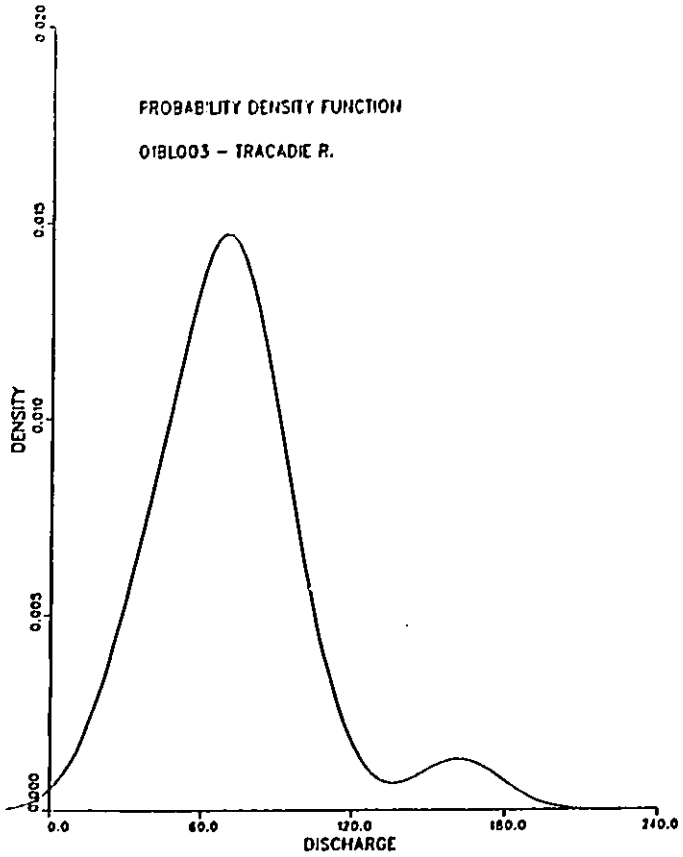

PROBABILITY DENSITY FUNCTION
01BJ004 – EEL R.
DENSITY
0.05
0.04
0.03
0.02
0.01
0.00
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
DISCHARGE

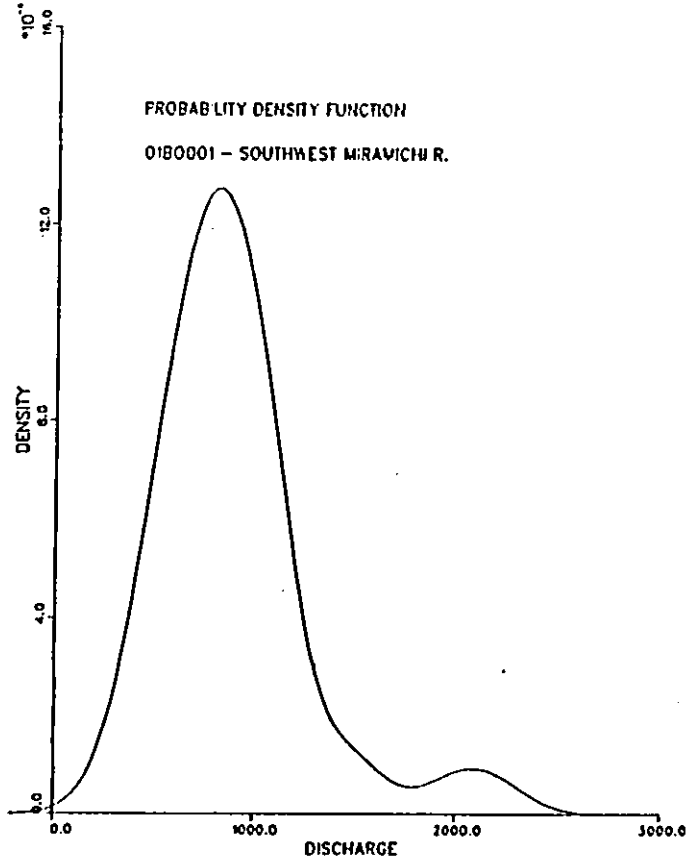

PROBABILITY DENSITY FUNCTION
01BJ007 – RESTIGOUCHE R.
*10⁻⁴
3.0
6.0
4.0
2.0
0.0
DENSITY
0.0
1500.0
3000.0
4500.0
DISCHARGE

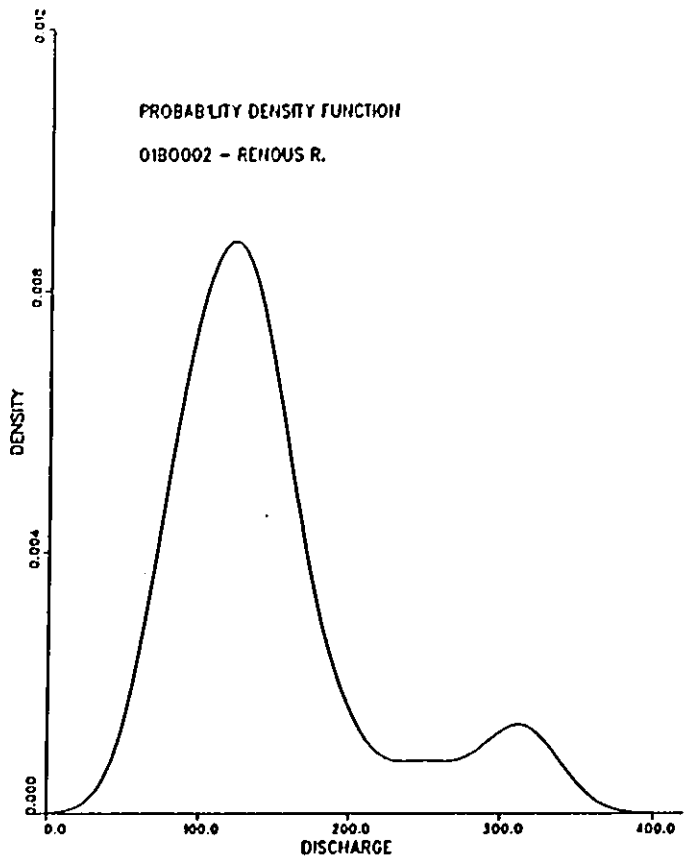

PROBABILITY DENSITY FUNCTION
01BK004 – NEPISIQUIT R.
*10⁻³
3.0
2.0
1.0
0.0
DENSITY
0.0
200.0
400.0
600.0
800.0
1000.0
DISCHARGE

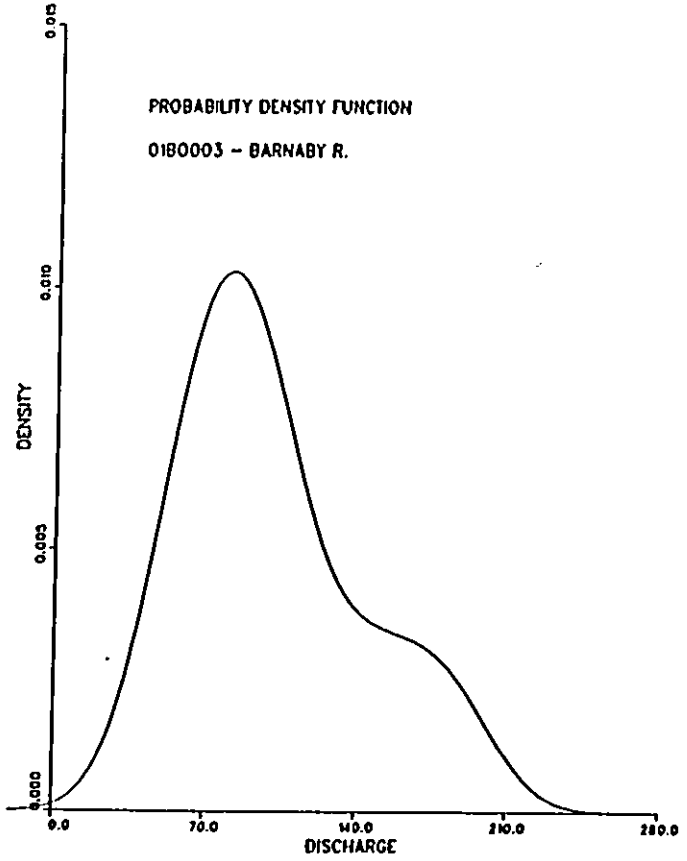

PROBABILITY DENSITY FUNCTION
01BL001 – BASS R.
DENSITY
0.04
0.03
0.02
0.01
0.00
0.0
50.0
100.0
150.0
DISCHARGE

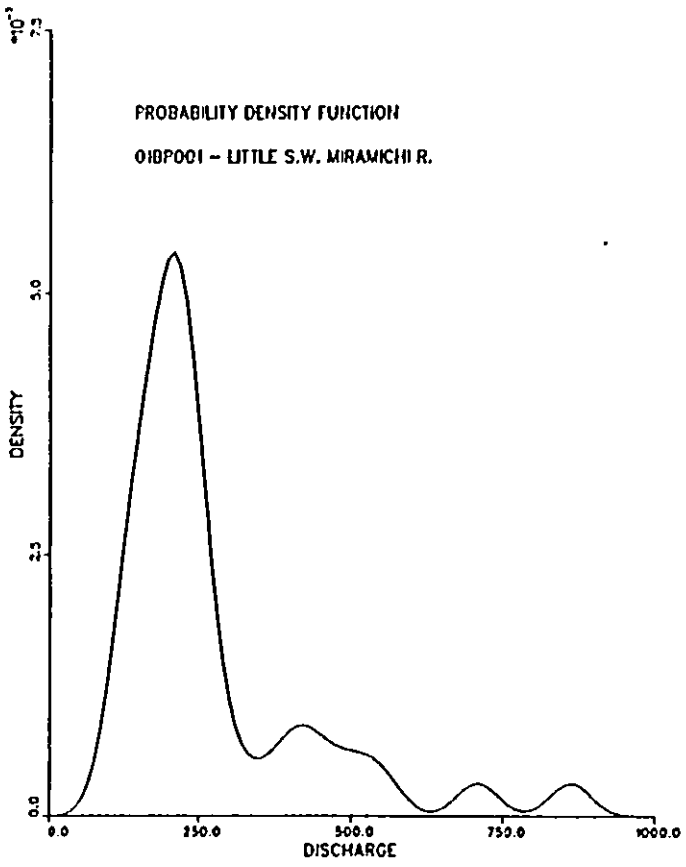

PROBABILITY DENSITY FUNCTION
01BL002 – SOUTHWEST CARAQUET R.
DENSITY
0.04
0.03
0.02
0.01
0.00
0.0
50.0
90.0
120.0
DISCHARGE

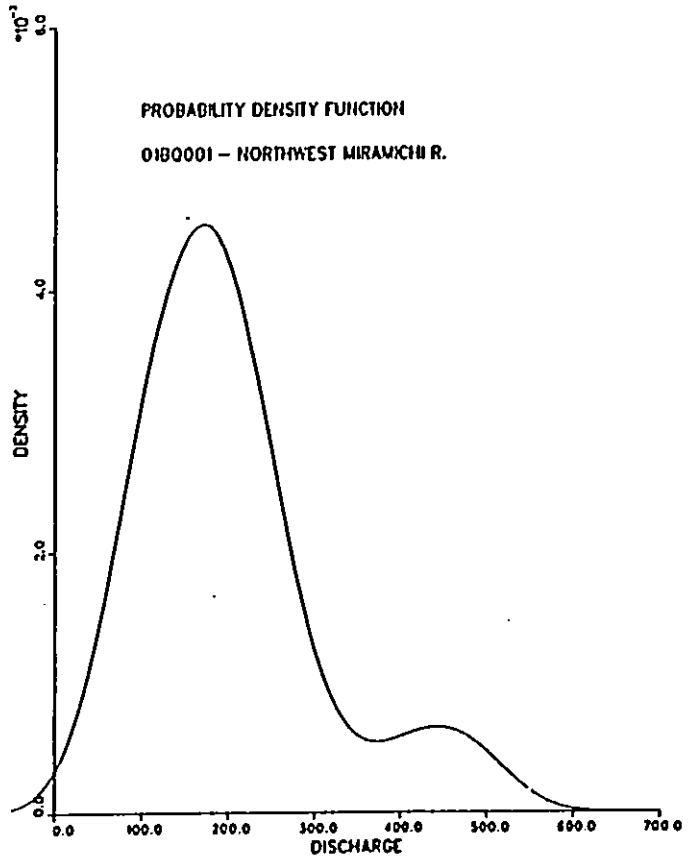

PROBABILITY DENSITY FUNCTION
01BL003 – TRACADIE R.
DENSITY
0.020
0.015
0.010
0.005
0.000
0.0
60.0
120.0
180.0
240.0
DISCHARGE

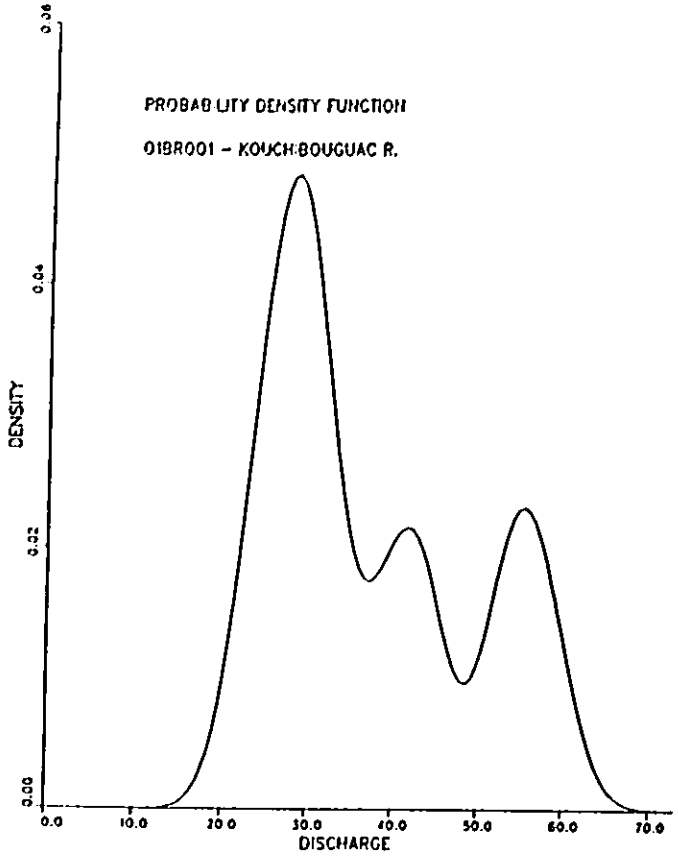

PROBABILITY DENSITY FUNCTION
01BO001 – SOUTHWEST MIRAMICHI R.
DENSITY
*10⁻⁴
16.0
12.0
8.0
4.0
0.0
0.0
1000.0
2000.0
3000.0
DISCHARGE

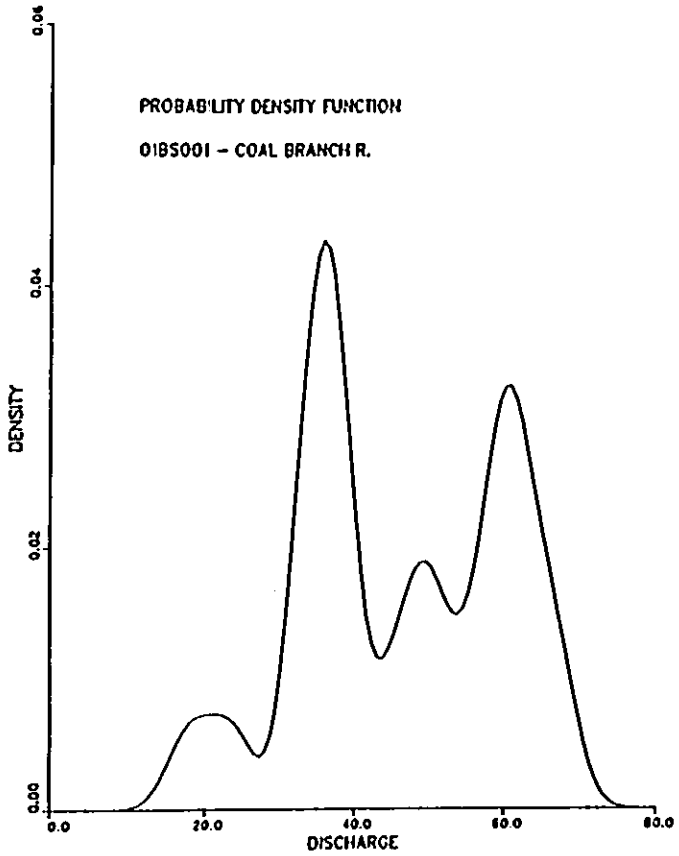

PROBABILITY DENSITY FUNCTION
01B0002 - RENOUS R.
DENSITY
0.012
0.008
0.004
0.000
0.0
100.0
200.0
300.0
400.0
DISCHARGE

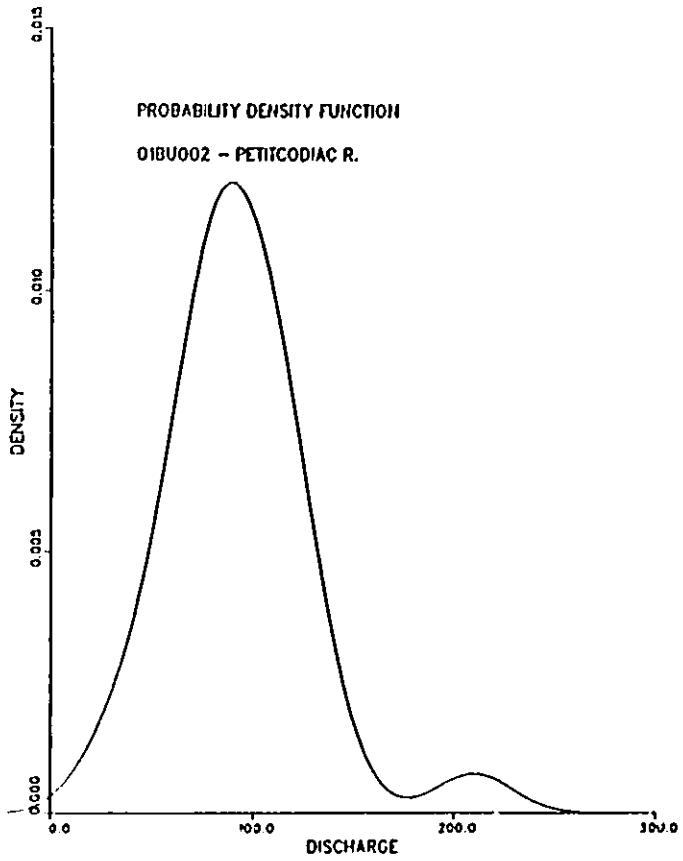

PROBABILITY DENSITY FUNCTION
01B0003 - BARNABY R.
DENSITY
0.015
0.010
0.005
0.000
0.0
70.0
140.0
210.0
280.0
DISCHARGE

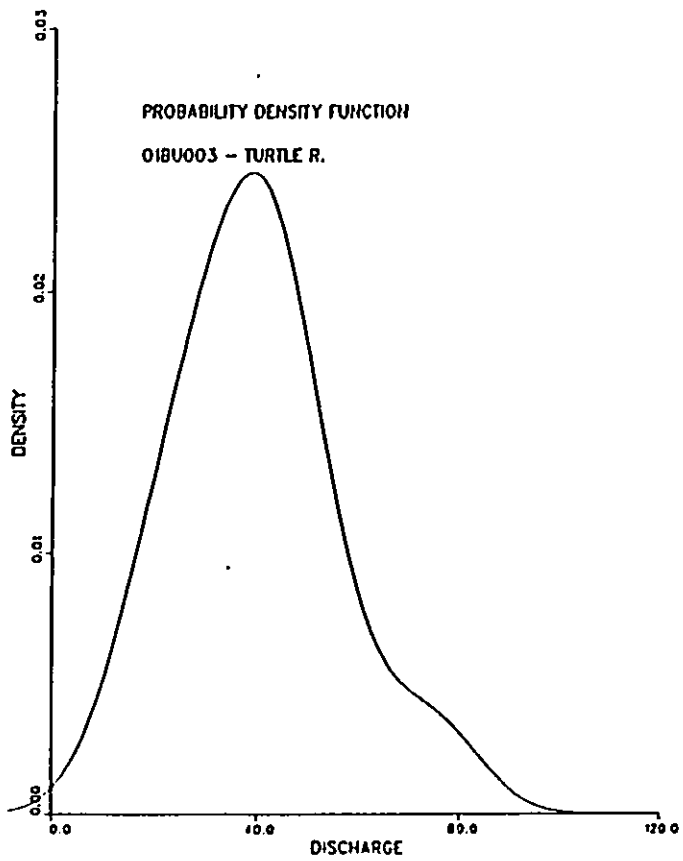

PROBABILITY DENSITY FUNCTION
01BP001 – LITTLE S.W. MIRAMICHI R.
*10⁻³
7.5
5.0
2.5
0.0
DENSITY
0.0
250.0
500.0
750.0
1000.0
DISCHARGE

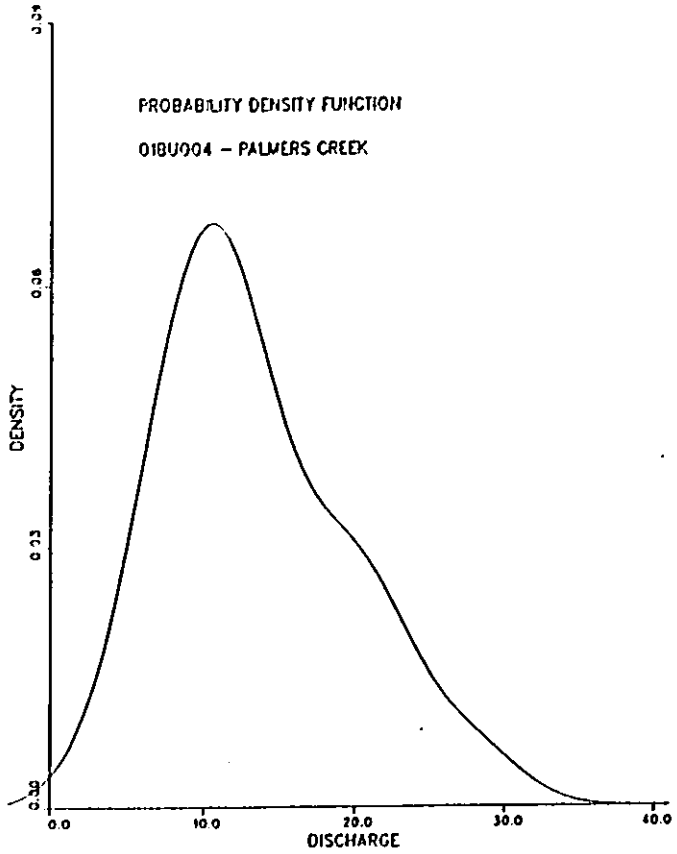

PROBABILITY DENSITY FUNCTION
01BQ001 – NORTHWEST MIRAMICHI R.
*10⁻³
6.0
4.0
2.0
0.0
DENSITY
0.0
100.0
200.0
300.0
400.0
500.0
600.0
700.0
DISCHARGE

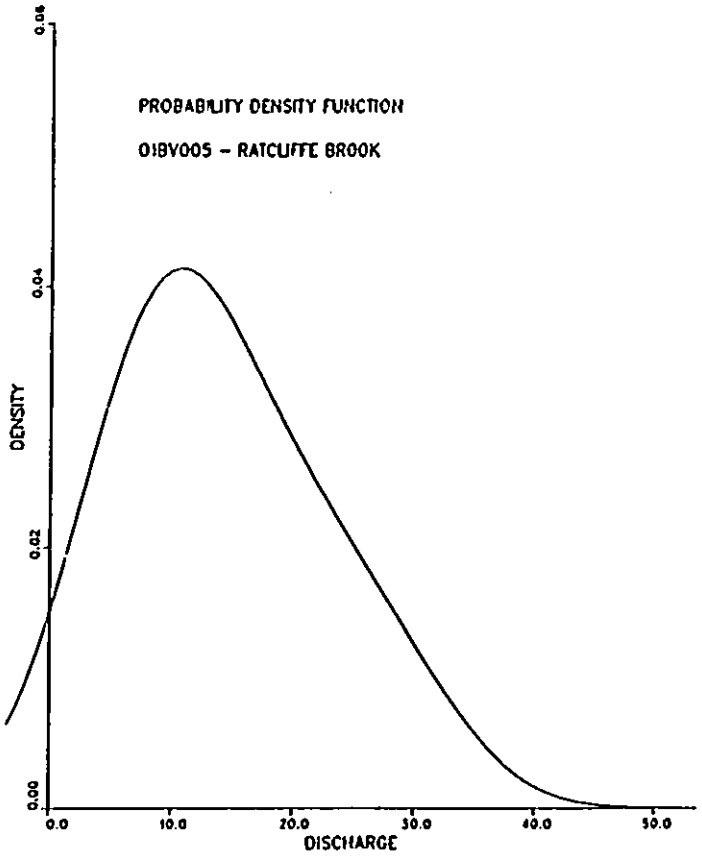

PROBABILITY DENSITY FUNCTION
01BR001 – KOUCHIBOUGUAC R.
DENSITY
0.00
0.02
0.04
0.06
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
DISCHARGE

