MODELLING FOR FLOOD FLOWS

A THESIS

submitted in fulfilment of the
requirements for the award of the degree
of
DOCTOR OF PHILOSOPHY
in
HYDROLOGY

By
NARENDRA KUMAR GOEL

DEPARTMENT OF HYDROLOGY
UNIVERSITY OF ROORKEE
ROORKEE-247 667 (INDIA)

September, 1990

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled 'MODELLING FOR FLOOD FLOWS' in fulfilment of the requirement for the award of the Degree of Doctor of Philosophy submitted in the Department Of Hydrology of the University is an authentic record of my own work carried out during a period from March 1986 to September, 1990 under the supervision of Dr. Satish Chandra and Dr. S.M. Seth.

The matter embodied in this thesis has not been submitted by me for the award of any other degree.

N.K. Goel

(N.K. Goel)

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

(S.M. Seth)
Scientist 'F'
National Institute of Hydrology
Roorkee - 247667, INDIA

(Satish Chandra)
Director
National Institute of Hydrology
Roorkee - 247667, INDIA

Roorkee, Dated Sept. 18, 1990.

The Ph.D. Viva Voce Examination of Shri N.K. Goel, Research Scholar has been held on 16.th April, 1991

(S.M. Seth) (Satish Chandra) (External examiner)

TO MY TEACHERS

ABSTRACT

Flood frequency analysis plays an important role in the estimation of design flood for various hydraulic structures. In the past years, attempts have been mainly concentrated for the analysis of flood peak only which provides a very limited assessment of a flood event as risk is formulated in terms of flood peaks only. Hydrological phenomena like flood flows always appear as multivariate events and characterized by various components such as volume and duration etc. besides flood peak. In many aspects of water resources planning and management the information about volume and duration etc. besides flood peak of critical flood events is essential. This in turn requires the probability of whole event rather than the probability of only peak discharge. In spite of their importance very limited research has been conducted in the past for the study of flood duration and volume characteristics of flood flows and for the study of flood event as a whole.

The objective of the preset work is to study various flood characteristics such as flood peak, volume, duration and other characteristics, and their interdependence for a typical flood series, development and validation of the methodology for bivariate stochastic modelling of flood flows using synthetically generated data and application to real data. It also includes the study to ascertain the suitability of daily flow generation models for Indian rivers (where the flow is concentrated in five months of monsoon season) and to use the selected model for generation of long term data for validation of the methodology.

The noted developments in the area of flood frequency analysis using annual flood series (AFS) and partial duration series (PDS), modelling of flood features other than peak discharge and synthetic generation of daily flows have been reviewed.

The methodology for stochastic modelling of flood flows has been developed using partial duration series approach with bivariate normal distribution function as the parent distribution function for two dependent variables such as flood volume and flood peak. The application of bivariate normal distribution assumes normal marginal distributions which is usually not the case with flood peaks, volumes and durations. The normalization of the variables required for applying bivariate approach, has been done using two step power transformation (TSPT). The TSPT makes the C_s of a given series as 0 in the first step and C_k equal to 3 in second step. Using the bivariate normal density function the marginal, conditional and joint distribution functions have been derived, and used to estimate the return periods of following events.

(i) exceedance of at least one of the values (x,y) in a year,

(ii) exceedance of both x and y in a year,

(iii) exceedance of $y|x$ in a year,

(iv) exceedance of x (univariate) in a year and

(v) exceedance of y (univariate) in a year.

For the above events x is the value of flood volume (X) and y is value of flood peak (Y).

The methodology for stochastic modelling of flood flows developed in this study has got the flexibility of selecting (i) any threshold level which can be fixed from engineering consideration as long as the Poissonian assumption for number of flood events in a year is satisfied, (ii) any volume of a flood event which can also be decided on the basis of probable maximum precipitation and by adopting a suitable loss and (iii) any flood peak magnitude.

The daily discharge data of river Narmada at Garudeshwar site (catchment area 88000 km^2) from 1949-79 have been used. The preliminary analysis of data includes conventional flood frequency analysis of annual flood series and analysis of (i) number of flood events in a year, (ii) flood peaks, volumes and durations, (iii) largest flood peaks, volumes and durations and (iv) timing of largest flood peaks, for different threshold levels with emphasis on the study of their distribution functions.

Based on literature review the two schemes namely, linear autoregressive and shot noise models have been used for generation of daily flows. The purpose of this generation is (i) to ascertain the suitability of these models for generation of flows in the situation when major portion of the flow is occurring during five months of monsoon season and (ii) to validate the methodology developed for bivariate stochastic modelling of flood flows. The computer programmes for these models have been developed after incorporating suitable modifications in the schemes and their performance inter-compared. The inter-comparison of two approaches indicates that (a) the linear autoregressive model gives better reproduction of overall statistical parameters of daily flows and flood peaks above a particular threshold level and (b) shot noise model (modified) gives overall better reproduction of statistical parameters of remaining flood characteristics. As such modified shot noise model was selected for further generation of long term (1500 years) daily flows for Narmada at Garudeshwar for validation of the methodology.

After achieving satisfactory validation, the methodology was applied to daily discharge data of Narmada at Garudeshwar. The capabilities of the methodology for dealing with univariate as well as bivariate modelling of flood characteristics has been clearly established as also illustrated with typical examples.

A number of SUBROUTINES were developed in FORTRAN IV language and used for carrying out data processing, preliminary analysis of various flood characteristics, daily flow models, and validation and application of the methodology. The important subroutines have been given in the thesis for general use. Though this study has focussed on flood flows, the developed methodology can be extended to other dependent hydrological variables, such as drought related characteristics, sediment yield and runoff and many others, also. It is hoped that this study would contribute in the areas of flood frequency analysis and daily flow generation.

ACKNOWLEDGMENTS

It is my proud privilege to express my sincere gratitude to Dr.Satish Chandra, Director, National Institute of Hydrology, Roorkee, who suggested the field for the present study and rendered guidance and all possible help in carrying out the study.

Dr.S.M.Seth, Scientist 'F', NIH, has been my main source of guidance. His capacity to handle variety of jobs, untiring efforts and the patience to listen and suggest to improve the work, are gratefully acknowledged.

I also take this opportunity to thank Dr.B.S.Mathur, who always wished the early completion of this thesis and provided all possible help, working environment and suggestions to attain this.

Thanks are due to Dr.Subhash Chander, Professor, IIT Delhi, for the encouragement and useful suggestions for the work.

The author is thankful to Dr.V.P.Singh, Professor, Louisiana State University, USA, Dr.C.Cunnane, University College Galway, Ireland and Dr.J.D.Salas, Professor, Colorado State University, USA, for useful discussions.

I also wish to express sincerest thanks to my colleagues Dr.D.K.Srivastava, Dr.Ranvir Singh, Dr.Deepak Kashyap, Dr.D.C.Singhal and Mr.Himanshu Joshi for the useful discussions and for making real good working environment in the Department.

Sincere appreciation is extended to Mr.Ajay Singhal, Lecturer, Civil Engineering Department, U.O.R. and Mr.Sanjay Mohan Mittal for providing help in word processing and data processing, because of which the thesis is in present shape.

The author is indebted to his friends, research scholars of Department of Hydrology, other staff members of the Department and many other persons who have contributed in the thesis in one way or the other.

But foremost, to his wife, Amita, daughters, Shilpi and Shivangi, for their sacrifices throughout the study, to his parents and other family members, for their encouragement, goes the author's deepest gratitude.

N.K Goel

(N.K. Goel)

MODELLING FOR FLOOD FLOWS

CONTENTS

List of Tables

List of Figures

List of Symbols

Symbols having a common meaning throughout are defined here. Other locally used symbols are defined wherever they occur.

AFS : Annual flood series

C_k : Coefficient of kurtosis

C_s : Coefficient of skewness

C_v : Coefficient of variation

CDF : Cumulative distribution function

EV1 : Extreme value type 1 distribution

GEV : General extreme value distribution

LAR : Linear autoregressive model

LLG : Log logistic distribution

LN : Log normal distribution

LPIII : Log Pearson type 3 distribution

ML : Maximum likelihood

MOM : Method of moments

MSNM : Modified shot noise model

PDF : Probability density function

PDS : Partial duration series

POT : Peak over threshold model

PTIII : Pearson type 3 distribution

PWM : Probability weighted moments

POME : Principle of maximum entropy

r_1 : Lag one serial correlation coefficient

S.D. : Standard deviation

SNM : Shot noise model

T : Return period

TCEV : Two component extreme value distribution

TSPT : Two step power transformation

Wak-4 : Wakeby 4 parameter distribution

Wak-5 : Wakeby 5 parameter distribution

CHAPTER I

INTRODUCTION

1.1 General

Since time immemorial floods have been causing loss of human lives, valuable property and crops in most parts of the world and India is no exception to this. The flood damage can be reduced significantly by proper management through various structural measures such as dams, levees, embankments etc., by taking various non-structural measures and by proper design of various hydraulic structures. The estimation of design flood and flood characteristics is thus crucial for flood management and control.

The current methods of estimating design flood for hydraulic structures mainly include the probable maximum flood (PMF) approach and flood frequency analysis approach. The PMF approach suffer from the major disadvantage of being entirely subjective and of having no probability level, Kite (1977). This sometimes leads to disastrous consequences since non technical people tend to believe that maximum flood cannot exceed this certain limit of PMF. As such the use of flood frequency analysis approach plays an important role in the design flood estimation for various structures.

In the past two to three decades, attempts in the flood frequency analysis area, have been mainly concentrated on fitting one probability distribution or the other to the annual flood series and carrying out flood frequency analysis. This provides a very limited assessment of the flood event as risk is formulated in terms of flood peak magnitude only. Hydrological phenomena like flood flows always appear as multivariate events and characterized by various components such as volume, duration etc. besides flood peak. In many aspects of water resources planning and management the information about the magnitude, duration and volume of the critical flood events is essential. This requires the probability of whole flood event rather than the probability of only peak discharge. In spite of its importance, very limited attention has been paid in the past for the study of flood event as a whole. The present work is a step ahead in this direction.

1.2 Objective and Scope of the Work

The objective of the present work is to (i) study various flood characteristics, such as flood peak, volume and duration, and their interdependence (ii) develop the methodology for multivariate stochastic modelling of flood flows, (iii) validate the methodology using synthetically generated data and (iv) application of the methodology to real data.

The multivariate modelling though increases the possibilities of practical application, yet requires more and more data as number of variables to be considered in the analysis increases. Considering the data limitations and mathematical complexities involved in multivariate modelling the scope of the present work is limited to bivariate modelling only. This involves consideration of only two variables at a time among flood peak, volume and duration by analyzing daily flow data. Besides this, the work would also include study of other characteristics of flood flows such as number of flood events in a year, timing of the flood peaks, flood peaks, volumes and durations, largest flood peaks, volumes and durations in different years and timing of largest flood peaks for different levels of threshold.

For validation of the methodology synthetic daily flows would be generated using appropriate procedure. For this purpose, the scope of work would be limited to comparison of two approaches of daily flow generation, namely time series analysis approach and application of shot noise model after incorporating appropriate modifications for monsoon behavior. The selected approach would be used for generation of long term data for validation of the methodology. The daily flow data of river Narmada at Garudeshwar covering the period from 1949 to 1979 would be used in the study.

1.3 Outline of Contents

Chapter II is devoted to brief review of literature in the area . Considering the scope of the study, the review concentrates mainly on noted developments in the areas of flood frequency analysis using annual flood series (AFS) and partial duration series (PDS), modelling of flood features other than peak discharge and synthetic generation of daily flows.

Chapter III presents the formulation of the methodology along with the basic considerations for defining the flood characteristics and the details of the normalization method required in the application of the methodology.

Chapter IV provides relevant details of the river basin, earlier studies of flood estimation in Narmada basin particularly for Sardar Sarovar dam, the data used ,its processing and preliminary analysis. The preliminary analysis includes conventional flood frequency analysis of AFS and analysis of following flood characteristics (i) number of flood events in a year (ii) timing of flood peaks (iii) flood peaks, volumes and durations (iv) largest flood peaks, volumes and durations in different years and (v) timing of largest flood peaks for different threshold levels.

Chapter V gives details of synthetic generation of daily flows. The comparison of two schemes of daily flow generation, namely linear autoregressive and shot noise models is presented and discussed. The comparison is made on the basis of reproduction of flood related characteristics in the generated data.

The validation of the methodology of multivariate modelling of flood flows is described in Chapter VI. This chapter also gives details of suitability of the methodology for analysis of historical data and results of application to Narmada at Garudeshwar.

Chapter VII presents conclusions drawn from the study and suggestions for further work.

CHAPTER II

REVIEW OF LITERATURE

2.1 General

The flood frequency analysis methods were introduced about eighty years back and since then interests in this area have waxed and waned several times. Numerous papers, covering various aspects of flood frequency analysis, have appeared in literature. A lot of work has been done and is currently being done in this area by many researchers throughout the world. A close look at various papers reveals that most of the work, reported in literature, is regarding the frequency analysis of peak discharges using either annual flood series (AFS) approach or partial duration series (PDS) approach. Very few attempts have been made for the study and analysis of other characteristics of flood such as volume and duration etc. of flood events and their interrelationships.

The survey of literature covering the first part, i.e., frequency analysis of peak discharges has been presented in section 2.2. Section 2.3 covers review of literature regarding analysis of flood characteristics other than the peak discharge. The synthetically generated data plays an important role in simulation studies, in comparing the performance of various methods and for the validation of new techniques proposed for various aspects of frequency analysis. Keeping this in view, a brief review of models developed for synthetic generation of daily flows, is given in section 2.4. The conclusions drawn from the review presented in sections 2.2, 2.3 and 2.4 have been summarized in section 2.5.

2.2 Frequency Analysis of Peak Discharges

In flood frequency analysis, in general, the sample data is used to fit the probability distribution. The distribution, in turn, is used to extrapolate from recorded events to design events either graphically or analytically after estimating its parameters. There are essentially two approaches of carrying out flood frequency analysis. One corresponds to the streamflow annual flood series (AFS), in which the largest annual peak discharges of each year of record are considered, and the other approach corresponds to streamflow partial duration series (PDS), in which all the discharges above a particular threshold are considered. In both the approaches, the research has mainly concentrated on fitting one probability distribution or the other to the given series and its parameter estimation procedures. Some studies have also been made involving normalization of the given series irrespective of the parent distribution of the series and using the normalized series for flood frequency analysis.

The other fields of flood frequency analysis such as regional flood estimation and homogeneity, use of historical data in flood frequency analysis, use of geomorphological unit hydrograph (GUH) in regional analysis, Bayesian analysis and simulation and behavior analysis etc. have been investigated by many researchers. Cunnane (1987) has traced selected developments under a number of different headings in his classical paper on 'Review of Statistical Models for Flood Frequency Analysis'. He classifies the historical review in these areas under three phases; an early phase of problem definition, a second phase of technical development beginning with the introduction of extreme value theory and a final phase in which development of models based on the outcome of simulation experiments, took place. The review of work in all these areas is beyond the scope of the present study. However, the outline of development of flood frequency analysis as given by Cunnane (1987) is presented in Tables 2.1 and 2.2.

Keeping in view the objective and scope of the present study the noted developments in the areas of AFS and PDS models are presented in subsequent subsections. The work reported in literature for the comparison of AFS and PDS models is also reviewed and discussed in this section.

2.2.1 Annual Flood Series (AFS) models

An AFS is the sequence of largest annual peak discharge of each year of record. Although annual maximum instantaneous peak discharge values should constitute the AFS, the use of maximum daily mean discharges is satisfactory for large basins, (Goel, 1989; NERC, 1975) since their time of concentration generally exceeds one day. The development of AFS models are reviewed in the light of (a) the form of distribution and estimation procedures and (b) normalization methods.

Table 2.1*

Outline of development of flood frequency analysis methods in three phases under selected headings

THREE PHASES OF DEVELOPMENT 1.

Early Phase	Recognition that there is no such thing as a single design flood but rather a choice of different return period floods depending on circumstances.
Middle Phase	Introduction of Extreme Value theory, algebraic development of alternative parametric model types and estimation schemes.
Recent Phase	Development of models which have been stimulated by the outcome of simulation experiments

ANNUAL MAXIMUM MODEL DEVELOPMENT 2.

Fuller (1914)	Concept of return period
Foster (1924)	Theoretical frequency curves
Hazen (1932)	Large body of empirical knowledge
Gumbel (1941)	Deductive EV knowledge
Kimball (1946)	ML estimation
Langbein (1949)	Linked AM and POT
Thomas (1948) Adamowski (1985)	Distribution-free methods
Chow (1951)	Frequency Factor K_T
Jenkinson (1955, 1969)	GEV Distribution
Kaczmarek (1957)	Study of Standard Errors
Matalas et al. (1975)	Condition of Separation
Houghton (1978)	Wakeby Distribution
Rossi et al (1984)	TCEV Regional Model

POT MODEL DEVELOPMENT 3

Langbein (1949)	Linked AM and POT
Borgman (1963) Shane & Lynn('64) Bernier (1967a)	POT Model as Renewal Process
Todorovic (1970) Todorovic & Zelenhasic (1970) Todorovic & Rousselle (1971)	Generalised Probability Model
Cervantes (1983) Kavvas (1982 a,b,)	Trigger-type cluster models
Cunnane (1979) Ashkar and Rousselle(1983a,b)	Test of Poisson Rates of occurence

USE OF TIME SERIES MODELS 4

Quimpo (1967) Hall and O'Connell (1972) O'Donnell, Hall and O'Connell (1972)	Seasonal means and AR Residuals
Weiss (1973) NERC (1975,1.2.9)	Shot-noise Models

COMPARISON OF AM AND POT MODELS 5

Langbein (1949) Chow (1950) Takeuchi (1984)	Theoretical comparison of T_{AM} and T_{POT}
Beran and Nozdryn-Plotnicki (1977) Yevjevich and Taesombut (1978) Taesombut and Yevjevich (1978) Takeuchi (1984)	Empirical Comparison of T_{AM} and T_{POT}
Cunnane (1973) Yevjevich and Taesombut (1978) Taesombut and Yevjevich (1978) Tavares and da Silva (1983) Rosjberg (1985)	Comparison of efficiency of AM and POT estimates

REGIONAL FLOOD ESTIMATES 6

Fuller (1914) Hazen (1932)	Station Year Method
Dalrymple (1960)	Index flood and distribution of standardised variate $X = Q/\bar{Q}$, and choice of homogeneous regions
Cole (1965)	Application of above to UK data
Nash and Shaw (1965)	Relation of regional average dimensionless moments to catchment characteristics
USWRC (1977)	Map of Regional Skew
NERC (1975 1.2.6) Beable and McKercher (1982)	Regional average dimensionless order statistics
Wallis (1980)	Regional average standardised PWM's.

Contd.......

Table 2.1* (contd.)

AM DISTRIBUTIONS RECOMMENDED FOR USE	7
(a) on basis of regional studies	
Benson (1968) USWRC (1967,1977,1981)	LP3
NERC (1975)	GEV
Houghton (1978)	Wakeby
Wu and Goodridge (1976)	P3/LN
McMahon & Srikanthan (1981a)	LP3
Hua (1985)	P3
(b) on basis of simulation results	
Hosking et al (1985a)	GEV/PWM
Wallis and Wood (1985)	Wakeby/PWM
Rossi et al (1984)	TCEV

PARAMETER ESTIMATION	8
Kimball (1946)	ML for EV1
Chow (1953)	Least Squares
Gumbel (1958)	Geometric mean of two least squares lines
Jenkinson (1969)	ML and Sextiles for GEV
Greenwood et al (1979)	PWM for distributions expressible in inverse form
Wallis (1980)	Regionally averaged standardised PWM's for index flood method
Rossi et al (1984)	Iterative ML procedure for regional TCEV

STANDARD ERRORS	9
See Table 2.2	

SIMULATION AND BEHAVIOUR ANALYSIS	10
Nash and Amorocho (1966) Lowery and Nash (1970)	EV1 quantile standard error
Matalas & Wallis (1973)	Pearson 3 estimation
Wallis et al. (1974)	Properties of small sample statistics
Matalas et al (1975)	Regional skewness condition of separation
Landwehr et al (1978)	Relationships between real and log space statistics
Rossi et al (1984) Hosking et al (1985a) Wallis and Wood (1985) Lettenmaier et al (1985) Lettenmaier & Potter (1985) Arnell & Gabrielle (1985)	Statistical properties of regional flood estimation procedures. Selection of some regional parameter values.

BAYESIAN ANALYSIS	11
Bernier (1967b)	Non-informative prior distribution
Cunnane and Nash (1971/74)	Informative Regional prior distribution
Davis et al (1972)	Bayesian decision theory for flood alleviation design
Wood and Rodriguez-Iturbe (1975a)	Informative conjugate Regional prior distributions for flood protection design
Kuczera (1982a)	Linear empirical Bayes procedures

DECISION RELATED PROCEDURES	12
Matalas and Wallis (1972) Slack et al (1975)	Study on usefulness to decision-making of prior information about distribution type.
Davis (1972) Wood and Rodriguez-Iturbe (1975) Kuczera (1982 a,b; 1983)	Use of Bayesian decision theory

USE OF HISTORICAL DATA	13
Benson (1950)	Graphical Procedure
Chen et al. (1975) Hua (1985) Luo (1985)	Use of historical floods in China
Pang (1985)	Floods in early Chinese history
Sutcliffe (1985)	Historical floods on River Nile
Baker (1985) Xu and Ye (1985)	Palaeoflood hydrology and geomorphological methods of quantifying historical floods
Leese (1973)	Censored estimation techniques
Condie and Lee (1983)	Censored estimation and Monte Carlo evaluation
Hosking et al. (1985a)	Question the statistical wisdom of using historical data
Hosking and Wallis (1984) Stedinger and Cohn (1985)	Evaluations of usefulness of historical data

REGIONAL HOMOGENEITY	14
Langbein (1947) Dalrymple (1960)	Homogeneity test - based on plot of T against record length.
Mosley (1981) Wiltshire (1985) Wiltshire (1986 a,b,c) Acreman and Sinclair (1986)	Question the use of geographical regions and seek homogeneous regions in catchment characteristic space.

* *Reproduced from: Cunnane, C. (1987), 'Review of Statistical Models for Flood Frequency Estimation', in V.P.Singh (Ed.), 'Hydrologic Frequency Modeling', D.Reidel Publishing Company, pp.89-90.*

Table 2.2(*)

Selected References to investigations into sampling properties of quantile estimates

Model Distribution	At-site	Regional	Reference
Extreme Value	*		Kaczmarek (1957)
Type 1	*		Nash and Amorocho (1966)
	*		Lowery and Nash (1970)
	**		Landwehr et al. (1979a)
		§	Greis and Wood (1981)
	*		Fiorentino and Gabrielle (1984)
	**	§§	Lettenmaier et al. (1985)
	**		Lettenmaier and Potter (1985)
Lognormal	*		Kaczmarek (1957)
	*		Sangal and Biswas (1970)
	*		Burges et al. (1975)
	**	§§	Kuczera (1982b)
	**		Stedinger (1980)
	**	§§	Lettenmaier and Potter (1985)
Pearson Type 3	*		Matalas and Wallis (1973)
	*		Bobee (1973)
Log Pearson 3	*		Condie (1977)
	*		Nozdryn-Plotnicki and Watt (1979)
	*		Hoshi and Burges (1981)
	*		Phien and Hsu (1984)
	**	§§	Wallis and Wood (1985)
General Extreme Value (GEV) and Log EV1	*		Jenkinson (1969)
		§§	Hosking et al. (1985a)
	*		Hosking et al. (1985b)
	**	§§	Wallis and Wood (1985)
	**	§	Lettenmaier et al. (1985)
		§§	Arnell and Gabrielle
Wakeby	*		Landwehr et al. (1979b,c)
		§	Wallis (1980)
		§§	Hosking et al. (1985a)
		§§	Wallis and Wood (1985)
		§§	Arnell and Gabrielle (1985)
TCEV		§§	Arnell and Gabrielle (1985)

* Parent and assumed model distributions are the same.

** Also tests model distribution under different parent distribution assumption(s) (Robustness test).

§ and §§ For regional cases correspond to * and ** in at-site estimation

(*) *Reproduced from: Cunnane, C. (1987), 'Review of Statistical Models for Flood Frequency Estimation', in V P Singh (Ed.), 'Hydrologic Frequency Modeling', D.Reidel Publishing Company, pp.91.*

Form of Distribution and Estimation Procedures

The first successful attempt to interpret flood flows in terms of probability was made by Fuller (1914) who stated that the mean annual flood was approximately proportional to 0.8 power of the drainage area and that flood flows above the mean followed an exponential tailed probability law. The work of Fuller was supported by Hazen (1914) who constructed the normal and lognormal probability papers for plotting of floods. Foster (1924) gave the theoretical frequency curves. Hazen (1932) proposed the use of lognormal probability paper for graphical fitting of data.

In 1941, Gumbel published the first of a number of papers in the application of Fisher Tippett theory of extreme values to flood frequency analysis. The use of extreme value theory was further extended by many other hydrologists and researchers. Kimball (1946) gave maximum likelihood estimates of parameters of the distribution of maximum values. Thomas (1948) and Chow (1951) gave the concept of frequency factor and distribution free methods. Benson (1950) developed a synthetic 1000 years record of peak floods based on a straight line plotting on the extreme value probability paper.

The concept of General Extreme Value (GEV) distribution was given by Jenkinson (1955) who also applied it to annual maximum and minimum values of meteorological elements. The efficiency of the estimation of floods with a given return period for extreme value type 1 (EV1) and lognormal distributions was given by Kaczmarek (1957). Kendall (1959) discussed the relationship between the risk of occurrence of an event in a given period of time and its return period. Riggs (1961) derived a relation between the magnitude, design period in years and probability of not exceeding that magnitude in the design period from a cumulative frequency curve.

Nash and Amorocho (1966) showed that the extrapolation of magnitude frequency relationship obtained from finite samples is not too hazardous when the form of the frequency distributions is known for the population of floods. They made a plea for research to establish, if possible, the true form of the frequency distributions of floods.

Water Resources Council, WRC (1966) describes the methods most commonly used by Federal agencies for making frequency studies of runoff at individual streamflow stations and provides an extensive list of applications of frequency analysis methods.

WRC (1967, 77, 81) adopted the log Pearson type III distribution (of which lognormal is a special case) to achieve standardization of flood frequency procedures used by Federal agencies in USA.

Benson (1968) as the chairman of the work group on flow frequency methods, Hydrology Committee, WRC, studied the most commonly used methods of flood frequency analysis and compared the results by applying these methods to a selected group of long record representative sites in different parts of the country. He showed that there are large differences in the predicted floods when different distributions are assumed particularly for larger recurrence intervals.

Jenkinson (1969) gave the method of maximum likelihood and the method of sextiles for parameter estimation for GEV distribution.

CWC (1972) recommended the use of EV1 distribution for Indian rivers. Singh and Sinclair (1972) proposed mixed distribution which is a combination of two distributions. Such a distribution contains a mixture parameter which determines the relative frequency of occurrence of events from each population. This model has been proposed to be used in situations when there is clear physical distinction between the two distinct types of events in the flood series.

NERC (1975) adopted GEV distribution to achieve standardization of flood frequency procedures. As reported, goodness of fit of seven distributions was tested for 28 stations in Great Britain and 7 stations in Ireland. The results of these tests showed that the performance of Pearson type III (PT III) and log Pearson type III (LP III) distributions was sensitive to the formulation of the tests. The GEV distribution was more stable and for this and other reasons it was recommended as first choice of annual flood peaks.

Matalas et al. (1975) and Wallis et al. (1977) showed that in nearly all the 14 regions subdividing the United States a condition of ´separation´ existed which can neither be explained by LP III distribution or by drawing regional skew maps. The separation was observed while comparing the historical skewness and the simulated skew values derived using other distributions. It was observed that the natural data consistently displays more inherently unstable skews than the one derived from various statistical distributions. Wallis et al. (1977) suggested spatial mixing of skewness coefficient as a possible reason for 'separation'.

To explain some of the 'separation effects', the Wakeby distribution was introduced by Houghton (1978 a) as the grand parent of distributions because of its ability to generate flows which mimic most conventional hydrologic distributions if the parameters are chosen correctly. Houghton (1978 b) noted the similarity of the Wakeby to the older Fuller flood model $x = a+b[(T-1)]^c$ where $T = 1/(1-F)$. One of the most attractive features of the Wakeby distribution is that the right and left hand tails of the distribution can be modelled separately. Secondly, this distribution is able to explain some of the 'separation effects' (Matalas et al. 1975) which many other distributions cannot. In conjunction with the Wakeby distribution, Houghton (1978 b) also proposed the incomplete means estimation procedure. This procedure uses no moments higher than the first, resulting in the bias of the estimates.

Method of least squares, method of moments and method of maximum likelihood are the commonly used parameter estimation techniques for most of the distributions. Greenwood et al. (1979) suggested a more elegant probability weighted moments (PWM) approach for estimating the parameters of the distributions which can be expressed in inverse form. An application of PWM technique in regional flood frequency was proposed by Wallis (1980). The PWM method works well for situations where records are extremely short and streamflow samples are highly skewed and kurtotic.

The major drawback of PWM method used to be that it cannot be applied to distributions whose inverse forms are not available. In a major breakthrough, very recently, Jing et al. (1989) have overcome this drawback. They have shown that the functional domain of PWM method can be extended to the function inexpressible in inverse form also. They have derived expressions relating probability weighted moments to parameters of several distributions inexpressible in inverse form such as normal, log-normal, Gamma and Pearson type III distributions.

Now a days another method of parameter estimation based on principle of maximum entropy (POME) is gaining importance in hydrology field. The POME has widely been used in telecommunication (Shannon, 1948; Jaynes, 1957). Sonuga (1972, 76) used it for hydrologic frequency analysis and rainfall runoff modelling. Amorocho and Espildora (1973) used it for assessment of uncertainty in hydrologic systems and models. Singh and associates (Singh and Krstanovic, 1986, 1987; Singh et al. 1985) have applied it for derivation of frequency distributions, parameter estimation and sediment and water quality modelling.