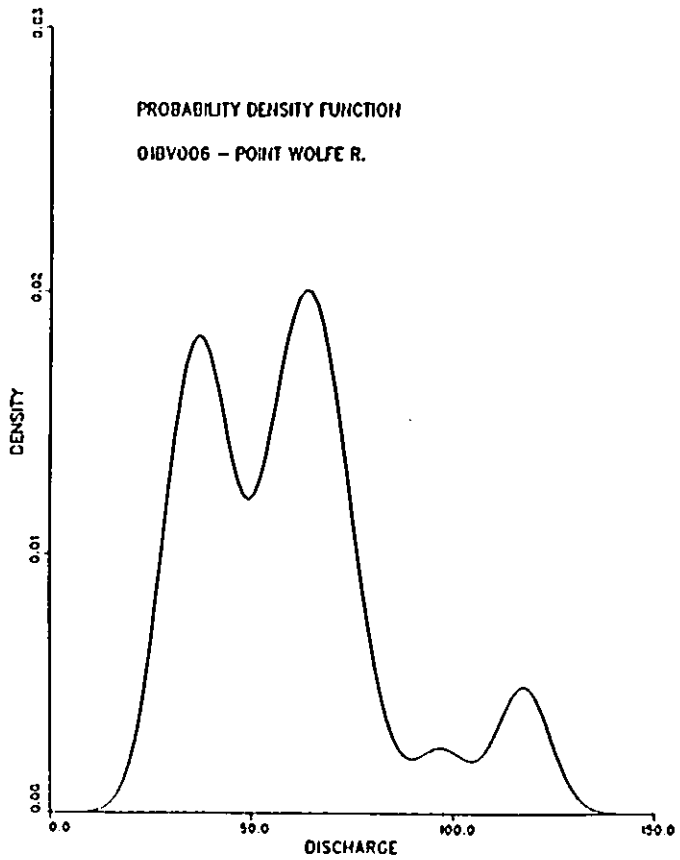

PROBABILITY DENSITY FUNCTION
01BS001 – COAL BRANCH R.
DENSITY
0.00
0.02
0.04
0.06
0.0
20.0
40.0
60.0
80.0
DISCHARGE

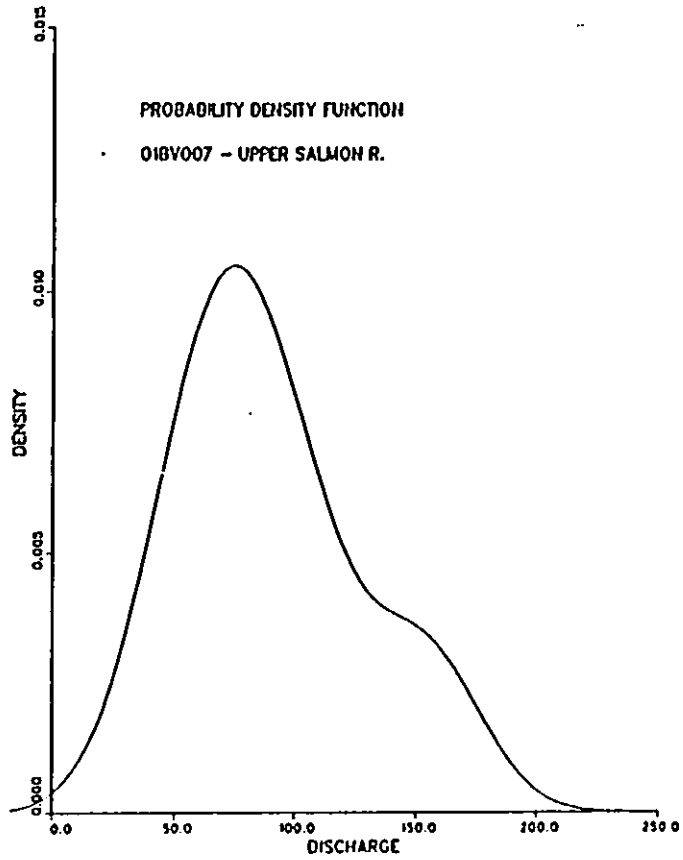

PROBABILITY DENSITY FUNCTION
01BU002 – PETITCODIAC R.
DENSITY
0.015
0.010
0.005
0.000
0.0
100.0
200.0
300.0
DISCHARGE

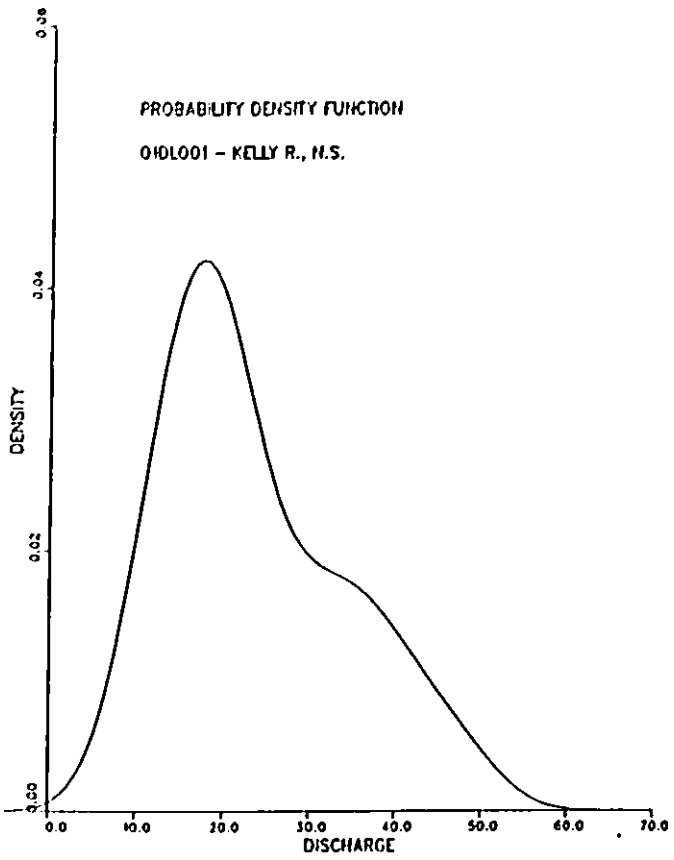

PROBABILITY DENSITY FUNCTION
01BU003 – TURTLE R.
DENSITY
0.03
0.02
0.01
0.00
0.0
40.0
80.0
120.0
DISCHARGE

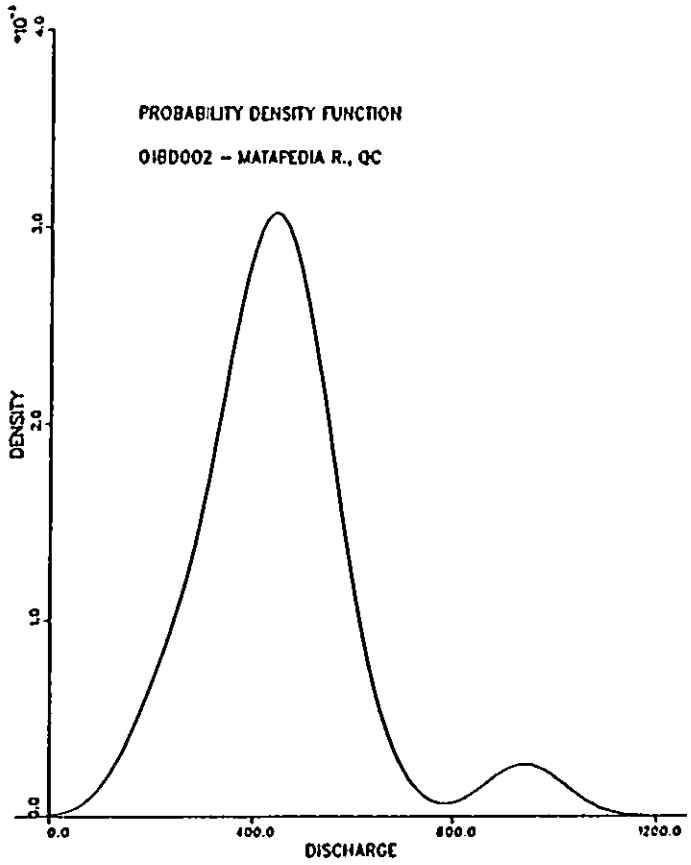

PROBABILITY DENSITY FUNCTION
01BU004 – PALMERS CREEK
DENSITY
0.00
0.02
0.04
0.06
0.0
10.0
20.0
30.0
40.0
DISCHARGE

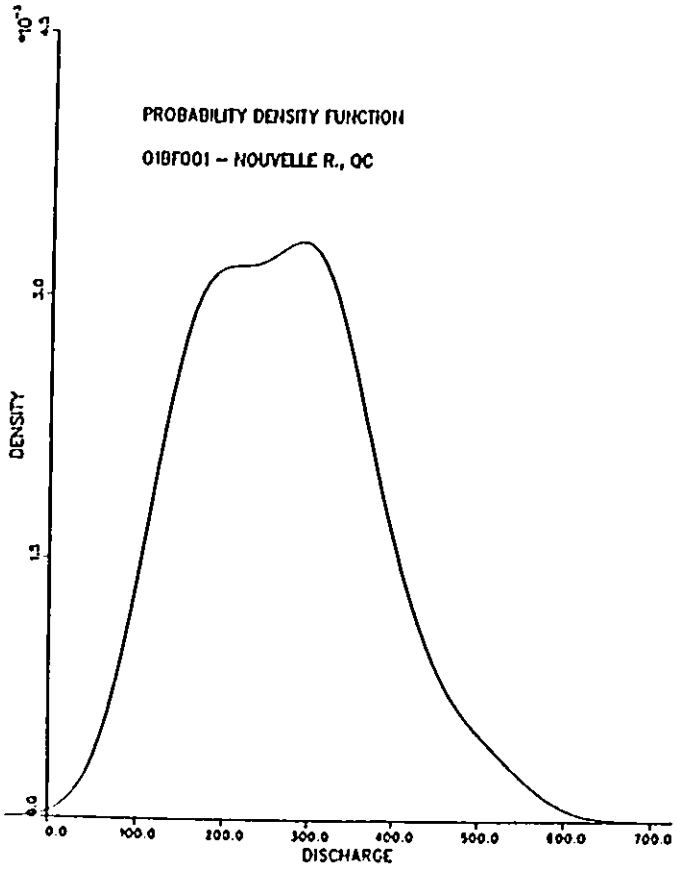

PROBABILITY DENSITY FUNCTION
01BV005 – RATCLIFFE BROOK
DENSITY
0.00
0.02
0.04
0.06
0.0
10.0
20.0
30.0
40.0
50.0
DISCHARGE

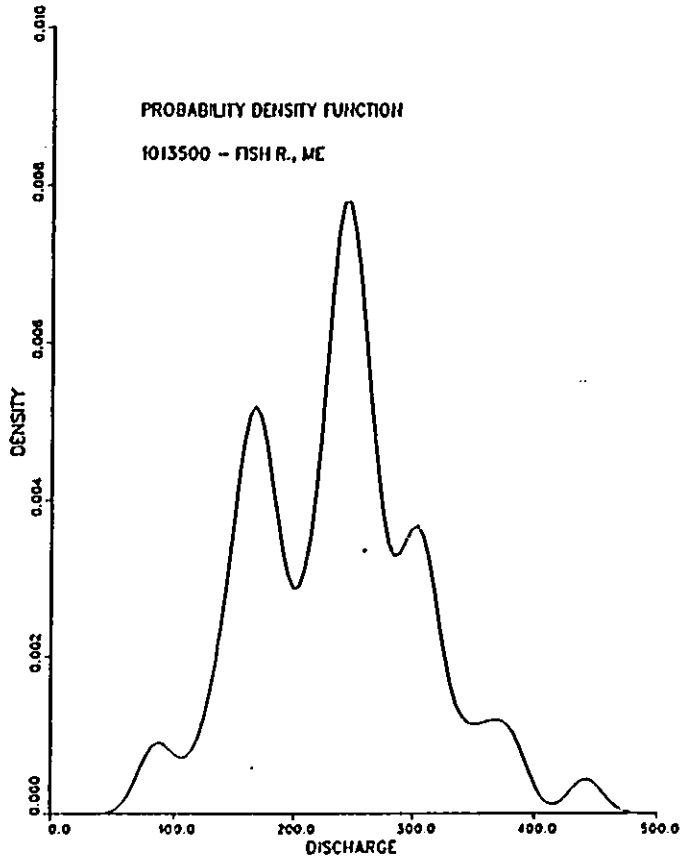

PROBABILITY DENSITY FUNCTION
01BV006 – POINT WOLFE R.
DENSITY
0.03
0.02
0.01
0.00
0.0
50.0
100.0
150.0
DISCHARGE

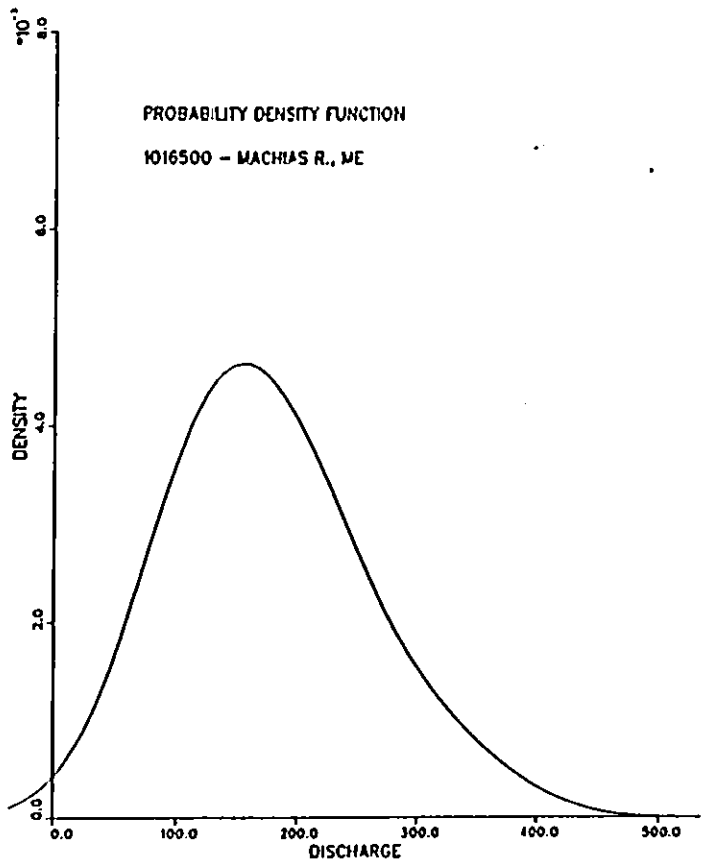

PROBABILITY DENSITY FUNCTION
01BV007 – UPPER SALMON R.
DENSITY
0.015
0.010
0.005
0.000
0.0
50.0
100.0
150.0
200.0
250.0
DISCHARGE


PROBABILITY DENSITY FUNCTION
01DL001 – KELLY R., N.S.
DENSITY
0.00
0.02
0.04
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
DISCHARGE


PROBABILITY DENSITY FUNCTION
01BD002 – MATAPEDIA R., QC
DENSITY
0.0
1.0
2.0
3.0
4.0
0.0
400.0
800.0
1200.0
DISCHARGE


PROBABILITY DENSITY FUNCTION
01BF001 -- NOUVELLE R., QC
DENSITY
DISCHARGE
0.0
100.0
200.0
300.0
400.0
500.0
600.0
700.0


PROBABILITY DENSITY FUNCTION
1013500 -- FISH R., ME
DENSITY
DISCHARGE
0.000
0.002
0.004
0.006
0.008
0.010
0.0
100.0
200.0
300.0
400.0
500.0


PROBABILITY DENSITY FUNCTION
1016500 – MACHIAS R., ME
DENSITY
$*10^{-3}$
8.0
6.0
4.0
2.0
0.0
0.0
100.0
200.0
300.0
400.0
500.0
DISCHARGE

Appendix B

NUMERICAL RESULTS

List of Tables

Table B.1: Parametric Flood Frequencies

Station	Return Period (Yrs)				
	2	10	20	50	100
01AD003	201	329	378	441	489
01AF003	218	359	410	473	519
01AG002	32.7	48.6	55.1	64.0	70.9
01AG003	918	1500	1740	2090	2360
01AH005	39.3	67.6	79.4	95.4	108
01AJ003	243	410	475	559	622
01AJ004	88.8	147	171	205	231
01AJ010	80.1	123	137	153	165
01AJ011	33.4	61.0	72.7	88.9	102
01AK001	37.2	59.8	68.4	79.5	88.0
01AK005	5.84	10.5	12.7	16.0	18.7
01AK007	50.6	87.8	104	126	143
01AK008	69.2	101	111	123	131
01AL002	291	556	675	844	982
01AL004	1.10	2.27	2.97	4.14	5.22
01AM001	115	243	322	455	580
01AN001	8.66	13.5	15.2	17.5	19.1
01AN002	199	263	289	323	349
01AP002	141	199	217	238	252
01AP004	228	434	528	663	774

Table B.1: Parametric Flood Frequencies (continued)

Station	Return Period (Yrs)				
	2	10	20	50	100
01AP006	77.2	127	149	180	205
01AQ001	64.4	131	166	220	266
01AQ002	227	372	432	512	575
01AR006	26.0	39.9	45.3	52.4	57.9
01AR008	11.3	22.2	26.9	33.3	38.5
01BC001	560	932	1070	1260	1400
01BE001	331	547	636	755	849
01BJ001	71.7	107	120	135	147
01BJ003	111	154	168	184	195
01BJ004	25.7	43.0	52.0	65.6	77.5
01BJ007	1430	2490	2940	3560	4060
01BK004	344	636	763	937	1080
01BL001	38.6	66.4	79.3	98.0	113
01BL002	32.0	63.1	76.3	94.6	109
01BL003	68.0	111	127	147	162
01BO001	827	1360	1560	1840	2050
01BO002	123	238	303	409	505
01BO003	90.1	162	196	247	290
01BP001	220	465	600	812	1000
01BQ001	176	324	388	474	543

Table B.1: Parametric Flood Frequencies (continued)

Station	Return Period (Yrs)				
	2	10	20	50	100
01BR001	33.4	56.9	69.5	S9.0	106
01BS001	46.6	63.7	6S.6	74.2	77.9
01BU002	S9.7	136	152	173	1SS
01BU003	3S.1	60.4	6S.1	77.7	S4.7
01BU004	12.4	21.9	25.7	30.6	34.5
01BV005	12.3	26.0	32.2	41.0	4S.2
01BV006	52.S	96.4	117	14S	174
01BV007	S2.3	142	167	202	230
01DL001	20.9	39.4	4S.3	61.5	72.6
01BD002	434	649	727	S25	S96
01BF001	249	40S	471	555	619
1013500	232	330	361	39S	423
1016500	162	2S2	333	403	460

Table B.2: Nonparametric Flood Frequencies

Station	Return Period (Yrs)				
	2	10	20	50	100
01AD003	209	350	3S7	42S	454
01AF003	223	374	447	504	532
01AG002	32.9	4S.3	56.0	70.S	73.4
01AG003	906	14S0	1570	1640	1690
01AH005	41.0	69.5	S3.9	97.5	104
01AJ003	241	444	491	533	557
01AJ004	90.S	162	196	219	231
01AJ010	S1.5	13S	153	169	1S0
01AJ011	33.6	62.0	S4.9	91.5	94.3
01AK001	37.9	59.4	69.6	S4.4	114
01AK005	5.95	11.4	12.9	14.4	15.1
01AK007	50.1	S9.5	101	115	120
01AK00S	70.0	106	114	119	122
01AL002	295	619	710	822	S67
01AL004	1.12	2.55	3.32	3.62	3.74
01AM001	121	282	356	455	472
01AN001	8.85	14.0	16.6	19.2	20.4
01AN002	204	276	296	317	331
01AP002	139	204	216	232	242
01AP004	244	453	513	569	601

Table B.2: Nonparametric Flood Frequencies (continued)

Station	Return Period (Yrs)				
	2	10	20	50	100
01AP006	79.5	132	165	185	195
01AQ001	65.2	142	198	221	335
01AQ002	227	364	452	679	719
01AR006	25.1	37.8	41.5	45.9	63.7
01AR008	11.7	22.5	32.4	36.7	38.7
01BC001	575	973	1160	1340	1430
01BE001	341	570	633	693	727
01BJ001	72.7	111	126	143	154
01BJ003	112	161	178	200	212
01BJ004	27.6	43.7	47.9	52.2	54.9
01BJ007	1470	2650	2980	3300	3490
01BK004	360	660	724	789	828
01BL001	40.0	64.0	74.5	111	116
01BL002	32.6	74.9	90.5	98.0	102
01BL003	69.7	109	142	168	177
01BO001	840	1350	1740	2130	2250
01BO002	130	260	308	332	344
01BO003	93.5	172	189	205	214
01BP001	219	502	673	839	872
01BQ001	182	351	437	491	518

Table B.2: Nonparametric Flood Frequencies (continued)

Station	Return Period (Yrs)				
	2	10	20	50	100
01BR001	33.S	56.4	59.0	61.3	62.S
01BS001	45.S	64.2	66.6	6S.9	70.2
01BU002	90.9	135	154	20S	222
01BU003	3S.S	63.7	73.3	S1.5	S5.9
01BU004	12.5	23.1	26.1	29.1	30.S
01BV005	13.3	27.S	31.6	35.5	37.9
01BV006	57.9	9S.4	116	122	125
01BV007	S4.S	152	16S	1S2	191
01DL001	21.7	40.S	45.3	49.6	52.0
01BD002	441	625	815	964	1010
01BF001	263	413	463	517	547
1013500	238	327	366	396	437
1016500	170	297	336	377	402

Table B.3: Regression Results for 2 Year Flood

$$\log Q = a + b \log DA$$

Region	Coefficients		ISE		No. of
	a	b	Par.	Np	Stns.
PW	-0.336	0.871	3.85	3.56	53
C	-0.510	0.930	1.47	3.82	10
N	-0.529	0.924	3.25	1.82	18
E	-0.486	0.924	2.15	1.80	12
S	-0.131	0.819	4.28	4.02	13
HB	-0.346	0.876	4.07	3.58	44
NM	-0.688	0.974	2.86	1.66	16
EM	-0.476	0.919	2.36	9.34	10
SM	-0.135	0.846	4.28	4.93	8

Table B.3: Regression Results for 2 Year Flood (continued)

$$\log Q = c + d \log DA + e \log MAP$$

Region	Coefficients			ISE		No. of
	c	d	e	Par.	Np	Stns
PW	-5.246	0.919	1.572	3.33	1.87	53
C	-1.116	0.933	0.195	1.40	3.28	10
N	-3.215	0.927	0.888	3.26	2.27	18
E	0.915	0.914	-0.455	2.17	1.99	12
S	-9.123	0.890	2.855	2.91	3.54	13
HB	-5.849	0.932	1.759	3.40	2.68	44

PW - province-wide; C - central; N - northwestern; E- eastern; S - southern; HB - homogeneous bimodal region; NM - northwestern modified; EM - eastern modified; SM - southern modified.

Table B.4: Regression Results for 10 Year Flood

$$\log Q = a + b \log DA$$

Region	Coefficients		ISE		No. of
	a	b	Par.	Np	Stns.
PW	-0.043	0.850	4.37	4.14	53
C	-0.169	0.909	1.93	3.25	10
N	-0.327	0.928	3.87	3.20	18
E	-0.166	0.875	1.74	2.66	12
S	0.154	0.809	4.28	4.02	13
HB	-0.049	0.852	4.41	3.96	44
NM	-0.476	0.975	3.76	0.68	16
EM	-0.168	0.877	1.92	1.13	10
SM	0.183	0.817	3.99	1.52	8

Table B.4: Regression Results for 10 Year Flood (continued)

$$\log Q = c + d \log DA + e \log MAP$$

Region	Coefficients			ISE		No. of
	c	d	e	Par.	Np	Stns
PW	-6.792	0.915	2.160	3.30	2.88	53
C	-4.357	0.931	1.351	1.39	2.99	10
N	-3.243	0.931	0.965	3.87	1.40	18
E	-0.214	0.875	0.016	1.74	1.00	12
S	-8.998	0.882	2.906	3.41	3.95	13
HB	-7.075	0.925	2.245	3.18	2.11	44

PW - province-wide; C - central; N - northwestern; E- eastern; S - southern; HB - homogeneous bimodal region; NM - northwestern modified; EM - eastern modified; SM - southern modified.

Table B.5: Regression Results for 20 Year Flood

$$\log Q = a + b \log DA$$

Region	Coefficients		ISE		No. of
	a	b	Par.	Np	Stns.
PW	0.040	0.844	4.61	4.37	53
C	-0.072	0.905	2.36	2.82	10
N	-0.263	0.926	3.78	2.40	18
E	-0.066	0.856	1.76	1.31	12
S	0.213	0.818	4.71	4.46	13
HB	0.024	0.849	4.49	4.02	44
NM	-0.395	0.968	3.79	3.65	16
EM	-0.069	0.858	2.01	0.93	10
SM	0.224	0.826	3.80	4.55	8

Table B.5: Regression Results for 20 Year Flood (continued)

$$\log Q = c + d \log DA + e \log MAP$$

Region	Coefficients			ISE		No. of
	c	d	e	Par.	Np	Stns
PW	-7.204	0.914	2.319	3.40	3.03	53
C	-4.577	0.928	1.453	1.78	3.04	10
N	-3.369	0.929	1.028	3.66	1.34	18
E	-0.803	0.860	0.240	1.77	1.43	12
S	-7.566	0.879	2.470	3.58	3.71	13
HB	-7.253	0.924	2.326	3.15	2.74	44

PW - province-wide; C - central; N - northwestern; E- eastern; S - southern; HB - homogeneous bimodal region; NM - northwestern modified; EM - eastern modified; SM - southern modified.

Table B.6: Regression Results for 50 Year Flood

$$\log Q = a + b \log DA$$

Region	Coefficients		ISE		No. of
	a	b	Par.	Np	Stns.
PW	-0.005	0.881	4.13	3.92	45
C	-0.041	0.916	2.48	***	8
N	-0.155	0.913	3.22	1.66	16
E	-0.038	0.871	2.89	2.77	11
S	0.061	0.900	3.94	2.46	10
HB	-0.009	0.882	3.87	3.32	37
NM	-0.272	0.952	3.07	1.14	14
EM	-0.062	0.884	3.16	4.40	9
SM	0.172	0.853	1.83	6.86	6

*** - no convergence of cross-validation approach

Table B.6: Regression Results for 50 Year Flood (continued)

$$\log Q = c + d \log DA + e \log MAP$$

Region	Coefficients			ISE		No. of
	c	d	e	Par.	Np	Stns
PW	-6.594	0.930	2.124	3.14	2.83	45
C	-6.753	0.954	2.159	1.58	2.11	8
N	-4.222	0.906	1.356	3.15	1.05	16
E	0.289	0.867	-0.105	2.90	2.19	11
S	-5.922	0.940	1.909	3.88	1.87	10
HB	-6.156	0.930	1.980	2.94	2.61	37

PW - province-wide; C - central; N - northwestern; E- eastern; S - southern; HB - homogeneous bimodal region; NM - northwestern modified; EM - eastern modified; SM - southern modified.

Table B.7: Regression Results for 100 Year Flood

$$\log Q = a + b \log DA$$

Region	Coefficients		ISE		No. of
.	a	b	Par.	Np	Stns.
.					
PW	0.036	0.877	3.98	3.83	45
C	-0.027	0.918	2.41	2.52	8
N	-0.094	0.903	2.92	0.95	16
E	-0.023	0.874	2.88	2.82	11
S	0.135	0.890	4.10	2.44	10
.					
HB	0.011	0.883	3.69	3.17	37
.					
NM	-0.193	0.936	2.83	2.63	14
EM	-0.041	0.884	3.27	3.05	9
SM	0.191	0.854	1.64	1.29	6

Table B.7: Regression Results for 100 Year Flood (continued)

$$\log Q = c + d \log DA + e \log MAP$$

Region	Coefficients			ISE		No. of
.	c	d	e	Par.	Np	Stns
.						
PW	-6.897	0.928	2.235	2.99	2.12	45
C	-6.696	0.956	2.145	1.54	2.05	8
N	-4.876	0.894	1.595	2.77	1.07	16
E	0.436	0.868	-0.147	2.90	2.10	11
S	-4.429	0.920	1.456	3.70	2.12	10
.						
HB	-6.060	0.931	1.955	2.75	2.16	37

PW - province-wide; C - central; N - northwestern; E- eastern; S - southern; HB - homogeneous bimodal region; NM - northwestern modified; EM - eastern modified; SM - southern modified.

Table B.8: ISE's for Linear Simulations

Linear Regression

Simulation	Sample Size				
No.	10	20	30	40	50
1	10.63	7.24	10.24	8.24	8.96
2	12.35	9.81	9.33	9.19	7.64
3	5.27	7.63	9.56	10.32	8.45
4	6.03	10.88	7.41	9.99	9.47
5	7.08	7.51	6.51	9.39	9.25
6	7.89	7.64	9.42	8.33	7.33
7	8.16	7.39	8.65	10.20	9.22
8	9.41	7.47	8.62	11.20	8.67
9	4.57	7.01	9.42	7.87	8.06
10	8.49	10.51	9.61	9.70	8.00

Table B.5: ISE's for Linear Simulations(continued)

Nonparametric Regression

Simulation	Sample Size				
No.	10	20	30	40	50
1	10.46	5.71	9.48	6.27	7.50
2	11.56	10.55	8.70	7.60	6.83
3	3.04	6.98	8.37	9.51	7.73
4	5.98	9.07	7.08	8.43	9.29
5	6.64	6.84	6.41	9.09	6.56
6	7.25	6.29	8.72	7.90	6.59
7	4.54	6.74	7.88	10.26	8.45
8	10.26	7.05	8.39	9.32	7.96
9	1.65	9.23	8.82	6.96	7.27
10	6.97	10.28	8.34	6.12	7.27

Table B.9: ISE's for Nonlinear Simulations

Linear Regression

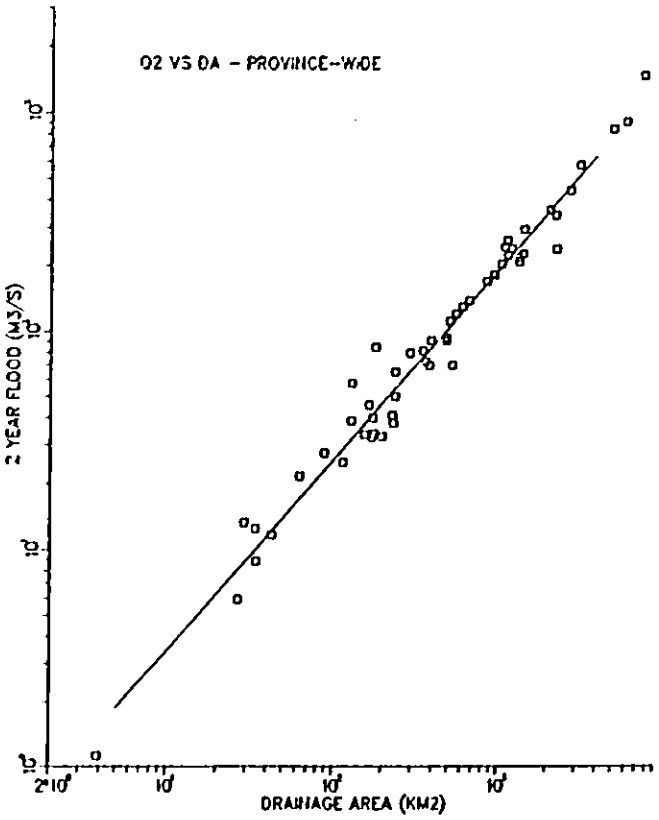
Simulation	Sample Size				
No.	10	20	30	40	50
1	9.66	9.18	10.67	9.41	8.75
2	7.66	10.82	9.48	10.25	10.57
3	11.75	9.86	8.24	9.27	9.89
4	7.27	8.19	11.02	10.59	9.80
5	7.07	8.82	8.83	9.31	9.91
6	5.86	6.84	13.55	10.42	11.73
7	9.14	7.85	10.14	8.64	9.92
8	7.50	9.05	9.39	9.63	8.59
9	9.96	10.61	9.31	8.83	11.43
10	8.64	9.68	7.76	10.66	10.43

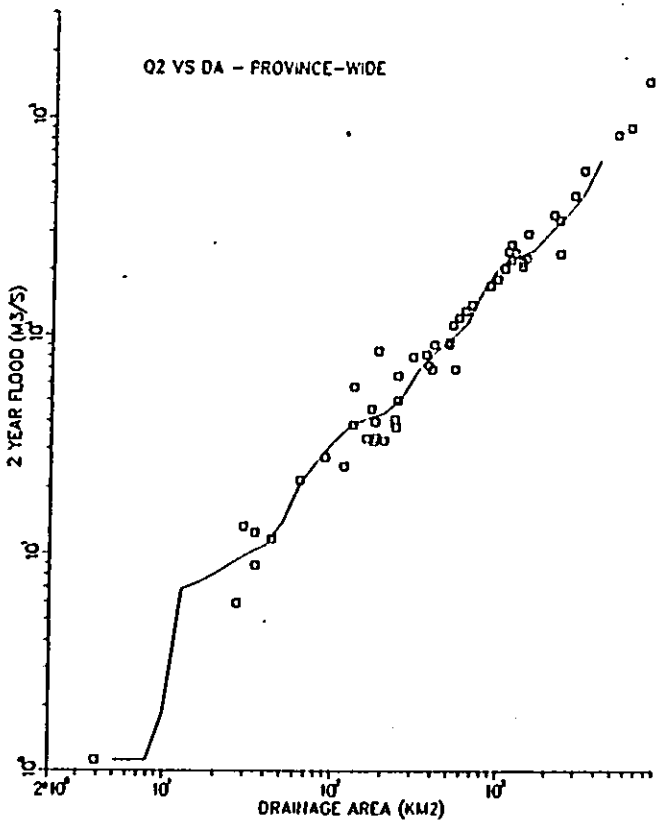
Table B.9: ISE's for Nonlinear Simulations(continued)

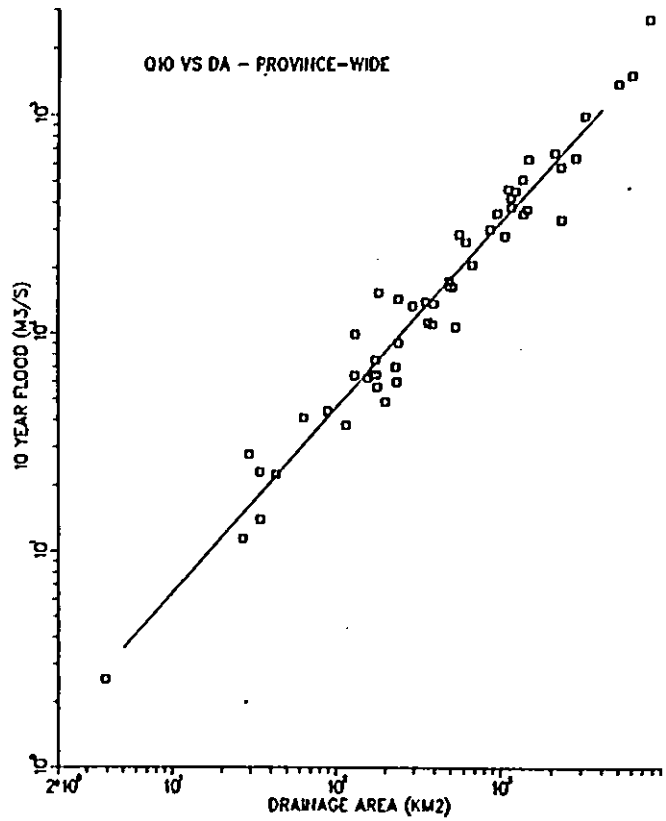
Nonparametric Regression

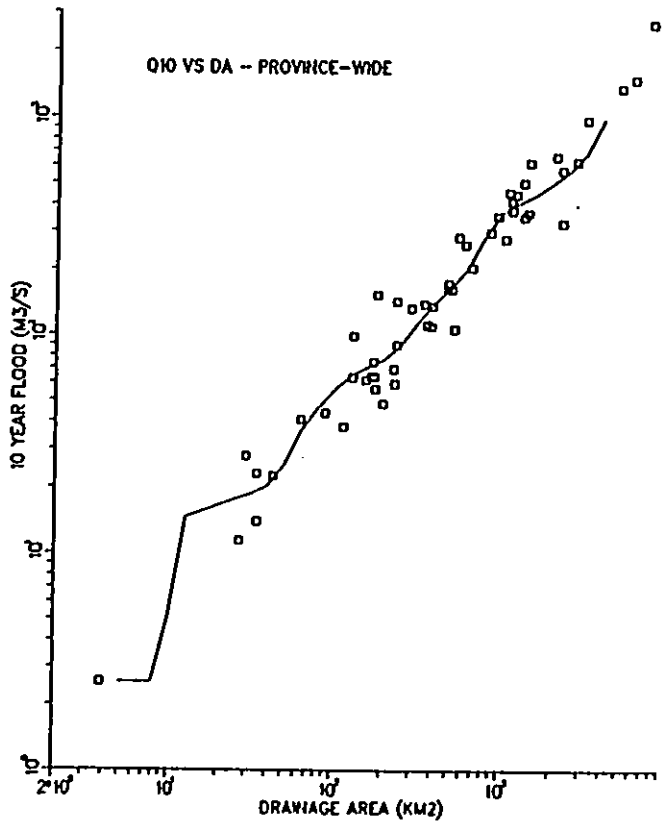
Simulation	Sample Size				
No.	10	20	30	40	50
1	8.97	7.61	10.82	8.60	8.16
2	8.66	9.62	8.74	9.05	9.90
3	11.31	8.45	6.82	8.80	7.45
4	6.20	5.78	9.86	4.84	9.30
5	8.31	7.36	9.75	7.61	10.63
6	6.16	6.65	13.88	8.84	10.74
7	8.03	5.52	7.99	7.33	8.56
8	8.80	8.28	9.88	8.85	7.88
9	9.55	9.35	9.15	8.50	11.16
10	7.23	7.89	5.61	9.90	9.74

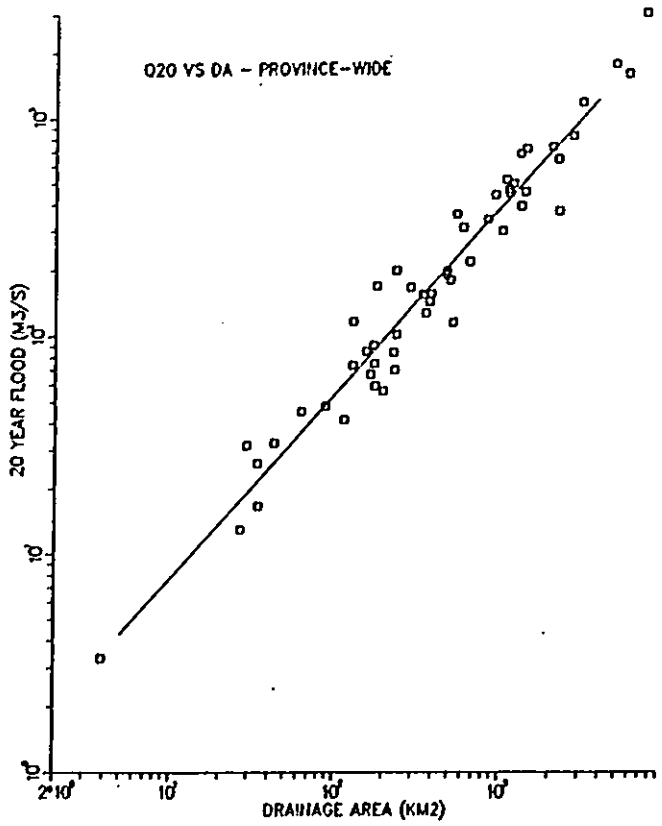
REGRESSION PLOTS

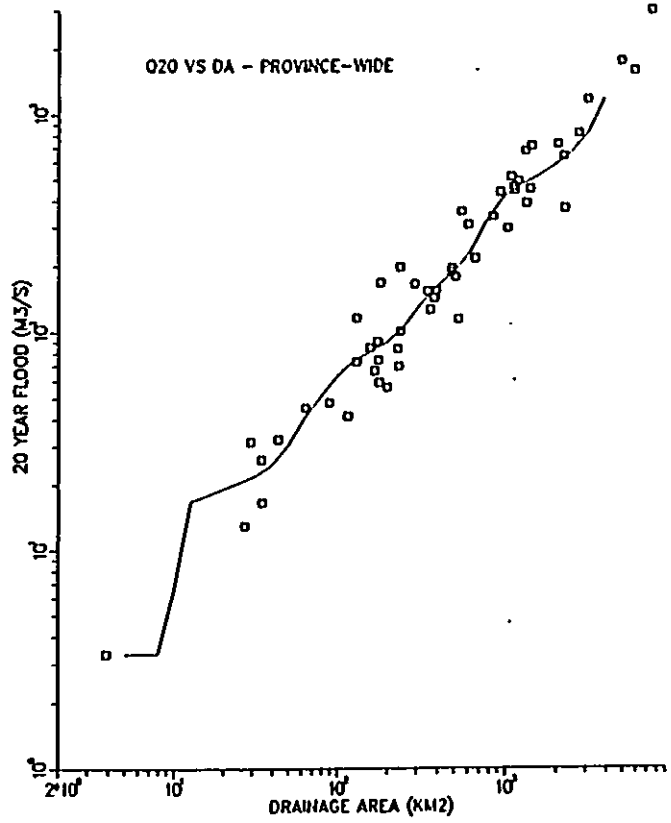

Q2 VS DA - PROVINCE-WIDE
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

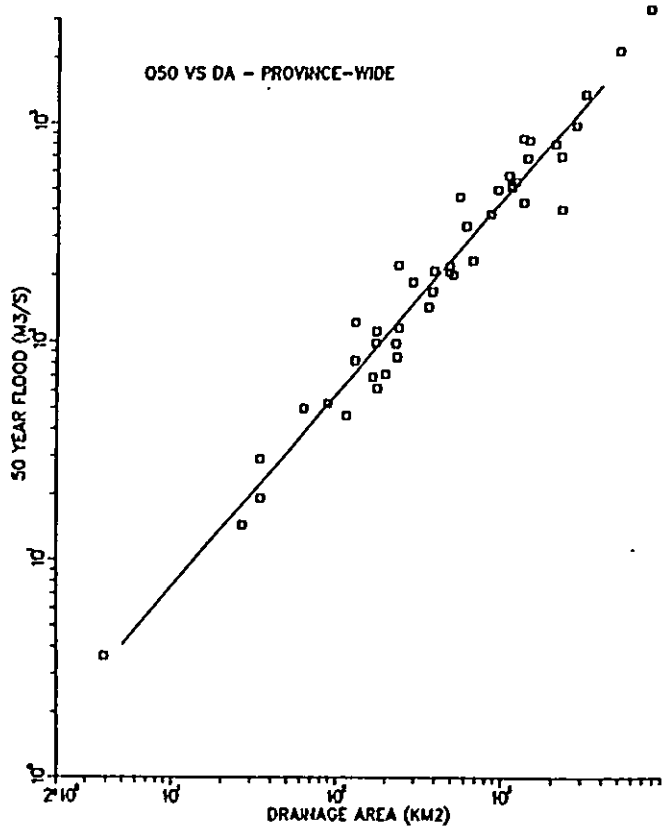

Q2 VS DA - PROVINCE-WIDE
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

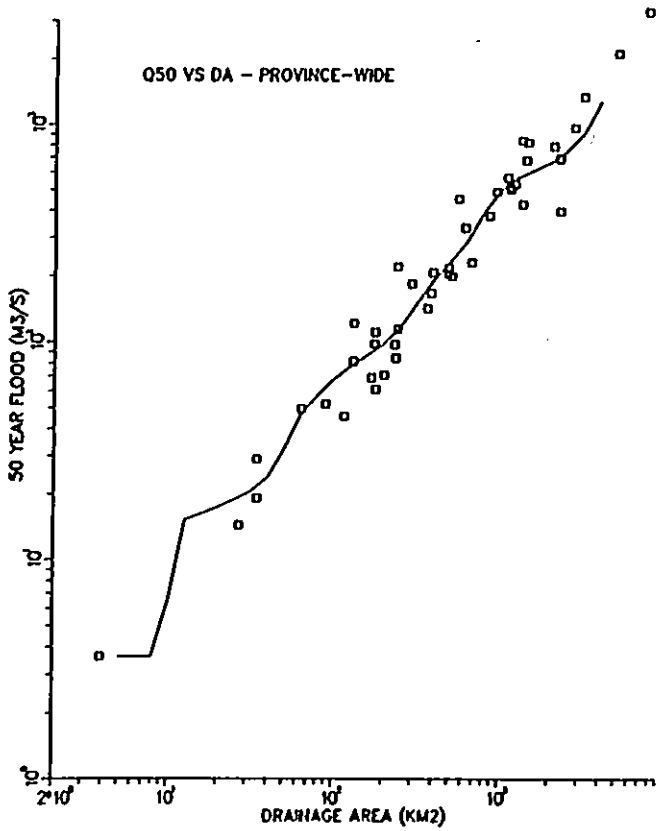

Q10 VS DA – PROVINCE-WIDE
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

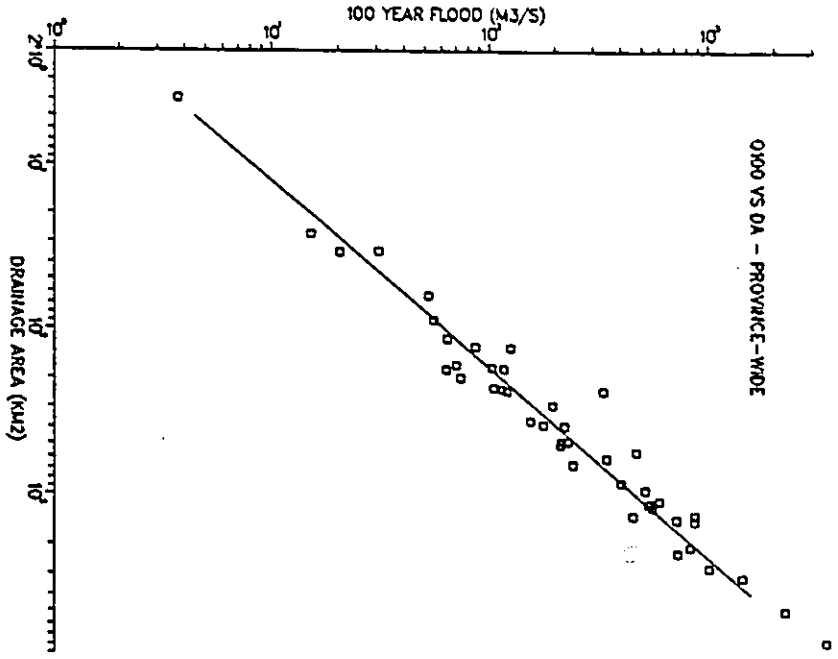

Q10 VS DA -- PROVINCE-WIDE
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

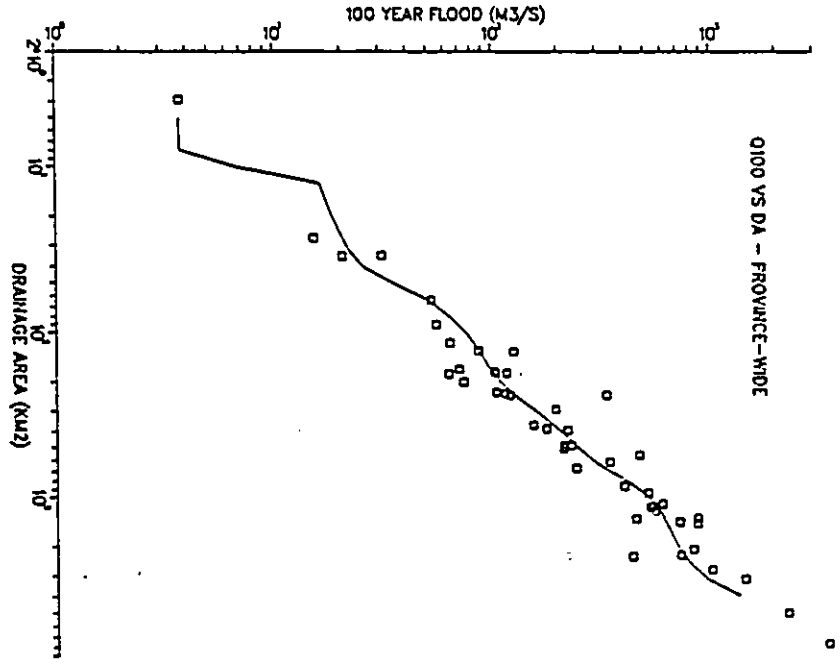

Q20 VS DA - PROVINCE-WIDE
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

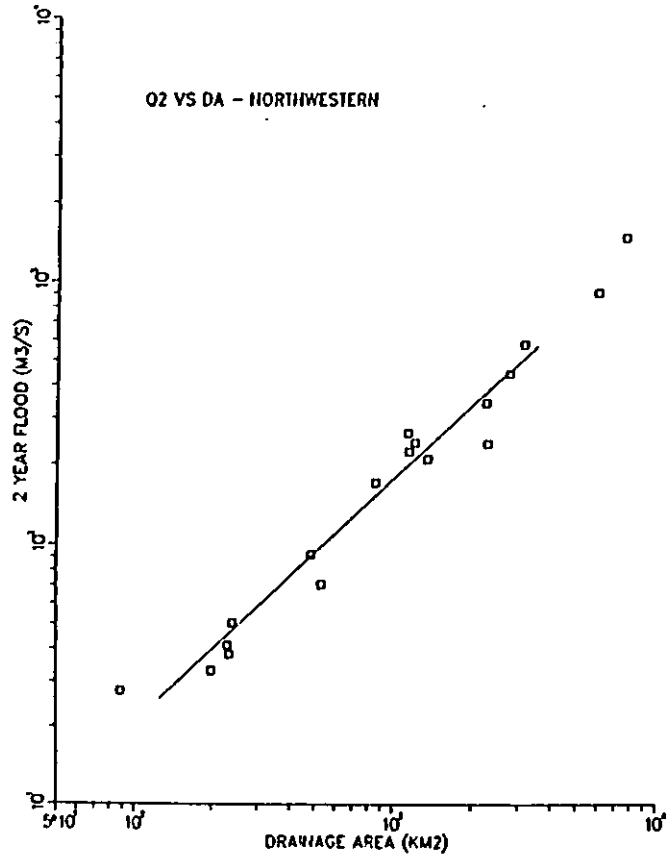

Q20 VS DA - PROVINCE-WIDE
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

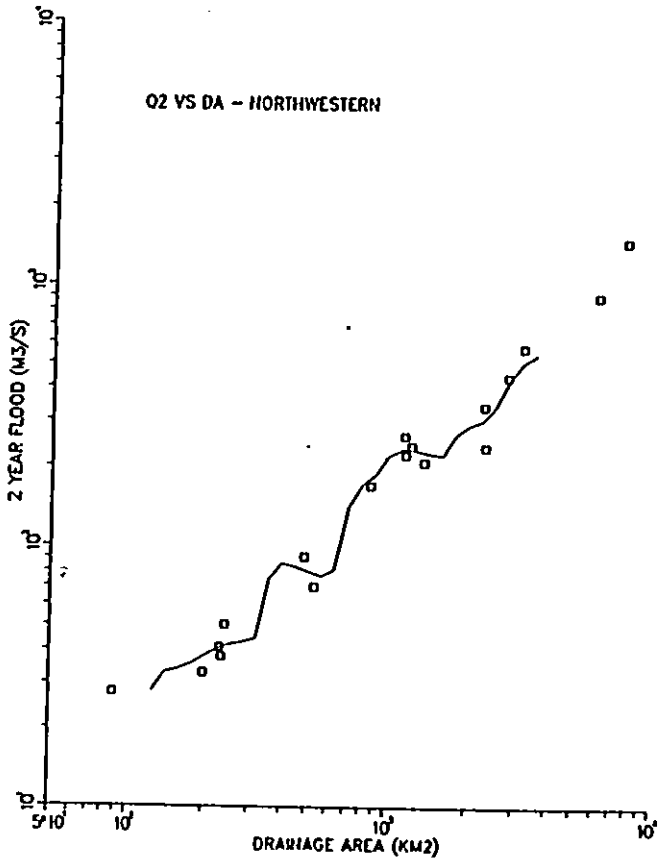

Q50 VS DA – PROVINCE-WIDE
50 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

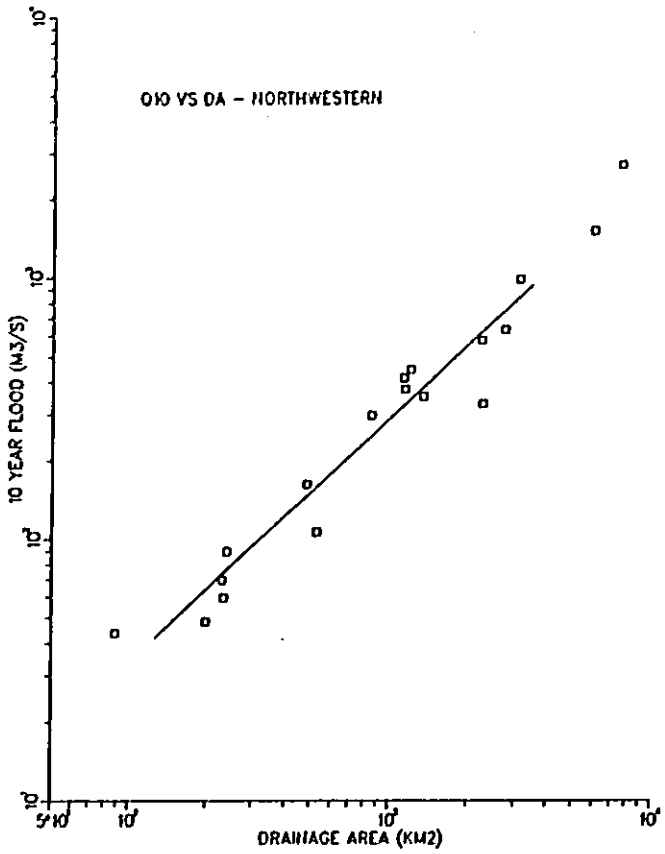

Q50 VS DA – PROVINCE-WIDE
50 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