Rao (1980) used the method of mixed moments to estimate parameters of log Pearson type III distribution.

Rossi et al. (1984) offered the two component extreme value distribution (TCEV) as a good distribution for modelling annual floods in Italy. The TCEV is motivated on the premise that floods above a particular threshold come from two independent processes, each occurring according to a compound Poisson process and having exponentially distributed exceedances. With such a structure for floods above a threshold, annual floods have a TCEV distribution. Use of TCEV is supported by the observed 'separation effect' in estimated skews and by the observed distribution of reduced largest floods. Beran et al. (1986) explored the basic geometry and moments of the TCEV distribution.

Ahmad et al. (1988) demonstrated that log - logistic distribution (LLG) has many properties well suited for modelling flood frequency data. They compare the performance of the log - logistic distribution with GEV, three parameter lognormal (LN3) and PT III distributions. On the basis of empirical distribution function tests they conclude that (i) the LLG provides a better fit than the GEV for reasonably long flood series (> 24 years in length) at individual sites and (ii) that it provides a better fit than the GEV, LN3 and PT III distributions on a regional basis.

In a classical paper on 'Methods and Merits of Regional Flood Frequency Analysis' Cunnane (1988) concludes that 'At present the WAK/ PWM regional procedure is the best available at site regional flood quantile estimating procedure'. This is supported by the work of Kuczera (1982 b) who found that four parameter Wakeby distribution (WAK-4) outperformed the LP III and log EV1 distributions as at site quantile estimators. In the regional context also the five parameter Wakeby distribution (WAK-5), estimated from regional averaged standardized PWMs and denoted by WAK/ PWM has been consistently found to be a better quantile estimator than GEV/ PWM (Hosking et al., 1985 a) or LP III / USWRC (Wallis and Wood, 1985) or TCEV (Arnell and Gabrielle, 1985).

Normalization Methods

Data arising from various situations form their own distributions. Thus instead of fitting a known distribution to the data it is better to find the probable distribution of the population series from the data itself which is a difficult computational task. Alternatively, the data could be reconstituted by some suitable transformation such that the transformed series follow a particular distribution. Transformation of the series to follow normal distribution has caught the attention of many research workers, though there have been attempts to transform the series to follow other distributions, eg. EV1 distribution by Seth et al. (1986) also.

Bethlahmy (1977) has suggested the use of SMEMAX transformation to normalize the skewed data. This method transforms a skewed series using the smallest median and the largest value of the given series. The resulting series has the difference between the largest value and the median value equal to that between median value and the smallest value. This is a necessary but not sufficient property to normalize the transformed series. The resulting series can still have appreciable skewness or kurtosis or both, even though median value is equidistant from the smallest and the largest value.

Chander et al. (1978) suggested the use of Box-Cox transformation (1964) for normalization of the given series and flood frequency analysis. The authors expressed satisfaction with Box-Cox transformation for estimating flood quantiles on the basis of 15 long records examined by them. The Box-Cox transformation makes the coefficient of skewness of the transformed series equal to zero. For kurtosis correction, the authors have proposed the use of table based on the work of Box and Tiao (1973) and Tiao and Lund (1970).

Seth et al. (1983) compared the efficiencies of several normalization procedures and found the use of Box-Cox transformation satisfactory for normalization of the annual flood series.

Gupta et al. (1989) proposed two step power transformation (TSPT) for normalization of the series and applied it to AFS of 17 Indian rivers satisfactorily. The TSPT rectifies some of the operational difficulties with regard to the use of correction factor for kurtosis. At present TSPT seems to be the most suitable normalization procedure as it preserves C_s and C_k of normal distribution in the transformed series and avoids use of any table.

2.2.2 Partial Duration Series (PDS) models

In partial duration series (PDS) approach, all the floods that exceed a certain threshold level Q_b are considered. If there are M such floods in N years of record then the average rate of occurrence, λ , will be equal to M/N. In PDS approach, once the truncation level is defined the flood events can be analyzed in terms of the number of events per year, peak discharges, durations and volumes. Most of the research reported in literature considers only number of events per year and the peak discharges associated with these events. The brief chronological review of literature in this area is given in the following paragraphs.

Bergman (1961) presented a simplified technique for computing the probability that a near extreme occurrence of a physical phenomenon would exceed a selected value. The limiting distribution of the maximum term in a sequence of independent identically distributed random variables was first analyzed by Berman (1962) who showed that the limiting distribution of the maximum term was a mixture of distributions.

Shane and Lynn (1964) found the distribution function of the flood peaks in a certain time interval (0,t) assuming their number to be time homogeneous Poisson process. The flood magnitudes were assumed to follow an exponential distribution.

Zelenhasic (1970) and Todorovic and Zelenhasic (1970) considered the case where the number of exceedances follows a non homogeneous Poisson process with the sequence of flood discharges kept independently identically distributed (iid). Zelenhasic (1970) comments that the determination of the distribution function of flood exceedances is purely a problem of estimation and according to the present state of art there are no theoretical grounds that indicate the form of the distribution. Todorovic and Rousselle (1971) and Rousselle (1972) extended the work of Zelenhasic (1970) to non identically distributed exceedances by dividing the water year into different seasons and keeping the iid assumption only within the season being considered.

Todorovic (1978 a) presented stochastic models of extreme flows and their application to design. He also explained various assumptions made in the formulation of the models. North (1980) proposed time dependent stochastic model of floods based on non homogeneous Poisson process for the occurrences and time dependent exponential distribution for the magnitudes.

In the application of PDS approach, it is occasionally observed that successive exceedances are correlated and to reduce this correlation some investigators (WRC, 1976; Cunnane, 1979) tend to impose certain restrictions on the inter-arrival times of the flood events. Ashkar and Rousselle (1983 a) showed analytically how such restrictions interfered with the underlying hypothesis of the Poisson process commonly used to model flood counts, and cautions against imposing such restrictions that may render this simple and appealing model inapplicable.

While commenting on the truncation level (threshold) used in PDS models, Ashkar and Rousselle (1983 b) concluded that both the Poisson distribution as a model for flood frequency and exponential distribution as a model for flood magnitudes would remain so with any higher level of truncation also. A great degree of freedom is thus left to the user to chose the truncation level. If the Poisson and exponential distributions are to be used then the choice of the base level should be made primarily on mathematical grounds rather than on economic and engineering considerations.

2.2.3 Comparison of AFS and PDS models

There is classical dilemma in flood frequency analysis regarding the use of AFS or/and PDS models, and hence several investigators have made efforts to compare and relate the two types of models. The AFS and PDS models can be related on the basis of value of flood discharge Q for a given return period T, i.e., Q (T) in the parent process and the value of Q (T) evaluated by the two models. The models can be related in other way also i.e, by relating T given by various models for a particular Q.

Langbein (1949) showed that when T is small T_{AFS} differs appreciably from T_{PDS} and hence from T but differs by only half year at large values of T. He related the two return periods by the following formula

$$T_{AFS} = 1/(1 - \exp(-1/T_{PDS})) \qquad (2.1)$$

Chow (1950) discussed Langbein's formula, and pointed out that the difference between T_{AFS} and T_{PDS} evaluated by the relative difference $(T_{AFS} - T_{PDS})/T_{AFS}$ is less than 5% for $T_{PDS} \geq 10$ years and greater than 10% for $T_{PDS} < 5$ years. Chow further stated that ' in ordinary engineering practice a five percent difference is tolerable and that, the two methods give essentially, identical results for intervals greater than about 10 years'.

Takeuchi (1984) evaluated Langbein's theory (1949) and suggested an alternative derivation procedure. The resultant formula given by him is identical to Langbein's but the condition to be satisfied for the formula to hold good is replaced by a new, more relaxed condition. He confirmed the validity of Langbein formula (1949) and Chow ´s discussion (1950).

Cunnane (1973) compared the statistical efficiency of Q(T) obtained by AFS model and PDS model on the basis of variance of Q(T) obtained by the two models. He showed analytically that, under commonly used assumptions and for return periods greater than about 10 years, the PDS estimate of annual maximum discharge Q (T) for a given return period (T) had smaller sampling variance than the AFS model estimate for $\lambda \geq 1.65$.

NERC (1975) describes use of the Poisson and the negative Binomial distributions to model the number of exceedances and the exponential distribution to model the magnitude of exceedances for 26 hydrometric stations from all over UK. It was found that in most of the cases the Poisson and exponential distributions fit reasonably well the series for number and magnitude of annual exceedances respectively. However these distributions cannot be taken as granted as they were clearly rejected for some stations. As a final recommendation NERC (1975) suggests that for small samples of length less than 10 years the PDS method should be used to estimate the mean annual maximum discharges.

Taesombut and Yevjevich (1978) estimated the probability distribution of annual maximum flood peaks by using a combination of probability distributions for number and magnitude of flood peaks that exceed a certain threshold based on data of 17 stations from United States. They concluded that in the range of truncation levels with an average number of exceedances from one to four, the independence assumptions underlying the use of PDS method could be accepted. The mixed Poisson or the Poisson distributions and mixed exponential or exponential distributions gave the best results in modelling the annual number and magnitude of exceedances. The authors confirmed the results of Cunnane (1973) based on synthetic data generated by a daily flow model. As a basic conclusion Taesombut and Yevjevich (1978) state that ' when the model of partial flood series is developed with assumptions for its derivation supported by data from low truncation levels, the partial flood series is more efficient or more useful in estimating annual flood peaks than the annual flood series, especially in the case of small sample sizes'.

Tavares and Da Silva (1983) showed that the Cunnane's (1973) equation for Var (Q_T) for a given return period could underestimate (or overestimate) the variance observed in a simulation study, if the average number of annual exceedances was higher (or smaller) than 2. It was shown that there was significant lower estimation variance for PDS if λ was greater than 2, and this reduction of the estimation variance increased with return period and λ. They proposed negative exponential law with constant serial auto- correlation for auto- correlated annual exceedances and showed that estimation variance of Q (T) increases with increase in ρ_1.

Rosbjerg (1985) further extended the work of Tavares and Da Silva (1983) by introducing a correction factor in the variance formula to reduce the deviations between theoretical and Monte Carlo generated samples. He also developed a formula for estimation of variance of T year flood for dependent case and showed that it was in fine agreement with Monte - Carlo based variance calculations.

Correia (1983) applied the PDS method to daily flow records from 12 stations in Portugal for truncation levels corresponding to λ equal to 0.5 to 4. He compared estimates of annual peak discharges for return periods from 5 to 1000 years with results obtained from the EV1, Lognormal, PT III and LP III and GEV distributions applied to annual flood series. The Poisson and exponential distributions were found adequate to model the number of annual exceedances and the magnitude of exceedances, respectively. He found that PDS method with λ from 2 to 3 gave the best results by fitting the observed annual maximum daily flows and was significantly more efficient in estimating the annual peak discharges for the return periods considered in the study. The use of PDS method was recommended principally for short records.

Goel et al. (1987) compared the efficiencies of AFS and PDS models on the basis of exact theoretical and approximate theoretical approaches using data of river Narmada at Mortakka (India). On the basis of exact theoretical approach, the sampling variance of Q (T) by PDS model (for any value of λ) was shown to have lesser sampling variance than AFS model if the return period was less than 11 years. For any return period the PDS estimate of Q(T) had smaller sampling variance than that of AFS if λ was at least 1.65. However, on the basis of approximate theoretical approach, in the range of λ studied (1.0 to 2.437) the sampling variance given by AFS was smaller than that of PDS.

Buishand (1989) derived an expression for asymptotic variance of quantile estimates. He shows that the method of maximum likelihood leads to slightly biased quantile estimates, and unbiased quantile estimates can be obtained from the minimum variance unbiased (MVU) estimates.

2.3 Modelling of Other Characteristics of Flood

In all the studies mentioned in previous sections, only peak discharge of every flood event has been considered, and flood risk is simply characterized in terms of the annual flood for a given return period. The other flood features such as duration and volume are not considered although they are important characteristics to be considered in the design of many flood control systems. Of the numerous papers written on flood analysis very few deal with the analysis of flood features other than peak discharge. Some important work in this area has been reviewed as given in following paragraphs.

Todorovic (1971) used the PDS approach together with the mathematical assumptions of Todorovic and Zelenhasic (1970) to derive exact expression for time of occurrence of extreme flood in selected time. Todorovic and Woolhiser (1972) applied this theory to two rivers of USA and found good agreement between observed and theoretical distributions.

Gupta et al. (1976) extended the work of Todorovic and Woolhiser (1972) and developed the expression for the joint distribution function of the largest flood peak and its time of occurrence. They also derived the distribution function of the time of occurrence of the largest flood for the two rivers in USA and modified the expression valid for independently identically distributed exceedances to non identically distributed exceedances.

Todorovic (1978 b) presented three stochastic models based on PDS approach. These models varied only in assumptions concerning properties of exceedances of threshold level. He determined the distribution of time of occurrence of the largest exceedance and derived distribution function of the largest flood volume in a time interval (Ø,t).

Ashkar and Rousselle (1982) studied the multivariate and marginal distributions of flood magnitudes, duration and volume for three stations in Quebec. They considered hydrograph of flow above a particular threshold as a triangle and showed that assumption was not unrealistic.

A theoretically more general model was developed by Kavvas and co-workers (Kavvas and Delleur, 1975; Kavvas, 1982; and Kavvas, et al., 1983) treating flooding as the clustering phenomenon and its mechanisms as centers of clusters of flood peaks. This model is not used because of its mathematical complexity (Krstanovic and Singh, 1987).

Krstanovic and Singh (1987) used the principle of maximum entropy to derive a multivariate stochastic model for flood analysis. By specifying appropriate constraints in terms of covariances, variances and cross covariances, multivariate Gaussian and exponential distributions were derived. As a special case the bivariate process of flood peaks and volumes was investigated for the following three cases: (i) the peaks and volumes are independent and occur the same number of times, (ii) the number of peaks is more than that of volumes in the same time interval and (iii) peaks and volumes exhibit dependence. Marginal distributions of flood characteristics were obtained first with no restriction imposed and then with assumptions of independent occurrences and a high threshold value. They also obtained the conditional distribution of flood volume for the given peak. In the model special emphasis was given to the matrix of Lagrange multipliers.

Correia (1987) derived the joint distribution of flood peaks and durations using PDS approach assuming (i) independence between them and (ii) linear dependence between the two variables. For the independent case the marginal distributions of flood peaks and durations were assumed as exponential. For the dependent case he assumed the conditional density function h $(x|y)$ as normal with variance σ^2 and mean $\mu(X) = AY + B$. For volume, the triangular relationship between volume, peak and duration was assumed. On the basis of results of application to twelve Portuguese rivers, he reported satisfaction over the use of partial duration series approach for the multivariate characterization of flood peak, duration and volume. However, he emphasized the need of additional research before final conditions could be formulated with respect to the general use of multivariate partial duration series.

The approach given by Correia (1987) seems to be promising but the general applicability of the assumption of conditional density function as normal is somewhat doubtful.

Sackl and Bergmann (1987) used bivariate normal distribution as the parent bivariate distribution function for flood peaks and volumes of direct runoff after transforming the marginal distributions of both variables into normal distribution using transformation proposed by Schroder (1969). They fitted and tested the bivariate normal distribution using the equi- lines of probability density function (ISO - PDF lines). According to the authors their approach offers various possibilities of probability interpretation and the conditional distributions of one variable can be computed for a constant value of the other variable which shows the probability by which the variable is equalled or exceeded for the fixed variable.

The use of bivariate normal distribution for flood peaks and volumes seems to be quite appealing. There is scope of (i) improvement in the normalization procedure used by the authors and (ii) extension of the methodology using partial duration series.

The studies of Zelenhasic and Salvai (1987) and Singh and Krstanovic (1987) need special mention in this review. Though, these are not directly related to flood analysis, yet the concepts presented and discussed in these studies have been very useful in the formulation of methodology of the present work. Zelenhasic and Salvai (1987) have presented a method of completely describing and analyzing the stochastic process of streamflow droughts. Singh and Krstanovic (1987) applied the principle of maximum entropy to derive a stochastic model for sediment yield from upland watersheds. By maximizing the conditional entropy subject to certain constraints, they derived probability distribution of sediment yield conditioned on the probability distribution of direct runoff volume. It was shown by the authors that the joint distribution of sediment yield and runoff, obtained by the application of principle of maximum entropy and subject to constraints in terms of variances and covariances, was bivariate normal distribution.

2.4 Generation of Daily Flows

As stated earlier, though for flood frequency analysis, the annual maximum instantaneous discharges constitute the AFS, the use of annual maximum daily discharges is satisfactory for large basins (Goel, 1989; NERC, 1975). Therefore the synthetically generated daily flows play an important role in comparing the performance of various models (Taesombut and Yevjevich, 1978) and for the validation of new techniques proposed for various aspects of flood frequency analysis. Their role in simulation studies is well known. This section gives a brief review of some important developments in the area of synthetic generation of daily flows.

The work in the area of daily flow generation started with the pioneering work of Quimpo (1967). The model proposed by Quimpo (1967) and Payne et al. (1969) used autoregressive time series model based on Gaussian distribution. These models did not yield fully satisfactory results as the models were based on a statistical analysis which only used the classical stationary co-variance spectrum estimators, and ignored the statistical properties of the rising and recession limbs of the hydrograph.

Yakowitz (1973) proposed another daily flow model based on nonlinear autoregressive model. This model was operationally successful in modelling zero flows, and also steep rising and falling behavior. However, this model lacked in mathematical tractability (Kumar, 1982).

Weiss (1973, 1977) showed that Gaussian autoregressive models cannot reproduce the rapid rises and slow recessions observed in the streamflow records of daily flow intervals. He suggested a new model known as 'Shot Noise Model' which has a built in capability to model the ascension recession behavior. Making an analogy to the theory of the unit hydrograph, he also proposed the double shot noise model, based on filtered Poisson process, as a model for the continuous streamflow records. He used one shot noise process for the surface runoff and other shot noise process for the ground water discharge so as to approximate the nonlinear behavior of the watershed system. His double shot noise model was the summation of these two processes which he assumed to be independent of each other. O'Connell (1974) gave further details of application of shot noise models in synthetic hydrology.

NERC (1975) applied the 'Shot Noise Model' to daily mean discharges from Avon at Evesham in a trial of the flood reproducing properties of the model. Three methods to incorporate the skewness into the model were also tried. Based on the application of these methods to this data, it was found that the simplest shot noise model with exponentially distributed jump heights was the only model which could be used to fit the flows.

Treiber and Plate (1975) assumed that the input rainfall pulses to the watershed system were a white noise sequence. They obtained the watershed system transfer function for the daily flows under this assumption. Later in their work, they assumed a Markov structure for the rainfall pulses and used these Markovian pulses as the input series.

O'Connell and Jones (1979) describe three types of models viz. linear autoregressive model, shot noise models and nonlinear autoregressive models which have been developed in the UK for the stochastic simulation of daily flows.

Kottegoda and Horder (1980) used an alternating renewal process for the occurrence of rainy and dry days. They verified their renewal process assumption by the correlation coefficients of the lengths of adjacent wet and dry spells. They showed that these coefficients are nearly zero. They made the assumption that the rainfall amounts on successive days are independent. Using this independence assumption and also assuming stationarity for the daily flows, they constructed a time invariant system transfer function from a stationary sample covariance function of the daily flows. Later, they generalized the system transfer function to a function which depends on the state of the flow when the rainfall pulse occurs. They then used regression equations to obtain effective rainfall amounts from the actual rainfall. Finally, the daily flows were obtained from a convolution of the effective rainfalls with the variable transfer function. The daily flow model was compared to the historical records in terms of the annual maxima of the observed flows, and the synthetic flows that are generated by the model. The results of this comparison are not too satisfactory (Kavvas and Delleur, 1984).

Kelman (1980) proposed a stochastic model for the description and generation of daily flows. The basic assumption is that the rising and falling limbs of the hydrographs ought to be modelled separately due to the fact that they translate different physical processes. The rising limb is mostly due to factors external to the watershed. On the other hand, the falling limb is mostly governed by the emptying water from the watershed. The model assumes the conceptual representation of the watershed as two linear reservoirs. Any sequence of recession discharge is then a stochastic output from these two reservoirs. His model worked well for the fast decreasing recession limbs with high peak discharge, but yielded unsatisfactory results when the peak discharge was low. The significant contribution of Kelman (1980) with respect to the statistical analysis of daily flows was that he used time varying mean, standard deviation and first lag correlation coefficient estimators in order to draw inferences about the behavior of the first and second moments of the daily flow process. He also computed the recession curves of the observed daily flow series but only within the stationary domain. His use of constant decay coefficients resulted in unsatisfactory modelling of the recession limbs, corresponding to low peak discharges.

Kavvas and Delleur (1984) have given very useful comments on the works of Quimpo (1967), Payne et al. (1969), Weiss (1973), Treiber and Plate (1975), Kottegoda and Horder (1980) and Kelman (1980). The periodic statistical analysis of daily streamflow data in Indiana, USA, by them shows that (a) the persistence properties of daily flows depend on the storage state of the basin at the specified time origin of the flow process; (b) the daily streamflow process is time irreversible; (c) the probability distribution of the daily hydrograph peak inter arrival time depends both on occurrence time of the peak from which the inter arrival time originates and on the discharge exceedance level; and (d) if the daily streamflow process is modelled as the release from a linear watershed storage, this release should depend on the state of the storage and on the time of the release as the persistence properties and the recession limb decay rates were observed to change with the state of the watershed storage and time. Therefore, a time varying reservoir system needs to be considered if the daily streamflow process is to be modelled as the release from a linear watershed storage.

Vandewiele and Dom (1989) proposed another non-Gaussian multi component model for modelling river flows. In the model, recession and base flow components are subtracted from the observed flow before deseasonalization. This results in the possibility of modelling sharp rises and slow decreases, as are observed in flow time series with a sufficiently small time base. The remaining part of the flow called random shocks are then deseasonalized in order to model them as a stationary second order Markov process with non normal transition probabilities. This allows the modelling of zero random shocks. Deseasonalizing is modelled by a truncated Fourier series, so that the number of parameters remains small. They applied this model on a weekly time basis to the Meuse catchment (20,000 sq. km) upstream of Liege in Belgium and France. The comparison of 1000 year simulated series properties with that of the observed series was satisfactory even on the basis of properties of hydrological interest such as return periods, which were not modelled explicitly.

2.5 Summing up

Based on this review of literature related to AFS models, PDS models, comparison of AFS and PDS models, modelling of flood characteristics other than peak discharge and generation of daily flows the following broad conclusions can be drawn.

AFS Models

There is no general agreement among the hydrologists for the use of any particular distribution and its parameter estimation technique. It is very difficult to discard any distribution and a parameter estimation technique without extensively testing it and at the same time it will be very much time consuming and expensive to try all the available techniques for a particular site. Wakeby /PWM, GEV /PWM, LPIII /PWM and EV1 /PWM techniques cover most of the major schools of thought for at site frequency analysis of annual flood series.

In the area of application of normalization procedures in flood frequency analysis, the two step power transformation is an improvement over original Box-Cox transformation for normalization of a given series since it preserves C_s and C_k of normal distribution in the transformed series and avoids use of any table.

PDS Models

Besides typical features of PDS approach, data limitations in many cases have led to significant developments in this area. Using partial duration series, a number of models, ranging from the simplest up to one where peak magnitudes vary with season are available.

The determination of the distribution function of flood exceedances is purely a problem of estimation and according to the present state of art there are no theoretical grounds that indicate the form of the distribution (Zelenhasic, 1970).

The choice of the threshold level should be made primarily on mathematical grounds rather than on economic or engineering considerations if the Poisson distribution as a model for flood frequency and exponential distribution as a model for flood magnitudes are considered (Ashkar and Rousselle, 1983 b).

Comparison of AFS and PDS Models

The comparison indicates two important points.

(i) Under commonly used assumptions and for return periods greater than 10 years, the PDS estimate of annual maximum discharge Q (T) for a given return period T has smaller sampling variance than AFS model estimate for $\lambda \geq 1.65$ (Cunnane, 1973).

(ii) For small samples ($N \leq 10$), the PDS model should be used to estimate the mean annual maximum discharges (NERC, 1975).

Modelling of Other Characteristics of Flood

Very few attempts have been made in the past for modelling and analysis of flood characteristics other than peak discharge and for multivariate modelling of flood characteristics.

The use of bivariate distribution has been considered adequate for modelling two components of flood events at a time as the third variable among peak, volume and duration can be estimated with reasonable accuracy by assuming triangular relationship (Correia, 1987; USDA, 1957) between these variables.

Keeping in view the objective and scope of the present study, the use of bivariate normal distribution along with PDS approach appears to be quite appealing. Such approach would be mathematically less complex and at the same time, would provide a useful methodology which has not been tried earlier.

Generation of Daily Flows

The synthetically generated flows play an important role in simulation studies by providing alternate samples. The synthetically generated data are also useful in comparing the performance of various models and for validation of new techniques proposed for various aspects of flood frequency analysis, where daily flows are used for obtaining flood series, as is the case with large basins.

For daily flow generation in case of perennial rivers, only two models viz. linear autoregressive and shot noise models are suitable. The review has not indicated any evidence wherein the performance of these two models has been compared explicitly on the basis of flood related characteristics in the generated daily flow data.

It would be worth while to consider such a composition also as a part of the present study. The conclusions drawn from literature review were taken into consideration in formulation of the methodology for multivariate modelling of flood flows, which is presented in the next chapter.

CHAPTER III

DEVELOPMENT OF THE METHODOLOGY

3.1 General

The methodology for stochastic modelling of flood flows is presented in this chapter. The methodology is based on partial duration series approach and involves use of bivariate normal distribution as the parent bivariate distribution function for two dependent variables such as flood volumes and flood peaks. The basic considerations for defining a flood event and its various components are described in section 3.2. The bivariate normal distribution assumes normal marginal distributions which is usually not the case with data series of flood peaks, volumes and durations. Hence normalization of these variables is required before applying bivariate normal distribution. The two step power transformation (TSPT) as per the details given in section 3.3 has been adopted for normalization. The bivariate normal distribution function and further derivations required for its suitable and meaningful application to various components of flood flows are presented in section 3.4. Finally, the steps of the methodology are described in section 3.5.

3.2 Basic Considerations

The PDS approach which considers the flood discharges above a particular threshold, has been used throughout the development of the methodology. The flood features considered in the analysis are described as follows:

3.2.1 Main Features

For defining any flood event, the following flood features have been considered; viz. (i) Flood peak (FP), (ii) Flood duration (FD), (iii) Flood volume (FV) and (iv) Time of occurrence of flood peak (T_p). These are described as follows:

(i) Flood Peak (FP) : The flood peak (FP) is defined as the highest discharge above a particular threshold (Q_b) in a flood event.

For example in Fig. 3.1 (1), (2),.......(n-1) and n etc. are flood events. Flood event (1) starts at time T_{b1} and ends at time T_{e1}. Similarly flood event (n) is starting at time T_{bn} and ending at time T_{en}. These flood events are having highest discharges at times T_{p1}, T_{p2},........and T_{pn} etc. The flood peaks for these events are FP_1, FP_2,........and FP_n as shown in Fig. 3.1. If the discharges at T_{p1}, T_{p2},.......,T_{pn} etc. are Q_{tp1}, Q_{tp2},.......,Q_{tpn}, then flood peak for the n^{Th} event is defined as

$$FP_n = Q_{tpn} - Q_b \tag{3.1}$$

(ii) Flood Duration : The flood duration (FD) is defined as the duration for which a particular flood event remains above the threshold (Q_b). In Fig. 3.1 FD_1, FD_2,....,FD_n etc. are durations of flood events (1), (2),....,(n) respectively. Mathematically,

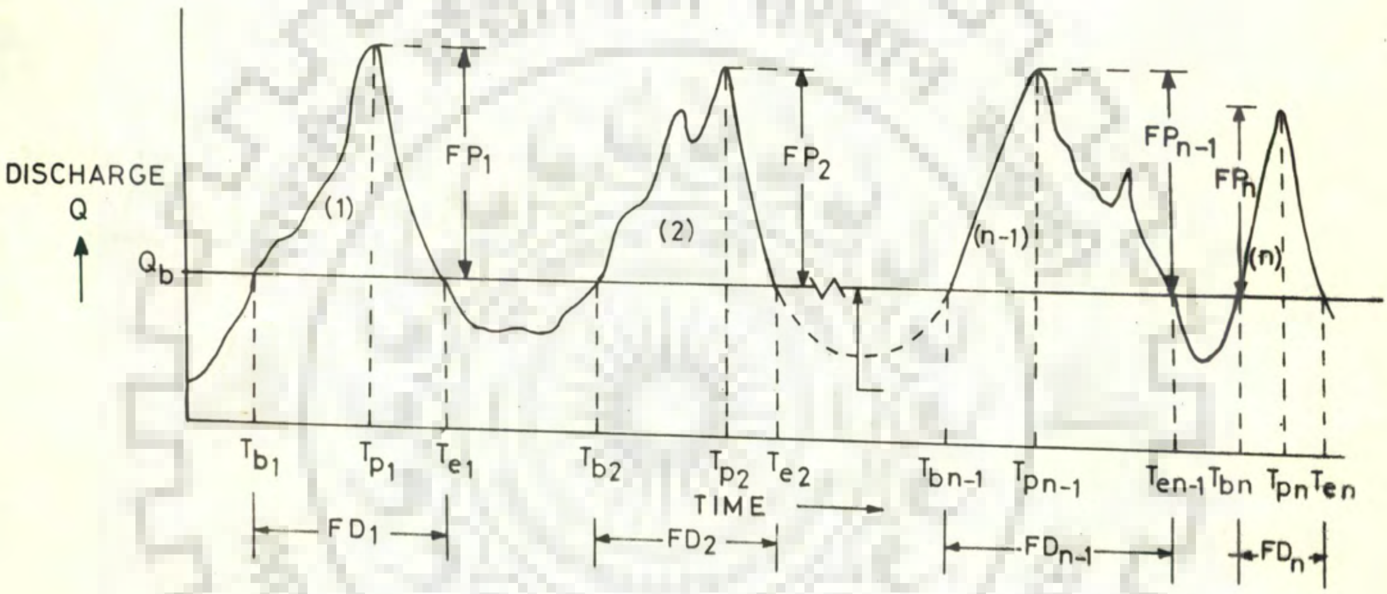

Fig. 3.1 Streamflow hydrograph showing various flood features

flood duration for the n^{Th} event FD_n is defined as

$$FD_n = T_{en} - T_{bn} \tag{3.2}$$

(iii) Flood Volume (FV) : The flood volume (FV) is the volume of water above a particular threshold (Q_b) for the event. Mathematically, the flood volume for the n^{Th} event, FV_n, is defined as

$$FV_n = \int_{T_{bn}}^{T_{en}} (Q_t - Q_b)dt \tag{3.3}$$

(iv) Time of Occurrence of Flood Peak (T_p) : As mentioned earlier, the time of occurrence of flood peak of n^{Th} event is T_{pn} and has been shown in Fig. 3.1.

3.2.2 Other Features

Considering the fact that there may be more than one flood events in a year, the following five features are also important.

(i) total number of flood events (k) in a year

(ii) largest flood peak in a year (FP_{lar})

(iii) largest flood volume in a year (FV_{lar})

(iv) largest flood duration in a year (FD_{lar})

(v) time of occurrence of largest flood peak (τ_{lar})

The subroutine DVPA has been developed by the author in FORTRAN IV for this study. It gives, for a given threshold value, number of flood events, corresponding flood peaks, volumes, durations and timing of flood peaks for different years. The subroutine is given in Appendix - I. Other important subroutines developed in the present study are also given in Appendix - I. The purpose of each subroutine and necessary instructions to use these subroutines are described in the subroutine in the form of comment statements.

The analysis and distribution functions of various variables, mentioned in this section, i.e., (i) number of flood events in a year, (ii) timing of flood peaks, (iii) flood peaks, volumes and durations (iv) largest flood peaks, volumes and durations in different years and (v) time of occurrence of largest flood peaks, for Narmada at Garudeshwar, are presented in chapter IV for various threshold levels.

In the next section the details of two step power transformation (TSPT) are presented. TSPT has been used for normalization of flood variables before applying bivariate normal distribution, described in section 3.4.

3.3 Two Step Power Transformation (TSPT)

Box and Cox (1964) suggested the following family of transformation for normality

$$y_i = \frac{x_i^{\lambda} - 1}{\lambda} \qquad \text{for} \quad \lambda \neq 0$$

$$y_i = \log x_i \qquad \text{for} \quad \lambda = 0 \qquad (3.4)$$

Here,

x_i = the variates of a given series,

y_i = the transformed normal variates and

λ = a constant of transformation.

Eq. (3.4) is a more general power transformation and the logarithmic, reciprocal and square root transformation are special cases of this. The constant λ is nonlinear and cannot be determined in a closed form. However it can be estimated by assigning various values to λ and choosing a value that makes the coefficient of skewness nearly zero. An increase or decrease in λ follows an increase or decrease in the coefficient of skewness. This trend is helpful in estimating the value of λ. Alternatively, Newton - Raphson method can be used to estimate the value of λ. The subroutine POWER1 developed for Box - Cox transformation uses Newton - Raphson method and is given in Appendix I.

The Box - Cox transformation though reduces the coefficient of skewness (C_s) to zero, yet is unable to make the coefficient of kurtosis (C_k) equal to 3 which leads to underestimation or overestimation of flood quantiles depending upon whether C_k is more than 3 or less than 3. The correction for C_k is envisaged through modulus transformation in two step approach as follows

$$Z_i = (|y_i - \bar{y}|)^{\nu} \qquad (3.5)$$

Where, ν is positive and Z_i is having same sign as $(y_i - \bar{y})$. The modulus transformation given by Eq. (3.5) depending upon the value of ν, equally stretches or compacts the transformed histogram of y_i obtained by Box -Cox transformation. Similar transformation has also been suggested by John and Draper (1980) and

Gupta et al. (1989).

With suitable value of γ, the C_k of Z series can be made equal to 3. In the modulus transformation it is evident that when $\gamma \to 0$, the C_k of Z series tends to be one and for $\gamma \to \infty$, C_k tends to be ∞. For $\gamma = 1$ the C_k of Z series will be same as that of y series obtained after Box -Cox transformation. Therefore, if C_k of y series is more than 3, γ will be between one and zero and for $C_k < 3$, γ will be more than one.

With the help of simple iterative algorithm the values of λ and γ can be obtained which will make C_s and C_k of x series as zero and 3 respectively.

The value of x for a given return period (T) can be obtained using back transformation as follows

$$Z_T = \bar{Z} + K_T \sigma_z \tag{3.6}$$

$$y'_T = (|Z_T|)^{1/\gamma} \tag{3.7}$$

Here y'_T has the sign of Z_T.

$$y_T = y'_T + \bar{y} \tag{3.8}$$

and

$$x_T = (y_T \lambda + 1)^{1/\lambda} \tag{3.9}$$

In Eq. (3.6) to (3.9),

x_T = variate x corresponding to T year return period

$\bar{Z}$ = mean of Z series

σ_z = standard deviation of Z series

K_T = normal reduced variate corresponding to prob. Of exceedance equal to 1/T.

The subroutine POWER2 gives the value of λ and γ for a given series. The subroutine requires the initial value of λ which is obtained from subroutine POWER1.

3.4 Bivariate Normal Distribution

The bivariate normal distribution, though quite old in statistics, has not been applied to hydrological problems quite extensively. In the foregoing text, the bivariate normal distribution function and further derivations required for its suitable application to various components of flood flows are presented.

3.4.1 The distribution

The bivariate normal density function for a bivariate (x,y) e.g. normalized flood volumes and flood peaks, with sample correlation coefficient r between x and y is given by

$$h(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\exp[-\frac{1}{2(1-r^2)}\{\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-2r\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\}] \tag{3.10}$$

where μ_x and σ_x are the mean and standard deviation of x and μ_y and σ_y of y. (Yevjevich, 1972).

The marginal p.d.f. of x can be obtained by integrating Eq. (3.10) with respect to y from $-\infty$ to $+\infty$ as

$$h(x) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{-(x-\mu_x)^2/(2\sigma_x^2)} \tag{3.11}$$

The marginal p.d.f. Of y can be obtained by integrating (3.7) with respect to x from $-\infty$ to $+\infty$ as

$$h(y) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{-(y-\mu_y)^2/(2\sigma_y^2)} \tag{3.12}$$

3.4.2 Further derivations for application to flood flows

Following the work of Singh and Krstanovic (1987) the conditional p.d.f. of y due to the knowledge of x can be expressed as

$$h(y|x) = h(x,y)/h(x) \tag{3.13}$$

Substituting $r = \sigma_{xy}/\sigma_x\sigma_y$ in (3.7), h(x,y) can be written as

$$h(x,y) = \frac{1}{2\pi(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)^{1/2}}$$

$$\exp\left\{-\frac{1}{2(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)}[\sigma_x^2(y-\mu_y)^2 + \sigma_y^2(x-\mu_x)^2 - 2\sigma_{xy}(x-\mu_x)(y-\mu_y)]\right\} \tag{3.14}$$

Substitution of (3.11) and (3.14) in (3.13) gives

$$h(y|x) = \frac{\sigma_x}{(2\pi(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2))^{1/2}}$$

$$\exp\left\{-\frac{\sigma_x^2(y-\mu_y)^2 - 2\sigma_{xy}(x-\mu_x)(y-\mu_y) + (x-\mu_x)^2(\sigma_{xy}^2/\sigma_x^2)}{2(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)}\right\} \tag{3.15}$$

The cumulative distribution function H(y|x) can be obtained by integrating with respect to y

$$H(y|x) = \frac{\sigma_x}{(2\pi(\sigma_x^2\sigma_y^2-\sigma_{xy}^2))^{1/2}}$$

$$\int_{-\infty}^{y} \exp\left\{-\frac{\sigma_x^2}{2(\sigma_x^2\sigma_y^2-\sigma_{xy}^2)}\left[(y-\mu_y)-\left(\frac{\sigma_{xy}}{\sigma_x^2}(x-\mu_x)\right)\right]^2\right\}dy \tag{3.16}$$

Making the following substitutions

$$\frac{\sigma_x}{(2(\sigma_x^2\sigma_y^2-\sigma_{xy}^2))^{1/2}} = c_1$$

$$c_1\left((y-\mu_y)-\left(\frac{\sigma_{xy}}{\sigma_x^2}(x-\mu_x)\right)\right) = Z/\sqrt{2}$$

and

$$dy = dz/c_1\sqrt{2}$$

$H(y|x)$ can be written as

$$H(y|x) = \int_{-\infty}^{Z} e^{-Z^2/2}dZ \tag{3.17}$$

which is the integral form of the standard normal distribution.

The joint distribution of X and Y can be obtained as

$$H(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} h(y|x)h(x)dxdy \tag{3.18}$$

Following the work of Zelenhasic (1970) and Correia (1987) the joint distribution function of X and Y when the time interval of one year is considered may be given by

$$F(x,y) = \sum_{k=0} H(x,y)^k . P(E_k) \tag{3.19}$$

where P(Ek) is the probability that k flood events occur in one year. F(x,y) represent the probability that in one year x and y are both not exceeded i.e.

$$F(x,y) = P(X \le x, Y \le y) \tag{3.20}$$

If the number of flood events per year is Poisson distributed, Eq. (3.19) may be written as

$$F(x,y) = \sum_{k=0} H(x,y)^k \frac{e^{-\lambda}\lambda^k}{k!} \tag{3.21}$$

$$= \frac{H(x,y)^0 e^{-\lambda}\lambda^0}{0!} + \frac{H(x,y)e^{-\lambda}\lambda}{1!} + \frac{H(x,y)^2 e^{-\lambda}\lambda^2}{2!} + \text{--------}$$

$$= e^{-\lambda}\left[1 + \frac{H(x,y)\lambda}{1!} + \frac{H(x,y)\lambda^2}{2!} + \text{--------}\right]$$

$$= e^{-\lambda} e^{H(x,y).\lambda}$$

or

$$F(x,y) = e^{\lambda H(x,y) - \lambda} \tag{3.22}$$

The probability that at least one of the values (x,y) is exceeded in one year is equal to 1 - F(x,y) and the return period associated with this event is

$$T(x,y) = 1./(1 - F(x,y)) \tag{3.23}$$

The return period associated with the exceedance of both x and y in one year is given by

$$T'(x,y) = 1./(1 - F_M(x) - F_M(y) + F(x,y)) \quad (3.24)$$

In the above expression $F_M(x)$ and $F_M(y)$ represent the marginal distributions of x and y being exceeded in one year. These marginal distributions can be computed by

$$F_M(x) = e^{\lambda H(x) - \lambda} \quad (3.25)$$

$$F_M(y) = e^{\lambda H(y) - \lambda} \quad (3.26)$$

where,

$$H(x) = \int_{-\infty}^{x} h(x)dx$$

or

$$H(x) = \int_{-\infty}^{x} \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \quad (3.27)$$

$$H(y) = \int_{-\infty}^{y} h(y)dy$$

or

$$H(y) = \int_{-\infty}^{y} \frac{1}{\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \quad (3.28)$$

The return period associated with the exceedance of x alone is given by

$$T(x) = 1. / (1 - F_M(x)) \quad (3.29)$$

The return period associated with the exceedance of y alone is given by

$$T(y) = 1. / (1. - F_M(y)) \tag{3.30}$$

The return period associated with the exceedance of y given x i.e. $(Y \geq y|X=x)$ is given by

$$T(y|x) = 1./(1.-F(y|x)) \tag{3.31}$$

where,

$$F(y|x) = e^{\lambda H(y|x)-\lambda} \tag{3.32}$$

3.4.3 Solution techniques

For the solution of integral of standard normal distribution in Eq. (3.17), (3.27) and (3.28) subroutine NDTR from Scientific Subroutine Package of IBM was tested and used. The subroutine has a maximum error of 7×10^{-7} and is based on the approximation of Normal distribution function given by Hastings (1955).

The integral for joint distribution of X and Y given in Eq. (3.18) has been solved using subroutine QG32 in which evaluation is done by means of 32 point Gauss Quadrature formula.

3.5 The Methodology

Based on the concepts given in earlier sections, the proposed methodology for stochastic modelling of flood flows consists of the following sequential steps.

1. Processing of daily discharge data for river site under consideration and selection of appropriate threshold level. The selection of threshold level requires that Poissonian assumption for number of flood events in a year is not violated.
2. Identification of flood events and their components above

selected threshold level using subroutine DVPA.

3. Normalization of flood volumes and peaks using TSPT (subroutine POWER2). The subroutine POWER2 will give the values of λ and γ for flood volume and flood peak series.

4. Use of bivariate normal distribution to estimate return periods of various events given by equations (3.23), (3.24), (3.29), (3.30) and (3.31).

The subroutine NKG was developed in FORTRAN IV. This subroutine performs step 3 and 4 for a given threshold level and is capable of handling any number of flood volume and flood peak values.

The methodology developed in this chapter has been validated using long term synthetically generated data in Chapter VI along with its application to Narmada at Garudeshwar. The details of data used and its preliminary analysis are presented in next chapter and the details of daily flow generation in Chapter V.

CHAPTER IV

DATA USED AND PRELIMINARY ANALYSIS

4.1 General

The chapter gives general description of Narmada river basin, earlier studies of flood estimation for Sardar Sarovar dam, details of data used, its processing and preliminary analysis. The preliminary analysis of data includes conventional flood frequency analysis of annual flood series (AFS) and analysis of various flood features for different threshold levels.

4.2 The Narmada Basin

General description of Narmada river basin and earlier studies of flood estimation for Sardar Sarovar dam are presented in this section, as follows:

4.2.1 General description

The River:

Narmada river is the fifth largest river of India and is known as the 'life line' of Madhya Pradesh. The Narmada river rises in the Amarkantak plateau of Maikala range in the Shahdol district of Madhya Pradesh at an elevation of 1057 m. above sea level. The river travels a distance of 1312 km before it falls into gulf of Cambay in the Arabian sea near Bharauch in Gujarat. It runs for about 1100 km in the state of Madhya Pradesh and rest in the states of Maharashtra and Gujarat.

<u>The Basin:</u>

The Narmada basin extends over an area of 98796 sq. km. and lies between longitudes 72° 32′E to 81° 45′E and latitudes 21° 20′N to 23° 45′N. The catchment area up to Narmada Sagar dam (under construction) is 61642 sq. km. and the area up to Sardar Sarovar dam (also under construction) is 88000 sq. km. The basin is bounded by Vindhyas in the north, by Maikala range in the east, Satpuras in the south and by the Arabian sea in the west. Most of the basin is at an elevation of less than 500 m above sea level. The index map of the basin is given in Fig. 4.1 with map of India in inset.

<u>The Climate:</u>

The climate of the basin is humid tropical ranging from sub humid in the east to semi- arid in the west with pockets of humid climates around higher hill reaches. The normal annual rainfall for the basin is 1178 mm. South west monsoon is the principal rainy season accounting for nearly 90% of the annual rainfall. About 60% of the annual rainfall is received during July and August months.

<u>Soils:</u>

The soils in Narmada basin are mainly black soils. The different varieties are deep, medium and shallow black soils. In addition to this, mixed red and black soil, red and yellow soil and skeletal soils are also observed in pockets. Of these deep black soil covers the major portion of the basin.

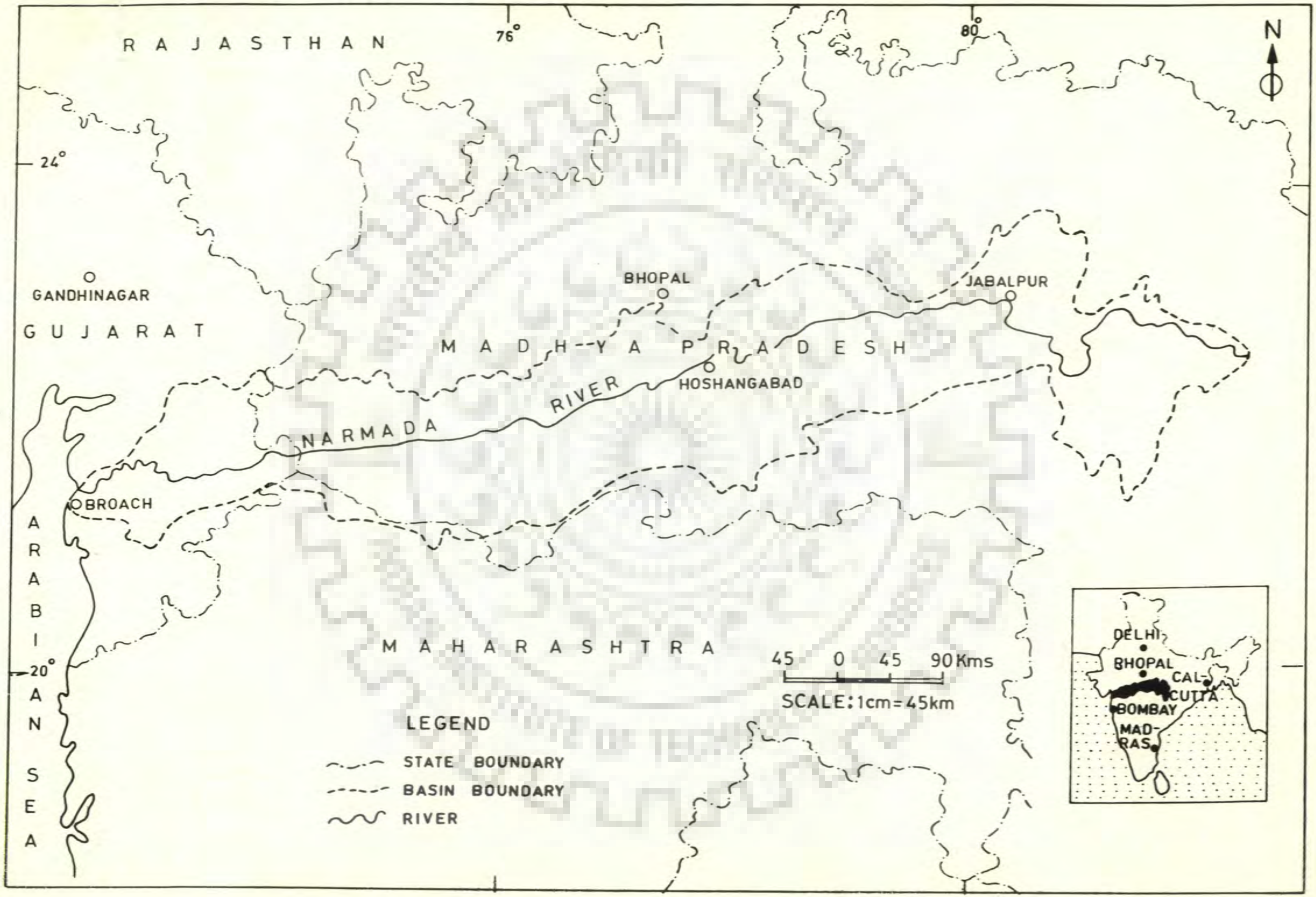

Fig. 4·1–Index map of Narmada river basin

Land Use:

About 35% of the area is under forest, about 60% under arable land and remaining under grass land, waste land etc.

4.2.2 Earlier studies of flood estimation

A number of attempts have been made by several agencies to estimate the design flood for Sardar Sarovar Dam. Various aspects of Sardar Sarovar dam project have been documented in Sardar Sarovar Narmada Nigam (1989). Some of the relevant studies of design flood estimation for this dam are briefly presented as follows:

The Central Water Power Commission (CWPC) carried out flood studies for Broach Irrigation Project in 1952 utilizing the Garudeshwar observed data for 3 years (1948-51). By adopting unit hydrograph method and design storm of 100 years frequency, peak of design flood was estimated as 57600 cumecs. Based on the use of empirical formula and this study, a peak of 58000 cumecs was recommended. In 1960, a Board of Consultants appointed to review the design flood at Navagam dam site recommended an increase by 10% which gave a design peak flood of 63000 cumecs. In 1961, IMD carried out fresh flood studies on the basis of up to date hydrograph and storm data. Considering July 1944 storm and peak flood discharges of 1949, 1950, 1959, 1961 and 1962 design peak flood arrived at was 64600 cumecs. In 1959, Government of Gujarat carried out flood frequency analysis with return periods of 1 in 1000 years and this gave a flood peak estimate of 46200 cumecs.

Gauge data at Broach railway bridge and observed gauge and discharge data at Garudeshwar were utilized in the study.

After 1962, there was a succession of unprecedented floods in the river Narmada. This set in process further rethinking on design floods. In 1974, CWPC carried out a flood frequency study for the project considering the observed gauge and discharge data up to 1973. The design flood having a return period of 1000 years worked out to be 86900 cumecs. After incorporating data up to 1977 design flood value of 85600 cumecs was estimated. The Sardar Sarovar Project prepared in Jan. 1980, therefore considered a flood with 1000 years return period as 86900 cumecs.

In view of the very large size of the catchment, variations in storm rainfalls and physical parameters of the basin, and the need to ensure safety of the dam under all possible adverse flood conditions, comprehensive assessment of design floods with the latest available approaches was considered urgent and essential. Accordingly rational flood hydrology studies for the basin were set in motion in consultation with Govt. of India at the Central Water and Power Research Station, Pune (CWPRS) and National Institute of Hydrology, Roorkee (NIH) in addition to the comprehensive studies taken up by the engineers of Govt. of Gujarat. These studies aimed at estimation of 1 in 1000 years flood (project design flood), 1 in 10000 years flood and PMF under most likely adverse hydro-meteorological conditions. All these studies had a common data base. The Dam Design Review Panel (DDRP) set up by Govt. Of Gujarat also reviewed the studies by various agencies. The CWPRS has carried out studies with deterministic approach using (i) OPSET model (Modified Stanford Watershed Model) and (ii) SSARR model. Statistical approach has also been used with various distributions such as EV1, LP III and PT III etc. NIH carried out studies using frequency analysis approach and deterministic approach using HEC 1 model by dividing the whole basin into 20 sub basins.

The PMF values estimated by various agencies ranged from 107600 cumecs to 221000 cumecs considering a moisture maximization factor of 1.35. The value of peak flood arrived at by NIH for 10000 years return period was 127500 cumecs.

These estimates indicate the range covered in design flood values in earlier studies and thus provide a rough guideline for selection of flood peak magnitudes for the present study.

4.3 Data Used For the Study:

4.3.1 The data

Daily discharge data of river Narmada at Garudeshwar site (downstream of Sardar Sarovar dam) from 1949-79 have been used in this study. The Garudeshwar site lies at a latitude of 21° 50′ N and longitude of 73° 58′ E. The zero of gauge is at 10.0 m above sea level. The cross section of river Narmada at Garudeshwar is shown in Fig. 4.2. The stage discharge relationship for the site was developed by NIH (1985). In developing this relationship the physical features like cross section, bed slope etc. were also taken into consideration. The relationship is given below:

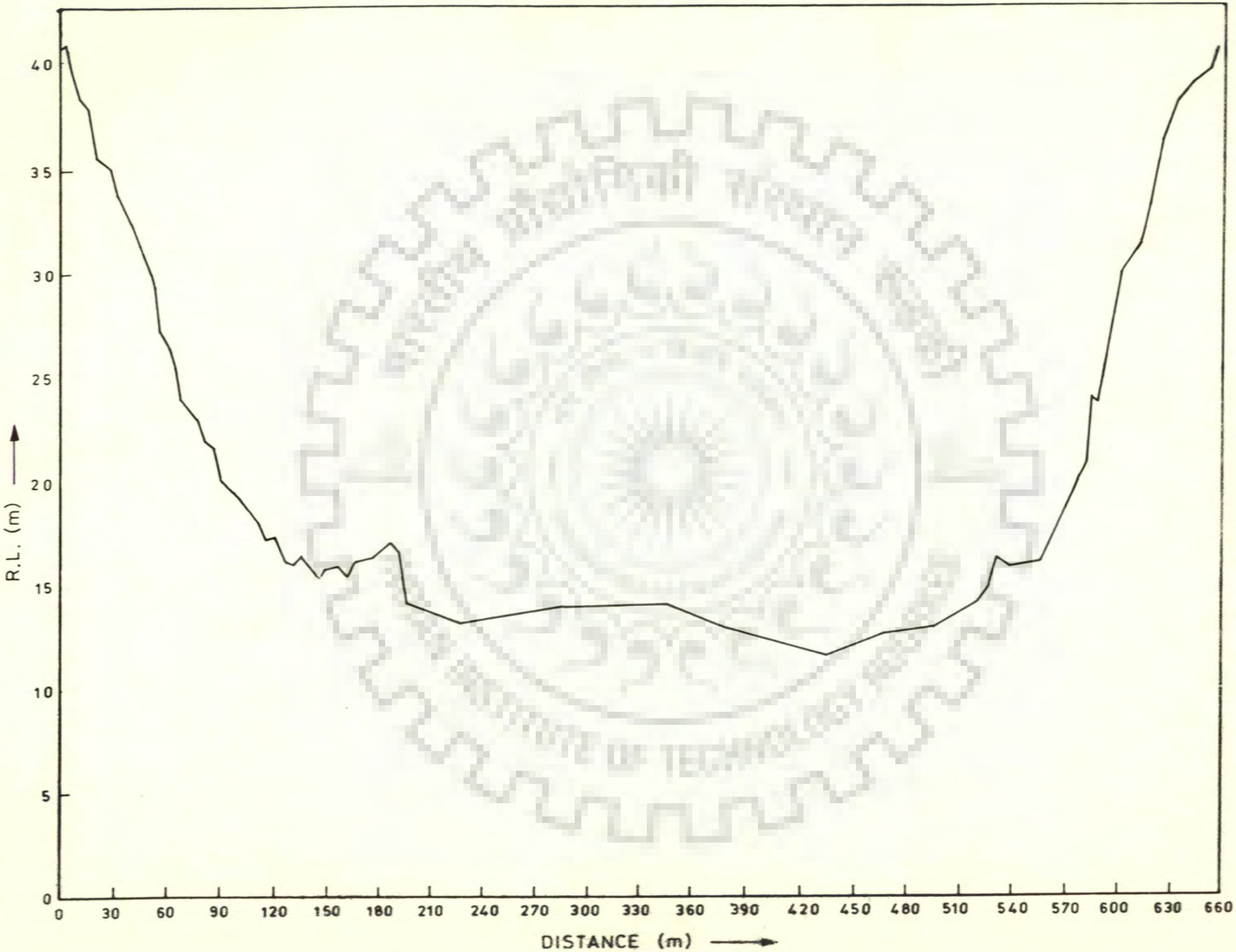

Fig. 4·2 – Cross section of river Narmada at Garudeshwar

$$Q = 250.0(H-11.7)^{1.66} \tag{4.1}$$

In Eq. (4.1), Q is discharge in cumecs and H is stage in meters above sea level.

4.3.2 Processing of daily flow data :

The regular gauging in the river at Garudeshwar started from July 1948. The daily discharge data compiled by Narmada Project Dam Design Circle, Gujarat, were available in manuscript form. The data was thoroughly checked for any inconsistent value, wrong entries and gaps. The discharge values were keyed in and checked through software developed by the author for any inconsistent value and punching error. This software included a simple subroutine to check number of values in a month, number of values in a year, conversion of unit etc. and a plotting subroutine to plot the discharge data. The gaps in the data were few in number and mostly in the lower range of flows. These were appropriately filled up by considering the trend of the rest of the data, evening and noon observations and hourly gauge data (wherever available). The processed data thus obtained was used for further analysis.

4.4 Preliminary Analysis:

The preliminary analysis using daily discharge data from 1949-79 for river Narmada at Garudeshwar, includes conventional flood frequency analysis of AFS and analysis of (i) number of flood events in a year, (ii) timing of flood peaks, (iii) flood peaks, volumes and durations, (iv) largest flood peaks, volumes and durations and (v) timing of largest flood peaks for different threshold levels. The analysis of above variables mainly include the study of distribution function of these variables. For testing the goodness of fit chi- square test at 5% significance level has been used. The details of this analysis are as follows:

4.4.1 Conventional flood frequency analysis using AFS:

Conventional flood frequency analysis with an objective to estimate various flood quantiles has been carried out. As pointed out in review of literature, still there is lot of controversy amongst hydrologists regarding the choice of a particular distribution and method of parameter estimation. It is futile and inadvisable to try all the distributions and parameter estimation techniques. Hence only some selected distributions and parameter estimation techniques have been used for flood estimation. These are listed below along with reasons for selection.

(i) Wakeby by PWM (recommended by Wallis and Wood, 1985 based on simulation studies and by Houghton, 1978a, on the basis of regional studies).

(ii) GEV by PWM (recommended by Hosking et al., 1985 on the basis of simulation results and NERC, 1975 on the basis of regional studies).

(iii) EV1 by PWM (Central Water Commission (India), 1972 recommended the use of EV1 distribution for Indian rivers; PWM was shown to be the robust method of parameter estimation for EV1 distribution by Goel and Seth, 1988)

(iv) LP III by method of moments (recommended by USWRC, 1967; 1977; 1981; and by Mc Mahon and Srikanthan, 1981)

The annual flood series (AFS) for 31 years was obtained by considering maximum daily discharge in a year as the annual flood peak.

Flood magnitudes of 2, 5, 10, 20, 50, 100, 200, 500, 1000 and 10000 years return period have been estimated by fitting each of the above distributions to the AFS. Both the forms (4 parameter as well as 5 parameter) of Wakeby distribution have been tried. The performance of these distributions has been judged on the basis of following three criteria:

(i) Average of relative deviation between observed and computed values of events (ADA),

(ii) Average of squares of relative deviation between observed and computed values of events (ADR) and

(iii) Efficiency (EFFI) of the distribution.