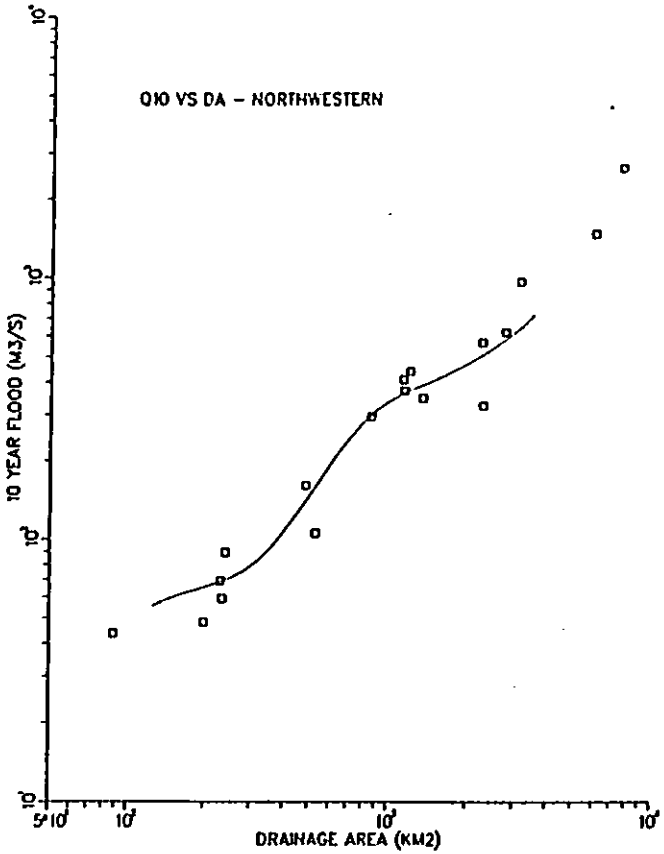

100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)
Q100 VS DA - PROVINCE-WIDE

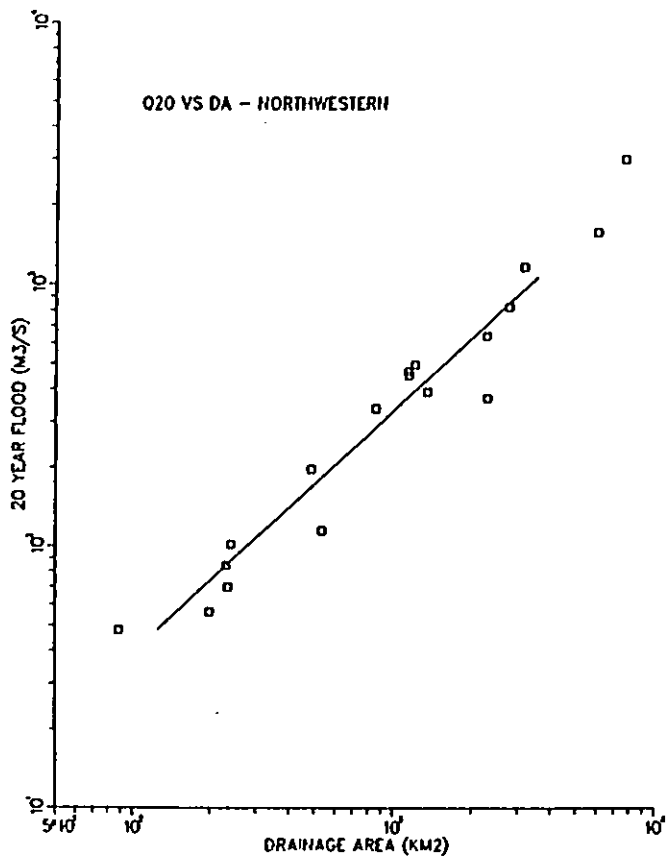

100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)
Q100 VS DA - PROVINCE-WIDE