ADA, ADR and EFFI have been computed using following relationships:

$$ADA = \sum_{i=1}^{N} \left| (Q_{i,obs} - Q_{i,comp})/Q_{i,obs} \right| \times 100/N \qquad (4.2)$$

$$ADR = \sum_{i=1}^{N} ((Q_{i,obs} - Q_{i,comp})/Q_{i,obs})^2 \times 100/N \qquad (4.3)$$

EFFI = (model variance/initial variance) x 100

$$= \frac{\sum_{i=1}^{N}(Q_{i,obs}-\bar{Q})^2 - \sum_{i=1}^{N}(Q_{i,obs}-Q_{i,comp})^2}{\sum(Q_{i,obs}-\bar{Q})^2} \times 100 \qquad (4.4)$$

In the above equations N is the no. of values, $Q_{i,obs}$ is the i^{th} value of observed discharge in AFS, $Q_{i,comp}$ is the i^{th} value of computed discharge in AFS and $\bar{Q}$ is mean of AFS.

Based on the above criteria floods of different magnitudes have been finally selected. The steps and results of analysis of 31 years AFS are as follows:

(i) The AFS (1949-79) was tested for randomness using Anderson's correlogram test (Anderson, 1941) and turning point test (Kottegoda, 1980) and was found to be random at 5% significance level. The AFS is plotted in Fig. 4.3. The presence of any rising or falling trend was checked using Kendall's rank correlation test (Kottegoda, 1980) and the series was found to be trend free.

(ii)The statistical parameters of AFS in original and log domain are given in Table 4.1.

Table 4.1

Statistical parameters of AFS for Narmada at Garudeshwar

(1949-49)

Max.	Min.	Mean	S.D.	C_s	C_k	r_1
Original series						
63800.000	10131.767	27790.760	13788.216	1.042	3.819	0.111
Log Transformed series						
11.064	9.223	10.121	0.479	0.154	2.502	0.158

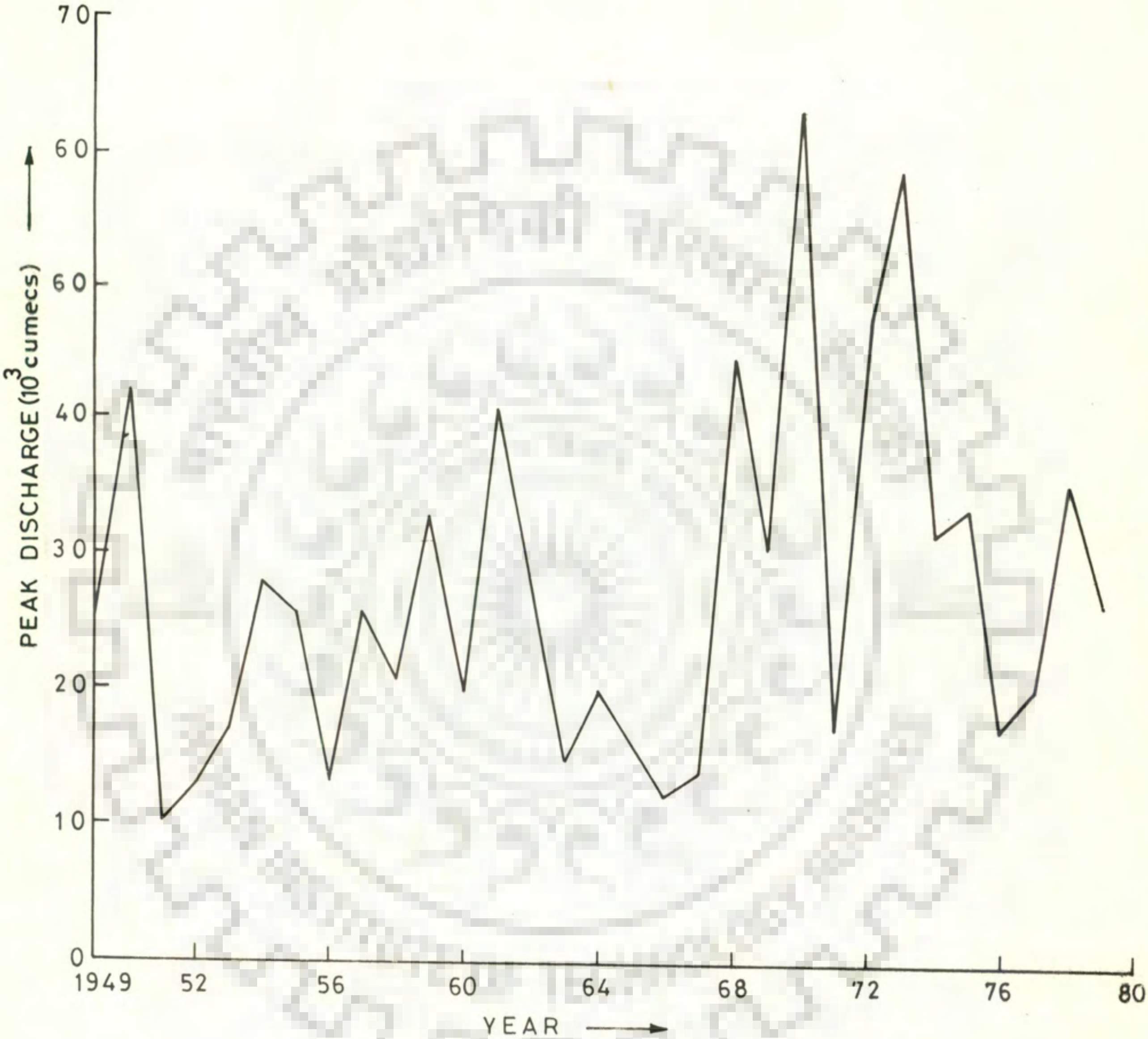

Fig. 4.3 - Plot of annual flood series (1949 - 79)

(iii) The probability weighted moments M_{1j0} which are used to estimate the parameters of Wak-4, Wak-5, GEV and EV1 distributions are computed by

$$M_{1j0} = E(X.F^{j}.(1-F)^{0}) \tag{4.5}$$

where,

F = prob. of non exceedance

= (m-0.35)/N (4.6)

m = rank of data if arranged in ascending order.

M_{100}, M_{110}, M_{120}, M_{130}, and M_{140} for the AFS are 27790.761, 17718.111, 13395.817, 10906.89 and 9264.504 respectively.

(iv)The parameters of Wak-4, Wak-5 ,GEV and EV1 distributions using PWMs and procedure given by Greenwood et al. (1979) and Hosking et al. (1984) are given in Table 4.2. In Table 4.2 u, α and k are location, scale and shape parameters respectively.

Table 4.2

Parameters of Wak-4, Wak-5, GEV and EV1 distributions

Distr.	Parameters
Wak-4	A=12084.824, B=56.932 C =-131304.420, D =-0.138
Wak-5	A= 5401.526, B=3.395, C =-307795.930, D =-0.048,
GEV	M=9544.909
EV1	u=20901.489, α =9846.809, k =-0.111
	u=21424.096, α =11030.068

(v) Using the parameters given in Tables 4.1 and 4.2 flood magnitudes for various return periods for Wak-4, Wak-5, GEV, EV1 and LP III distributions were computed and are given in Table 4.3.
(vi) ADA, ADR and EFFI values as given by Eq. (4.2) to (4.5) for various distributions are given in Table 4.4.

It can be seen from Table 4.4 that for Wak-4 distribution ADR is minimum and EFFI is maximum. The ADA is minimum in the case of Wak-5 but for Wak-4 also it is comparable. Hence on the basis of criteria of ADA, ADR and EFFI the flood quantiles given by Wakeby-4 distribution could be selected. These flood quantiles are given in Table 4.5.

4.4.2 Distribution of the number of flood events

The distribution of the number of flood events in a year plays an important role as the methodology developed in Chapter III considers simultaneous occurrence of flood events and their numbers.

The number of flood events in a year depends upon the threshold level, Q_b. The Poisson distribution assumption for number of flood events was tested for different levels of threshold based on chi- square test. For Poisson distribution the probability that exactly k flood events occur in a year i.e. $P(E_k)$ is given by:

$$P(E_k) = \lambda^k e^{-\lambda}/k! \tag{4.7}$$

Where, E_k is the event that exactly k flood events occur in a

Table 4.3

T years return period floods for Wak-4, Wak-5, GEV, EV1 and LP III distributions

Return period (T)	Q_T in cumecs for Wak-4	Wak-5	GEV	EV1	LP III
2	24056	24487	24585	25467	24564
5	38223	37767	36971	37969	37057
10	47811	47096	46073	46246	46258
20	56525	56102	55545	54186	55755
50	66837	67555	68984	64463	69064
100	73816	75891	80006	72164	79843
200	80159	83955	91871	79837	91324
500	87666	94211	108993	89961	107699
1000	92746	101677	123142	97612	121074
10000	106525	124768	178751	123014	172468

Table 4.4

ADA, ADR and EFFI values for different distributions

Crit.	Distr. Wak-4	Wak-5	GEV	EV1	LPIII
ADA	3.919	3.843	5.401	6.810	5.091
ADR	0.282	0.287	0.524	0.985	0.365
EFFI	97.948	97.406	95.606	97.655	97.301

Table 4.5

Various return period (T) floods Q_T in cumecs

T	2	5	10	20	50	100	200	500	1000	10000
Q_T	24056	38223	47811	56525	66837	73816	80159	87666	92746	106525

year and λ is the average number of flood events per year.

The computed and critical x^2 values at 5% significance level for Poisson distribution for number of flood events are presented in Table 4.6 for various threshold levels ranging from 700 cumecs to 20000 cumecs. Such a wide range has been taken to select the limits of threshold levels in which the methodology developed would be applicable. As mentioned in Chapter II, the following two points should be kept in mind while selecting the threshold level:

(a) The threshold level should be high enough to ensure that (i) number of flood events in a year is a non increasing function of threshold and (ii) the flood events are random,

(b) The threshold level on the other hand should not be very high, otherwise the number of events will be too less.

The Table 4.6, also gives lag one serial correlation coeff. (r_1) for flood peaks and volumes along with the 95% lower and upper confidence limits (r_1 =0.0). These are based on Anderson's correlogram test and are used to test the randomness of flood events.

It can be seen from Table 4.6, that for threshold level below 3000 cumecs the number of flood events are not decreasing with increase in threshold level and the flood events are not random for some of the threshold levels. From 4000 to 18000 cumecs both the requirements are satisfied. These results confirm the findings of Ashkar and Rousselle (1983) that ' the Poisson distribution assumption as a model for flood frequency should remain so with any higher level of truncation also'. For threshold levels

Table 4.6

Computed and critical x^2 at 5% significance level for Poisson distribution for number of flood events

Thres. cumecs	Total flood events	x^2comp.	x^2cri.	r1 of flood peak	r1 of flood volume	95% upper limit	95 % lower limit
700	71	9.85	12.6	0.046	-0.159	0.218	-0.246
800	76	13.46	12.6	-0.164	-0.270	0.211	-0.238
900	79	11.13	14.1	-0.141	-0.236	0.207	-0.233
1000	89	10.67	14.1	-0.102	-0.184	0.196	-0.219
2000	130	9.07	16.9	-0.092	-0.062	0.164	-0.179
3000	138	7.36	18.3	-0.184	-0.191	0.159	-0.174
4000	134	8.57	18.3	-0.148	-0.067	0.161	-0.176
5000	133	7.08	16.9	-0.081	-0.047	0.162	-0.177
6000	121	8.07	16.9	-0.007	0.052	0.169	-0.186
7000	117	3.72	16.9	-0.025	-0.009	0.172	-0.189
8000	107	5.75	15.5	-0.062	-0.019	0.180	-0.198
9000	100	11.89	15.5	-0.059	-0.016	0.185	-0.206
10000	91	2.05	14.1	0.007	0.051	0.194	-0.216
11000	76	7.32	12.6	-0.042	0.043	0.211	-0.238
12000	70	8.44	12.6	-0.090	0.017	0.219	-0.248
13000	62	11.87	12.6	-0.132	-0.025	0.232	-0.265
14000	56	4.63	11.1	-0.184	-0.045	0.243	-0.280
15000	50	8.87	11.1	-0.211	-0.071	0.256	-0.297
16000	47	7.79	11.1	-0.226	-0.093	0.264	-0.307
17000	43	7.60	9.5	0.027	0.098	0.275	-0.322
18000	37	5.51	9.5	-0.007	0.080	0.294	-0.349
19000	30	11.19	9.5	-0.075	0.022	0.323	-0.392
20000	28	14.81	9.5	0.021	0.173	0.333	-0.407

above 18000 cumecs, the number of flood events are even lesser than the length of the record. Above this threshold the Poissonian assumption is also violated based on chi- square test.

The computation of χ^2 for four typical threshold levels of 6000 cumecs (λ =3.903), 10000 cumecs (λ =2.935), 13000 cumecs (λ =2.0) and 18000 cumecs (λ =1.193) are shown in Table 4.7. The observed and corresponding theoretical (Poissonian) distribution for number of flood events for the same threshold levels are depicted in Fig. 4.4.

4.4.3 Timing of flood peaks

For most of the Indian rivers including Narmada floods occur during five months (June to October) of monsoon season. For river Narmada, the floods are mainly concentrated in the months of July, August and September.

4.4.4 Distribution of flood peaks, volumes and durations

The determination of the distribution function of flood peaks, volumes and durations is purely a problem of estimation and according to the present state of art there are no theoretical grounds that indicate the form of the distribution (Zelenhasic, 1970). Many investigators have found that the exponential distribution which is a special case of Gamma distribution, fits well the frequency distribution of flood peaks (Zelenhasic, 1970; Todorovic and Zelenhasic, 1970; Rousselle, 1972), flood durations (Correia, 1985, 1987) and flood volumes (Krstanovic and Singh, 1987). Hence the applicability of exponential distribution for

Table 4.7

Observed and corresponding theoretical (Poissonian) distribution for number of flood events per year

k	Q_b= 6000		Q_b =10000		Q_b =13000		Q_b =18000	
	f_{ob}	f_{th}	f_{ob}	f_{th}	f_{ob}	f_{th}	f_{ob}	f_{th}
0	0	0.6225	1	1.6463	5	4.1954	11	9.3975
1	1	2.4414	6	4.8326	10	8.3908	9	11.2163
2	5	4.7646	6	7.0929	6	8.3908	8	6.6936
3	8	6.1991	8	6.9404	6	5.5939	1	2.6630
4	6	6.0492	4	5.0934	0	2.7969	1	0.7946
5	8	4.7223	4	2.9903	2	1.1188	1	0.1897
6	1	3.0720	1	1.4630	2	0.3729		
7	1	1.7130	1	0.6135	0	0.1065		
8	0	0.8358	0	0.2251				
9	1	0.3625						
10	0	0.1415						
Σ	31	30.9267	31	30.89	31	30.966	31	30.9547
		λ=3.903		λ=2.935		λ=2		λ=1.193
		$\chi^2_{comp.}$=8.07		$\chi^2_{comp.}$=2.05		$\chi^2_{comp.}$=11.87		$\chi^2_{comp.}$=5.51
		$\chi^2_{cri.}$=16.9		$\chi^2_{cri.}$=14.10		$\chi^2_{cri.}$=12.6		$\chi^2_{cri.}$=9.49

In Table 4.7 k is number flood events in a year, f_{ob} and f_{th} are observed and theoretical absolute frequency, $\chi^2_{cri.}$ values refer to the 5% significance level.

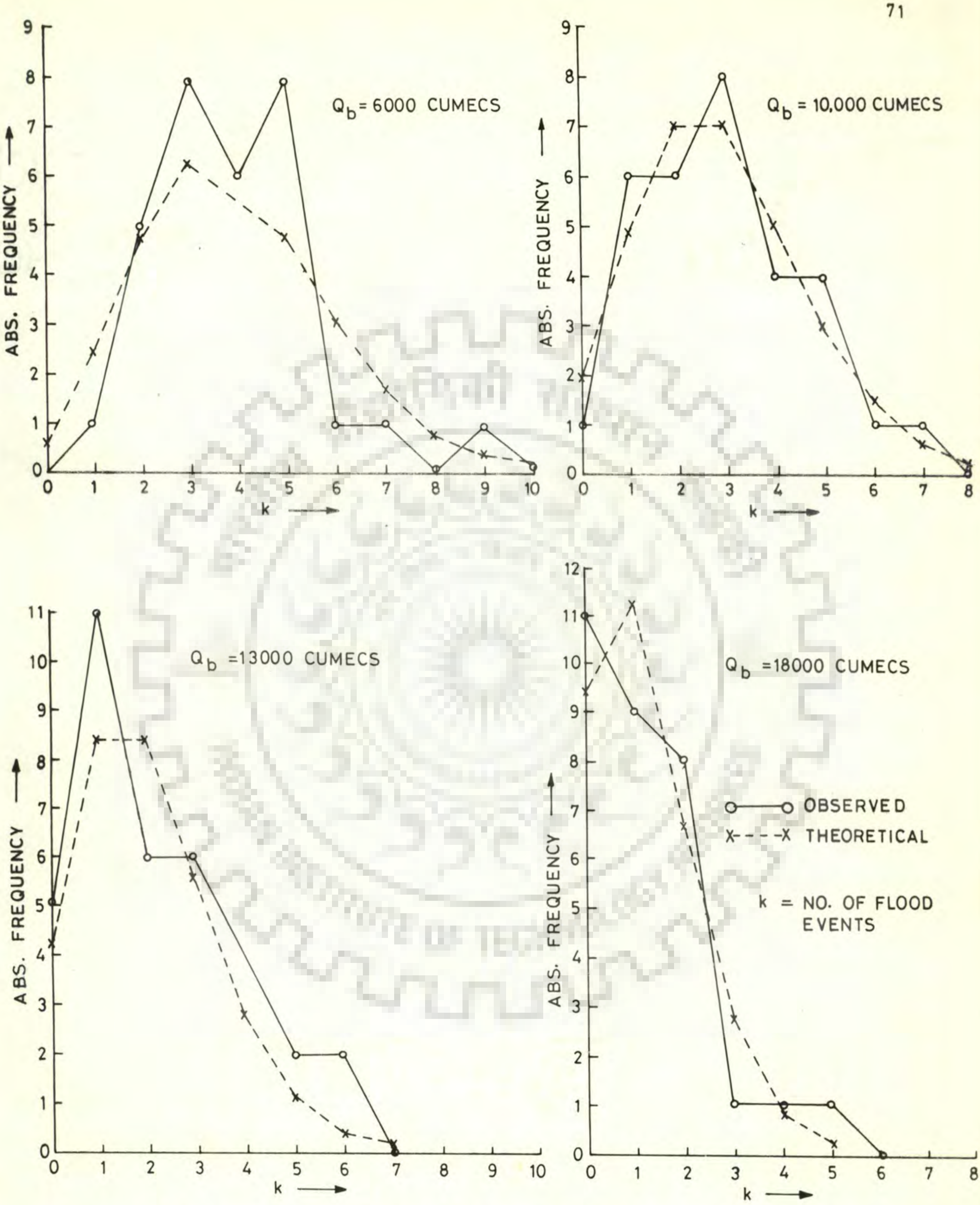

Fig. 4.4 Observed and theoretical (Poissonian) distributions for number of flood events in a year for different threshold levels (Q_b)

the flood peaks, volumes and durations was tested.

The theoretical (exponential) distribution for flood peaks, volumes and durations for a typical threshold level of 10000 cumecs are as follows:

$$H(FP) = 1. - \exp(-(FP-110.2)/10591.9) \quad (4.8)$$

$$H(FV) = 1. - \exp(-FV/19390.6) \quad (4.9)$$

$$H(FD) = 1. - \exp(-(FD-1.1)/1.7) \quad (4.10)$$

where, H(FP), H(FV), H(FD) are the functions of flood peaks (FP), volumes (FV) and durations (FD) respectively.

Similarly distribution functions for other threshold levels can be obtained using values of x_0 and β from Table 4.8.

The histogram of flood peaks, volumes and durations are plotted in Fig. 4.5 to 4.7 for a typical threshold of 10000 cumecs. The theoretical (exponential) and observed distribution functions of flood peaks, volumes and durations are plotted in Fig. 4.8 to 4.10 for the same threshold level.

The histograms (Fig. 4.5 to 4.7) and CDF plots (Fig. 4.8 to 4.10) indicate the applicability of exponential distribution for flood peaks, volumes and durations above 10000 cumecs threshold. This was further investigated on the basis of chi- square test. The computed and critical x^2 values at 5% significance level for exponential distribution are presented in Table 4.9 for different threshold levels ranging from 4000 to 18000 cumecs.

It can be seen from Table 4.9 that exponential distribution is fitting well the flood peaks, volumes and durations for most

Table 4.8

Values of location parameter (x0) and scale parameter (β) for exponential distribution for flood peaks, volumes and durations for different threshold levels

Thres. cumecs	λ	Flood peaks x0	Flood peaks β	Flood volumes x0	Flood volumes β	Flood dur. x0	Flood dur. β
4000	4.322	0.0	10521.6	0.0	36500.9	1.2	5.9
5000	4.290	0.0	10404.8	0.0	30226.0	1.0	4.6
6000	3.903	136.1	10699.2	0.0	27677.2	0.9	3.9
7000	3.774	123.6	10592.9	0.0	24109.2	1.3	2.7
8000	3.451	7.5	10630.4	0.0	22396.6	1.1	2.4
9000	3.225	52.1	10570.8	0.0	20481.8	1.3	1.8
10000	2.935	110.2	10591.9	0.0	19390.6	1.1	1.7
11000	2.451	1048.9	10603.5	0.0	20123.6	1.2	1.5
12000	2.258	865.5	10631.1	0.0	19077.6	1.1	1.4
13000	2.000	1125.9	10645.7	0.0	18852.8	1.0	1.4
14000	1.806	1205.3	10650.7	0.0	18285.1	1.0	1.3
15000	1.612	1313.0	10701.1	0.0	18039.8	1.0	1.2
16000	1.516	891.2	10743.2	0.0	16969.7	0.8	1.2
17000	1.387	753.8	10790.2	0.0	16435.5	0.7	1.2
18000	1.193	1149.7	10906.8	0.0	16946.2	0.7	1.2

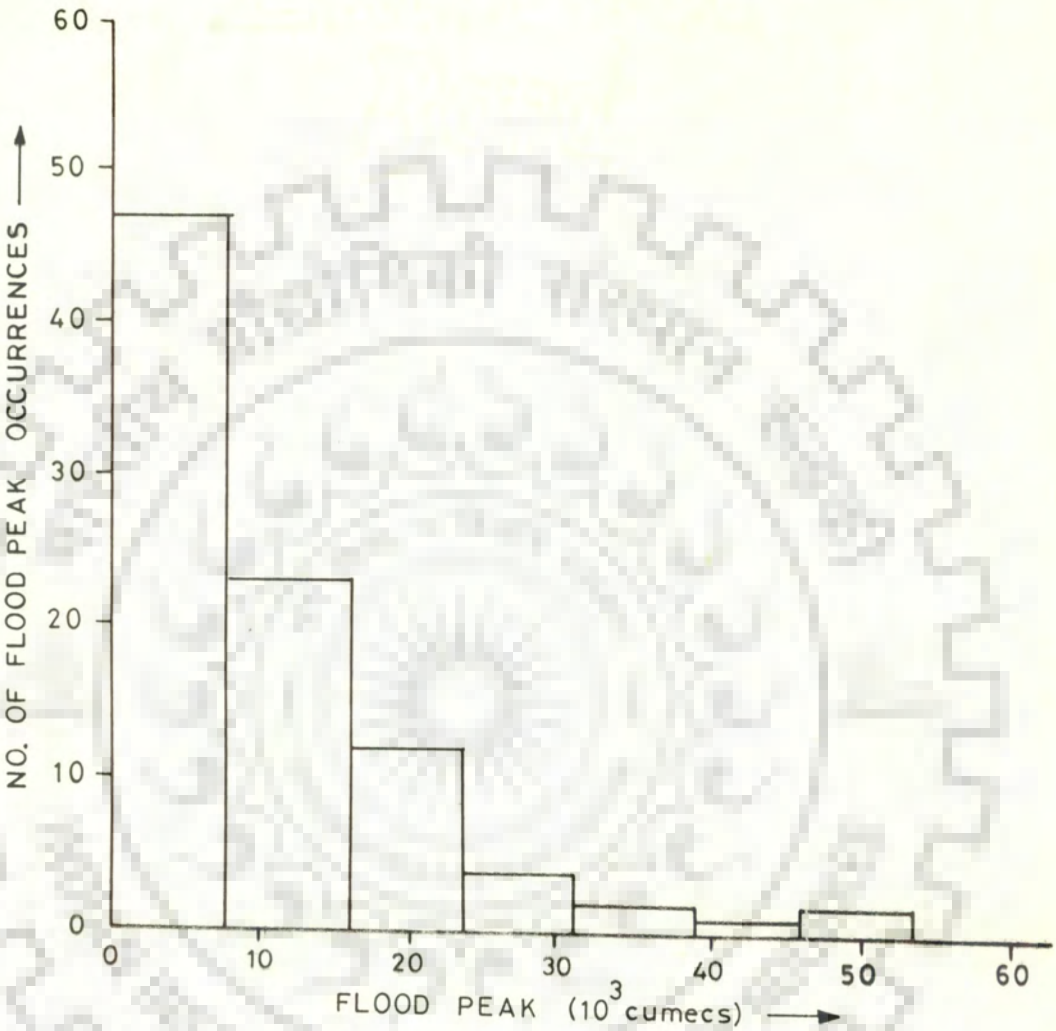

Fig. 4.5 - Histogram of flood peaks above 10,000 cumecs threshold

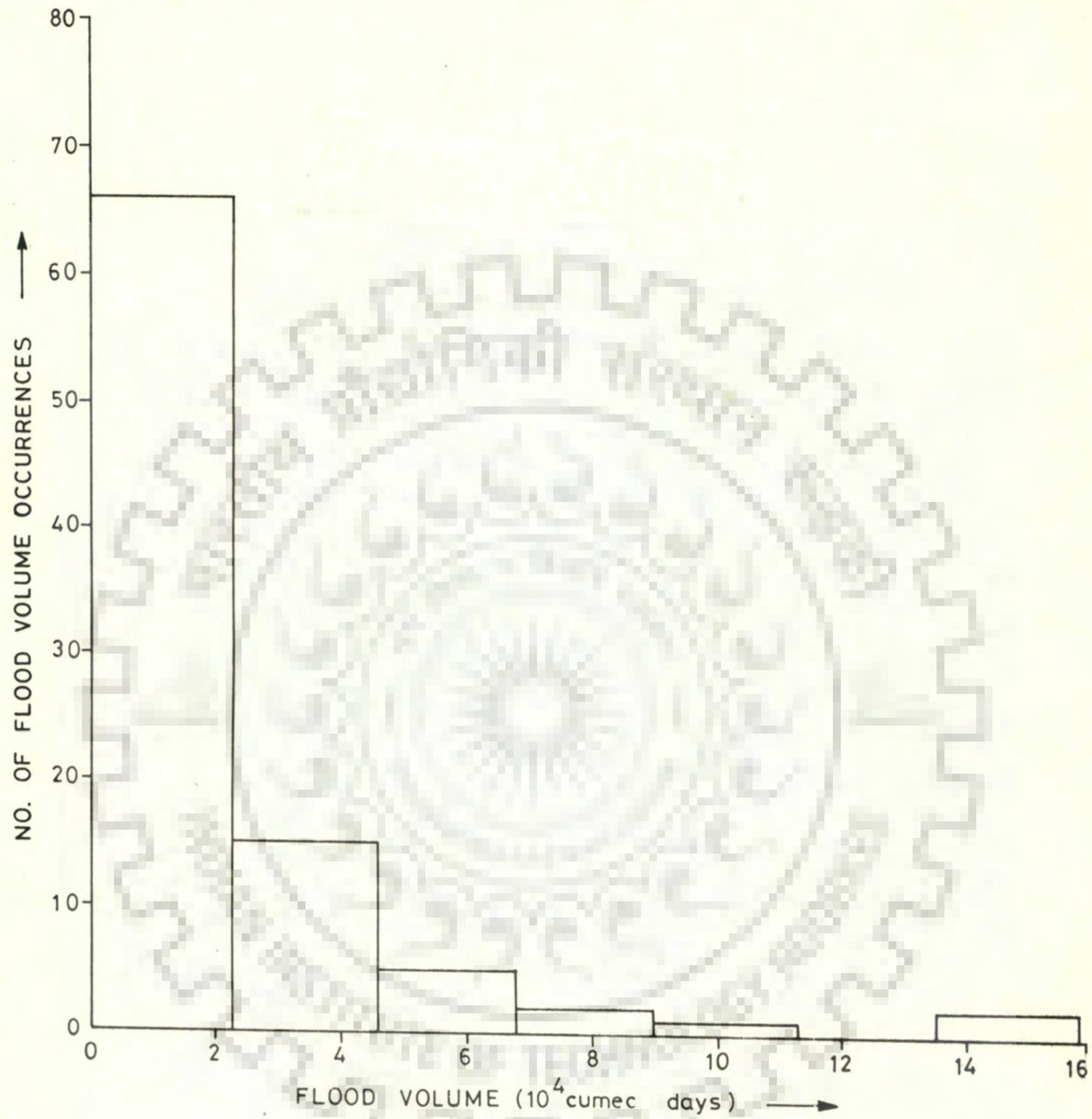

Fig. 4.6 Histogram of flood volumes above 10,000 cumecs threshold

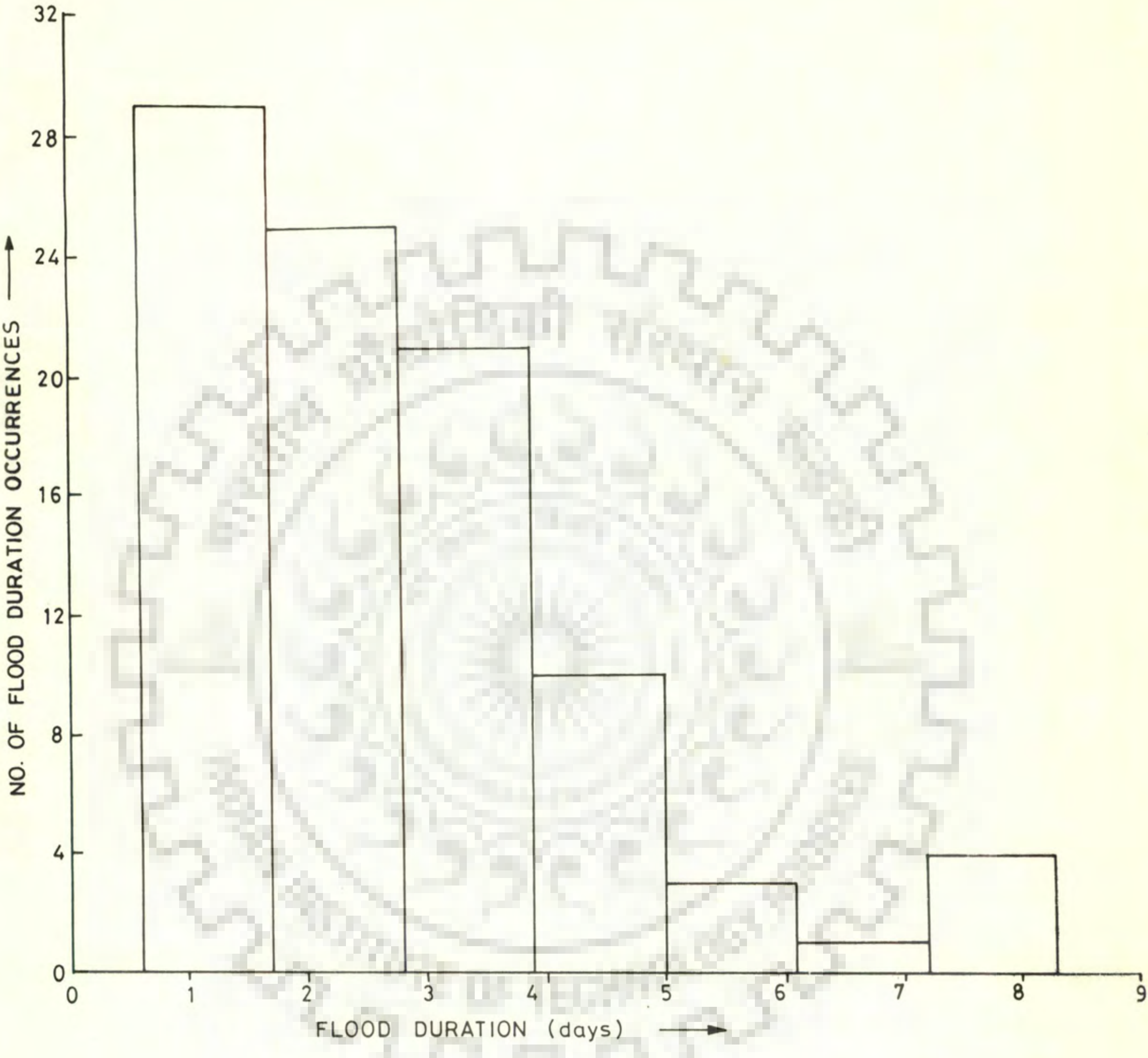

Fig. 4·7- Histogram of flood duratations above 10,000 cumecs threshold

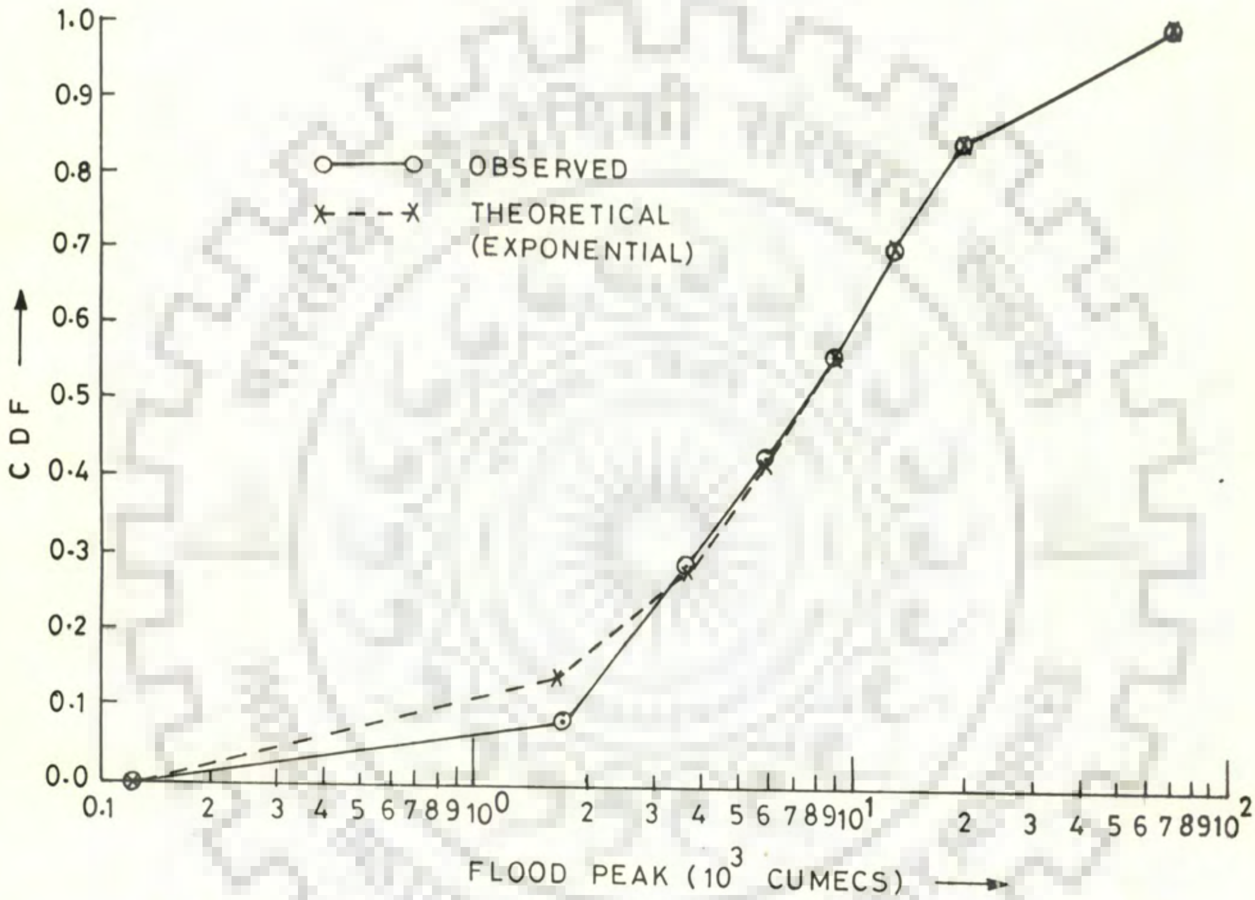

Fig. 4·8 - Observed and theoretical (exponential) distribution of flood peaks above 10,000 cumecs threshold

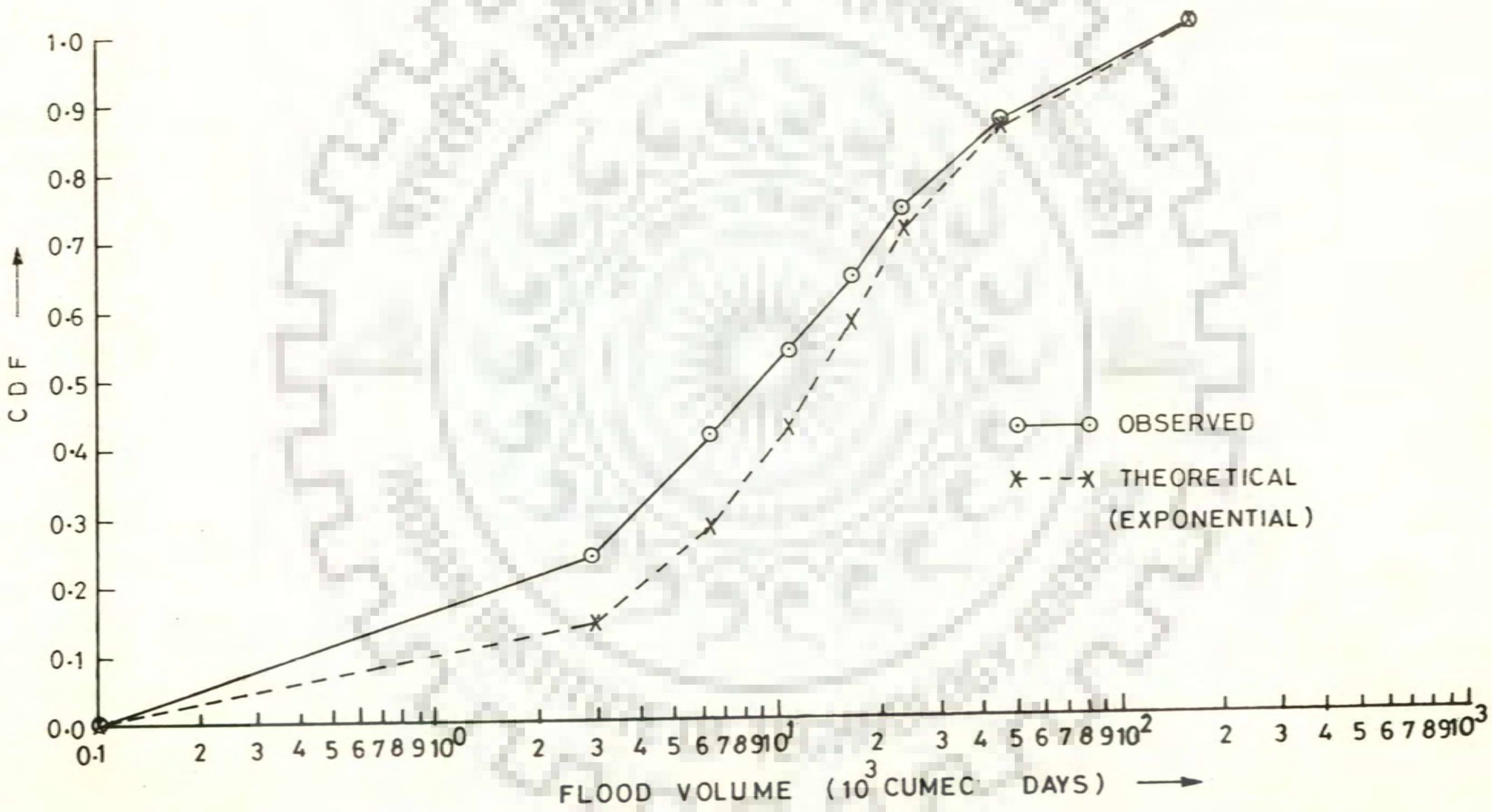

Fig. 4·9–Observed and theoretical (exponential) distribution of flood volumes above 10,000 cumecs threshold

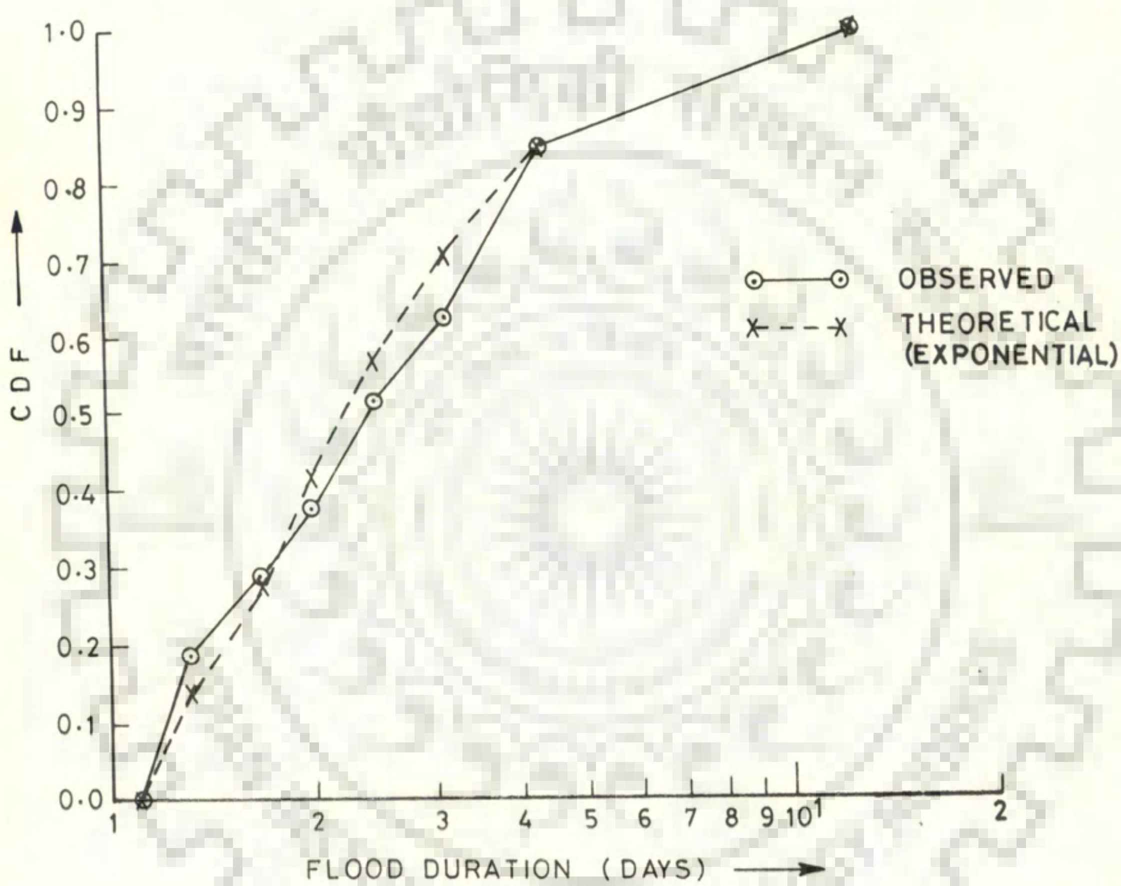

Fig. 4·10-Observed and theoretical (exponential) distribution of flood durations above 10,000 cumecs threshold

Table 4.9

Computed and critical χ^2 at 5% significance level for exponential distribution for flood peaks, volumes and durations for different threshold levels

Thres. cumecs	χ^2comp. flood peaks	flood volumes	flood dur.	χ^2cri.
4000	5.46	28.05*	4.38	11.10
5000	3.96	28.77*	10.45	11.10
6000	4.82	18.54*	4.36	9.49
7000	18.51*	13.29*	4.75	9.49
8000	12.78*	19.14*	4.41	9.49
9000	3.18	13.48*	6.54	9.49
10000	5.07	10.15*	9.53*	9.49
11000	1.92	7.69	15.00*	9.49
12000	1.80	4.50	14.00*	9.49
13000	3.41	4.74	12.70*	7.81
14000	0.78	6.32	6.14	7.81
15000	3.52	5.40	5.44	7.81
16000	3.93	9.89*	9.29*	7.81
17000	0.95	10.06*	8.20*	7.81
18000	1.43	5.64	2.08	7.81

Note: If χ^2comp. > χ^2cri., the value have been marked by * .

of the threshold levels. At the same time this fails either for peaks or volumes or durations or all the three for some of the threshold levels. From this analysis the general applicability of exponential distribution for flood peaks, volumes and durations for all the threshold levels is not confirmed and it again remains a mathematical problem to determine their distribution. Hence in the methodology, adopted for the study, a two step power transformation has been proposed to normalize these variables for further univariate and multivariate analysis in Chapter VI.

4.4.5 Distribution of largest flood peaks, volumes and durations

In a particular year there may be more than one flood event. The distributions of the largest flood peaks, volumes and durations are more important as compared to all the flood peaks, volumes and durations.

Zelenhasic (1970) derived the distribution function of largest exceedances, in a time interval, Ft(x), as given below:

$$F_t(x) = P(E_0^t) + \sum_{k=1}^{\infty} (H(x))^k . P(E_k^t) \qquad (4.11)$$

where,

H(x) is the distribution function of all the exceedances in given interval of time and P(E_k^t) is the probability that there will be k exceedances in the time interval.

Considering a fixed time interval of one year Eq. (4.11) can be written (Correia, 1987) as

$$F(x) = \sum_{k=0}^{\infty} \{H(x)\}^k . P(E_k) \quad (4.12)$$

Substituting the value of $P(E_k)$ from Eq. (4.7)

$$F(x) = \sum_{k=0}^{\infty} \{H(x)\}^k \frac{e^{-\lambda}\lambda^k}{k!} \quad (4.13)$$

or

$$F(x) = e^{\lambda . H(x) - \lambda} \quad (4.14)$$

In Eq. (4.14) $F(x)$ is the distribution function of largest flood exceedances, $H(x)$ is distribution function of all the flood exceedances and λ is average number of flood events occurring in a year.

Eq. (4.14) can be used to derive the distribution function of largest flood peaks, volumes and durations as follows:

$$F(FP_{lar}) = \exp(\lambda H(FP_{lar}) - \lambda) \quad (4.15)$$

$$F(FV_{lar}) = \exp(\lambda H(FV_{lar}) - \lambda) \quad (4.16)$$

$$F(FD_{lar}) = \exp(\lambda H(FD_{lar}) - \lambda) \quad (4.17)$$

The distribution functions of largest flood peaks, volumes and durations can be derived by obtaining functions of all the flood peaks, volumes and durations from Table 4.8. The computed and critical χ^2 at 5% significance level for distribution functions given by Eq. (4.15) to (4.17) are presented in Table 4.10 for different threshold levels ranging from 4000 to 18000 cumecs. Here the distribution functions of all the flood peaks, volumes and durations above a particular threshold level, have

Table 4.10

Computed and critical χ^2 at 5% significance level for largest flood peaks, volumes and durations for different threshold levels

Thresh. cumecs	χ^2comp. largest flood peaks	volumes	durat.	χ^2cri.
4000	5.45	8.03	12.54	14.1
5000	4.80	4.54	3.51	14.1
6000	7.38	4.54	4.16	14.1
7000	8.67	6.74	6.09	14.1
8000	8.67	6.74	9.32	14.1
9000	12.00	4.33	6.66	14.1
10000	13.33	4.33	2.66	14.1
11000	8.00	5.33	2.00	14.1
12000	9.27	8.58	9.96	14.1
13000	10.92	7.84	14.00	14.1
14000	19.40*	12.20	21.00*	14.1
15000	14.33*	13.50	18.50*	14.1
16000	13.95	12.21	14.82*	14.1
17000	16.61*	14.71*	14.71*	14.1
18000	16.00*	15.00*	12.00	14.1

Note: If χ^2comp. > χ^2cri., the values have been marked by * .

been assumed as exponential. The theoretical (Eq. (4.15) to (4.17)) and observed distribution functions of largest flood peaks, volumes and durations are plotted in Fig. 4.11 to 4.13 for a typical threshold of 10000 cumecs.

It can be seen from Table 4.10 that the distribution functions given by Eq. (4.15) to (4.17) for largest flood peaks, volumes and durations are fitting well the data for most of the threshold levels at 5% significance level. These also fail for some of the threshold levels. The possible reason for the failure for some the threshold levels may be the assumption of exponential distribution for flood peaks, volumes and durations above a particular threshold level. This can be taken care of by normalizing these variables before applying Eq. (4.15) to (4.17).

4.4.6 Distribution of the timing of largest flood peaks

To analyze the time of occurrence of largest flood peaks two parameter Gamma distribution was adopted ,following the work of Todorovic and Woolhiser (1972). This distribution was fitted to the time of occurrence of largest flood peaks and was found to fit well at 5% significance level for different levels of threshold. The mean, std. dev., and C_s of timing of largest flood peaks, τ_{lar} (counted from Jan. 1 in days) are given in Table 4.11 along with computed and critical χ^2 values for Gamma distribution for different threshold levels. It can be seen from the table that Gamma distribution is fitting well to the timing of largest flood peaks.

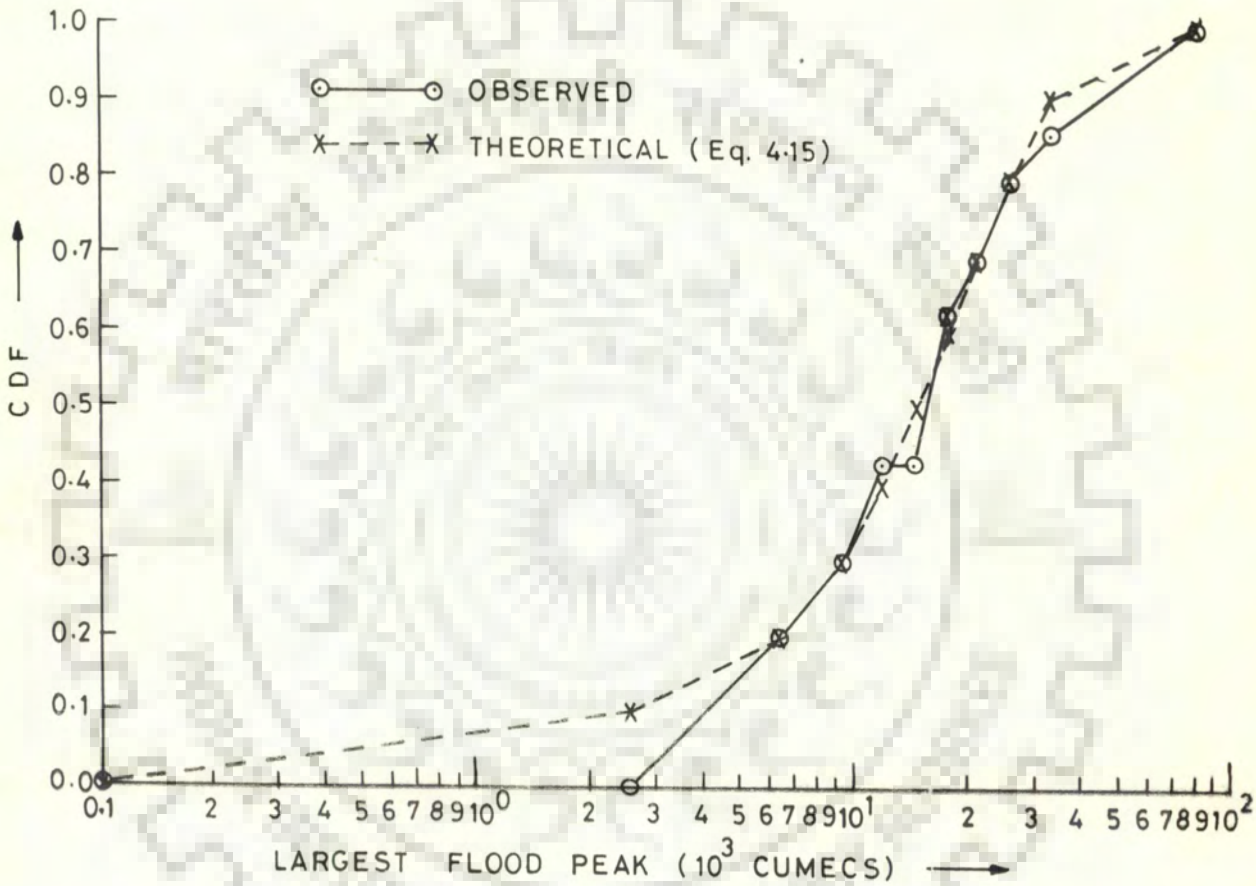

Fig. 4.11 – Observed and theoretical (Eq. 4.15) distribution of largest flood peaks above 10,000 cumecs threshold

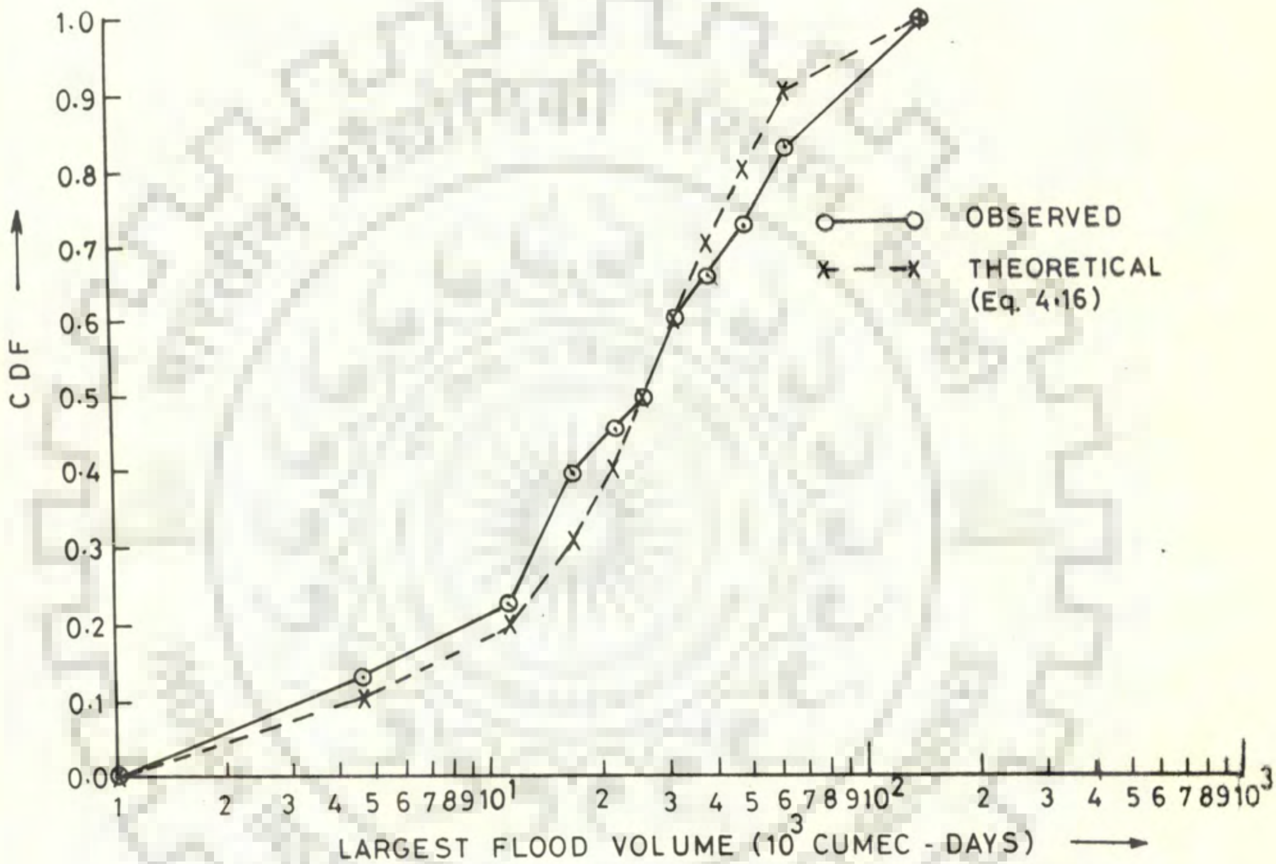

Fig. 4.12 Observed and theoretical (Eq.4.16) distribution of largest flood volumes above 10,000 cumecs threshold

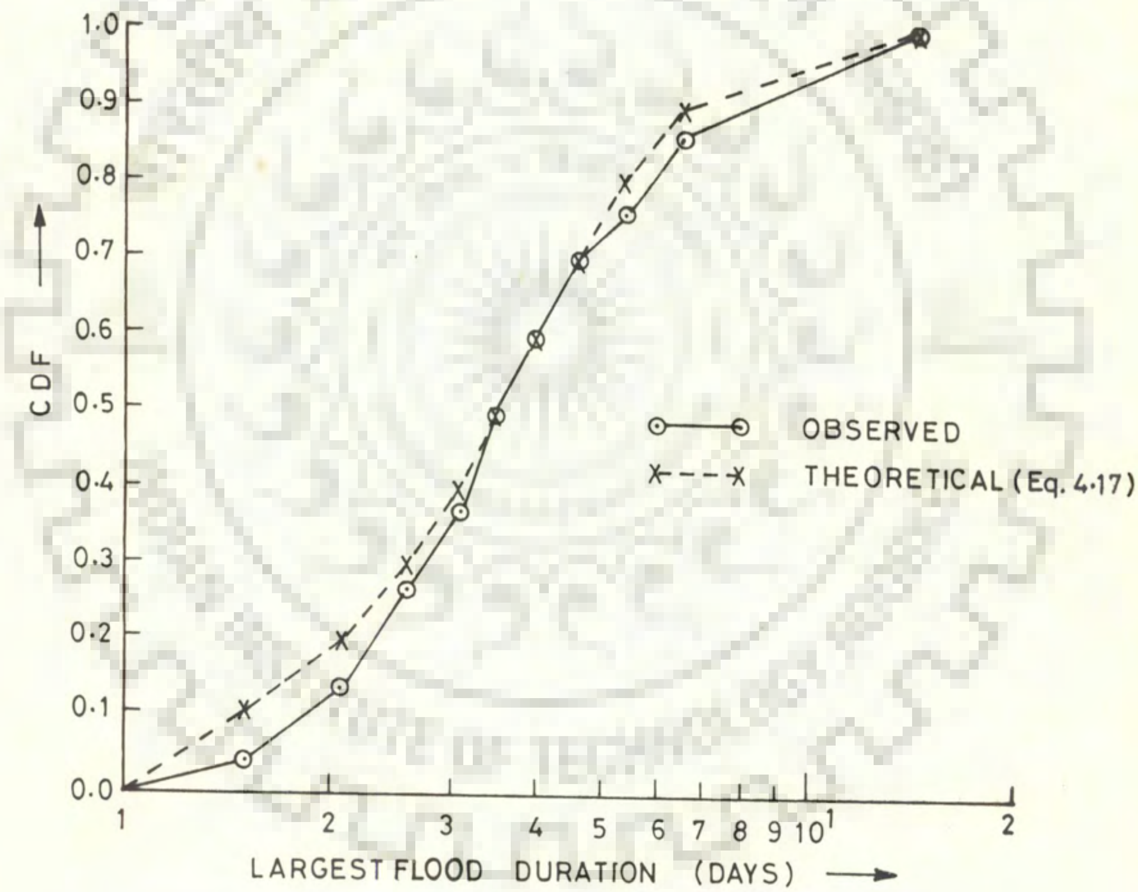

Fig. 4.13-Observed and theoretical (Eq. 4.17) distribution of largest flood durations above 10,000 cumecs threshold

Table 4.11

Computed and critical χ^2 at 5% significance level for time of occurrence of largest flood peaks for different threshold levels

Thres. cumecs	τ_{ler} mean	S.D.	Cs	χ^2 comp.	χ^2 cri.
4000 to 8000	235.4	20.5	-0.064	4.32	5.99
9000 to 11000	235.9	20.7	-0.124	5.33	5.99
12000	236.6	20.6	-0.202	5.99	5.99
13000	239.9	19.0	-0.304	5.92	5.99
14000	239.1	18.9	-0.237	4.80	5.99
15000	240.3	18.3	-0.309	4.75	5.99
16000	241.3	18.1	-0.418	2.86	5.99
17000	241.9	18.1	-0.474	3.52	5.99
18000	242.7	18.2	-0.600	3.50	5.99

For the threshold of 10000 cumecs the mean, and std. dev. of τ_{lar} are 235.867 and 20.649 days respectively. This gives the PDF of τ_{lar} as follows:

$$f(\tau_{lar}) = \frac{\tau_{lar}^{128.479} e^{-\tau_{lar}/1.8077}}{1.8077^{129.479}\Gamma(130.479)} \tag{4.18}$$

Similarly the PDF for other threshold levels can be obtained.