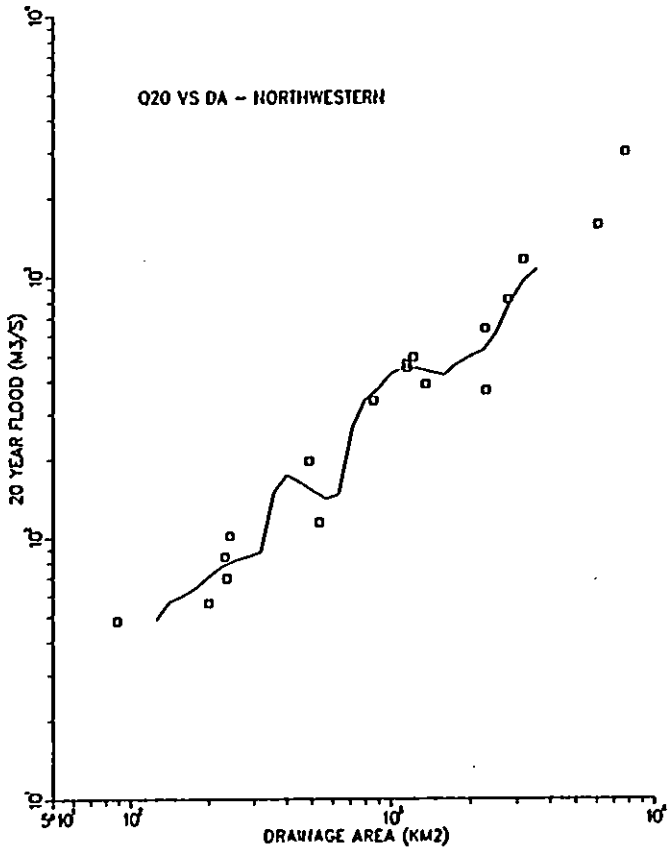

Q2 VS DA – NORTHWESTERN
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

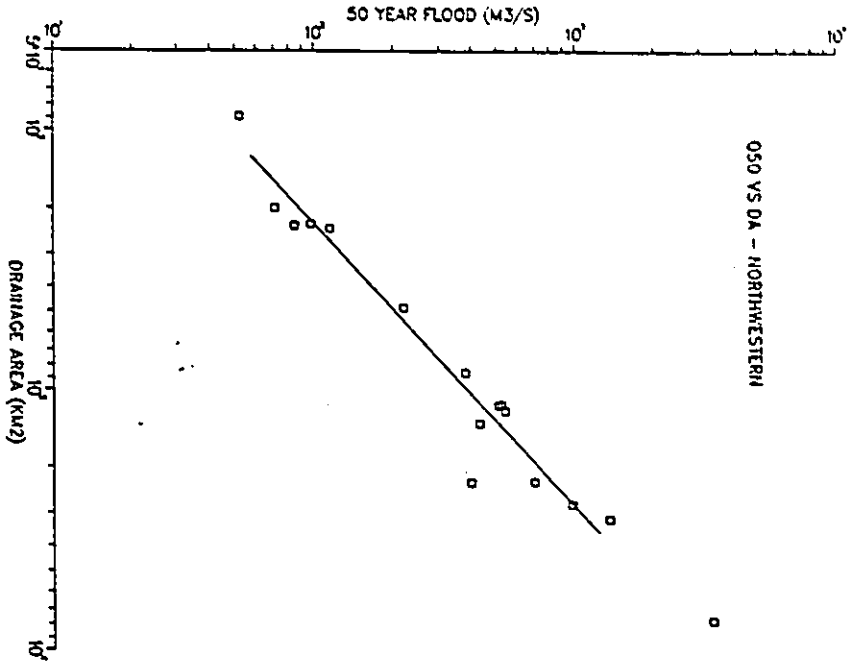

Q2 VS DA – NORTHWESTERN
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

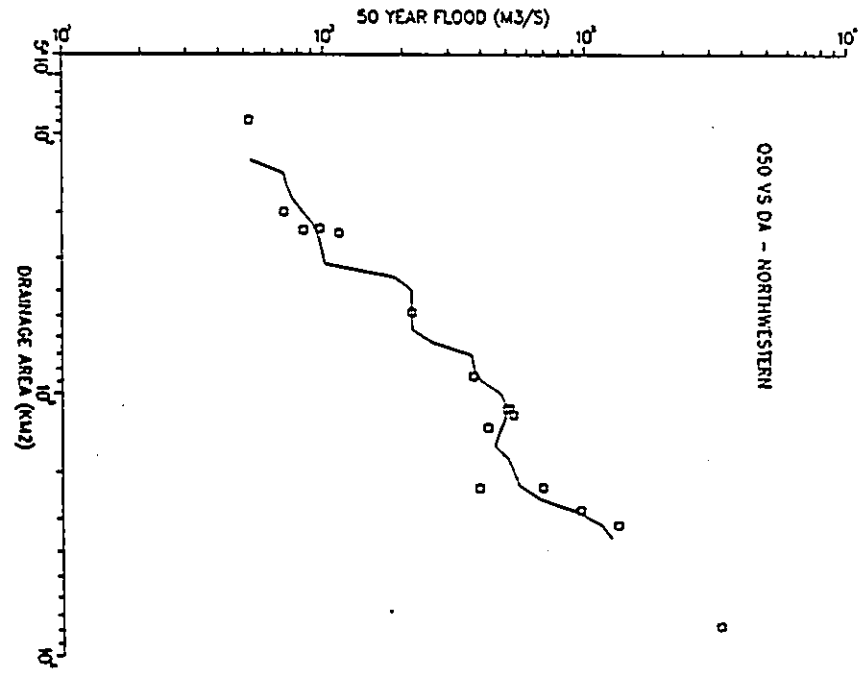

Q10 VS DA – NORTHWESTERN
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

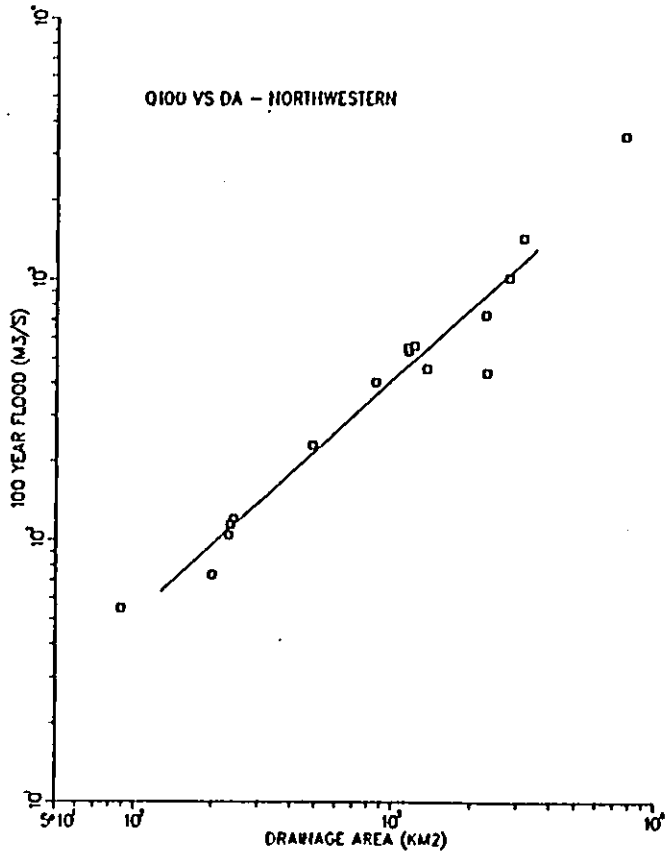

Q10 VS DA – NORTHWESTERN
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

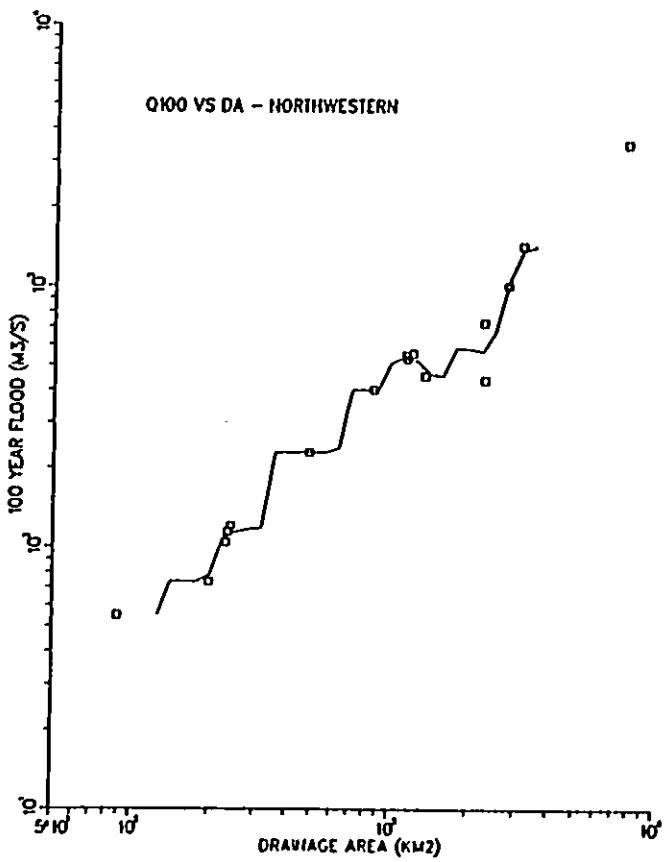

Q20 VS DA - NORTHWESTERN
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

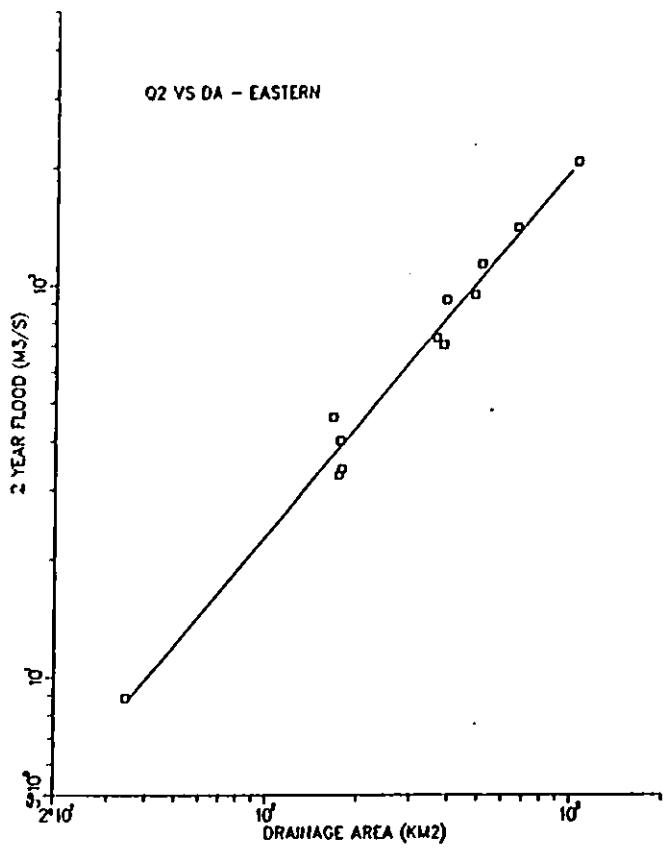

Q20 VS DA - NORTHWESTERN
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

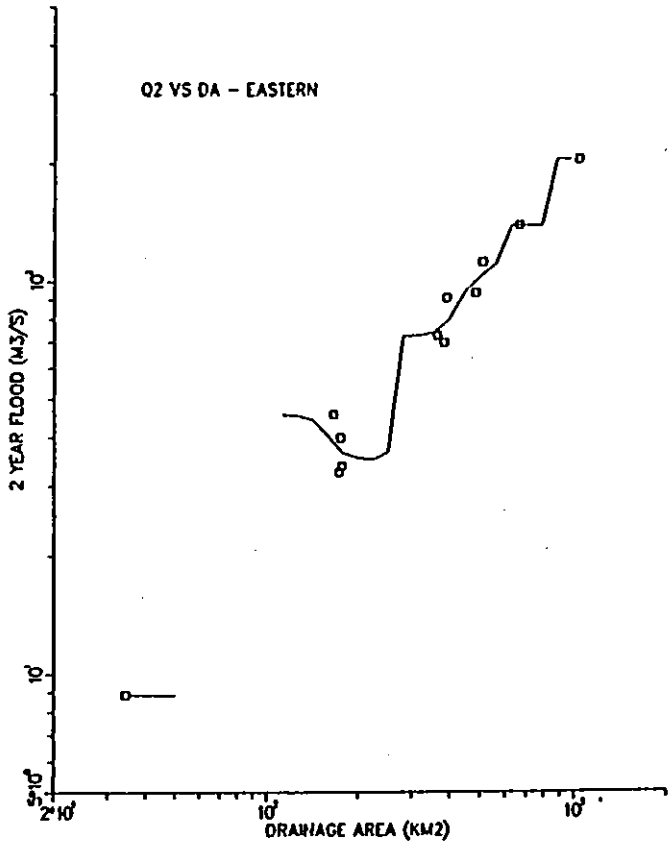

50 YEAR FLOOD (M3/S)
10⁰
10¹
10²
10³
5*10⁰
10¹
10²
5*10²
DRAINAGE AREA (KM2)
Q50 VS DA - NORTHWESTERN

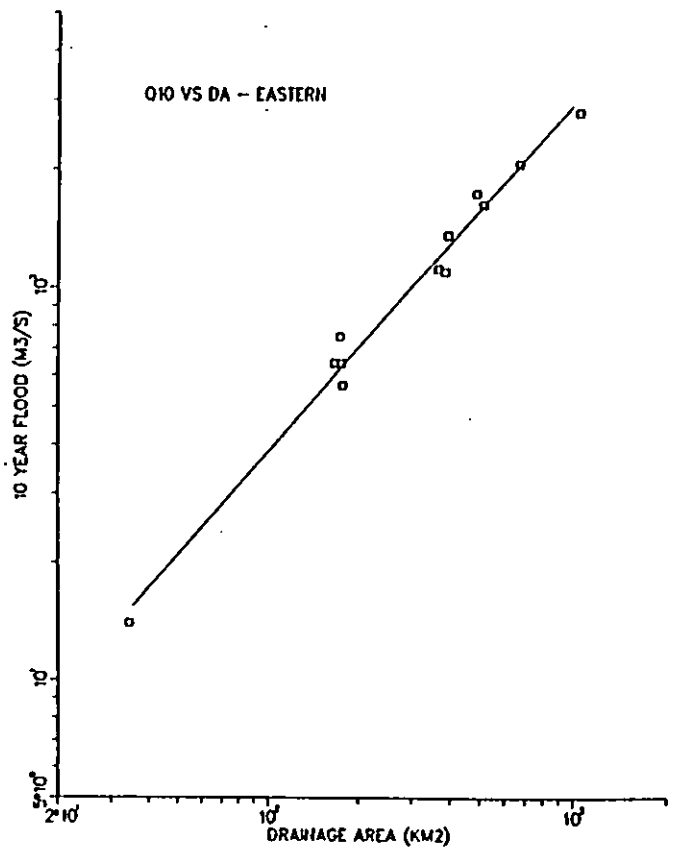

50 YEAR FLOOD (M3/S)
10⁰
10¹
10²
10³
5*10⁰
10¹
10²
5*10²
DRAINAGE AREA (KM2)
Q50 VS DA - NORTHWESTERN

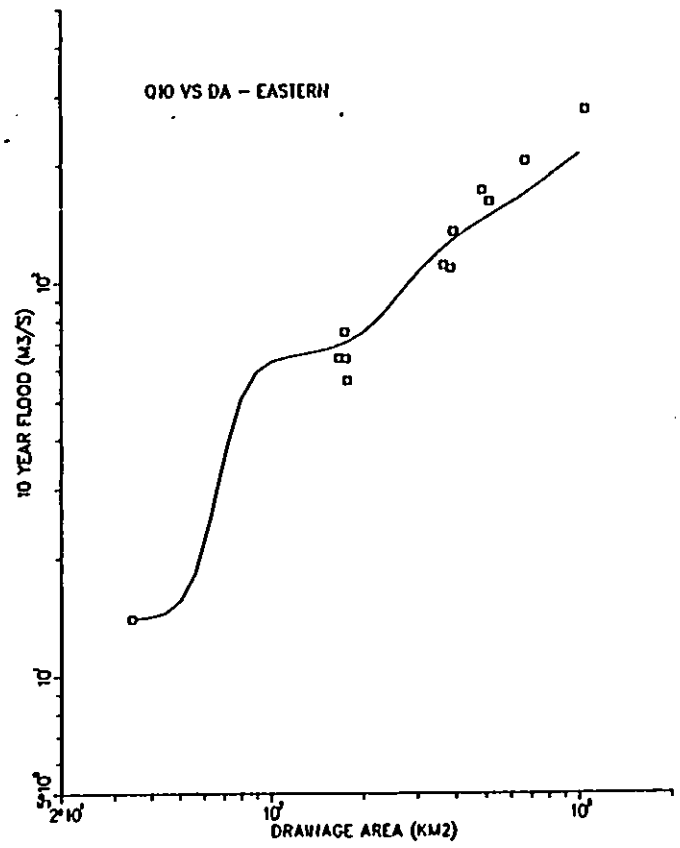

Q100 VS DA – NORTHWESTERN
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

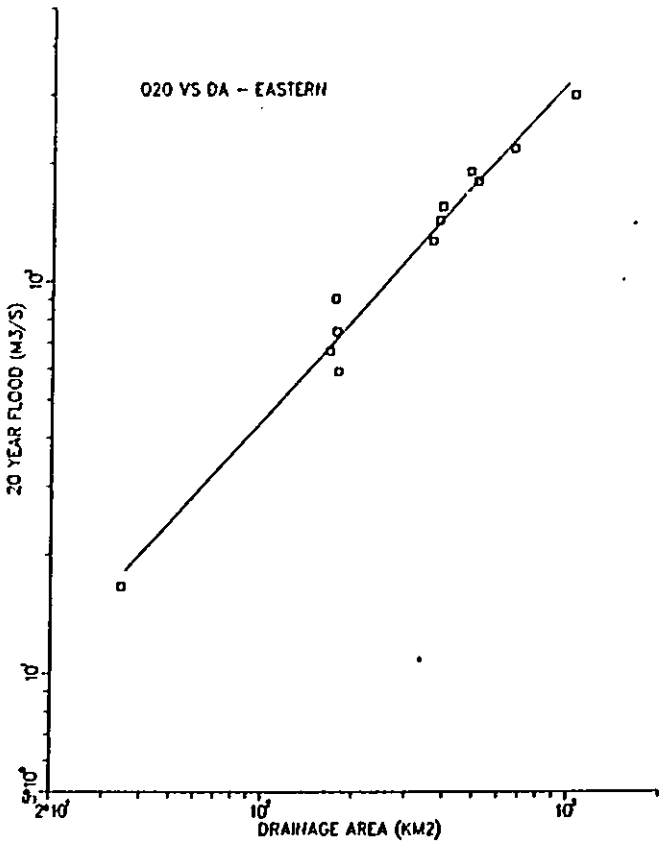

Q100 VS DA – NORTHWESTERN
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

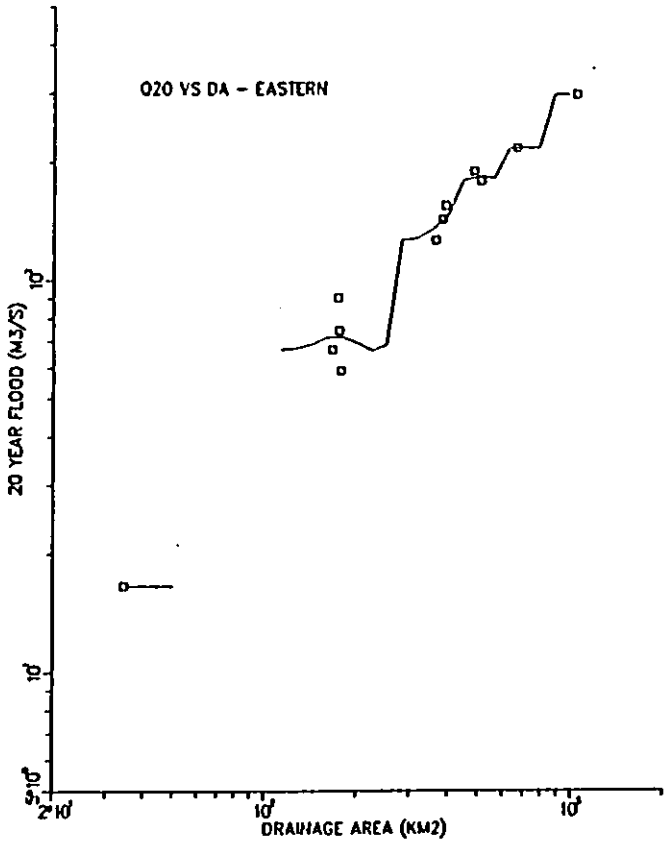

Q2 VS DA - EASTERN
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

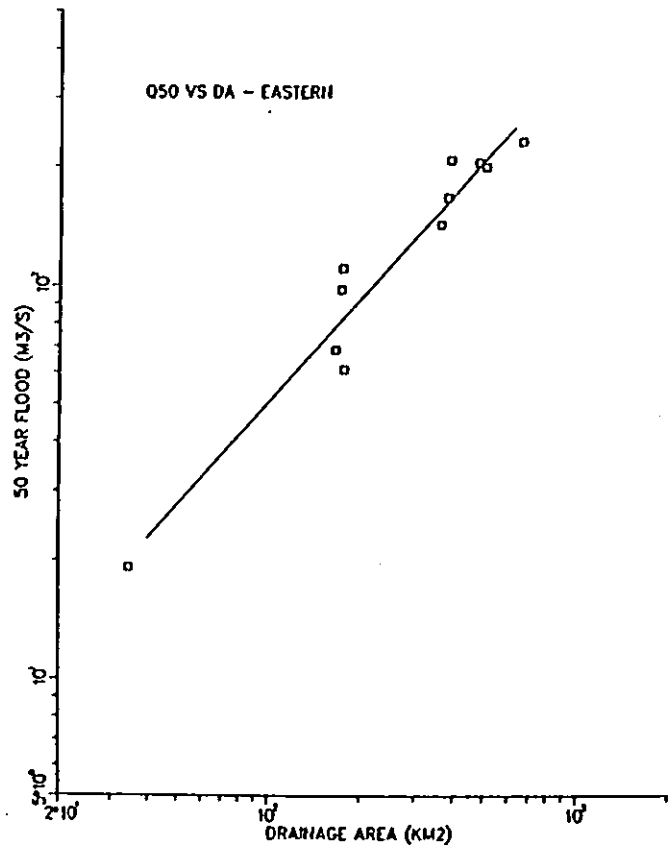

Q2 VS DA - EASTERN
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

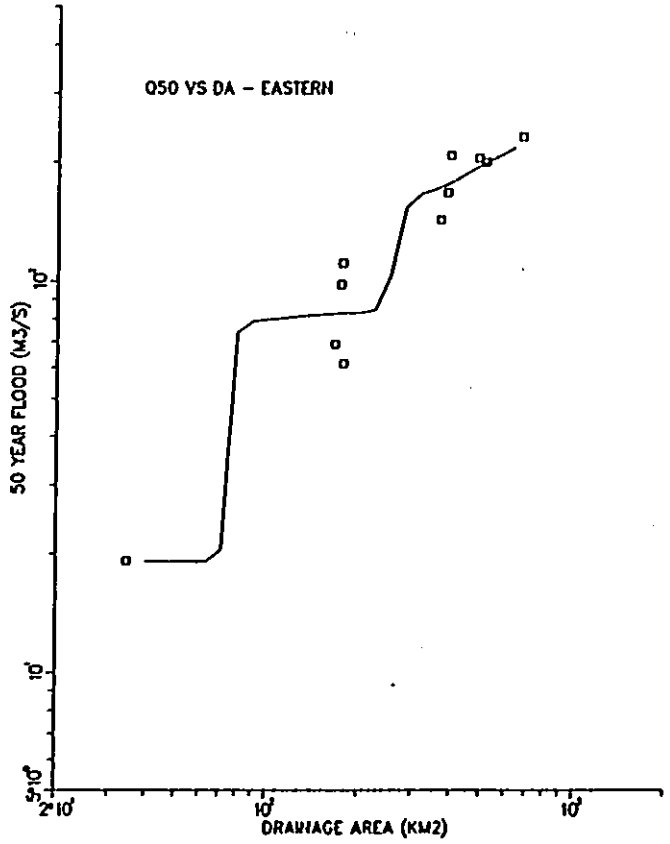

Q10 VS DA – EASTERN
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

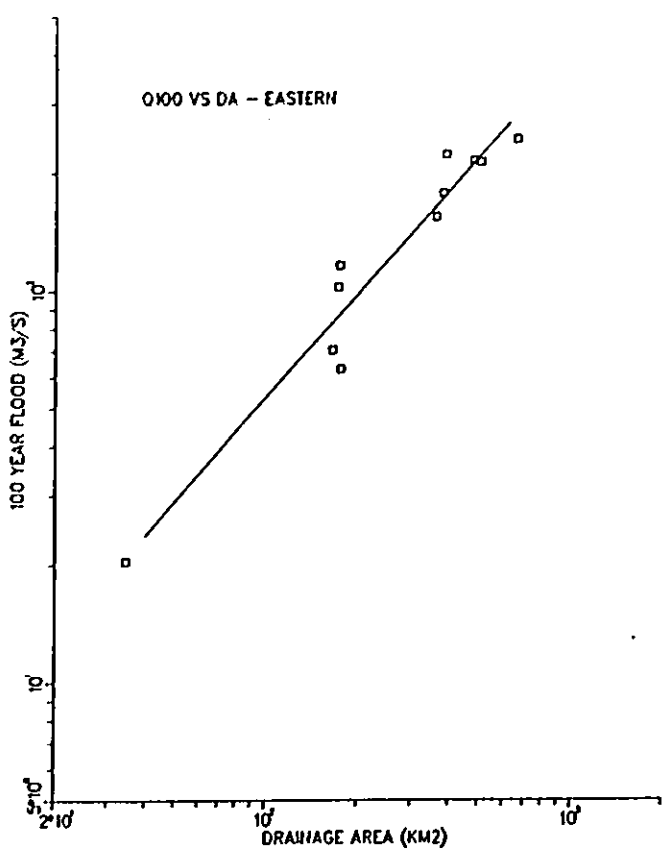

Q10 VS DA – EASTERN
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

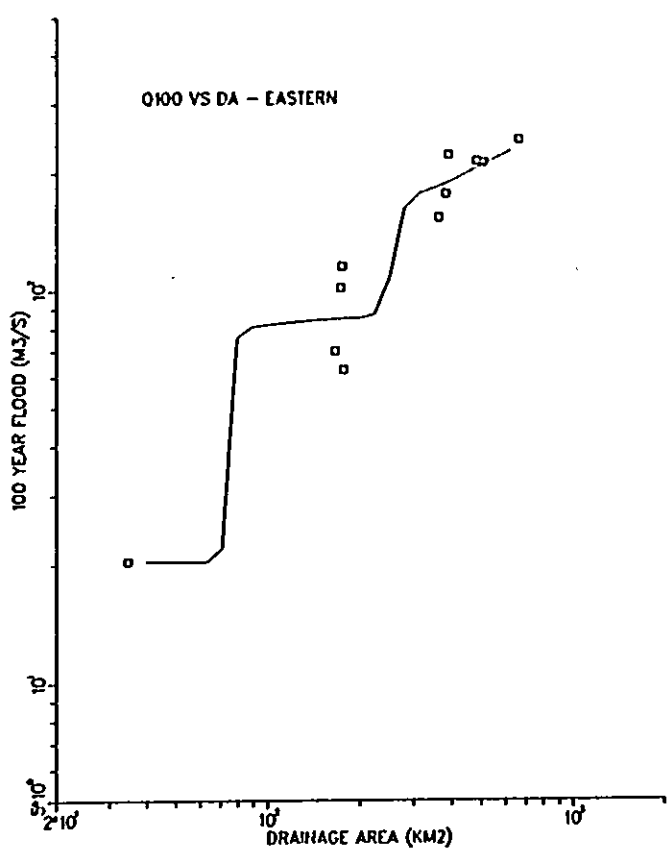

Q20 VS DA - EASTERN
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

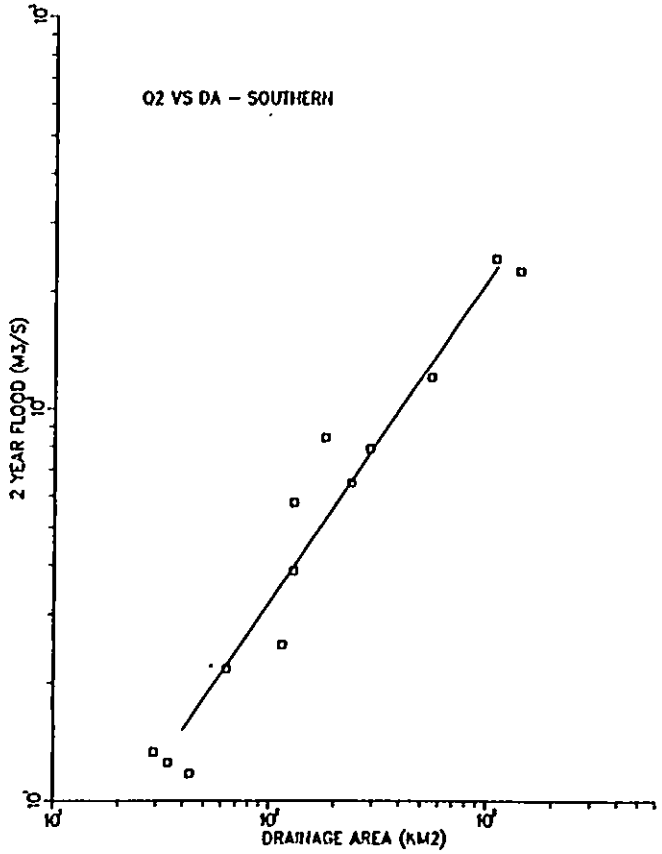

Q20 VS DA - EASTERN
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

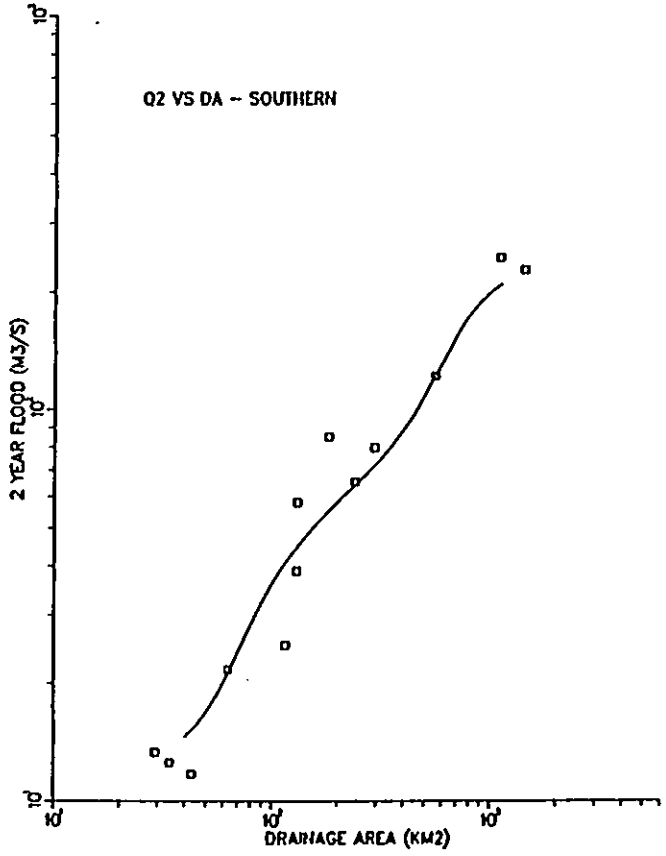

Q50 VS DA – EASTERN
50 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

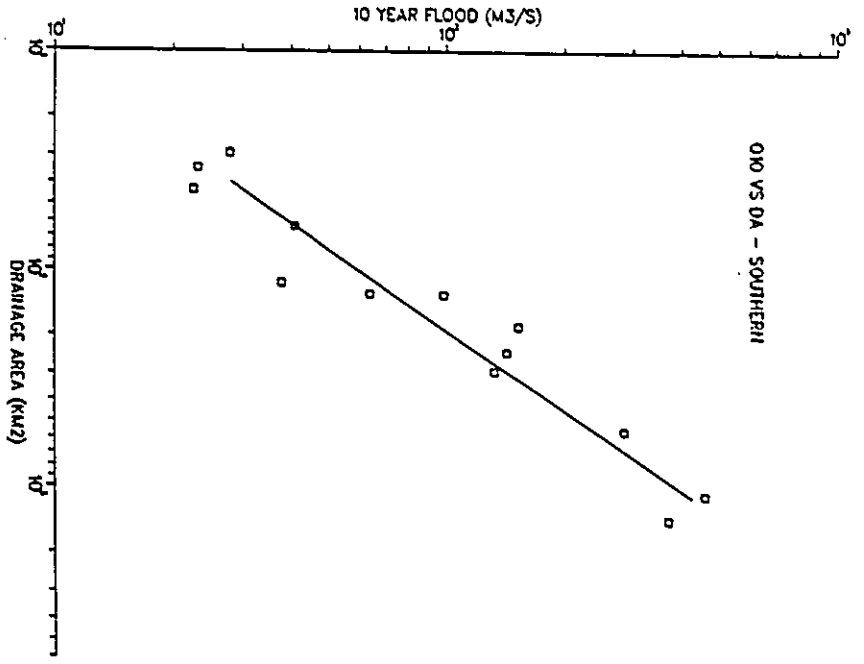

Q50 VS DA – EASTERN
50 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

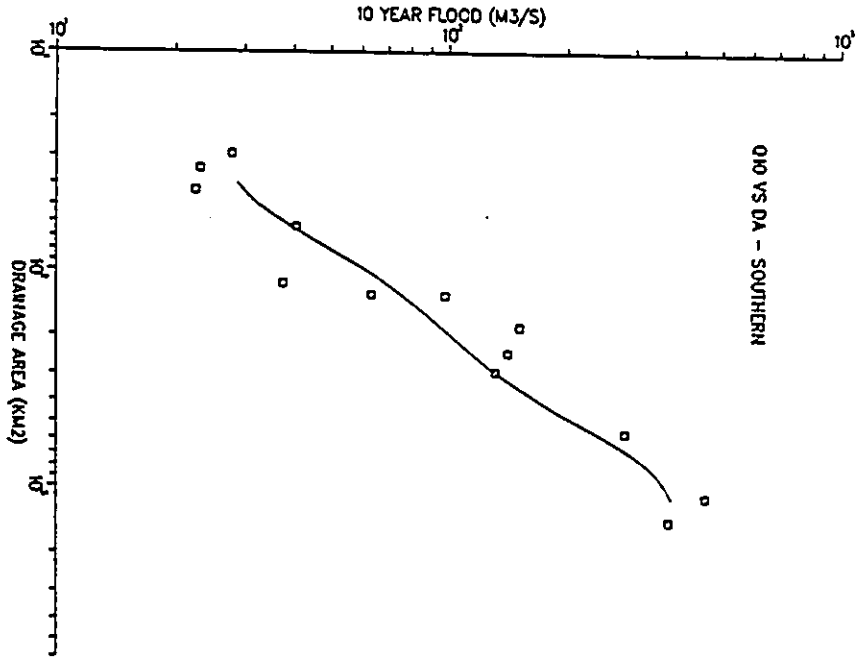

Q100 VS DA – EASTERN
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

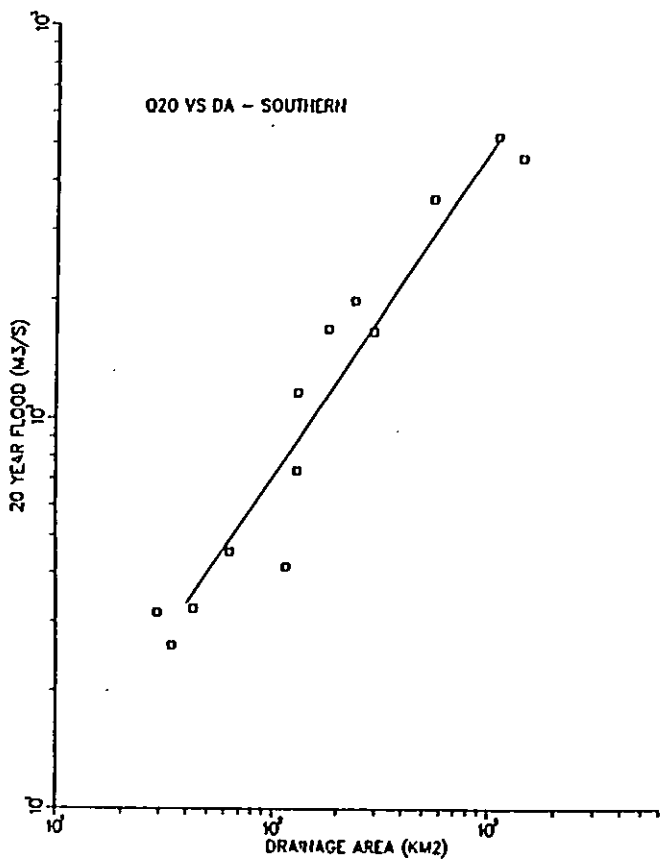

Q100 VS DA – EASTERN
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

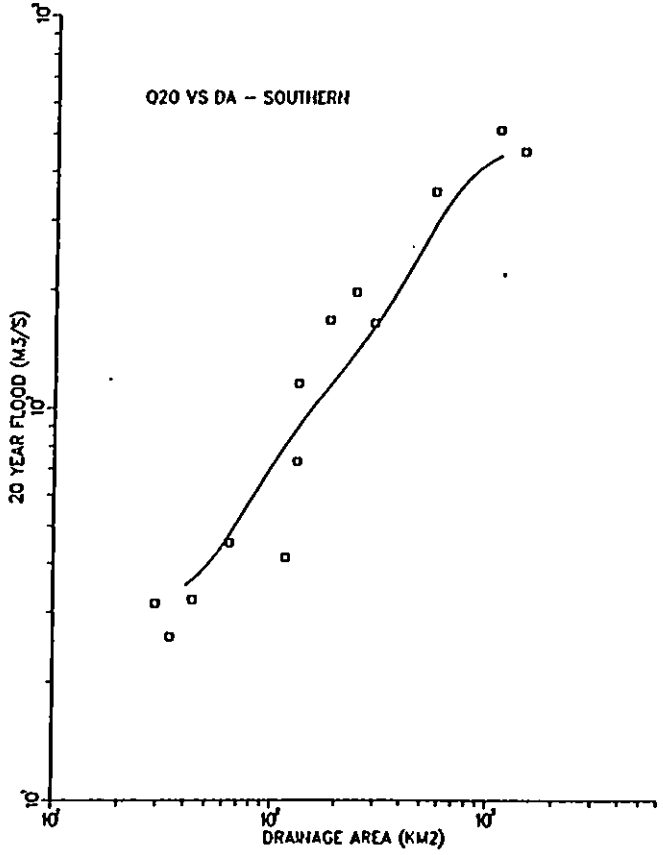