4.5 Inferences

General description of Narmada basin, data used and its processing, conventional flood frequency analysis of AFS and analysis of (i) number of flood events in a year, (ii) timing of flood peaks, (iii) flood peaks, volumes and durations, (iv) largest flood peaks, volumes and durations and (v) time of occurrence of largest flood peaks, have been presented in this chapter. On the basis of the analysis carried out in this chapter the following inferences can be drawn:

(i) For annual flood series of Narmada at Garudeshwar the Wakeby distribution performs better than GEV, EV 1 and LP III distributions.

(ii) Based on chi- square test at 5% significance level, the number of flood events in a year follow Poisson distribution. Once the applicability of Poisson distribution is established it remains applicable for any higher level of threshold also in the range 4000 to 18000 cumecs.

(iii) Exponential distribution fits well to the flood peaks, volumes, and durations for most of the threshold levels but the general applicability of this distribution for all the threshold levels is not confirmed by chi- square test at 5% significance level. Hence in the methodology, adopted for the study, a two step power transformation has been proposed to normalize these variables for further univariate and multivariate analysis in Chapter VI.

(iv) Two parameter Gamma distribution function fits well the time of occurrence of largest flood peaks for the threshold levels ranging from 4000 to 18000 cumecs.

CHAPTER V

GENERATION OF DAILY FLOWS

5.1 General

The methodology and application of two models of daily flow generation, viz. linear autoregressive and shot noise models are described in this chapter. The purpose of this study is (i) to ascertain the suitability of these models for generation of flows in the situation when major portion of flow occurs during 5 months of monsoon season, and (ii) to use the generated data for the validation of the methodology developed for stochastic modelling of flood flows. Synthetic data generation for river Narmada at Garudeshwar using linear autoregressive (LAR) model and shot noise model is described in section 5.2 and 5.3 respectively. The modified shot noise model (MSNM) which account for the different behavior of the river during monsoon and nonmonsoon seasons is also described in section 5.3. Section 5.4 describes the inter comparison of these models (MSNM and LAR) in preserving the statistical parameters of daily flows and other flood related characteristics in generated data. The conclusions drawn from this study are presented in section 5.5.

5.2 Linear Autoregressive Model

This section first describes the steps of the methodology used in linear autoregressive model to generate daily flows. The components of daily flow time series and generation of daily flows for river Narmada at Garudeshwar are then presented.

5.2.1 The Methodology

The classical approach to time series analysis comprises of decomposition of the series into deterministic component (trend and periodicity) and stochastic component. Various tests, e.g. Kendall's rank correlation test and regression test, etc. have been described in literature for identification of trend. Harmonic analysis has been traditionally employed to quantify periodic component in discrete time series. After removing the deterministic component, the resulting series is treated as stationary and described by linear autoregressive scheme. The model representing various components of time series is then used for data generation. The step wise procedure for such analysis, as given by Hall and O'Connell (1972), is as follows:

(i) Apply a logarithmic transformation to the flows denoted by X_t as:

$$Y_t = \log_e(X_t) \tag{5.1}$$

The logarithmic transformation stabilizes the seasonal fluctuations and also precludes the negative flows in the generated data.

(ii) Obtain a periodic representation of the seasonal fluctuations in the observed daily means $\bar{Y}_1$, $\bar{Y}_2$, ------, $\bar{Y}_{365}$ as

$$M_T = \bar{Y} + \sum_{k=1}^{s}\left\{A_k \cos\frac{2\pi kT}{365} + B_k \sin\frac{2\pi kT}{365}\right\} \tag{5.2}$$

where, $\bar{Y}$ denotes the overall mean, s is selected number of harmonics and coefficients A_k and B_k are given by

$$A_k = \frac{2}{365}\sum_{\gamma=1}^{365}\bar{Y}_\gamma \cos\frac{2\pi k\gamma}{365}$$

and

$$B_k = \frac{2}{365}\sum_{\gamma=1}^{365}\bar{Y}_\gamma \sin\frac{2\pi k\gamma}{365} \tag{5.3}$$

Though 182 harmonics can be used for smoothing a daily time series, yet more than 6 to 8 harmonics are seldom used (Yevjevich, 1976). The selection of the number of harmonics can also be made on the basis of P_{max} and P_{min} test or Periodogram test (Yevjevich, 1972b).

(iii)Obtain a periodic representation of the seasonal fluctuations in the daily standard deviations S_1, S_2,.......,S_{365} as:

$$S_T = \bar{S} + \sum_{k=1}^{s}\left\{C_k \cos\frac{2\pi kT}{365} + D_k \sin\frac{2\pi kT}{365}\right\} \tag{5.4}$$

where, C_k and D_k are defined by analogy with Eq. (5.3)

(vi) Obtain a standardized series as:

$$Z_t = (Y_t - M_T)/S_T \tag{5.5}$$

where, t = 365 j +T and j = Ø, 1, ..., (N-1) years.

Z_t series is approximately standardized. A further transformation is also done to standardize the series fully:

$$Z_t^* = \frac{Z_t - \mu_z}{\sigma_z} \tag{5.6}$$

where, μ_z and σ_z are overall mean and standard deviations of Zt series.

(v) Fit a multiple lag autoregressive model to the series Zt* .

$$Z_t^* = \sum_{i=1}^{m} a_i Z_{t-i}^* + \epsilon_t \tag{5.7}$$

Estimates of the coefficients a_i are derived by obtaining estimates of the first m auto correlation coefficients of Zt* series and then solving the linear Yule Walker equations. Tests on the model residuals ϵ_t are used to determine the order of the model. This description of the Zt series assumes that its auto correlation structure is nonseasonal.

(vi) Fit an appropriate probability distribution to the series of residuals ϵ_t. Normal, PT III and GEV distributions are generally tried in dealing with hydrologic time series.

(vii)Daily flows are then synthetically generated by reversing the above series of steps, starting by sampling independent random variates from the distribution ϵ_t.

5.2.2 The Components of daily flow time series

Daily flow data of 31 years (1949-79) for river Narmada at Garudeshwar have been used in the study. Various components of this time series of daily discharges (original domain) were analyzed as follows:

Trend: Kendall's rank correlation test (Kottegoda, 1980) was used to identify the presence of any rising or falling trend in the data and annual flows of 31 years were used for this purpose. The annual flows are plotted in Fig. 5.1. Based on Kendall's rank correlation test, the hypothesis of no trend in this data can be accepted at 95 % confidence level.

Periodicity: The harmonic analysis of daily means and standard deviations was carried out using the procedure described in section 5.2.1 and Fourier coefficients were obtained. As suggested by Yevjevich (1976) only 8 harmonics were used for daily means and standard deviations. When the harmonic means and harmonic standard deviations were computed from Eq. (5.2) and (5.4) respectively it was noticed that some of the ordinates were negative. Though the negative volume i.e. the sum of negative ordinates, was negligible. To avoid negative ordinates, the Fourier coefficients were adjusted as follows:

$$F_k' = \left(\frac{w}{2} x F_k\right) / \left(\frac{w}{2} + \frac{m.w}{2}\right) \tag{5.8}$$

where,

F_k'= adjusted Fourier coefficient,

w = number of seasons in a year; For daily flows w will be equal to 365.

m = minimum percentage to make the negative volume as zero. Generally m varies from 1% to 10% .

Fk = original Fourier coefficient.The variance accounted for the

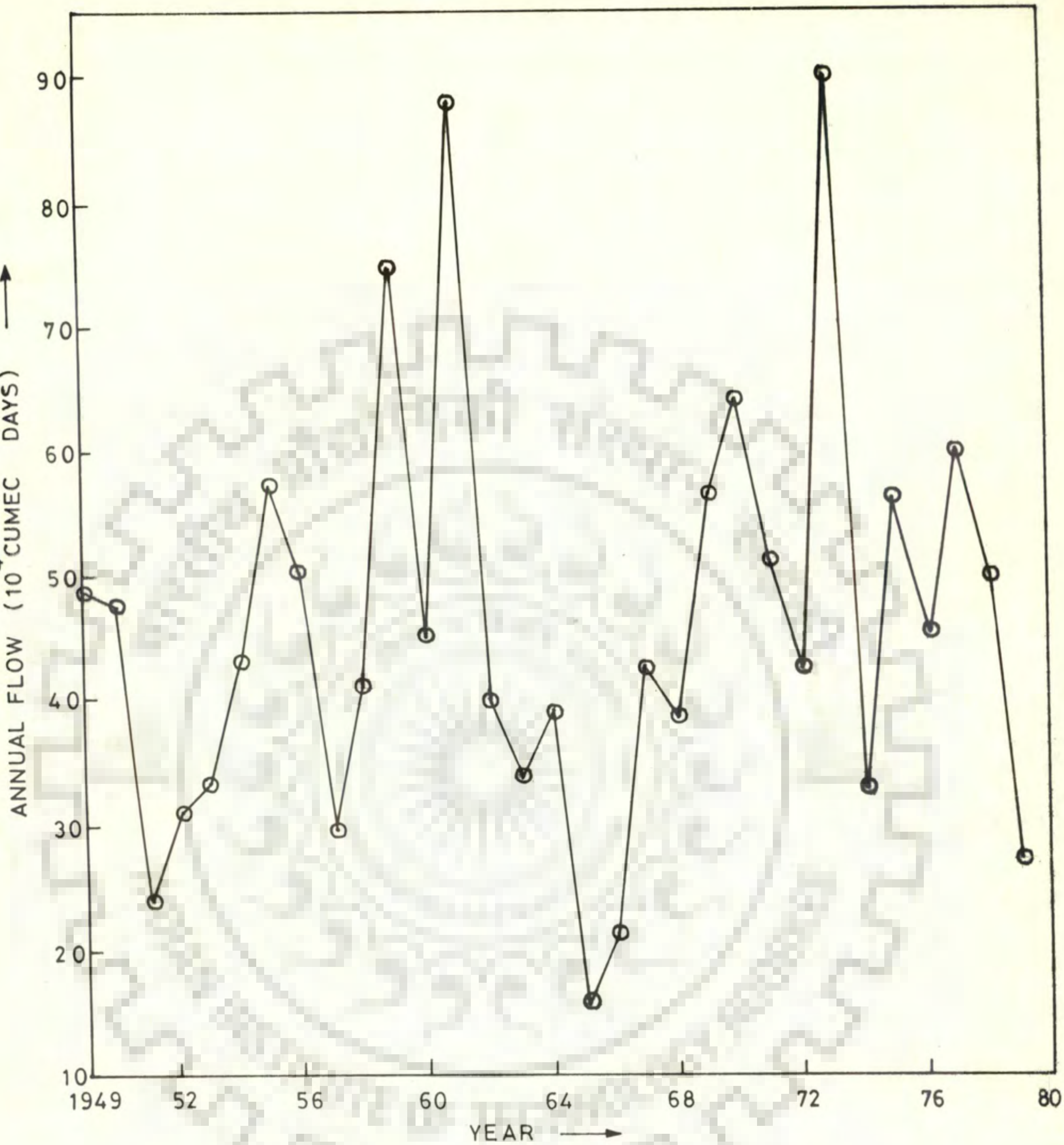

Fig. 5.1- Annual flows for river Narmada at Garudeshwar(1949-79)

jth harmonic is $0.5(A_j^2+B_j^2)$

The initial variances in daily means and daily standard deviation were computed as 3.77×10^6 and 5.14×10^6 respectively.

The adjusted Fourier coefficients, A_j and B_j, ratio of variance explained by each harmonic to the initial variance denoted by P_j and cumulative sum of P_j for first eight harmonics are given in Table 5.1 for daily means and standard deviations.

It can be seen from Table 5.1, that 8 harmonics are able to explain 88.8% of the variance in the daily means and 61.7% in daily standard deviations. The other harmonics of higher frequency were further not tried and were passed onto the stochastic component as noise. The selection of harmonics was not done on the basis of any formal test to avoid higher frequency (more than 8) in the daily standard deviations.

The unsmoothed daily means and superimposed harmonic means based on 8 harmonics are presented in Fig. 5.2, while the corresponding plots for standard deviations are shown in Fig. 5.3 .

<u>Dependent Stochastic Component:</u> The statistical parameters (mean, standard deviation, C_s, C_k, r_1, r_2, r_3) of the standardized series given by Eq. (5.5) are -0.042, 0.829, 7.07, 101.342, 0.732, 0.559 and 0.462 respectively. A further transformation (as per Eq. (5.6)) was done to standardize this series fully i.e. to have mean =0.0 and standard deviation =1.0. Different autoregressive models AR (1), AR (2), AR (3), etc. were fitted to this series and their coefficients of determination computed. The coefficients of determination for AR (1), AR (2), and AR (3)

Table 5.1

Harmonic analysis of daily means and standard deviations

Harmonic (j)	A_j	B_j	P_j	Cum.sum
Daily Means				
1	-1102.0	-1638.9	0.5160	0.5160
2	-419.7	1329.4	0.2571	0.7730
3	761.0	-262.0	0.0857	0.8587
4	-284.3	-250.6	0.0190	0.8777
5	-93.6	136.6	0.0036	0.8814
6	95.0	87.3	0.0022	0.8836
7	70.6	-114.7	0.0024	0.8860
8	-115.6	-35.1	0.0019	0.8879
Daily S.D.				
1	-1157.6	-1583.1	0.3739	0.3739
2	-395.7	1248.7	0.1668	0.5408
3	669.0	-129.2	0.0451	0.5859
4	-113.1	-305.7	0.0103	0.5962
5	-215.3	75.6	0.0051	0.6013
6	110.9	181.9	0.0044	0.6057
7	145.0	-181.0	0.0052	0.6109
8	-235.4	-42.0	0.0056	0.6165

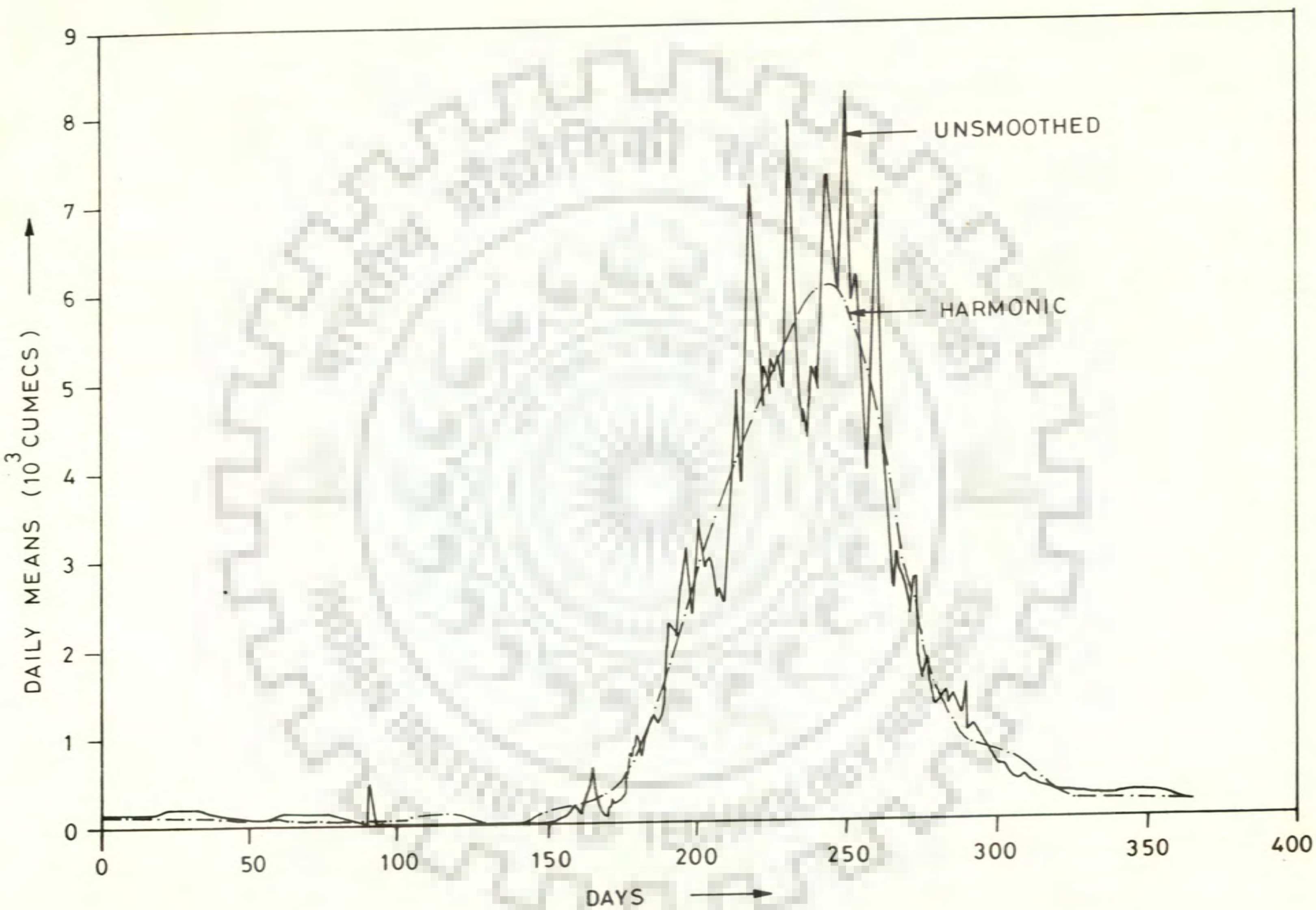

Fig. 5.2-Unsmoothed and harmonic daily means for flows of river Narmada at Garudeshwar

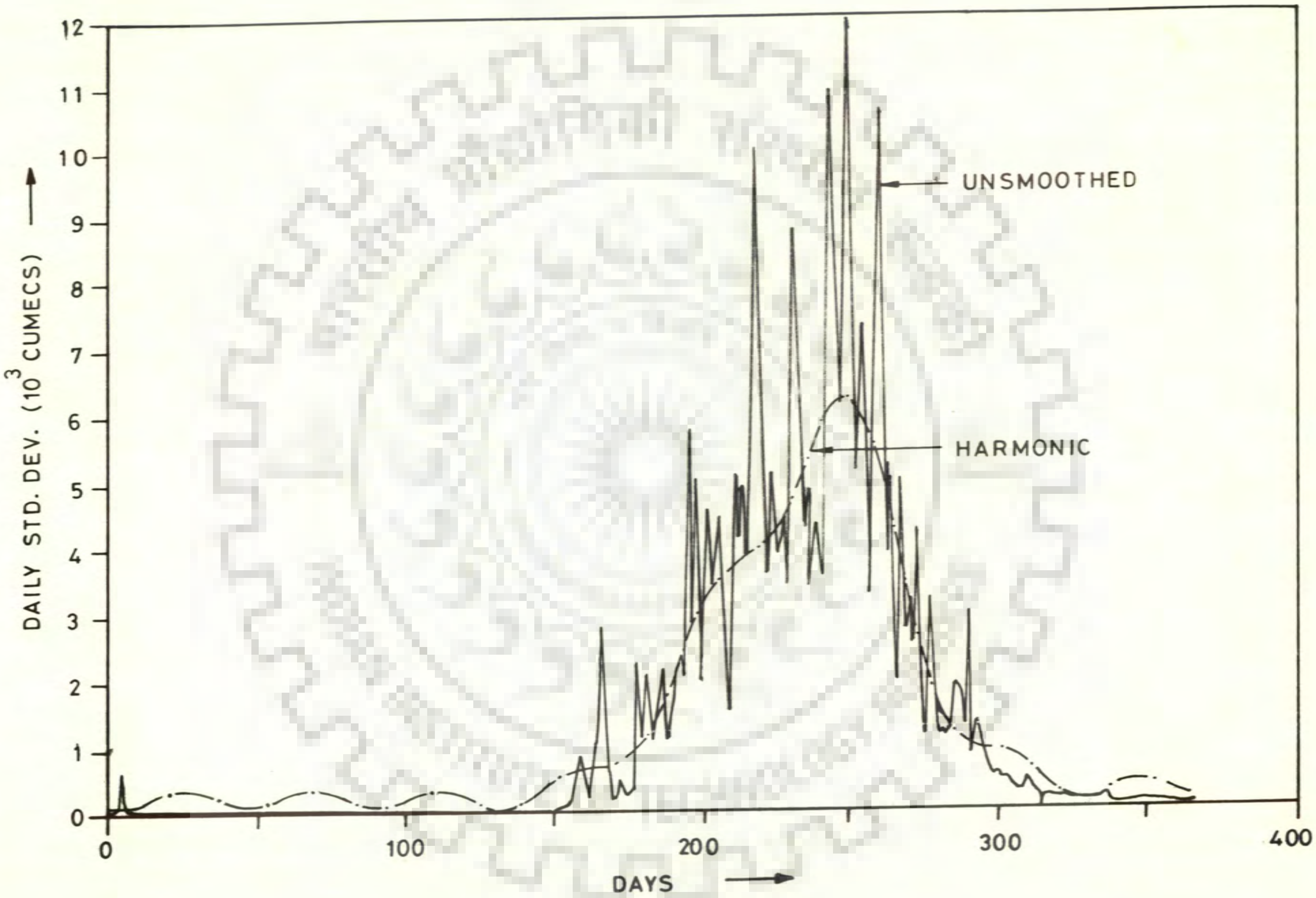

Fig. 5·3 - Unsmoothed and harmonic standard deviations for flows of river Narmada at Garudeshwar models were found to be Ø.5361, Ø.5372, and Ø.54Ø3 respectively. Adopting the criterion given by Yevjevich (1972 b) the AR (1) model was selected to remove the dependence from the stochastic component. The coefficient a_1 for the AR (1) model was evaluated as Ø.732, i.e. equal to r_1.

Independent Stochastic Component (ISC) : After removing the dependent stochastic components the statistical parameters (mean, standard deviation, C_s, C_k, r_1, r_2, r_3) of the ISC were obtained as -Ø.Ø, Ø.681, 8.154, 2Ø4.527, -Ø.Ø36, -Ø.Ø36, and Ø.Ø25 respectively. These parameters indicate that the ISC is highly skewed and will require a skewed distribution. For the present study it was decided to try normal, PT III and GEV as three possible alternatives.

5.2.3 Generation of Daily Flows

Daily flows are generated by reversing the series of steps explained in section 5.2.1.

The synthetic flows have been generated for the following two cases, (i) without logarithmic transformation in the beginning and exponentiation in the end, and (ii) logarithmic transformation in the beginning and exponentiation in the end. These will be referred as Case I and Case II respectively. For generation of ISC three distributions, namely: (i) Normal, (ii) PT III, and (iii) GEV, were tried. PT III and GEV distributions were taken as highly skewed series having C_s equal to 8.154 was to be modelled.

For generating the PT III distributed random numbers, Wilson-Hilferty transformation was used. This transformation converts normally distributed random numbers into PT III distributed random numbers. For generating GEV distributed random numbers the parameters (u, α and k) of GEV distribution were estimated from the series of ISC using PWM. These parameters were then used to convert uniformly distributed random numbers to GEV distributed random numbers. Subroutine GAUSS was used for generation of normally distributed random numbers and is given in Appendix I. Subroutine GAUSS requires uniformly distributed random numbers, generated using built in function RAN of DEC- 20 computer system.

Model Selection and Generation: 10 samples of 30 years length for each case were generated and their overall statistical parameters computed. The statistical parameters of the historical series and mean of the statistical parameters of 10 samples for each case are presented in Table 5.2.

It can be seen from Table 5.2, that reproduction of statistical parameters in Case II with normal distribution for ISC is better than other options tried. In case I with PT III distribution, even the overall mean is coming negative. The reason for this was further investigated by studying applicability of Wilson- Hilferty transformation for generation of PT III distributed random numbers. It was found that Wilson- Hilferty transformation is good only when C_s is around 2. The C_s of ISC is 8.154 and hence the transformation was not able to reproduce this C_s in generated ISC.

Based on Table 5.2, it can be concluded that out of various cases considered, the Case II (logarithmic transformation in the

Table 5.2

Statistical parameters of historical series and mean of the parameters of 10 samples generated by different cases of LAR models

Series	Max.	Min.	Mean	S.D.	C_v	C_s	C_k	r_1
Hist. series	63800	3	1249.6	3242.6	2.6	6.8	76.1	0.82
Case I Normal	2185	-10063	1191.5	2563.5	2.2	2.6	11.5	0.86
Case I PT III	46701	-130941	-3255.7	7513.7	--	-5.1	47.5	0.82
Case I GEV	24660	-354	1184.9	2029.6	1.7	2.8	14.0	0.93
Case II Normal	62589	2	1237.3	3168.6	2.6	6.7	79.0	0.95
Case II PT III	207512	4	1371.7	5303.8	3.9	16.6	489.3	0.87
Case II GEV	41293	4	1053.2	2345.2	2.2	5.5	59.2	0.95

Note: Case I represents analysis of time series in original domain.

Case II represents analysis of time series in log domain.

Normal, GEV and PT III are distributions of independent stochastic components.

beginning and exponentiation in the end) with ISC generated as normally distributed is the best one. The statistical parameters of individual samples along with the parameters of historical series are presented in Table 5.3.

Table 5.3

Statistical parameters of historical series and generated samples by selected linear autoregressive model

Series	Stat. Para.							
	Max.	Min.	Mean	S. D.	C_v	C_s	C_k	r_1
Hist. ser.	63800.0	3.1	1249.6	3242.6	2.59	6.79	76.1	0.82
Series 1	51485.0	3.6	1215.5	3139.1	2.58	6.32	62.1	0.95
Series 2	67234.6	0.8	1278.3	3337.6	2.61	6.87	78.0	0.94
Series 3	87969.6	1.0	1202.1	3139.0	2.61	9.26	165.7	0.95
Series 4	46309.1	3.1	1162.6	2756.3	2.37	5.39	46.3	0.94
Series 5	87832.0	0.8	1283.8	3681.4	2.86	8.58	120.4	0.94
Series 6	42156.5	0.6	1184.2	2774.0	2.34	5.24	41.5	0.95
Series 7	58140.7	1.8	1272.2	3366.7	2.64	6.70	70.2	0.95
Series 8	67417.3	2.6	1146.0	2956.8	2.58	7.14	91.2	0.95
Series 9	64483.6	1.3	1312.7	3460.7	2.63	6.46	65.1	0.94
Series 10	52816.8	0.4	1315.8	3074.6	2.33	5.38	48.9	0.94
Max. of 10 samp.	87969.6	3.6	1315.8	3681.4	2.86	9.26	165.7	0.95
Min. of 10 samp.	42156.5	0.4	1146.0	2756.3	2.33	5.24	41.5	0.94
Mean of 10 samp.	62589.0	1.6	1237.3	3168.6	2.55	6.74	78.9	0.94

It can be seen from this table, that all the samples are closely reproducing the overall statistical parameters. The maximum of these parameters (among samples) are higher than the historical parameters while the minimum are lower.

The ability of this model to reproduce the characteristics of AFS was further studied by generating two samples of 30 and 50 years length. The comparison of statistics of AFS of historical data and 30 and 50 years length samples is given in Table 5.4.

Table 5.4

Comparison of statistics of AFS of historical data and synthetic data generated by selected linear autoregressive model

Series	Stat. Para.							
	Max.	Min.	Mean	S.D.	C_v	C_s	C_k	r_1
Historical	63800	10132	27790.8	13788.2	0.50	1.0	3.8	0.11
AFS of 30 years generated data	51485	4552	16733.4	11754.0	0.70	1.6	5.3	-0.29
AFS of 50 years generated data	65696	4552	18271.5	13331.4	0.73	1.5	5.5	-0.25

It can be seen from Table 5.4 that the mean annual flood of AFS of synthetic data is lower while C_v is quite high as compared to AFS of historical data. The C_s and C_k are also higher. These indicate the poor performance of LAR model in reproducing the characteristics of AFS.

Further discussion on Table 5.3 and 5.4 is given in section 5.4, while discussing inter-comparison of LAR and MSNM models.

5.3 Shot Noise Model

The existence of rapid rise followed by slow recessions in the daily flows suggested another model called 'Shot Noise Model' in which a sequence of random disturbances occurs at random times. Each disturbance causes a rapid rise followed by a slow recession in the flow. This contrasts with linear autoregressive models where random disturbances occur at all the time points. In addition, linear Gaussian processes are time reversible, i.e. the statistical properties of the process are the same regardless of the direction in which time is measured, in contrast to the observed daily flows which have been found to be time irreversible owing to the classic asymmetric shape of the hydrographs.

The theory of shot noise model and testing of the model are described in section 5.3.1. The daily flow generation using modified shot noise model (MSNM) which account for different behavior of the river flows during monsoon and nonmonsoon seasons and requires much lesser number of parameters, is presented subsequently.

5.3.1 Theory of shot noise model

In shot noise model, the flows are treated as a series of impulses and decays. The model is based upon filtered Poisson process (Parzen, 1962) and was developed by Weiss (1973, 1977). The shot noise process is defined, in continuous time, as:

$$X(t)=\sum_{m=N(-\infty)}^{m=N(t)} w(t-\tau_m, y_m) \tag{5.9}$$

where, N(t) is a Poisson process having rate ν and the random variable y_m associated with the random time τ_m produces a pulse in the flow given by $w(t-\tau_m, y_m)$. One of the simplest filtered Poisson processes is obtained when $w(t-\tau_m, y_m)$ is linear and defined as:

$$\begin{aligned} w(t-\tau_m, y_m) &= y_m e^{-b(t-\tau_m)} && (t \geq \tau_m) \\ &= 0 && (t < \tau_m) \end{aligned} \tag{5.10}$$

The resulting process is referred to as a simple shot noise process, which can be synthesized through the following series of steps:

(i) Random event times ..., τ_{m-1}, τ_m, τ_{m+1}, ... are generated according to

$$P(N(t)=n)=\frac{e^{-\nu t}(\nu t)^n}{n!} \tag{5.11}$$

which means that the time between events is exponentially distributed with mean $1/\nu$.

(ii) Associated with the random times τ_{m-1}, τ_m, τ_{m+1} random jumps y_{m-1}, y_m, y_{m+1} are generated from an exponential distribution with mean θ.

(iii) Associated with the random times τ_{m-1}, τ_m, τ_{m+1} and random jumps y_{m-1}, y_m, y_{m+1} are pulses with values $y_{m-1}e^{-b(t-\tau_{m-1})}$, $y_m e^{-b(t-\tau_m)}$, $y_{m+1}e^{-b(t-\tau_{m+1})}$ at times t.

(iv) The continuous single shot noise process is defined as the sum of all contributing pulses at time t.

$$X(t)=\sum_{m=N(-\infty)}^{m=N(t)} y_m e^{-b(t-\tau_m)} \tag{5.12}$$

The steps (i) to (iv) are shown in Fig. 5.4.

The continuous single shot noise process with parameters ν, θ and b can be regarded as a model of continuous daily streamflow. The marginal distribution of the process is Gamma with density function

$$f(x)=\frac{(1/\theta)^{\nu/b}x^{(\nu-b)/b}e^{-x/\theta}}{\Gamma(\nu/b)} \tag{5.13}$$

O'Connell (1974) suggests that pulse function employed does not yield a very realistic model of continuous daily streamflow. However more realistic pulse functions increase the number of parameters and make the resulting process more mathematically intractable.

For application to daily flows, the single shot noise process can be defined as an averaged process over a period of one day, i.e.,

$$X_t=\int_{t-1}^{t} X(S)ds \tag{5.14}$$

As a result, the averaged process will not exhibit vertical jumps in the flow. The properties of X_t have been studied by Weiss (1973b) who suggested that the parameters ν, θ and b could

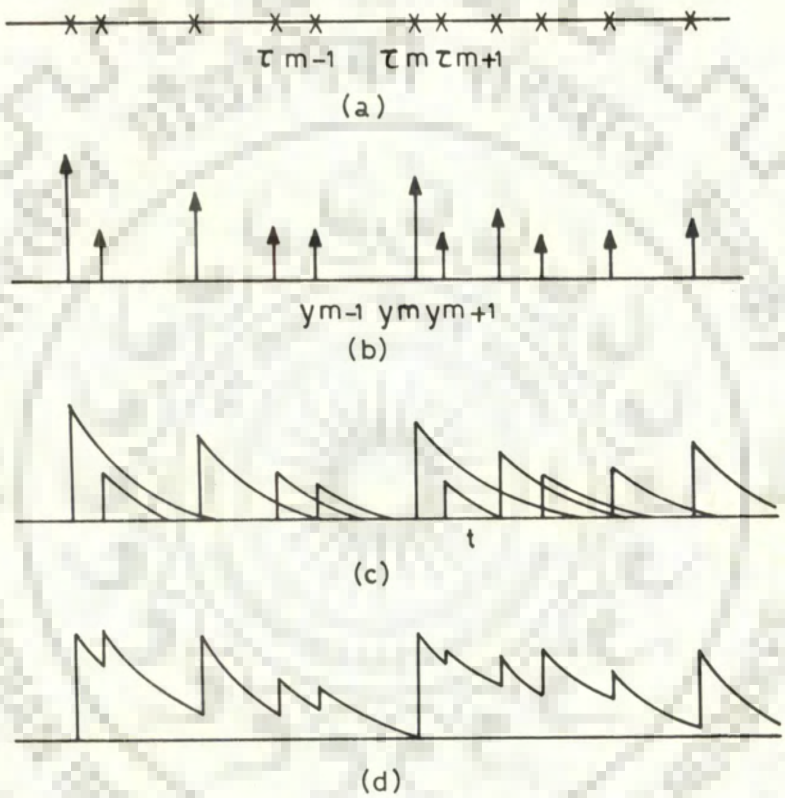

Fig. 5·4-(a) Events....τm,....from a Poisson process with rate ν. (b) Jumps....ym....from an exponential distribution with mean θ. (c) Pulses with values......ym $e^{-b(t-\tau_m)}$..,at time t. (d) Schematic plot of continuous single shot noise process.

be obtained from the following equations.

$$\mu = \frac{v\theta}{b} \tag{5.15}$$

$$\sigma^2 = \frac{2v\theta^2}{b^3}[b(1-e^{-b})] \tag{5.16}$$

$$C_s\sigma^3 = \frac{2v\theta^3}{b}\frac{3\left\{b-2(1-e^{-b})+\frac{1}{2}(1-e^{-2b})\right\}}{b^3} \tag{5.17}$$

$$\rho_1 = \frac{(1-e^{-b})^2}{2[b-(1-e^{-b})]} \tag{5.18}$$

Eq. (5.18) is used to determine b and Eq. (5.15) and (5.16) are used simultaneously for θ and v. To account for the seasonality in daily flows the parameters b, θ and v are estimated separately for each month of the year. The model would, thus, require (12x3)= 36 parameters.

5.3.2 Testing the model

For estimating the parameters of the model the monthly statics of daily flows was obtained and using Eq. (5.15), (5.16) and (5.18) the parameters of the model were calculated. The Newton- Raphson method was used to estimated 'b' from Eq. (5.18).

Table 5.5 gives the statistics and parameters θ , v and b for each month of the year. The monthly statistics and parameters are plotted in Fig. 5.5 and 5.6 respectively.

It can be seen from Table 5.5 and Fig. 5.6 that the variation in the parameters during nonmonsoon months (Nov. to May) is negligible as compared to monsoon months. This indicates the

Table 5.5

Monthly statistics of daily flows and parameters of shot noise model

Month	Mean	S.D.	C_s	Min	r_1	ν	θ	b
Jun.	301.3	687.4	3.67	10.2	0.797	0.060*	1756.9*	0.351*
July	2283.2	2703.8	2.58	224.4	0.708	0.326*	3807.3*	0.544*
Aug.	5270.6	5161.2	2.31	578.6	0.755	0.396*	5816.1*	0.437*
Sep.	4914.8	5848.9	2.62	527.2	0.799	0.219*	7786.5*	0.346*
Oct.	1173.1	1253.8	2.47	183.4	0.896	0.139*	1415.6*	0.167*
Nov.	341.5	210.5	1.38	74.5	0.951	0.194	133.1	0.076
Dec.	200.2	105.5	1.40	53.4	0.962	0.204	56.7	0.058
Jan.	139.6	66.6	1.88	43.8	0.970	0.202	32.3	0.047
Feb.	110.2	56.7	1.56	35.9	0.972	0.157	29.6	0.042
Mar.	77.4	37.9	1.14	23.8	0.969	0.192	18.9	0.047
Apr.	55.6	28.2	1.18	15.4	0.970	0.175	14.5	0.046
May	37.4	22.3	1.72	9.7	0.945	0.235	13.7	0.086
Nonmon.	137.2	75.3	1.46	9.7	0.963	0.187*	42.1*	0.057*

Note:* indicate parameter of Modified Shot Noise Model

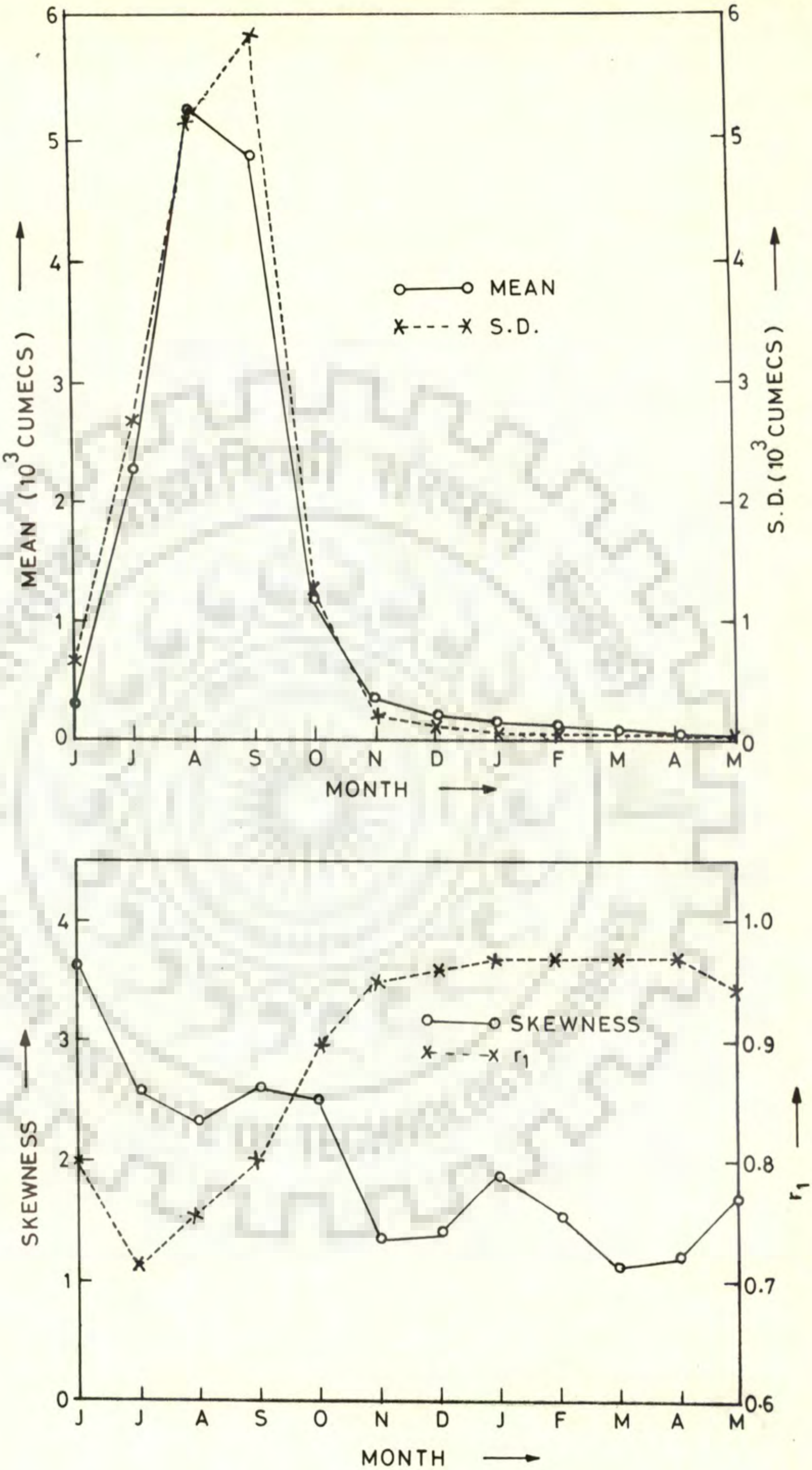

Fig. 5.5 - Monthly statistics of daily flows for river Narmada at Garudeshwar

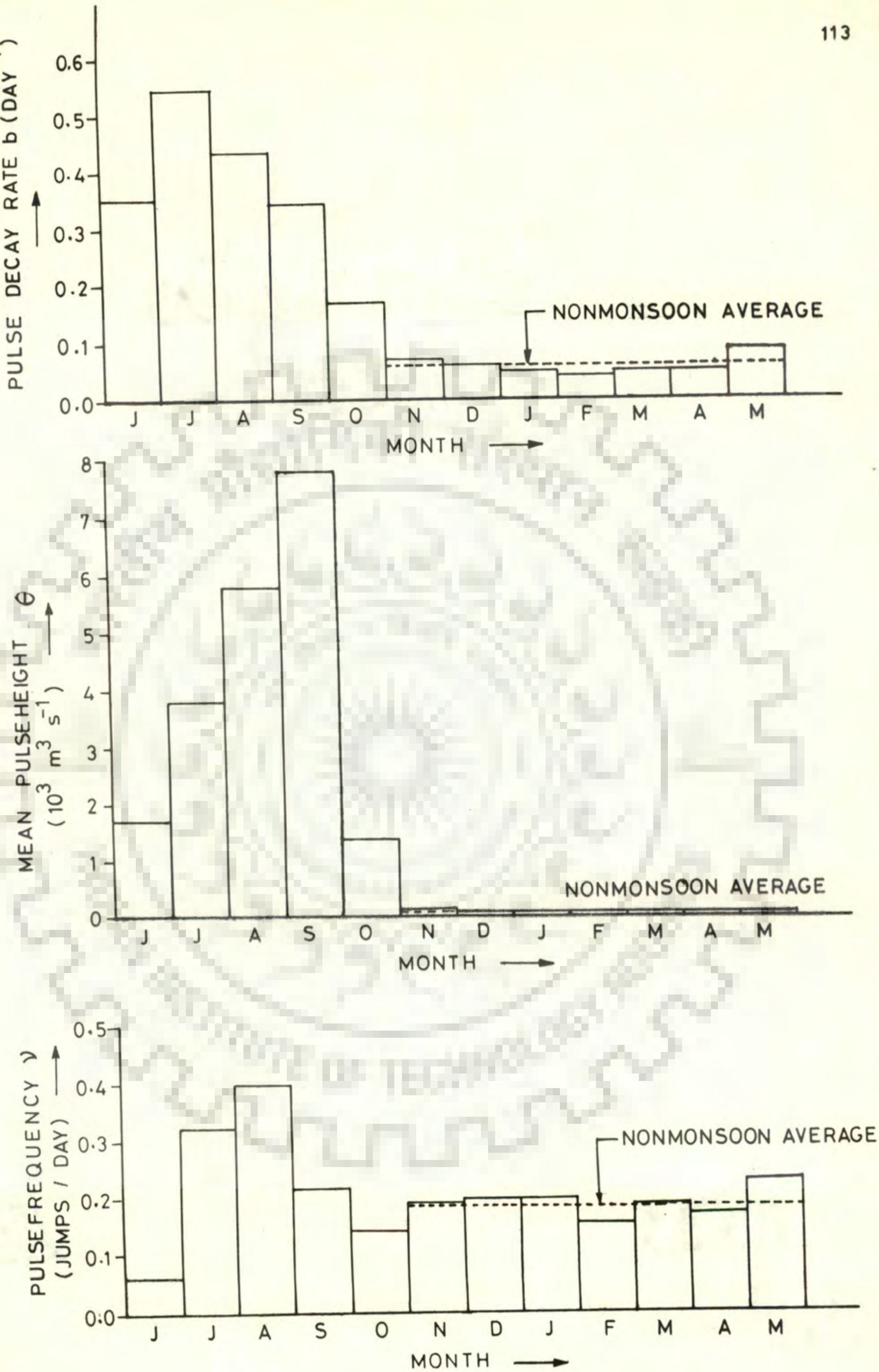

Fig. 5.6 - Parameters of shot noise model

possibility of lumping all the nonmonsoon months together. This way the shot noise model would require only 18 (3x5 +3=18) parameters without any significant effect on generated flows. This model has been designated as modified shot noise model (MSNM) in the following text.

In order to examine the possibility of incorporating the skewness in the model three methods (NERC, 1975; pp. 221-223) namely, (i) addition of constant base flow, (ii) use of Gamma-2 parameter distribution for jump heights, and (iii) Lomax distribution for jump heights were tried. All these methods require certain conditions for their application, which were not met by the data of Narmada at Garudeshwar. Hence exponential distribution was adopted for jump heights.

5.3.3 Generation of daily flows

The MSNM requires separate set of parameters for monsoon months (June to Oct.) and one set of parameters for all the nonmonsoon months (Nov. to May). It would, therefore, require only 18 (3x5+3 = 18) parameters. These parameters have been marked by * in Table 5.5.

The MSNM with exponentially distributed jump heights was used to generate 10 sequences of 30 years length. The statistical parameters of these sequences and historical series are given in Table 5.6.

It can be seen from Table 5.6 that synthetic sequences are closely reproducing the overall mean, standard deviation and Cv. However, Cs and Ck of generated data are much lower than that of

Table 5.6

Statistical parameters of historical series and generated samples by MSNM

Series	Stat. Para.							
	Max.	Min.	Mean	S.D.	C_v	C_s	C_k	r_1
Hist. ser.	63800.0	3.1	1249.6	3242.6	2.60	6.79	76.13	0.82
Series 1	46279.3	0.0	1172.2	2976.2	2.54	5.37	44.33	0.80
Series 2	56364.5	0.0	1261.8	3132.4	2.48	5.37	46.58	0.79
Series 3	51935.6	0.0	1307.6	3191.8	2.44	5.32	45.67	0.81
Series 4	43664.5	0.0	1237.9	3176.2	2.57	4.90	34.97	0.79
Series 5	49254.0	0.0	1292.5	3241.5	2.51	4.85	34.61	0.80
Series 6	36082.6	0.0	1218.3	2888.1	2.37	4.48	29.13	0.78
Series 7	50407.3	0.0	1098.7	2759.8	2.51	5.05	39.69	0.79
Series 8	46471.8	0.0	1280.3	3258.5	2.55	4.91	34.97	0.80
Series 9	52971.6	0.0	1283.0	3071.0	2.39	4.89	37.92	0.78
Series 10	41723.5	0.0	1282.2	3041.6	2.37	4.69	32.62	0.79
Max. of 10 samp.	56364.5	0.0	1307.6	3258.5	2.57	5.37	46.58	0.81
Min. of 10 samp.	36082.6	0.0	1098.7	2759.8	2.37	4.48	29.13	0.78
Mean of 10 samp.	47515.5	0.0	1243.5	3073.7	2.47	4.98	38.05	0.79

historical data. The highest flows in the synthetic sequences are also lower while the r_1 is comparable. The reproduction of overall parameters (Table 5.6) seems to be poorer than the LAR model (Table 5.3).

The behavior of AFS was further examined by generating two samples of length 30 and 50 years by MSNM. The statistics of AFS of historical data and synthetic data generated by MSNM are presented in Table 5.7.

Table 5.7

Comparison of statistics of AFS of historical data and synthetic data generated by MSNM

Series	Stat. Para.							
	Max.	Min.	Mean	S.D.	C_v	C_s	C_k	r_1
Historical	63800	10132	27790.8	13788.2	0.50	1.04	3.8	0.111
AFS of 30 years generated data	46279	12467	23995.5	8550.3	0.36	0.89	3.4	0.128
AFS of 50 years generated data	62734	1243	25962.9	11058.6	0.43	1.36	5.2	0.198

It can be seen from Table 5.7 that MSNM is closely reproducing the statistical parameters of AFS in the generated data of length 30 years and 50 years. The performance of MSNM (Table 5.7) seems to be better than LAR model (Table 5.4). The C_v

of AFS in generated data is lower as compared to AFS of historical data.

The further discussion on results presented in Tables 5.6 and 5.7 is given in next section where inter-comparison of LAR and MSNM models has been presented.

5.4 Inter-comparison of Linear Autoregressive and Modified Shot Noise Models for Daily Flow Generation

The LAR model though preserves the overall statistical parameters of daily flows yet does not reproduce the characteristics of AFS in the generated data. The MSNM preserves the characteristics of AFS but gives poorer performance in reproducing the overall statistical parameters of daily flows. To further explore the suitability of these models, the inter-comparison of these models has been attempted in this section. The inter-comparison has been made on the basis of reproduction of statistical parameters of (i) daily mean flcws (overall), (ii) AFS and (iii) flood related characteristics above a particular threshold. The details of comparison are as follows:

5.4.1 Reproduction of overall statistical parameters of daily flows

The mean of overall statistical parameters of daily flows of 10 samples of length 30 years each generated by LAR model and MSNM are presented in Table 5.8. This table has been derived from Table 5.3 and 5.6.

Table 5.8

Overall statistical parameters of daily flows generated by the two models

Para.	Hist. series	Mean of 10 samples. LAR	MSNM
Max.	63800.000	62589.000	47515.500
Min.	3.087	1.593	0.000
Mean	1249.600	1237.300	1243.500
S.D.	3242.600	3168.600	3073.700
C_v	2.595	2.558	2.470
C_s	6.794	6.740	4.982
C_k	76.131	78.998	38.048
r_1	0.822	0.945	0.792

The LAR model gives better reproduction of maximum flow, minimum flow, standard deviation, C_v, C_s and C_k as compared to MSNM. The MSNM is better in reproducing r_1. The overall mean in both the models is comparable. Based on reproduction of overall statistical parameters of daily flows, the performance of LAR model is found to be better than MSNM. However the performance of MSNM is also satisfactory.

5.4.2 Reproduction of statistical parameters of AFS

The statistical parameters of AFS from 2 samples of length 30 and 50 years generated by LAR model and MSNM are presented in Table 5.9 (combination of Table 5.4 and 5.7) along with parameters of AFS from historical data.

Table 5.9

Statistics of AFS of historical data and synthetic data generated by the two models

Stat. para.	Hist. series	LAR 30 years	MSNM 30 years	LAR 50 years	MSNM 50 years
Max.	63800.000	51485.000	46279.400	65695.600	62734.300
Min.	10131.800	4552.100	12467.300	4552.100	1243.300
Mean	27790.800	16733.400	23995.500	18271.500	25962.900
S.D.	13788.200	11754.000	8550.300	13331.400	11058.600
C_v	0.496	0.702	0.356	0.729	0.425
C_s	1.042	1.565	0.888	1.522	1.364
C_k	3.819	5.305	3.406	5.473	5.211
r_1	0.111	-0.291	0.128	-0.248	0.198

In 30 years sample MSNM is better reproducing mean annual flood, C_v , C_s and C_k as compared to LAR model. Almost similar picture emerges from 50 years sample. Broadly, it can be concluded that MSNM is better in reproducing the statistical parameters of AFS. The C_v of AFS of data generated by MSNM is low as compared to AFS of historical data.

5.4.3 Reproduction of flood characteristics above a particular threshold

Various flood features were explained in chapter III. In this section, the two models have been compared on the basis of reproduction of these flood characteristics. Both the models were used to generate the daily flows of 30 years length and only the flood events above a typical threshold of 10,000 cumecs were considered. This threshold was chosen for the sake of illustration and the analysis can be extended to other threshold levels also on the similar lines. The comparison of models has been made on the basis of reproduction of parameters of following characteristics:

(i) number of flood events

(ii) timing of flood peaks

(iii) flood peaks, volumes and duration, and

(iv) largest flood peaks, volumes and durations

Number of Flood Events

The statistical parameter of number of flood events above 10,000 cumecs threshold for historical data and data generated by two models are given in Table 5.10. The computed and critical χ^2 values for the Poisson distribution for number of flood events are also given in this table.

Table 5.10

Statistical parameters of number of flood events in a year above 10000 cumecs threshold from historical data and synthetic data generated by the two models

Stat. para.	Hist. series	LAR	MSNM	Better model
Max.	7	4	6	MSNM
Min.	0	0	1	Comp
Total events	91	39	101	MSNM
Mean	2.935	1.300	3.367	MSNM
S.D.	1.672	1.512	1.520	MSNM
Cv	0.569	1.160	0.451	MSNM
X^2 comp.	2.158	26.717	3.726	MSNM
X^2 cri.	19.700	14.100	21.000	

Note: Comp stands for comparable.

The number of flood events for different years follow Poisson distribution in historical data, which is exhibited only by MSNM. LAR model fails to reproduce this in the generated data, and it also fails to reproduce the total number of flood events in the generated data.

Timing of Flood Peaks

In the historical data all the flood peaks occur during five months of monsoon season. In the data generated by LAR and MSNM also, all the flood peaks are occurring in monsoon season. Based on this, it is difficult to establish the superiority of one model over the other.

Flood Peaks, Volumes and Durations

The statistical parameters of flood peaks, volumes and duration above 10,000 cumecs threshold from historical data and synthetic data of length 30 years generated by two models are given in Table 5.11. The table also gives computed and critical χ^2 values at 95% probability level for exponential distribution.

The LAR model is better in reproducing the parameters of flood peaks above 10,000 cumecs threshold as compared to MSNM. The LAR model though reproduces parameters of flood peaks, yet is unable to generate desired number of peaks, which is a serious draw back of LAR model.

The reproduction of statistical parameters of flood volumes and duration is much better in case of MSNM . The other details of comparison are given in Table 5.11.

Largest Flood Peaks, Volumes and Durations

Reproduction of statistical parameters of largest flood peaks in different years, i.e. AFS, has already been discussed in section 5.4.2. This section covers only the largest flood volumes and durations. The statistical parameters of largest flood volumes (annual flood volume series) and largest flood durations (annual flood duration series) from historical data and synthetic

Table 5.11

Statistics of flood peaks, volumes and durations above 10000 cumecs threshold from historical data and synthetic data of length 30 years generated by the two models

Stat. para.	Hist.	Flood peak LAR	MSNM	Better
Max.	53800.00	41485.00	36279.40	LAR
Min.	649.80	935.20	1058.30	LAR
Mean	10702.00	8631.30	7807.80	LAR
S.D.	10591.90	9618.60	6667.30	LAR
Cv	0.99	1.11	0.85	LAR
Cs	1.92	2.05	1.97	Comp
r1	0.01	-0.12	-0.08	Comp
χ^2 comp.	5.07	5.46	8.99	Comp
χ^2 cri.	9.49	7.81	9.49	
No. of events	91	39	101	MSNM

Table 5.11 Cont.

Stat. para.	Hist.	Flood volume LAR	MSNM	Better	Hist.	Flood dur. LAR	MSNM	Better
Max.	158007.8	536663.8	127388.8	MSNM	8.29	33.49	9.40	MSNM
Min.	546.9	537.3	547.8	MSNM	0.60	0.94	0.57	MSNM
Mean	19390.6	44689.3	13677.0	MSNM	2.77	6.55	2.50	MSNM
S.D.	27705.5	99432.0	22532.4	MSNM	1.69	7.45	1.77	MSNM
Cv	1.43	2.22	1.65	MSNM	0.61	1.13	0.71	MSNM
Cs	2.96	3.82	3.40	MSNM	1.36	2.26	1.61	MSNM
r1	0.05	-0.09	-0.05	Comp	0.09	-0.09	-0.08	Comp
χ^2 comp.	10.15	22.07	21.85	Comp	9.50	10.4	1.92	MSNM
χ^2 cri.	9.49	7.81	9.49		9.50	7.81	9.49	

data of length 30 years generated by two models are presented in Table 5.12. The Table also gives computed and critical χ^2 values at 95% probability levels for distribution functions of annual flood volume series and annual flood duration series represented by Eq. (4.16) and (4.17).

Table 5.12

Statistics of largest flood volumes and durations above 10000 cumecs threshold from historical data and synthetic data of length 30 years generated by two models

Stat. para.	Largest flood volume Hist.	LAR	MSNM	Better model	Larg. flood dura. Hist.	LAR	MSNM	Better model
Max.	158008	536664	127389	MSNM	8.29	33.49	9.40	MSNM
Min.	2630	2769	1301	Comp	1.34	1.74	0.90	MSNM
Mean	38366	97883	32412	MSNM	4.04	11.40	4.25	MSNM
S.D.	39135	140765	33646	MSNM	1.99	9.47	2.05	MSNM
C_v	1.02	1.43	1.03	MSNM	0.49	0.83	0.48	MSNM
C_s	1.79	2.37	1.77	MSNM	0.88	1.26	0.68	MSNM
χ^2comp.	4.33	9.12	4.33	Comp	2.66	16.50	4.67	MSNM
χ^2cri.	14.10	14.10	14.10		14.10	14.10	14.10	

It can be seen from Table 5.12 that the maximum and mean annual flood volume are much higher in data generated by LAR model. These are roughly 3 times of the corresponding values in historical data. These are quite close to historical parameters in case of MSNM. The reproduction of other parameters also is better in case of MSNM.

The statistical parameters of annual flood duration series are much better reproduced by MSNM. LAR gives a largest flood duration of 33.49 days which is much away from 8.29 days given by historical data and the realistic values of flood durations. Based on the reproduction of statistical parameters of annual flood volume and duration series, the MSNM is much better than LAR model.