Q2 VS DA – SOUTHERN
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

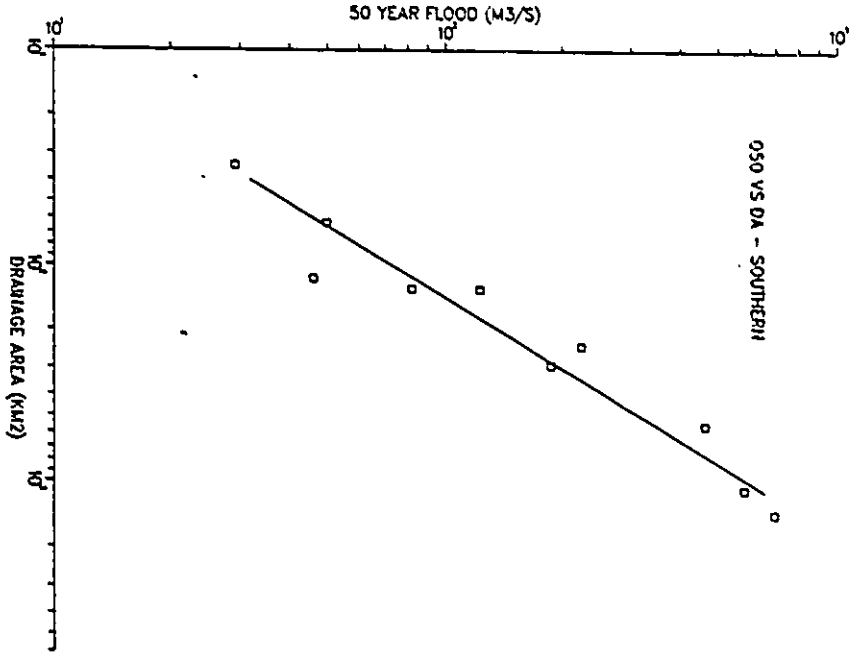

Q2 VS DA – SOUTHERN
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

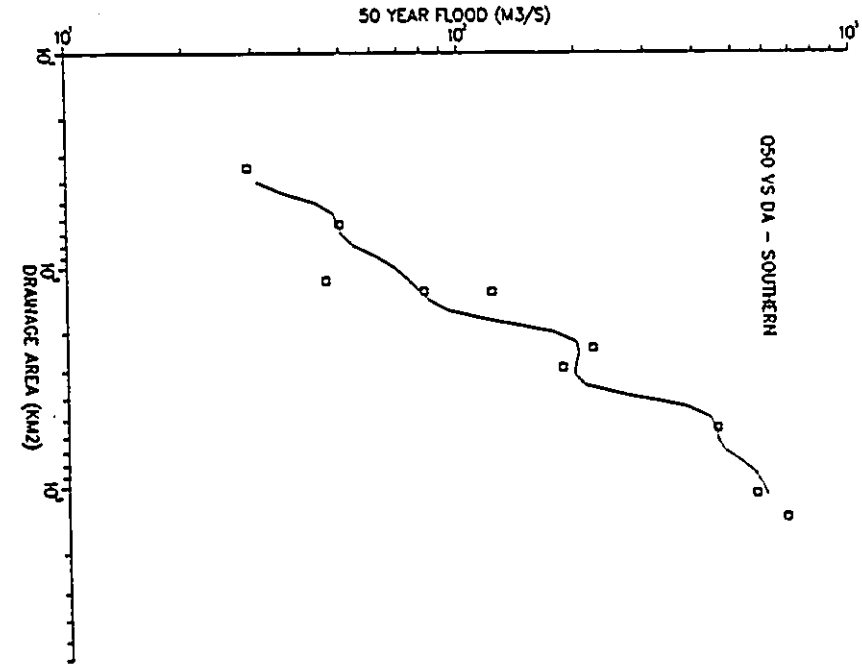

10 YEAR FLOOD (M3/S)
10¹
10²
10³
10⁰
10¹
10²
DRAINAGE AREA (KM2)
Q10 VS DA - SOUTHERN

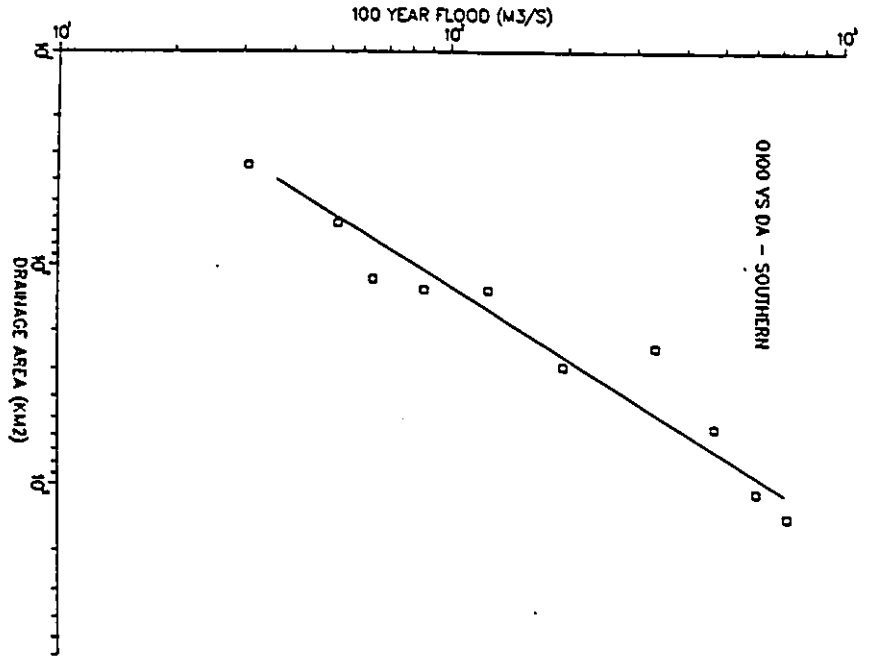

10 YEAR FLOOD (M3/S)
10¹
10²
10³
10⁰
10¹
10²
DRAINAGE AREA (KM2)
Q10 VS DA - SOUTHERN

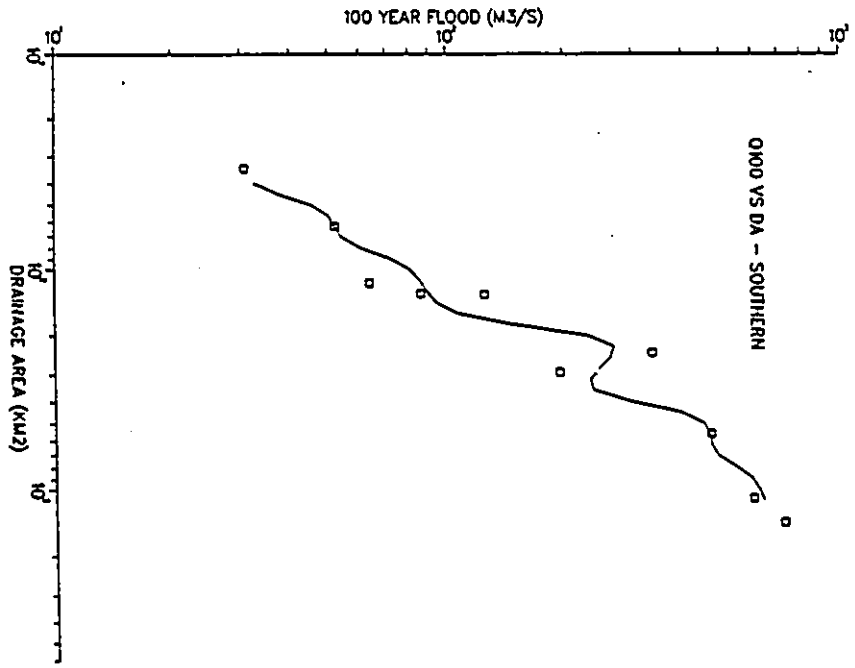

Q20 VS DA – SOUTHERN
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

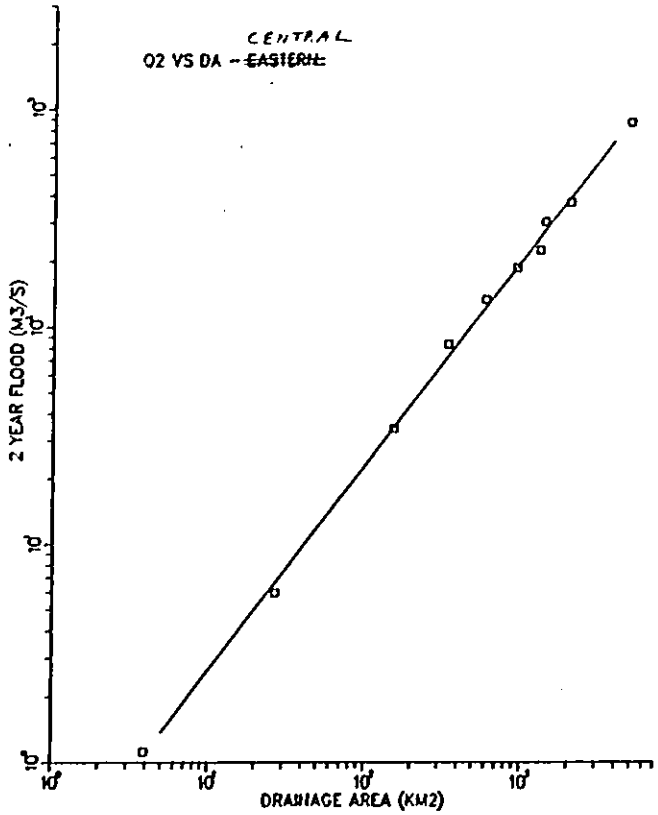

Q20 VS DA – SOUTHERN
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

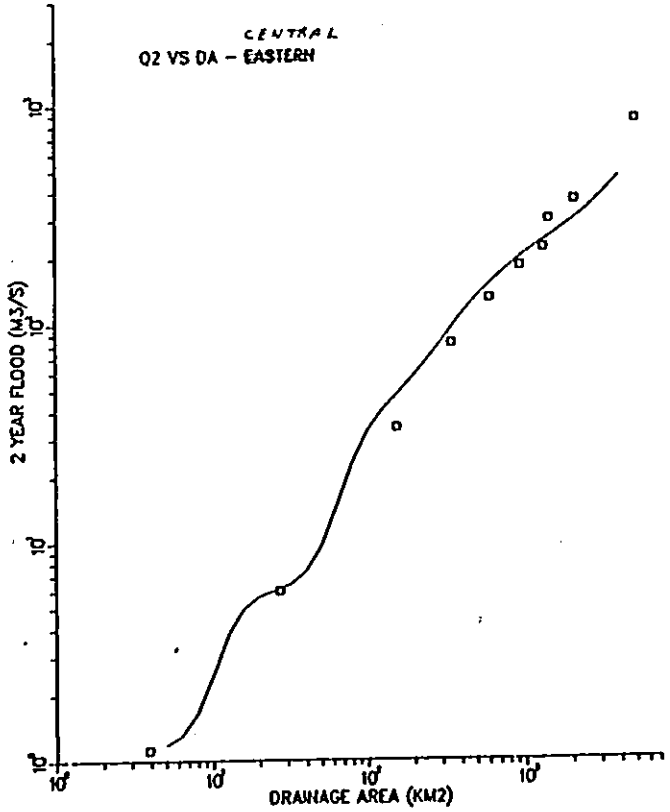

50 YEAR FLOOD (M3/S)
10¹
10²
10³
10⁰
10¹
10²
DRAINAGE AREA (KM2)
Q50 VS DA - SOUTHERN

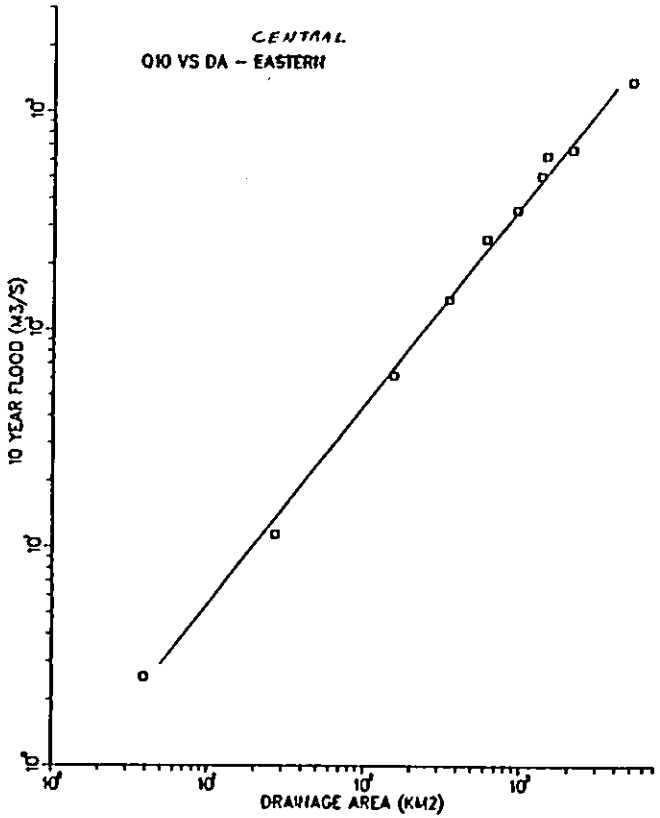

50 YEAR FLOOD (M3/S)
10¹
10²
10³
10⁰
10¹
10²
DRAINAGE AREA (KM2)
Q50 VS DA - SOUTHERN

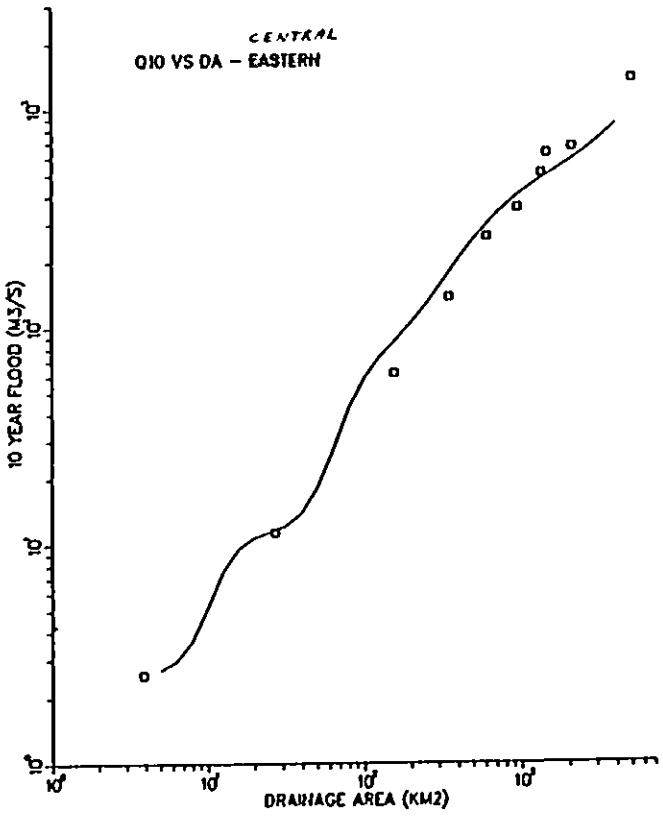

100 YEAR FLOOD (M3/S)
10¹
10²
10³
DRAINAGE AREA (KM2)
10¹
10²
10³
Q100 VS DA - SOUTHERN

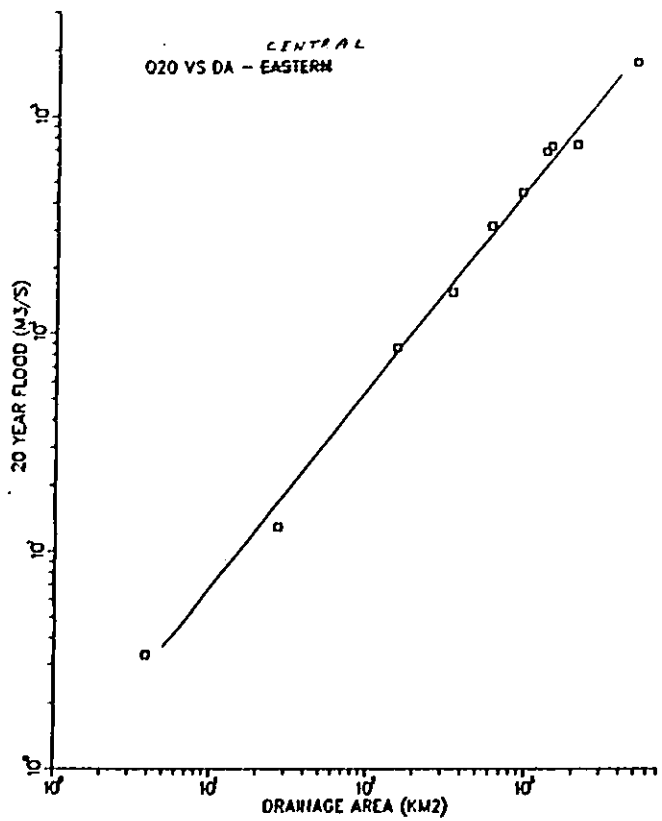

100 YEAR FLOOD (M3/S)
10¹
10²
10³
DRAINAGE AREA (KM2)
10¹
10²
10³
Q100 VS DA - SOUTHERN

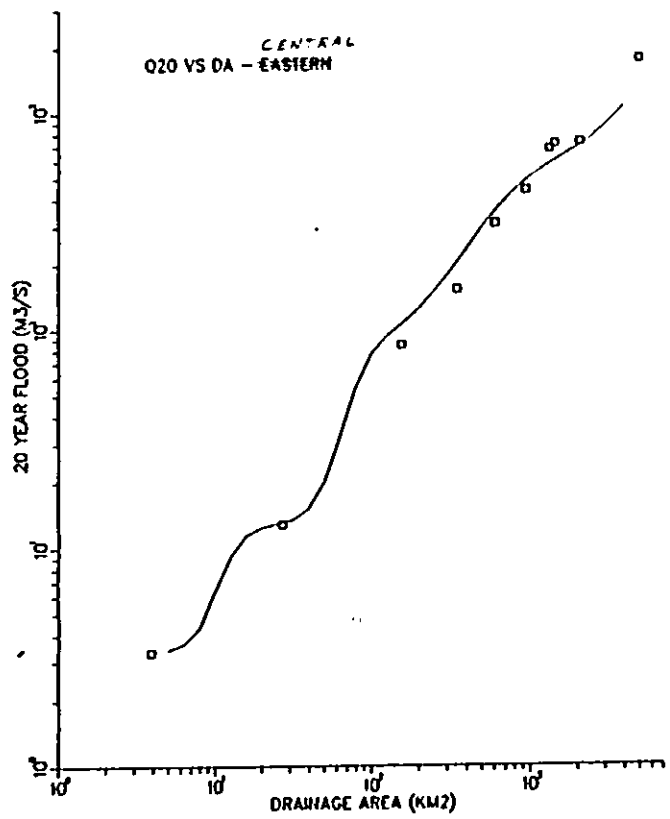

CENTRAL
Q2 VS DA – ~~EASTERN~~
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

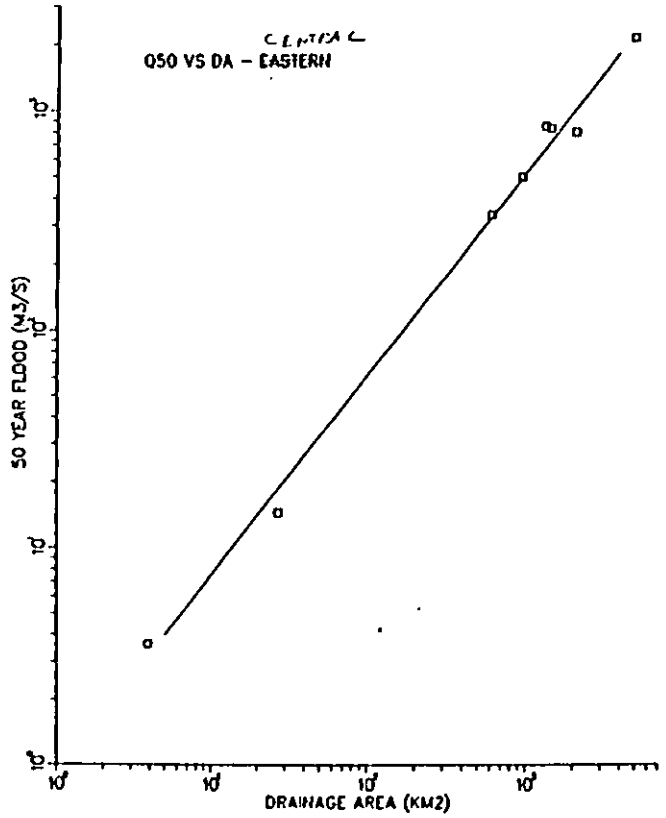

CENTRAL
Q2 VS DA – EASTERN
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


CENTRAL
Q10 VS DA – EASTERN
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


CENTRAL
Q10 VS DA – EASTERN
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


CENTRAL
Q20 VS DA – EASTERN
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


CENTRAL
Q20 VS DA – EASTERN
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)



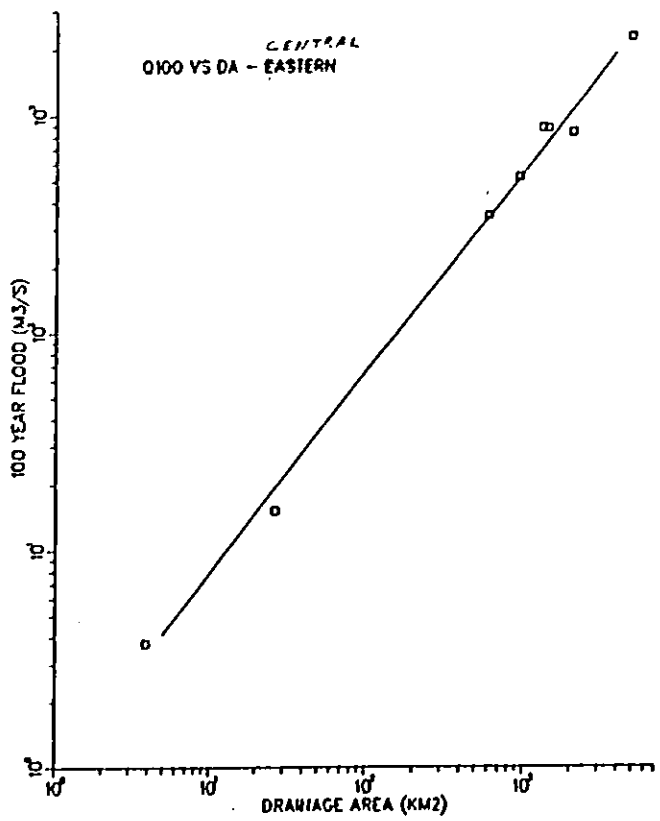
50 year flood

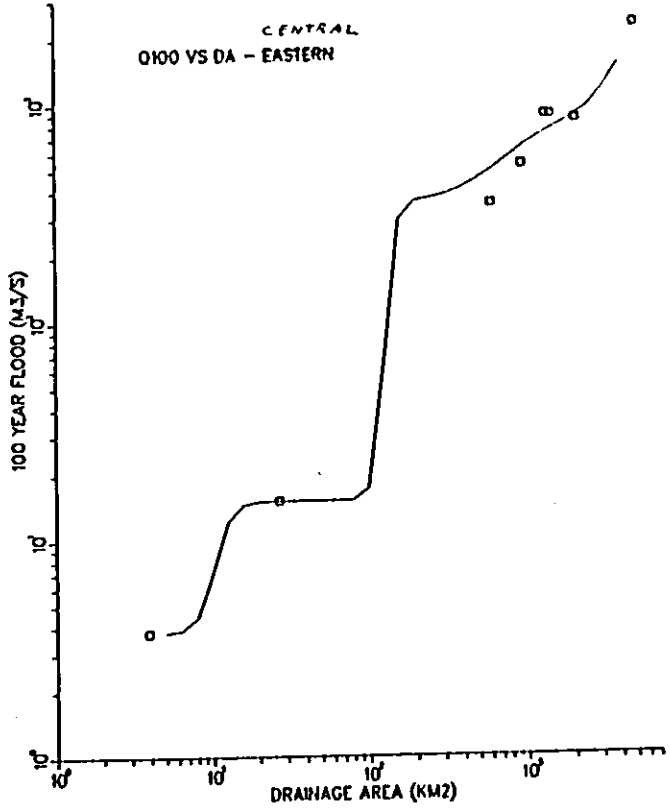
for Central region

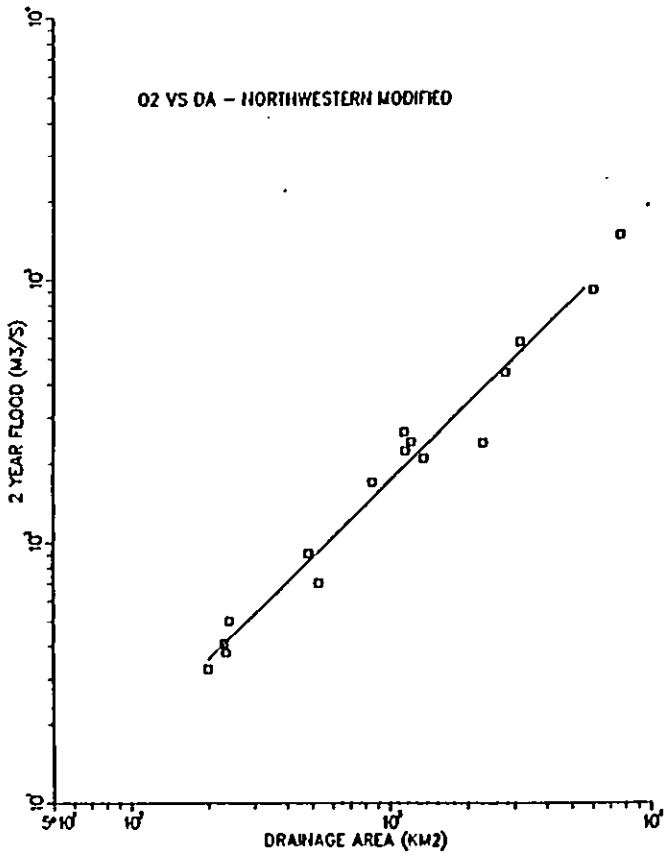
did not converge

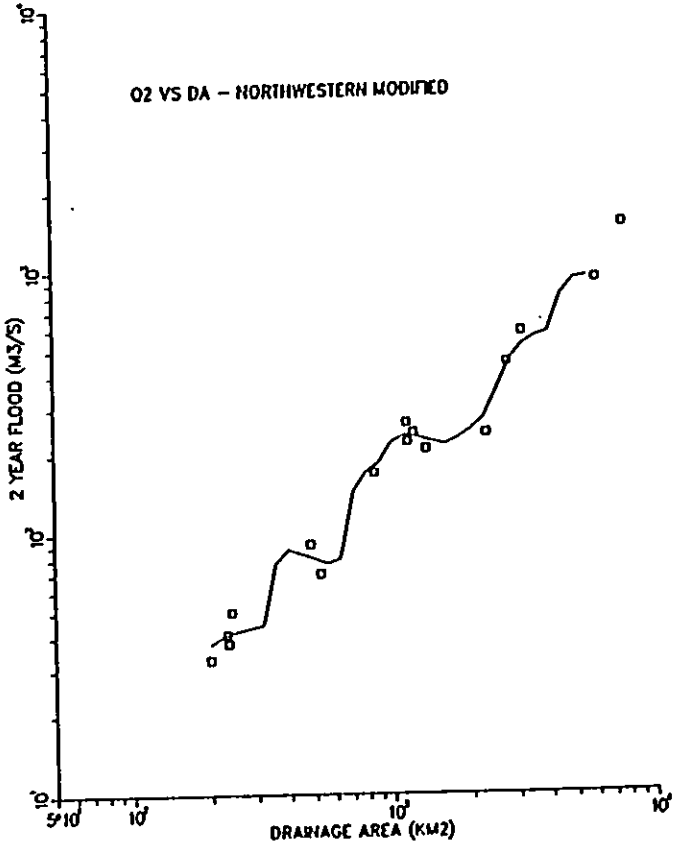
for nonparametric

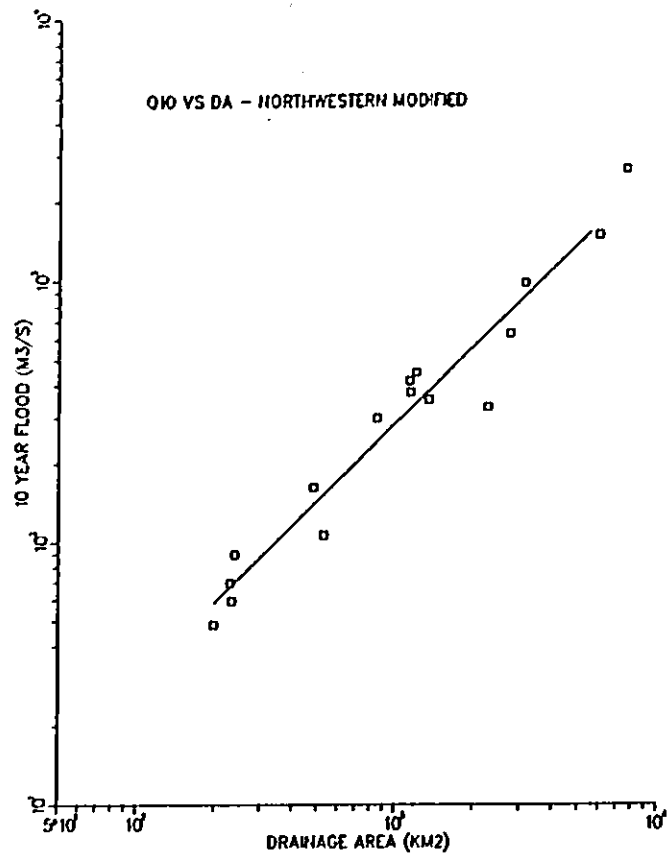
regression

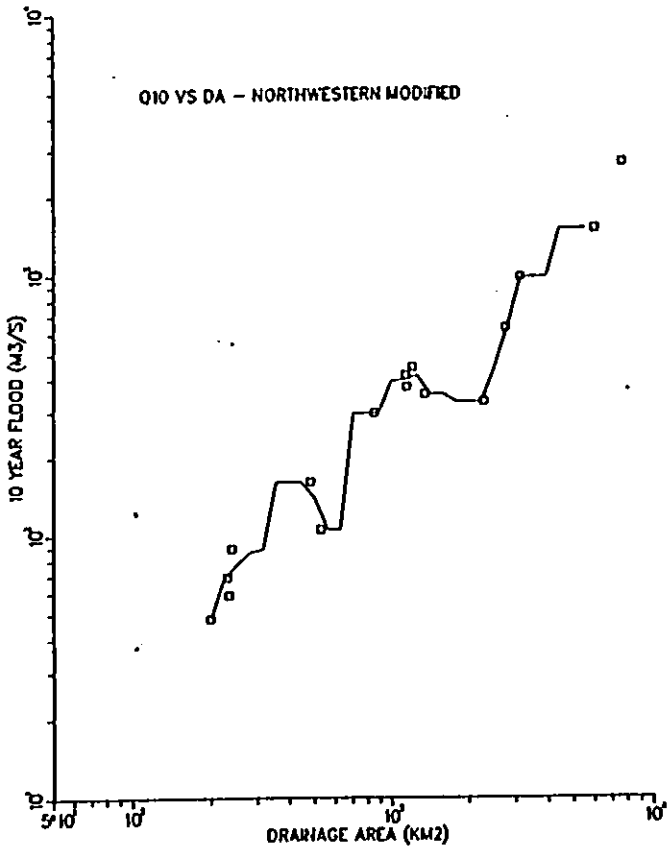

CENTRAL
Q100 VS DA - EASTERN
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

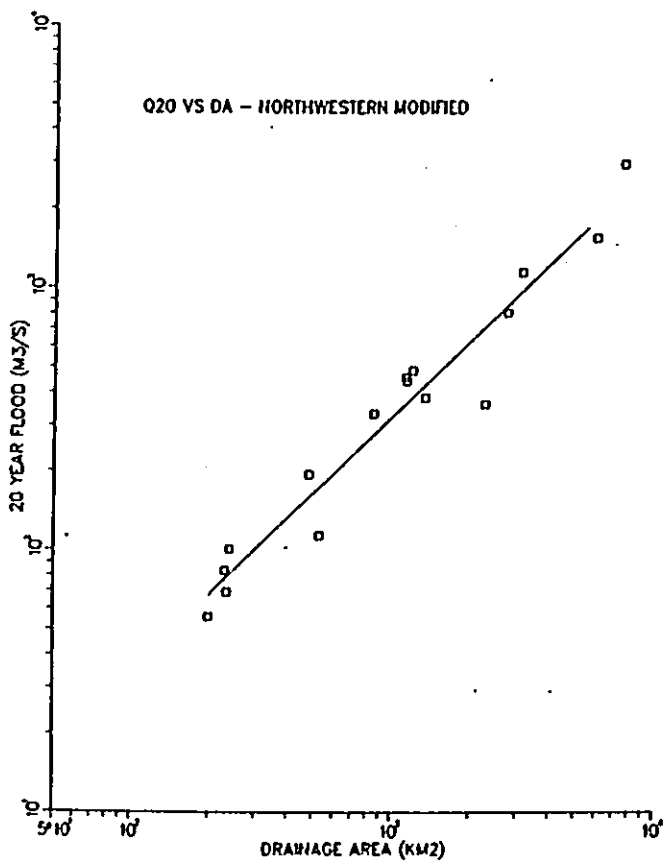

CENTRAL
Q100 VS DA - EASTERN
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

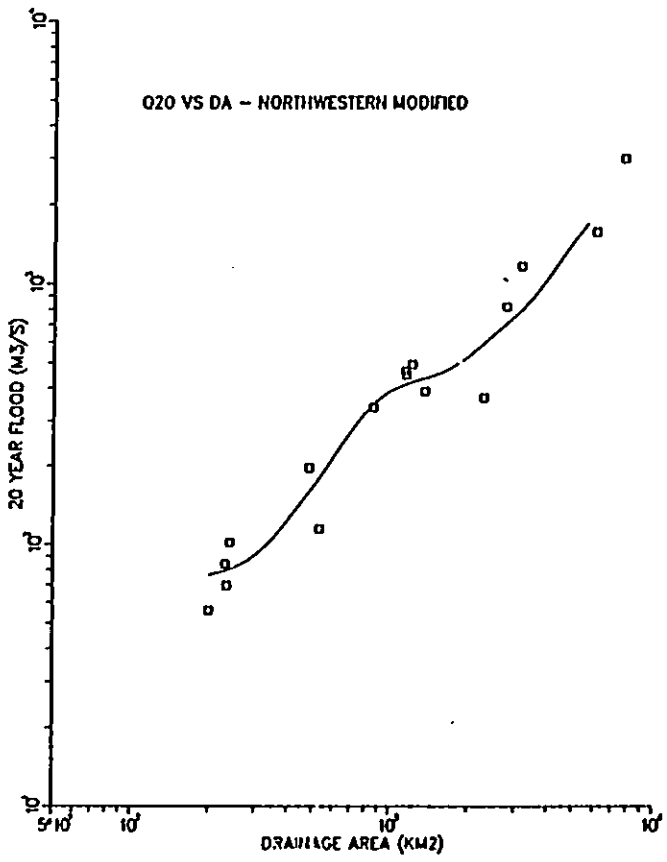

Q2 VS DA – NORTHWESTERN MODIFIED
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

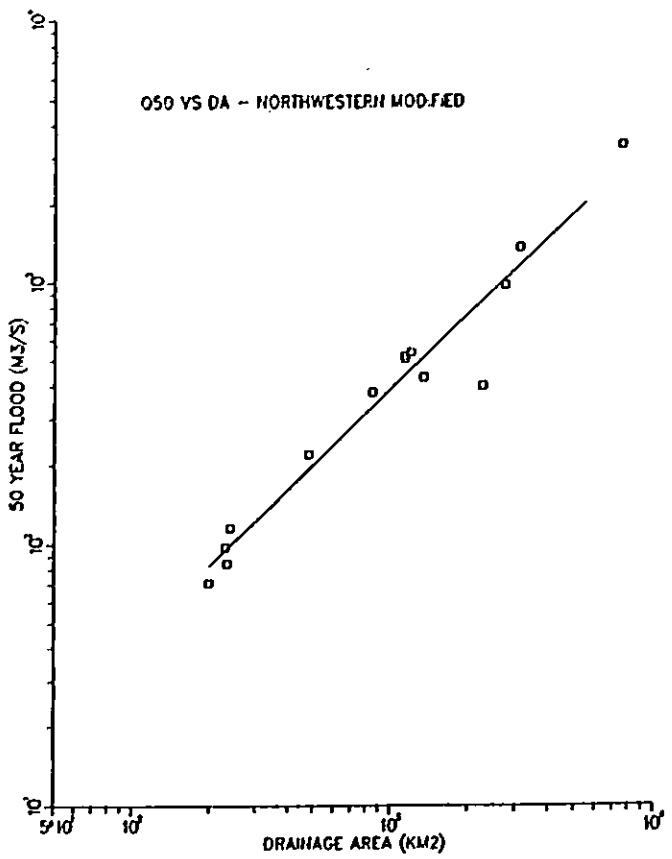

Q2 VS DA – NORTHWESTERN MODIFIED
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

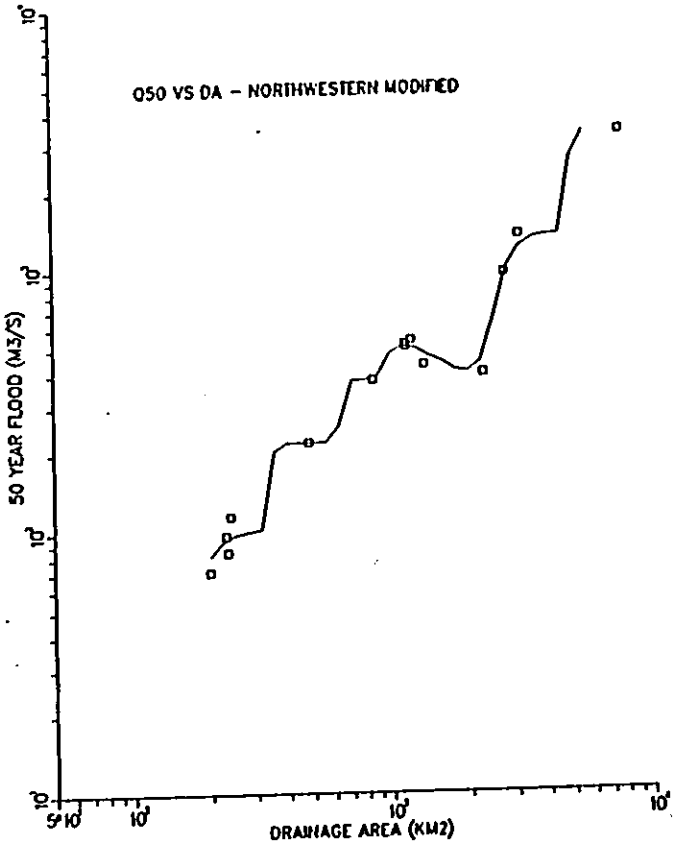

Q10 VS DA – NORTHWESTERN MODIFIED
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

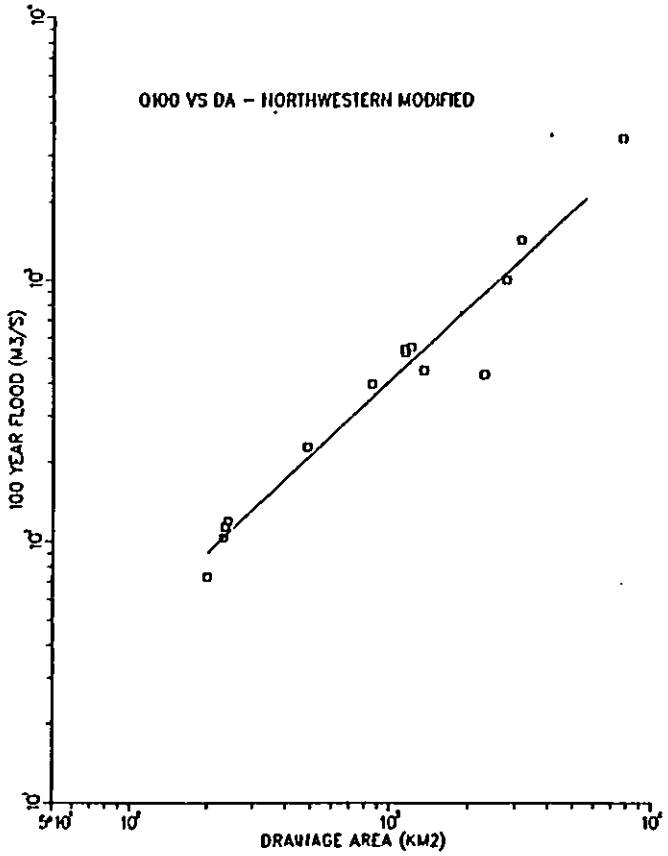

Q10 VS DA – NORTHWESTERN MODIFIED
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

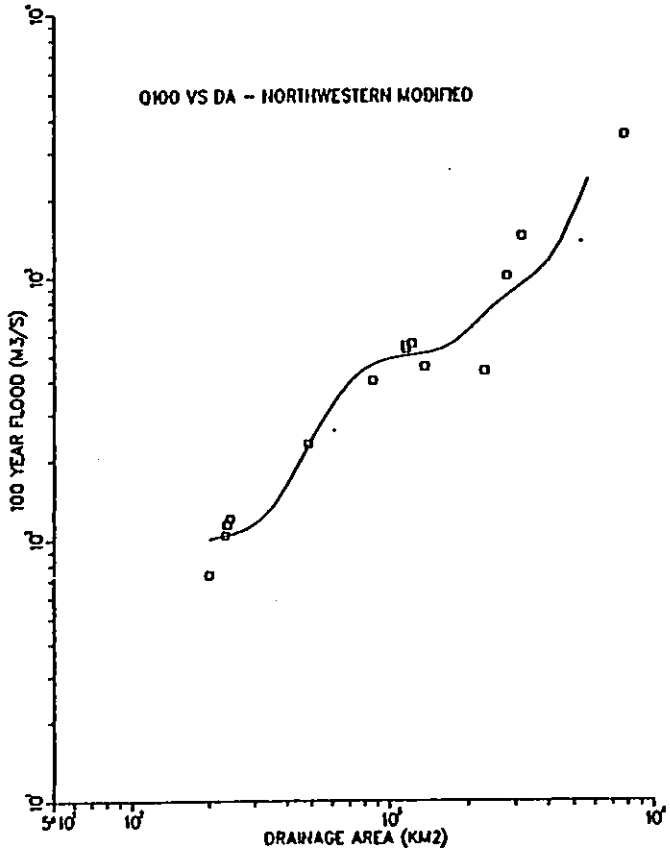

Q20 VS DA – NORTHWESTERN MODIFIED
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

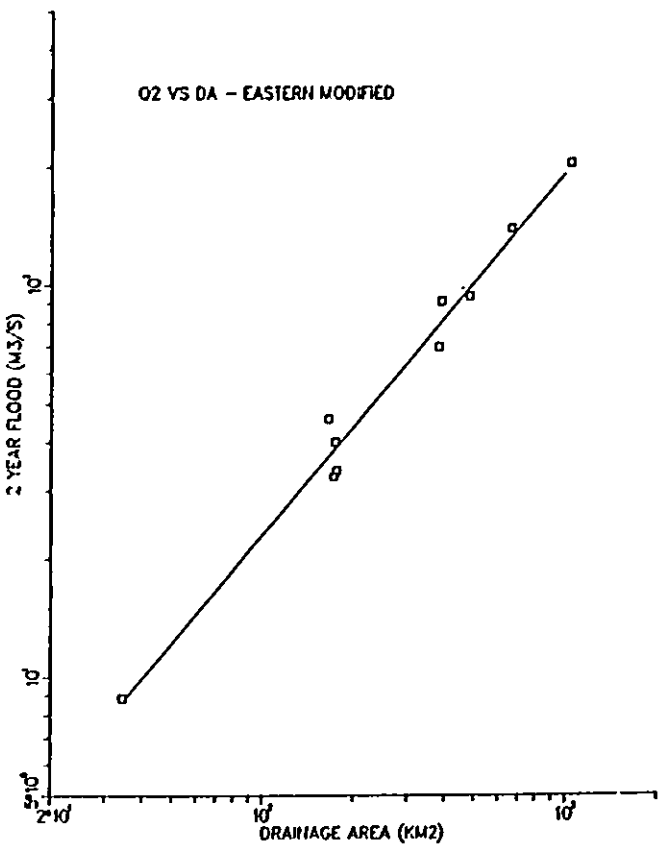

Q20 VS DA – NORTHWESTERN MODIFIED
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

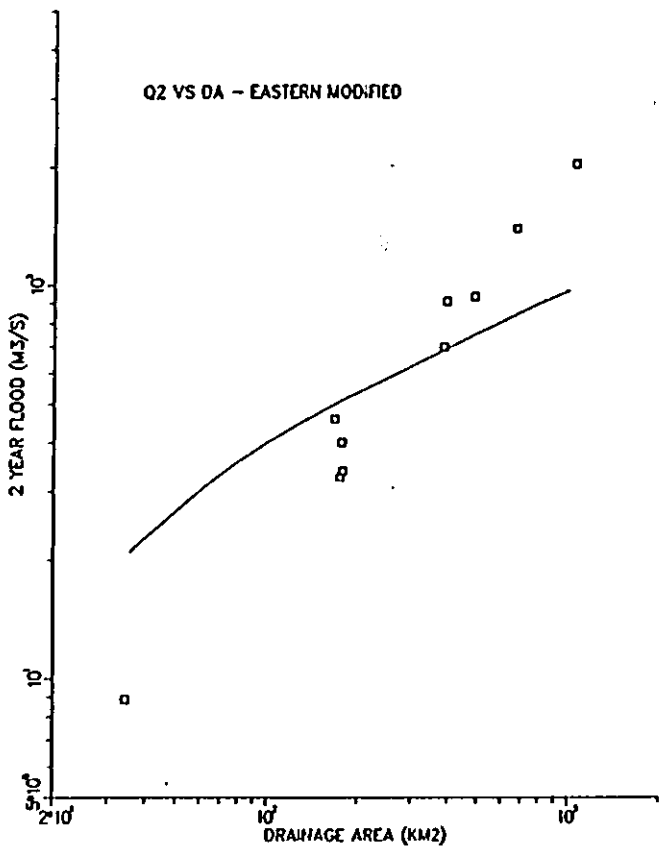

Q50 VS DA – NORTHWESTERN MODIFIED
50 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

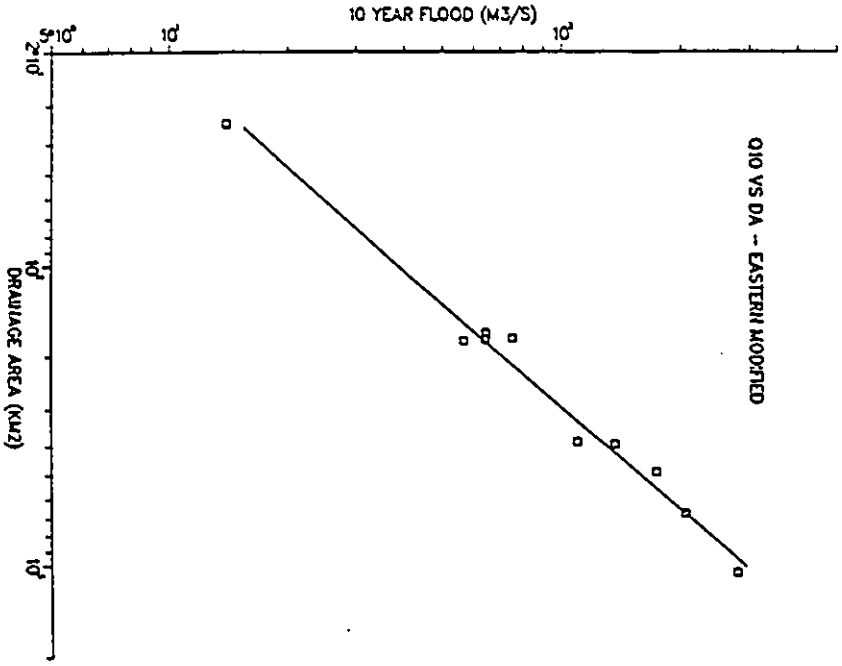

Q50 VS DA – NORTHWESTERN MODIFIED
50 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

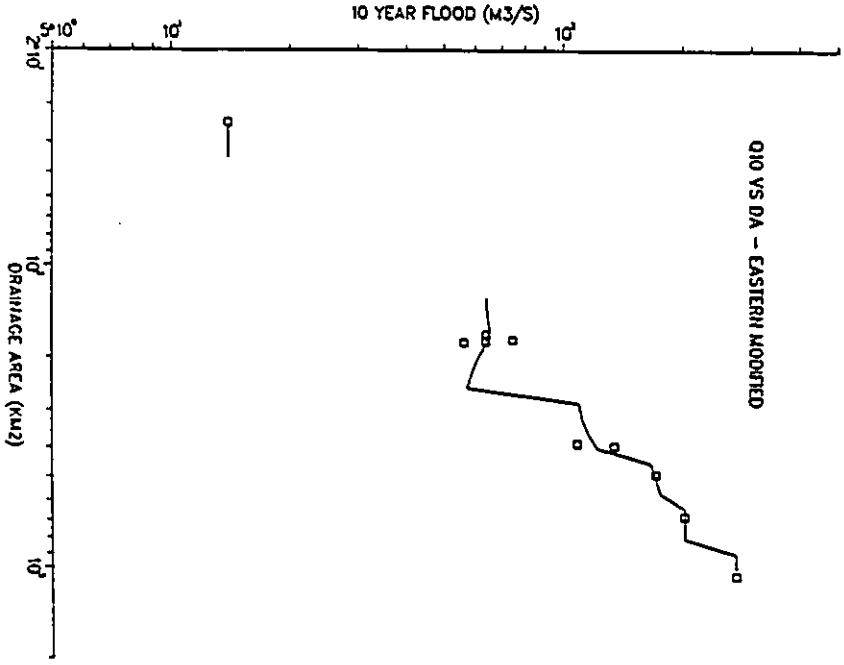

Q100 VS DA - NORTHWESTERN MODIFIED
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

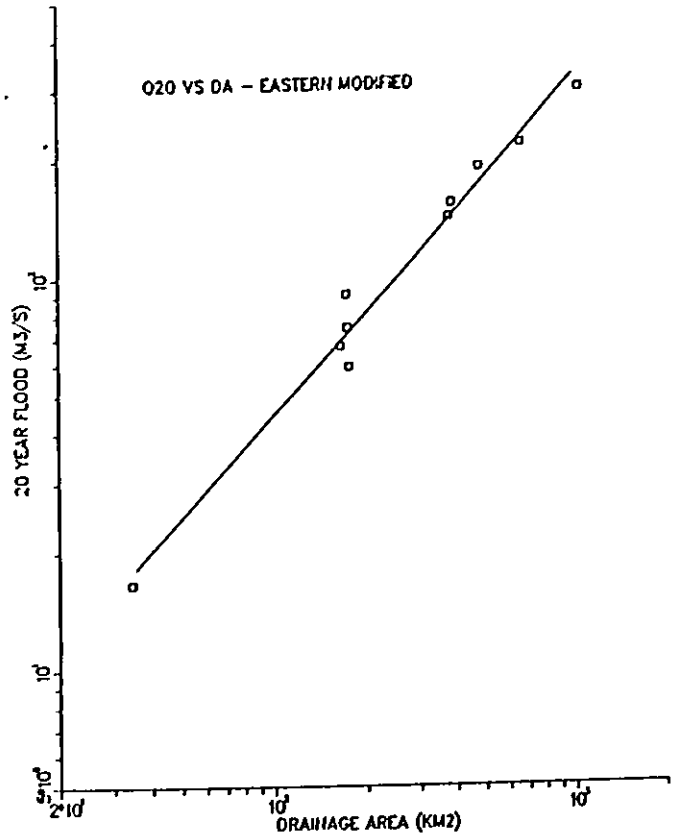

Q100 VS DA - NORTHWESTERN MODIFIED
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

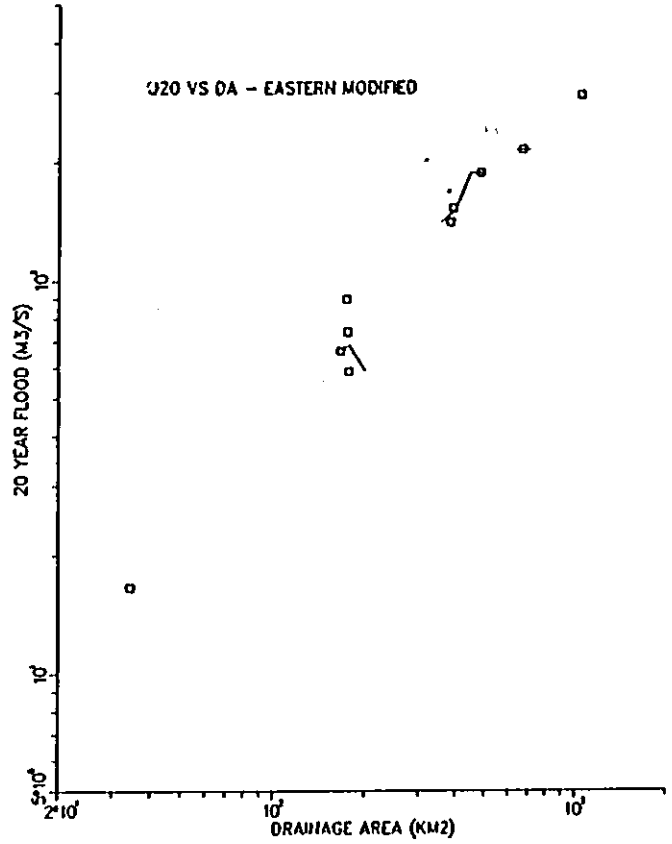

Q2 VS DA – EASTERN MODIFIED
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

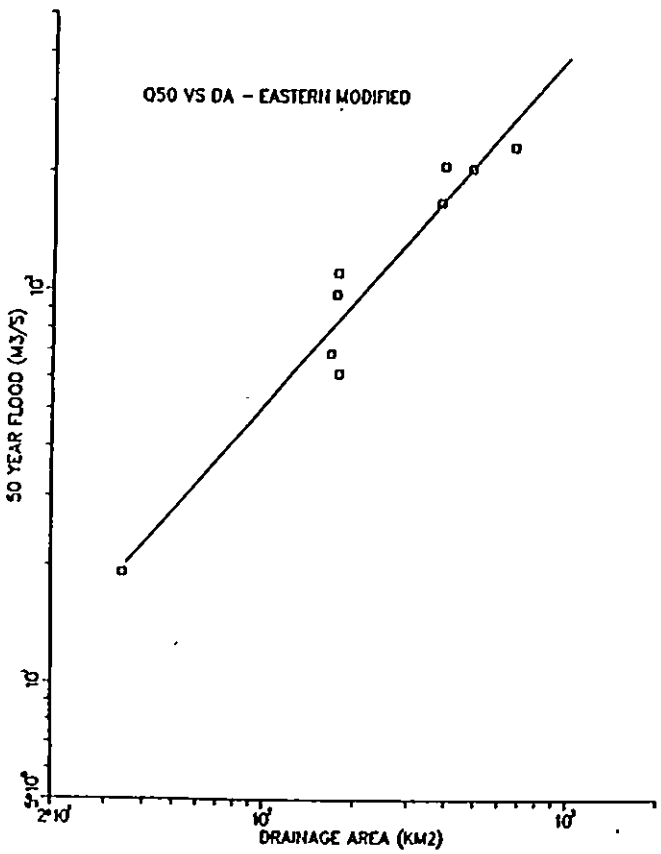

Q2 VS DA – EASTERN MODIFIED
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

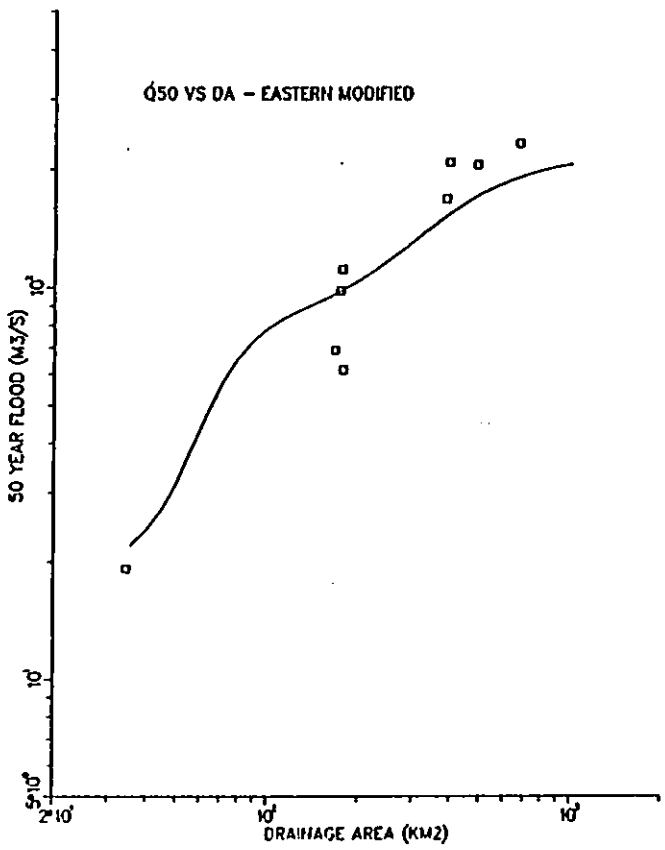

10 YEAR FLOOD (M3/S)
10⁰
10¹
10²
2x10²
10¹
10⁰
DRAINAGE AREA (KM2)
Q10 VS DA - EASTERN MODIFIED

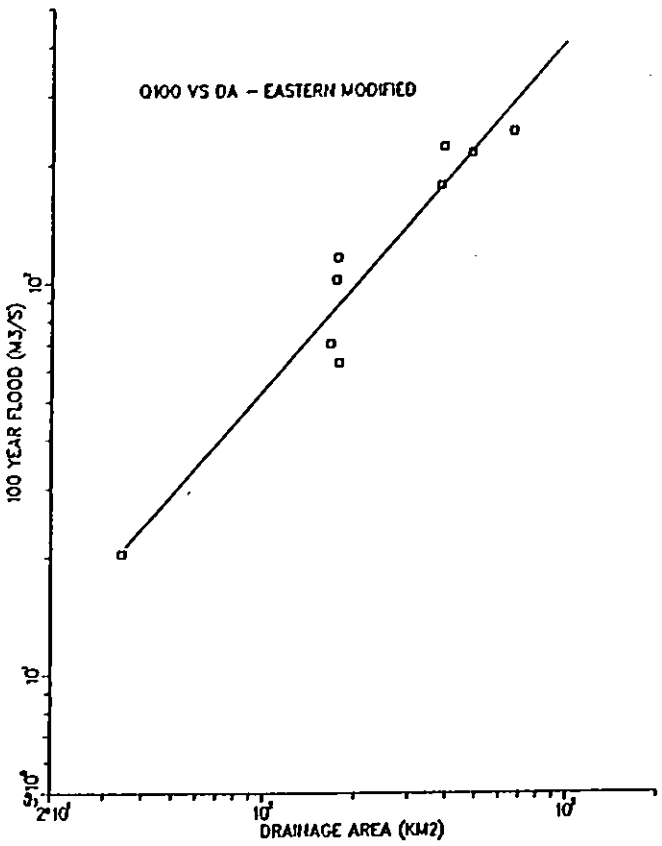

10 YEAR FLOOD (M3/S)
10⁰
10¹
10²
2x10²
10¹
10⁰
DRAINAGE AREA (KM2)
Q10 VS DA - EASTERN MODIFIED

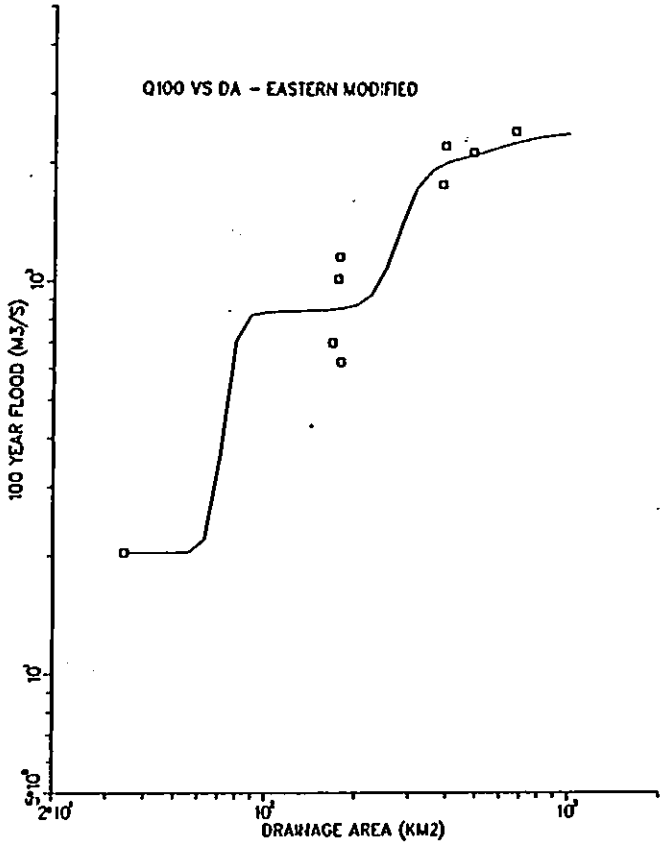

Q20 VS DA – EASTERN MODIFIED
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

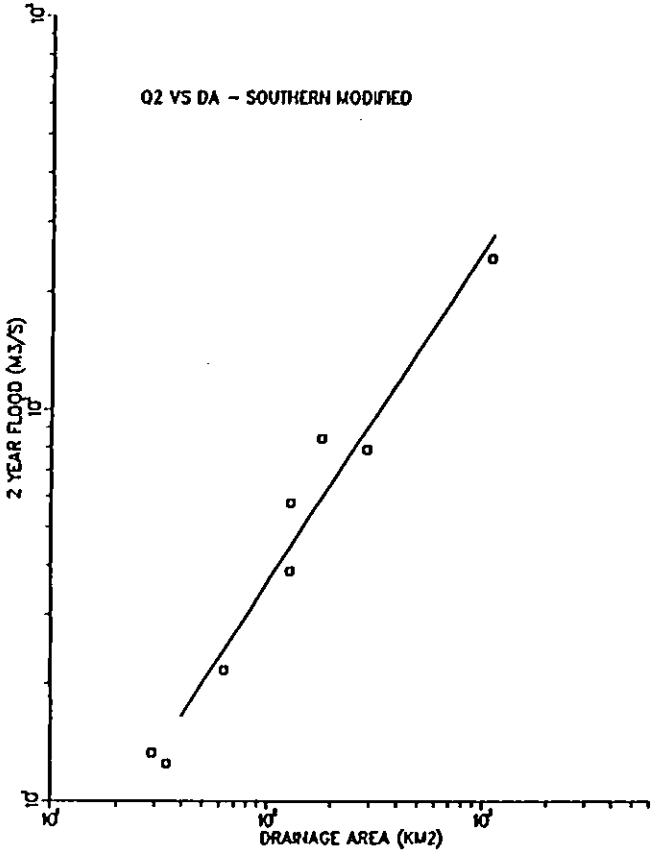

Q20 VS DA – EASTERN MODIFIED
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

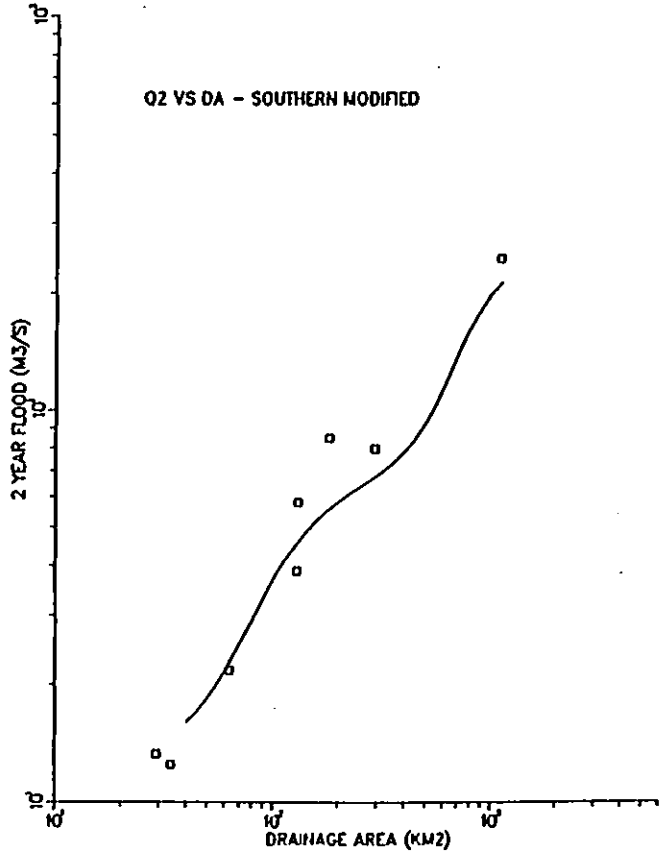

Q50 VS DA – EASTERN MODIFIED
50 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

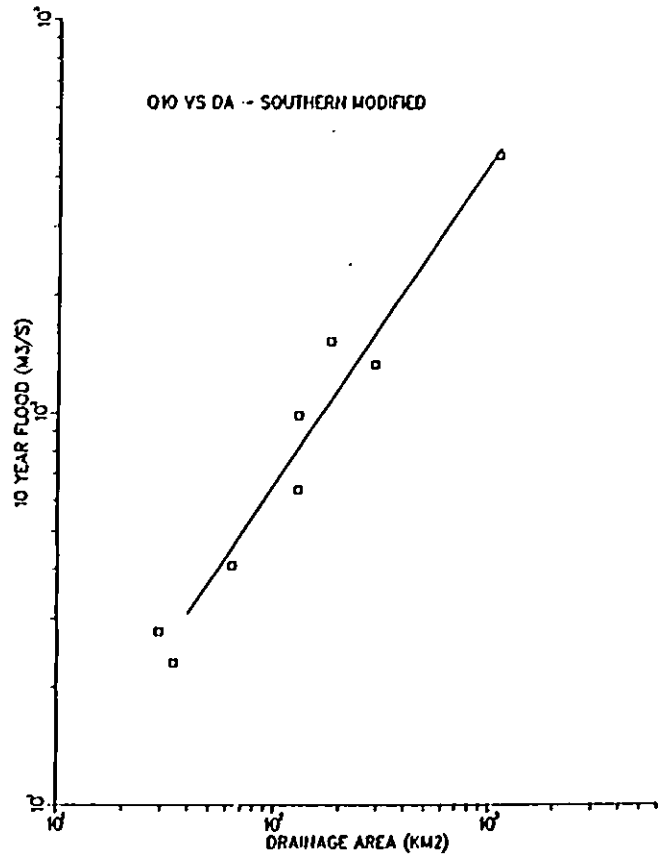

Q50 VS DA – EASTERN MODIFIED
50 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

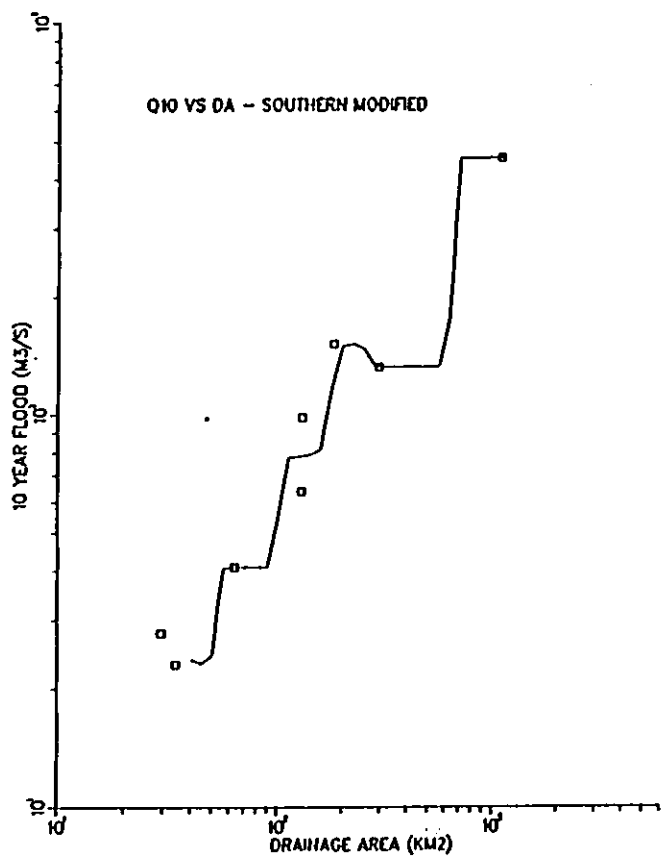

Q100 VS DA – EASTERN MODIFIED
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

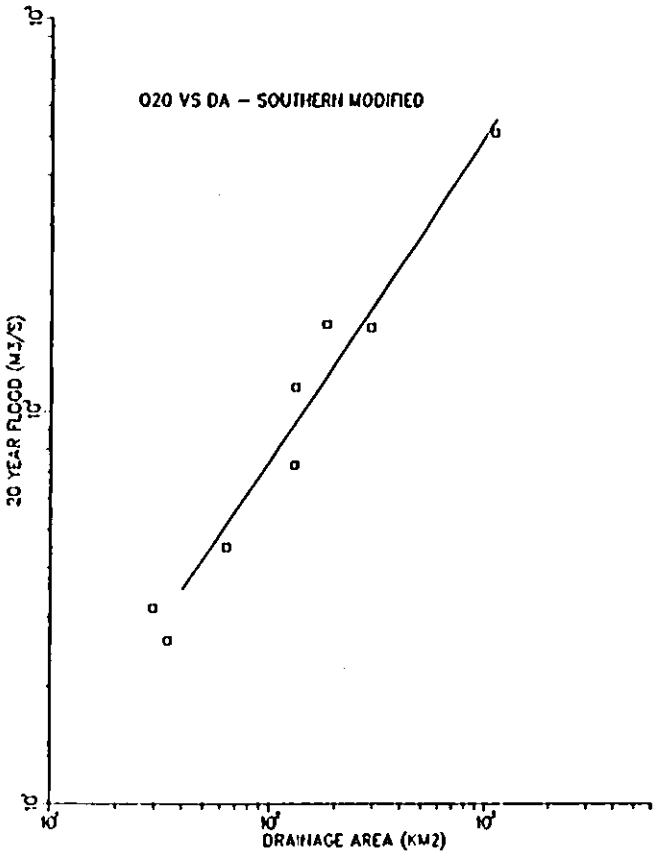

Q100 VS DA – EASTERN MODIFIED
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

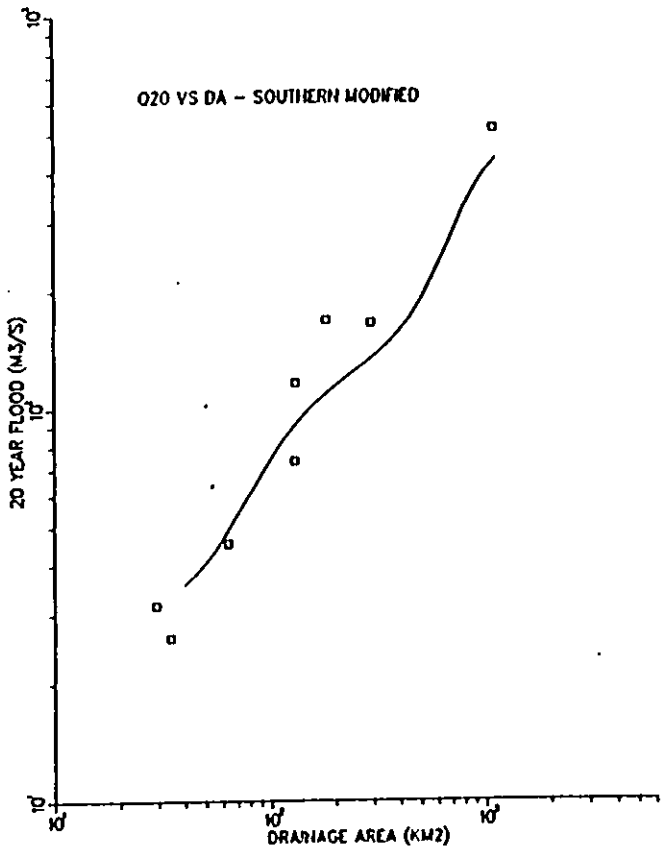

Q2 VS DA – SOUTHERN MODIFIED
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

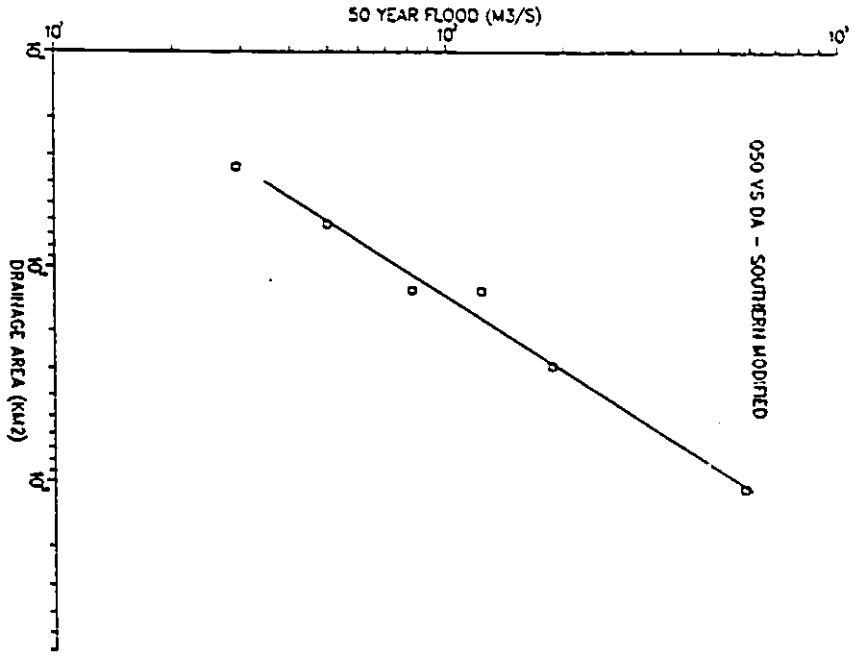

Q2 VS DA – SOUTHERN MODIFIED
2 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

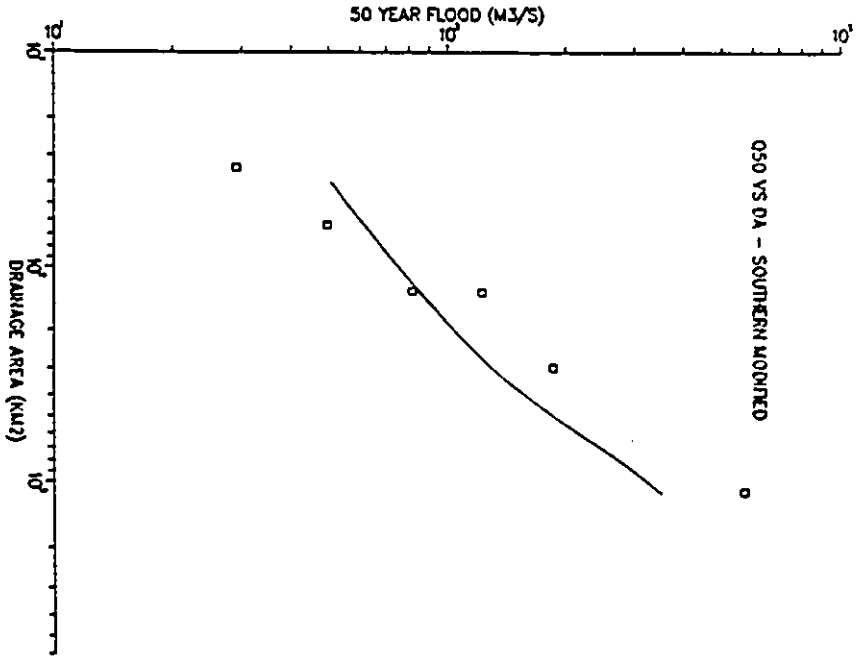

Q10 VS DA -- SOUTHERN MODIFIED
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


Q10 VS DA -- SOUTHERN MODIFIED
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


Q20 VS DA – SOUTHERN MODIFIED
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


Q20 VS DA – SOUTHERN MODIFIED
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


50 YEAR FLOOD (M3/S)
10¹
10²
10³
10⁰
10¹
10²
DRAINAGE AREA (KM2)
Q50 VS DA - SOUTHERN MODIFIED

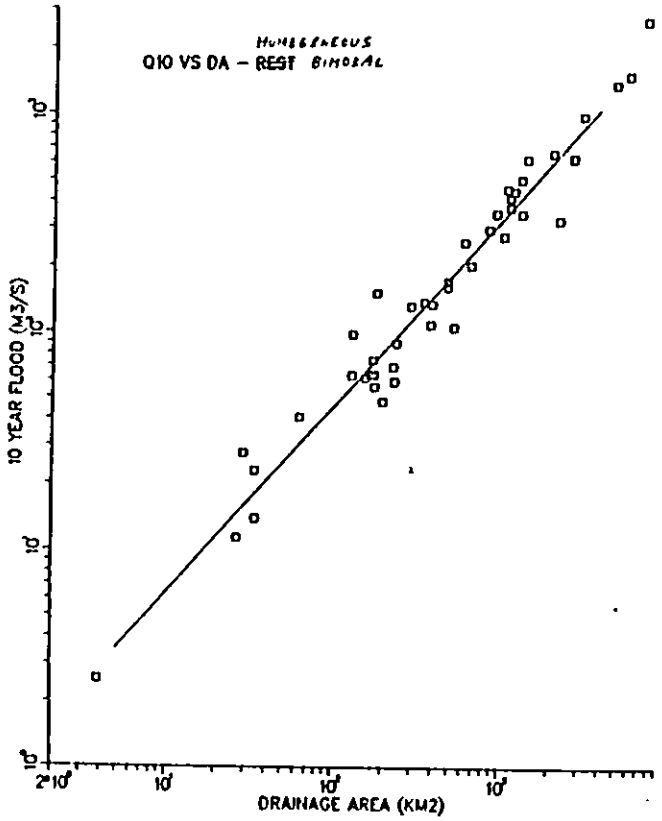

50 YEAR FLOOD (M3/S)
10¹
10²
10³
10⁰
10¹
10²
DRAINAGE AREA (KM2)
Q50 VS DA - SOUTHERN MODIFIED

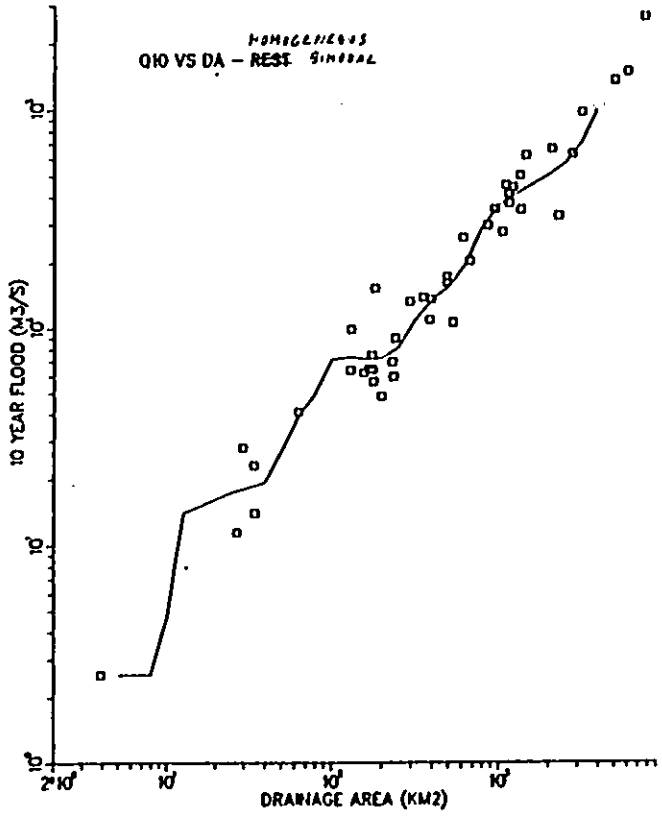
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)
Q100 VS DA - SOUTHERN MODIFIED

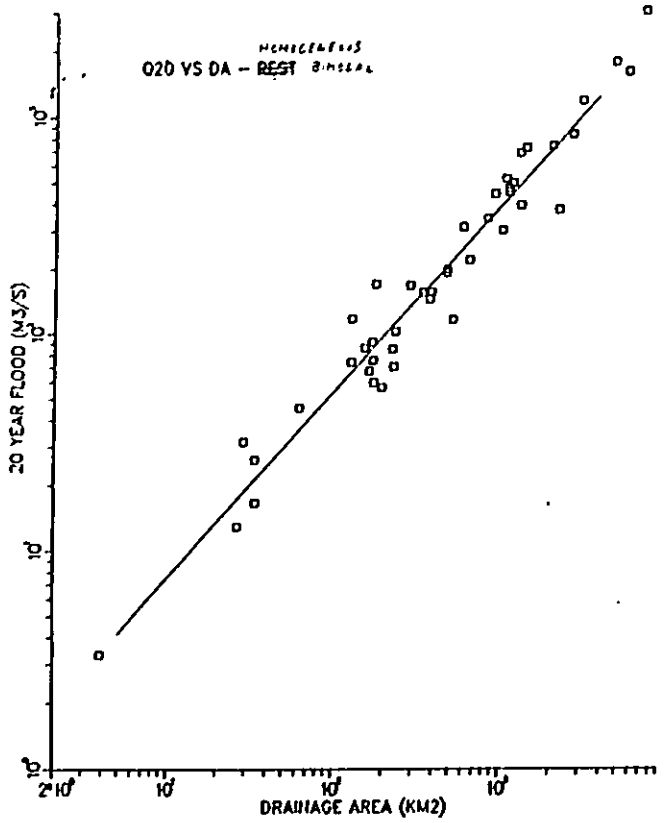
100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)
Q100 VS DA - SOUTHERN UNMODIFIED

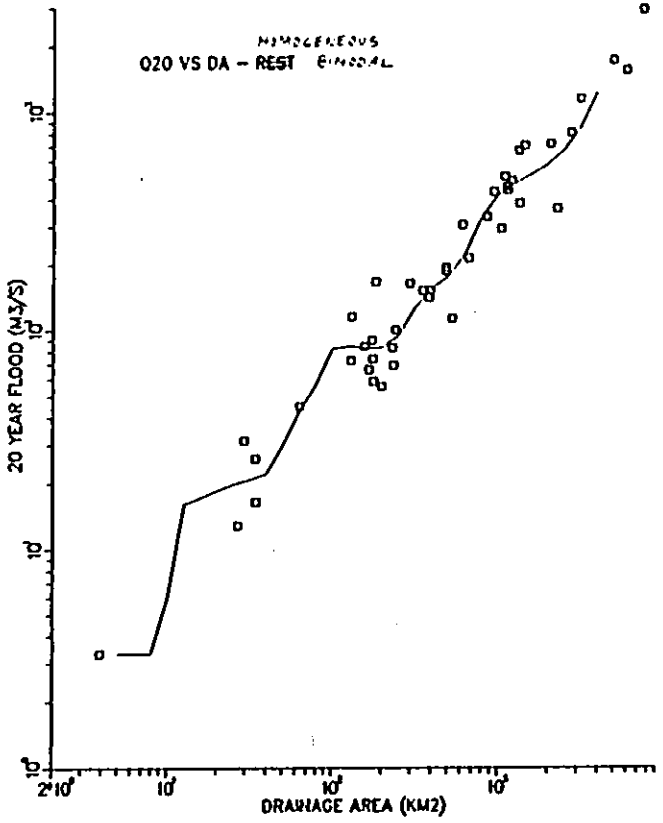
Q2 VS DA – REST of INDIA

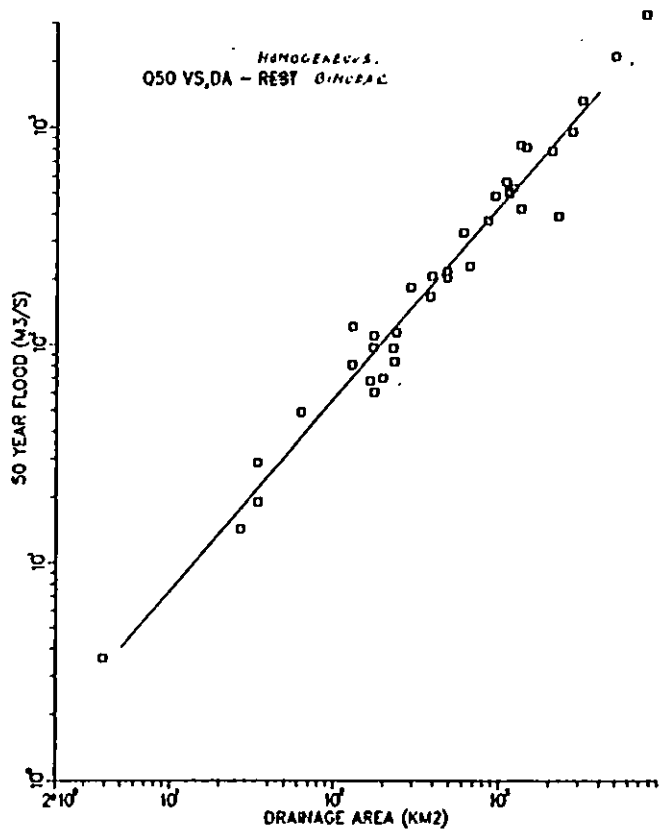
Q2 VS DA – WEST of INDIA

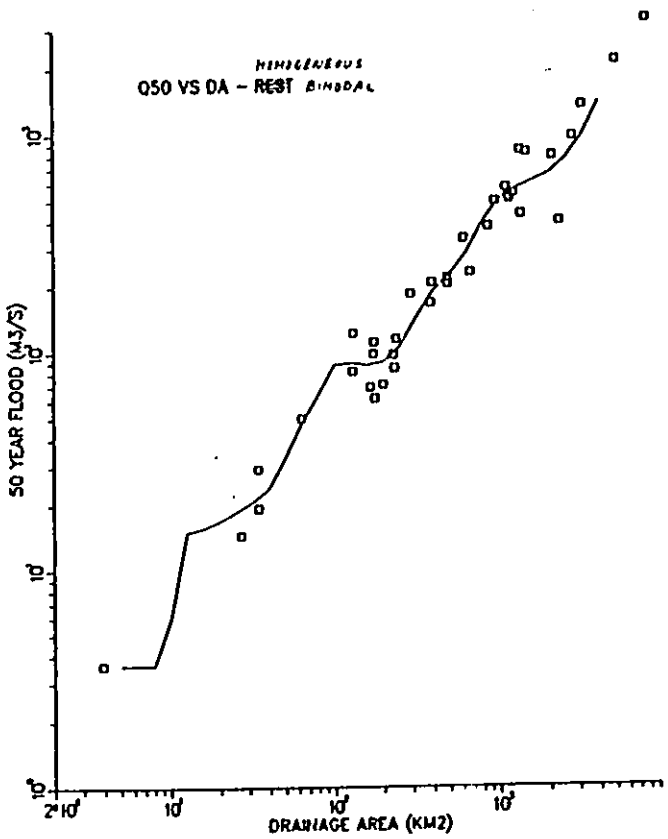

HOMOGENEOUS
Q10 VS DA – REST BIMODAL
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


HOMOGENEOUS
Q10 VS DA – REST SINGLE
10 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


HOMOGENEOUS
Q20 VS DA – REST BIMODAL
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


HOMOGENEOUS
Q20 VS DA – REST BIMODAL
20 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


HOMOGENEOUS
Q50 VS DA – REST BINODAL
50 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)


HOMOGENEOUS
Q50 VS DA – REST BINODAL
50 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)

100 YEAR FLOOD (M3/S)
DRAINAGE AREA (KM2)