The summary of inter-comparison based on various criteria is presented in Table 5.13.

Table 5.13

Summary of inter-comparison of two models

	Criterion	Better model
<u>Reproduction of statistical parameter of</u>		
(i)	Daily flows (overall parameters)	LAR
(ii)	Annual flood series	MSNM
(iii)	No. of flood events	MSNM
(iv)	Timing of flood peaks	Comparable
(v)	Annual flood duration series	MSNM
(vi)	Annual flood volume series	MSNM
(vii)	Flood peaks above 10000 cumecs thres.	Comparable
(viii)	Flood volumes above 10000 cumecs thres.	MSNM
(ix)	Flood durations above 10000 cumecs thres.	MSNM

Broadly it can be concluded from Table 5.13 that MSNM gives better reproduction of the statistical parameters of (a) number of flood events (b) annual flood series, (c) annual flood duration series, (d) annual flood volume series, (e) flood durations, and (f) flood volumes. The reproduction of overall statistical parameters of daily flows and flood peaks is also satisfactory in case of MSNM.

5.5 Inferences

The methodology and application of two schemes of daily flow generation viz. linear autoregressive (LAR) and shot noise models have been described in the previous sections. The suitability of these models for reproducing the flood related characteristics in the generated data has also been examined. The inferences drawn from the application and inter-comparison of the two models and the limitations of the study follow in subsequent text.

Linear Autoregressive Model

For LAR model Case II (logarithmic transformation in the beginning and exponentiation in the end) with normally distributed independent stochastic component was found to be better than other considered cases and distributions. The model closely reproduces the overall statistical parameters of daily flows.

The use of Wilson- Hilferty transformation for generating PT III distributed random numbers holds good only when C_s is around 2. The performance of this transformation deteriorates as C_s deviates from 2.

Shot Noise Model

The modified shot noise model (MSNM), which account for the different behavior of the river during monsoon and nonmonsoon seasons, requires only 18 parameters as compared to 36 parameters of the original shot noise model.

The three methods as suggested in NERC (1975) to incorporate C_s in the shot noise model were found to be unsuitable for the data of river Narmada at Garudeshwar.

Inter comparison of LAR Model and MSNM

The MSNM gives better reproduction of statistical parameters of: (i) number of flood events, (ii) annual flood series, (iii) annual flood duration series, (iv) annual flood volume series, (v) flood durations, and (vi) flood volumes.

The reproduction of overall statistical parameters of daily flows and flood peaks above a threshold of 10,000 cumecs is satisfactory in case of LAR model as well as MSNM. However, the LAR model gives comparatively better reproduction of overall statistical parameters.

Limitations

The inter-comparison of the two models has been made on the basis of only one sample of 30 years length and only one threshold level of 10,000 cumecs. In order to conclude about the general suitability and superiority of one of these models, further studies with a number of data sets for different sites would be necessary.

The MSNM has been used to generate long term data of 1500 years for validation of the methodology in Chapter VI.

CHAPTER VI

VALIDATION AND APPLICATION OF THE METHODOLOGY

6.1 General

The validation of the methodology of multivariate modelling of flood flows is described in this chapter. This chapter also gives details of suitability of the methodology for analysis of historical data and results of application for flood estimation for river Narmada at Garudeshwar. The usefulness of this type analysis is illustrated with the help of one example.

6.2 Validation of the Methodology

The methodology for stochastic modelling of flood flows has been validated using synthetically generated daily flows of 1500 years length. The daily flows have been generated using modified shot noise model (MSNM) which was found to be satisfactory in preserving the characteristics of flood flows in the generated data as discussed in the Chapter V.

For validation of the methodology use has been made of long term data of 1500 years instead of historical data because of the following two reasons: (i) the short term can be used to verify the return periods associated with various events only up to the length of data and (ii) the short term data may not have enough number of critical combinations of various events.

The details of the strategy adopted and the results of this study are presented in subsequent sections.

6.2.1 The strategy

From the long term generated data the return periods associated with various events have been obtained using counting procedure i.e. by the definition of return period. The return periods associated with the same events have also been obtained by applying the methodology to the long term data. The difference between the two estimates of return periods is used as a measure to decide about the applicability of a methodology. If the difference is small enough, the methodology is assumed to be validated, otherwise not. The drawback of this type of strategy is obviously the subjectivity involved in judging the difference as small or large.

6.2.2 Validation Results

The return periods of the following events have been estimated by two methods viz. counting procedure and by application of the methodology.

(a) exceedance of x (univariate) in a year,

(b) exceedance of y (univariate) in a year,

(c) exceedance of at least one of the values (x,y) in a year and

(d) exceedance of both x and y in a year.

For the above events x is the value of flood volume (X) and y the value of flood peak (Y).

The estimated return periods using counting procedure have been denoted by T while those estimated by the application of the methodology by T′. For estimating the return periods, in this validation exercise, the range of flood volumes (X) was taken from 60,000 cumecdays to 150,000 cumecdays with an increment of 6,000 cumecdays and flood peaks from 10,000 cumecs to 46,000 cumecs with an increment of 2000 cumecs. The threshold was kept as 10,000 cumecs. The above values were selected to cover a wider range of return periods associated with corresponding events.

The return periods associated with some selected events only are presented in Table 6.1 as an illustration since tabulation of all the results will be too voluminous. However return periods T and T′ for various events are plotted in (i) Fig. 6.1 for exceedance of x (univariate) in a year, (ii) Fig. 6.2 for exceedance of y (univariate) in a year, (iii) Fig. 6.3 for exceedance of at least one of the values (x,y) in a year and (iv) Fig. 6.4 for exceedance of both x and y in a year.

The difference of the two return periods i.e. bias in Table 6.1 is small enough and indicates that the return periods associated with the various events can be estimated quite accurately by applying the methodology. In Fig. 6.1 through 6.4 also the points for various return periods T and T′ fall on 45° line. However the values of T′ i.e. the return periods estimated by the methodology, are underestimated for events ($Y \geq y$) and ($X \geq x$ and $Y \geq y$), for higher return periods as shown in Figs. 6.2 and 6.4 respectively. The reason for this was investigated by comparing the Cv of historical peak discharges and peak discharges generated by MSNM. The value of Cv of peak discharges by MSNM was found to be somewhat lower. Because of this, MSNM was not able to produce enough variability in the generated peak discharges specially in the higher range. The underestimation of T′ seems to be

Table 6.1

Return periods for some selected events using counting procedure (T) and by applying the methodology (T′) to long term (1500 years) data

Event	T	T′	Bias (T′-T)
X ≥60000	6.9	7.6	0.7
X ≥90000	18.3	21.1	2.8
X ≥120000	46.9	51.3	4.4
X ≥150000	115.4	112.7	-2.7
Y ≥10000	1.4	1.5	0.1
Y ≥20000	4.0	4.3	0.3
Y ≥30000	13.9	15.8	1.9
Y ≥40000	55.6	55.3	-0.3
X ≥60000 OR Y ≥10000	1.4	1.4	0.0
X ≥90000 OR Y ≥20000	4.0	4.3	0.3
X ≥120000 OR Y ≥30000	13.2	15.8	2.5
X ≥150000 OR Y ≥40000	42.9	52.0	9.1
X ≥60000 AND Y ≥10000	6.9	7.6	0.7
X ≥90000 AND Y ≥20000	18.5	21.1	2.6
X ≥120000 AND Y ≥30000	57.7	52.9	-4.8
X ≥150000 AND Y ≥40000	125.0	113.2	-11.8

Note: X is flood volume in cumecdays and Y is flood peak in cumecs.

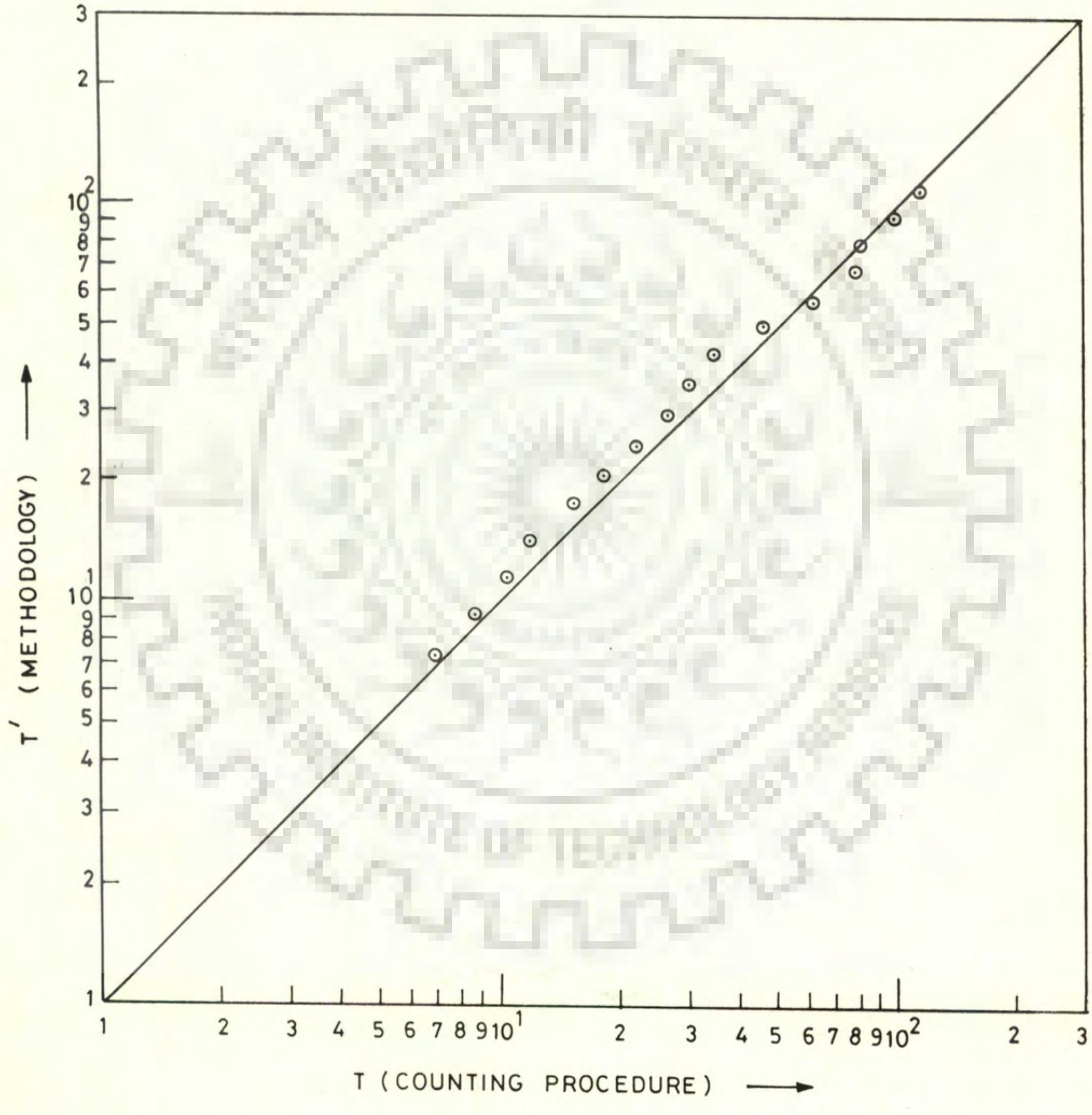

Fig 6·1 – Plot between T and T′ for (X ≥ x) in a year

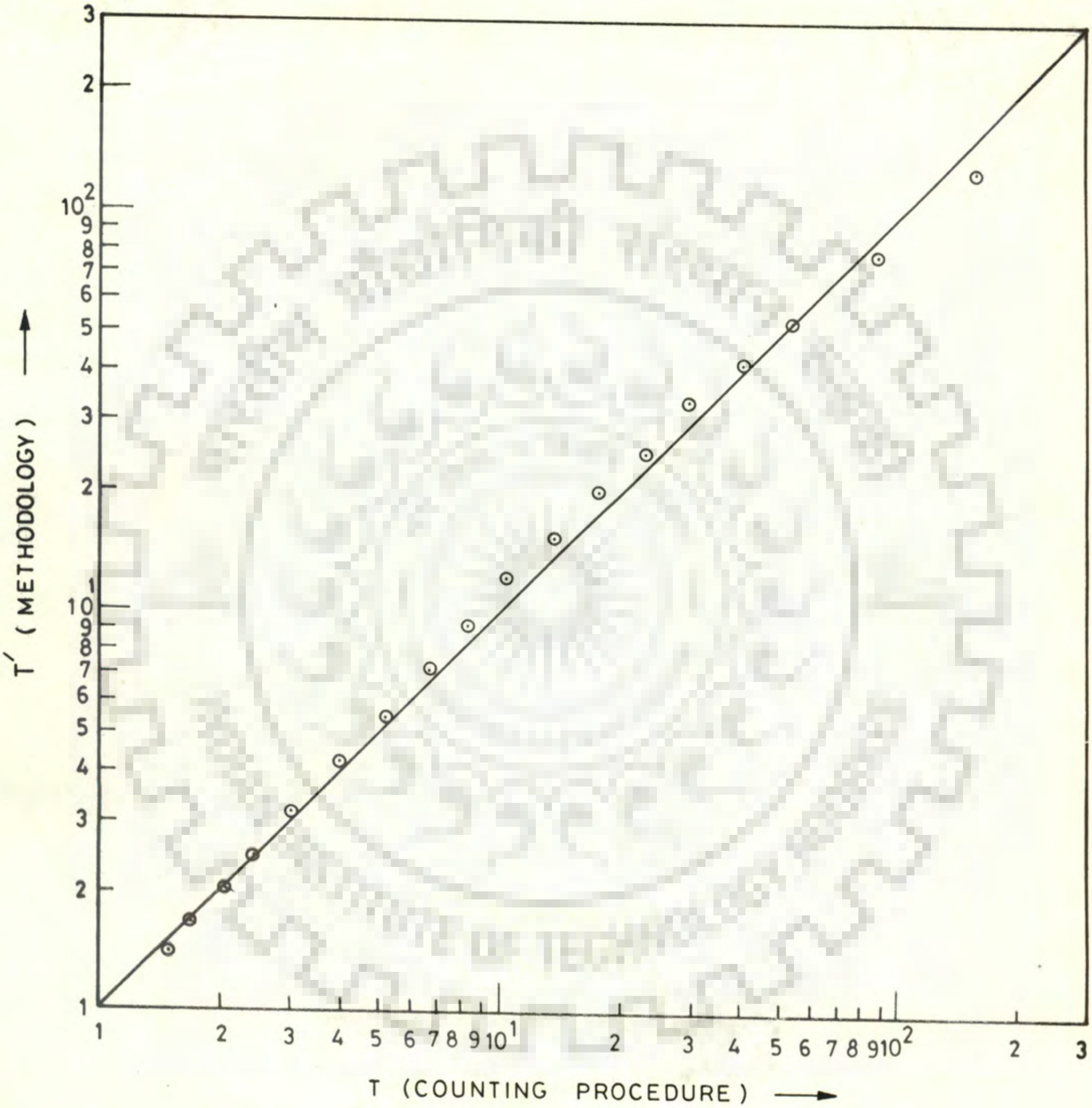

Fig. 6.2 - Plot between T and T′ for (Y ≥ y) in a year

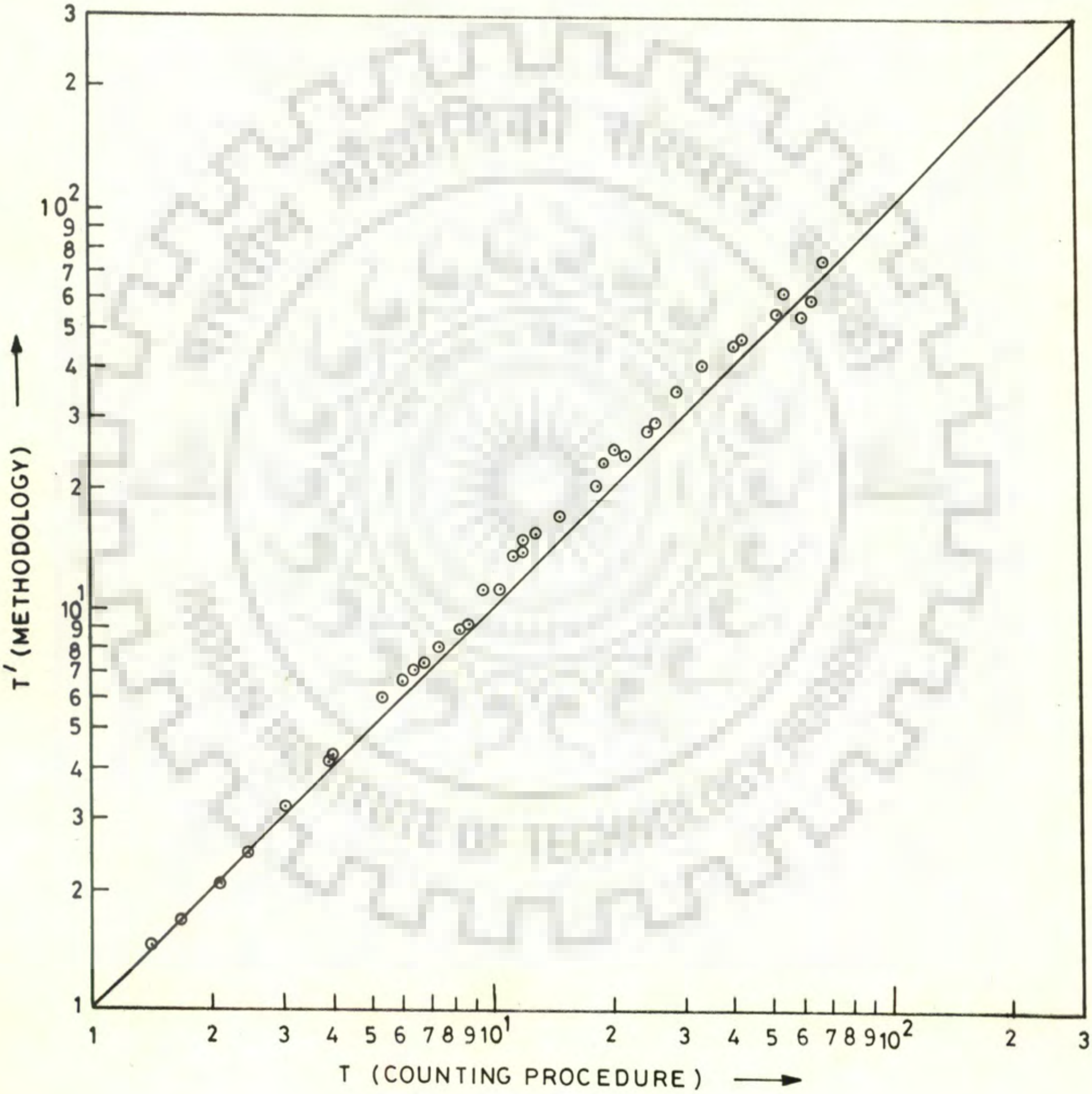

Fig. 6·3 – Plot between T and T′ for (X ≥ x or Y ≥ y) in a year

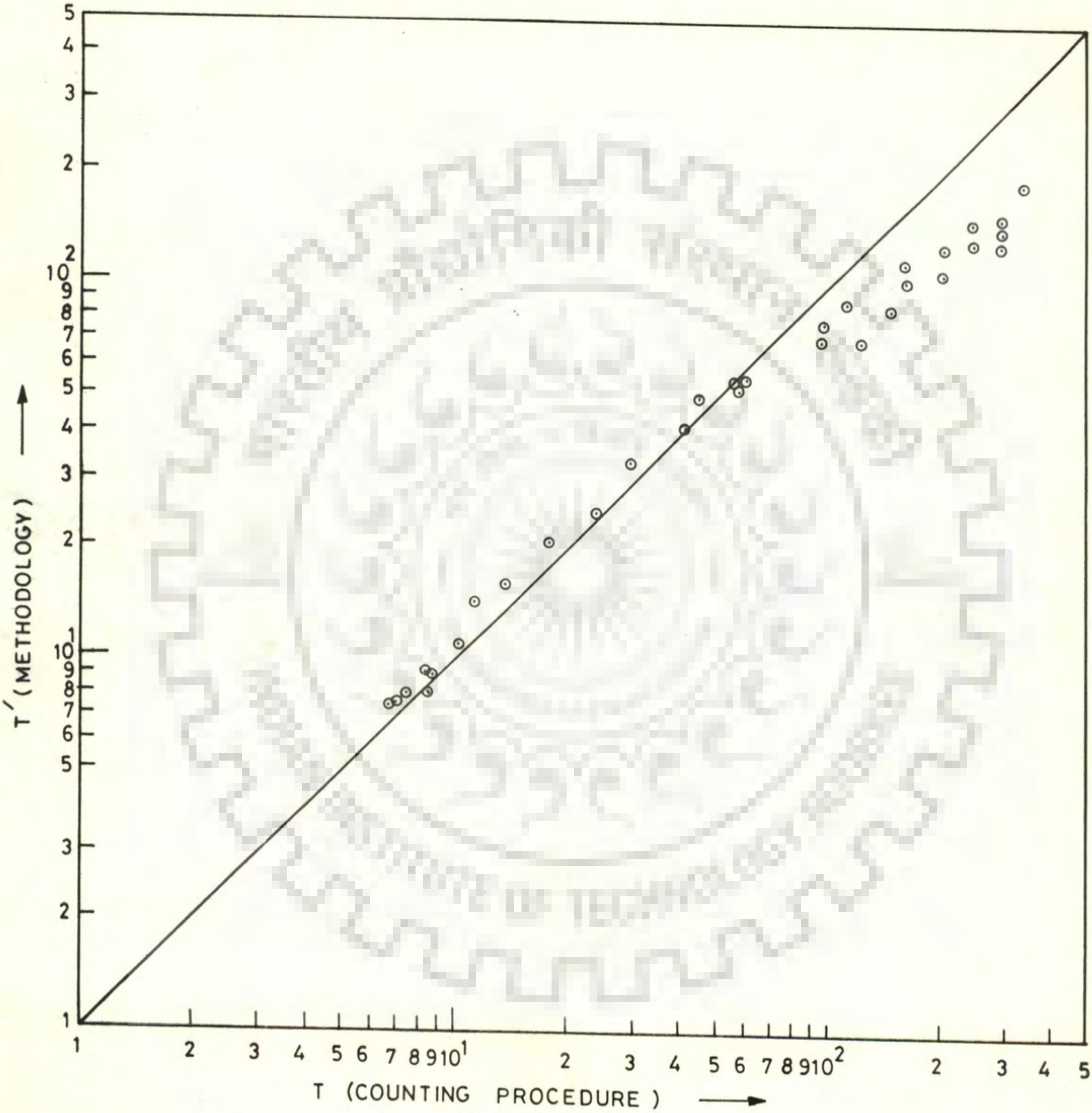

Fig. 6·4– Plot between T and T′ for (X ≥ x and Y ≥ y) in a year because of this limitation of MSNM.

Based on results as presented in Table 6.1 and Figs. 6.1 through 6.4 the methodology developed in the present study has been validated satisfactorily.

6.3 Suitability and Application of the Methodology

The methodology for stochastic modelling of flood flows has got the flexibility of selecting (i) any threshold level which can be fixed from engineering consideration as long as the Poissonian assumption for number of flood events in a year is satisfied, (ii) any volume of a flood event which can also be decided on the basis of probable maximum precipitation and by adopting a suitable loss rate and (iii) any flood peak magnitude.

The methodology was applied to daily discharge data of Narmada at Garudeshwar and the selection of various variables was made as follows:

Selection of Threshold Level

As mentioned in Chapters II and IV, the following points are to be considered while selecting the threshold level:

(a) The threshold level should be high enough to ensure (i) number of flood events in a year is a non increasing function of threshold and (ii) the flood events are random.

(b) The threshold level should be low enough so that λ is greater than or equal to 1.65 and

(c) The number of flood events in a year follow Poisson distribution.

It was shown in Chapter IV that for daily flow series of river Narmada at Garudeshwar, conditions a and c are met if the threshold level is in the range of 4000 to 18000 cumecs. However, for λ to be greater than or equal to 1.65, the threshold level should not be higher than 14000 cumecs. Keeping in view these considerations, for the application of the methodology, any threshold level between 4000 to 14000 cumecs can be selected.

For the sake of typical illustration the application of the methodology has been presented for the threshold level of 10000 cumecs (λ = 3.0).

Selection of Flood Volume

A number of attempts have been made by several agencies particularly Indian Institute of Tropical Meteorology, Pune and Dam Design Review Panel in consultation with the World Bank experts to determine the PMP for the basin (NIH, 1985). Some of the storms which have been considered for this purpose are detailed below.

1. 1970/73 storm with 1.35 uniform moisture maximization factor giving a PMP of 413.76 mm,
2. 1927 transposed storm centered in the catchment with 1.35 uniform moisture maximization factor giving a PMP of 634.7 mm and
3. 1926/27 storm progressing with 1.35 moisture maximization factor giving a PMP of 618.9 mm.

All the above storms were assumed to be of five days duration.

It is thus seen that the highest PMP depth, ever considered so far, is 634.7 mm. Assuming a loss rate of 0.75 mm/hr this is expected to give a effective rainfall depth of 544.7 mm. The volume of water corresponding to 544.7 mm comes out to be 554700 cumecdays (the catchment area being 88,000 km^2). In other words, the value of 5.6 x 10^5 cumecdays is an indication of the highest limit on flood volume which could be considered in the analysis from engineering point of view.

To cover a wider range, the flood volumes were varied from 100,000 to 560,000 cumecdays with an increment of 10,000 cumecdays.

Selection of Flood Peaks

As mentioned in Chapter IV, the PMF estimated by various agencies varied from 1.076 x 10^5 cumecs to 2.21 x 10^5 cumecs. The flood peaks were varied from 10,000 to 220,000 cumecs with an increment of 10,000 to cover a wider range of possibilities.

Applying the methodology (Subroutine DVPA and NKG), the return periods associated with the following events have been estimated for a threshold level of 10,000 cumecs.

(i) exceedance of x (univariate) in a year,

(ii) exceedance of y (univariate) in a year,

(iii) exceedance of $y|x$ in a year,

(iv) exceedance of at least one of the values (x,y) in a year and

(v) exceedance of both x and y in a year.

6.4 Results and Discussion

For the threshold level of 10,000 cumecs, there are 91 flood events. The mean, std. and Cs of the flood volume series are 19390.6, 27705.6 and 2.96. These parameters corresponding to flood peak series are 10702.0, 10591.9 and 1.92 respectively. The other parameters are given in Table 5.11 in Chapter V.

The values of λ and ν in TSPT to normalize flood volume series are 0.048 and 1.34. For flood peak series these values are 0.0883 and 1.22 respectively.

Using Eq. (3.6) to (3.9), 50, 100, 500, 1000 years return period flood volumes and peaks over a threshold of 10,000 cumecs were obtained. These are shown in Table 6.2. This table also gives the flood peaks for corresponding return periods obtained by the application of Wakeby - 4 parameter distribution to AFS for the sake of comparison.

Table 6.2

Various return period flood volumes and flood peaks over a threshold of 10000 cumecs

S. No.	Return period (years)	Flood volume (cumecdays)	Flood peak magnitude (cumecs) Above 10000 thres.	Absolute	Wakeby-4 distr. for AFS
1	50	146358	54414	64414	66387
2	100	177234	63070	73070	73816
3	500	258509	84482	94482	87666
4	1000	297884	94306	104306	92746

It can be seen from Table 6.2 that flood peak magnitudes for various return periods given by the methodology and by Wakeby - 4 parameter distribution are close enough. The deviation even for 1000 years return period flood is only 11%.

The return periods associated with the exceedance of various flood volumes and flood peaks are shown in Table 6.3. The same are plotted in Fig. 6.5 for flood volumes and in Fig. 6.6 for flood peaks respectively.

The return periods associated with events (i) $X \geq x$ or $Y \geq y$, (ii) $X \geq x$ and $Y \geq y$ and (iii) $Y \geq y \mid X = x$ for some selected flood volumes and flood peaks are shown in Table 6.4. The same are plotted in Figs. 6.7 to 6.9.

The developed methodology is capable of doing univariate as well as bivariate modelling of flood flows. The methodology offers various possibilities of probability interpretation which conventional flood frequency methods do not. Figs. 6.5 to 6.9 can be used to answer many of the additional questions required in the hydrologic design. The use of these figures is illustrated with the help of a typical example as explained below:

Example: Suppose the designer is interested in knowing the following quantities above 10000 cumecs threshold:

1. 1000 years return period flood volume.

In Fig. 6.5 1000 years return period flood volume will be corresponding to 0.001 prob. of exceedance as 298000 cumecdays.

2. The return period associated with flood volume $\geq$ 400,000 cumecdays.

Table 6.3

The return periods of various flood volumes and flood peaks above 10,000 cumecs threshold

Flood peak (10^3 cumecs)	Return period (years)	Flood volume (10^3 Cumecdays)	Return period (years)
10	1.5	100	15.8
20	2.8	110	20.5
30	6.3	120	26.5
40	15.0	130	33.9
50	34.8	140	43.1
60	78.5	150	54.5
70	171.2	160	68.4
80	361.5	170	85.5
90	740.9	180	106.3
100	1477.8	190	131.4
110	2876.0	200	161.6
120	5473.4	210	198.0
130	10205.1	220	241.5
140	18672.5	230	293.3
150to 220	≥ 30000	240	355.0
		250	428.0
		260	514.2
		270	615.7
		280	735.0
		290	874.7
		300	1038.0
		310	1228.3
		320	1449.7
		330	1706.7
		340	2004.3
		350	2348.3
		360	2745.0
		370	3201.9
		380	3727.0
		390	4329.5
		400	5019.0
		410	5808.5
		420	6708.5
		430	7735.0
		440	8905.7
		450	10230.0
		460 to 560	≥ 10000

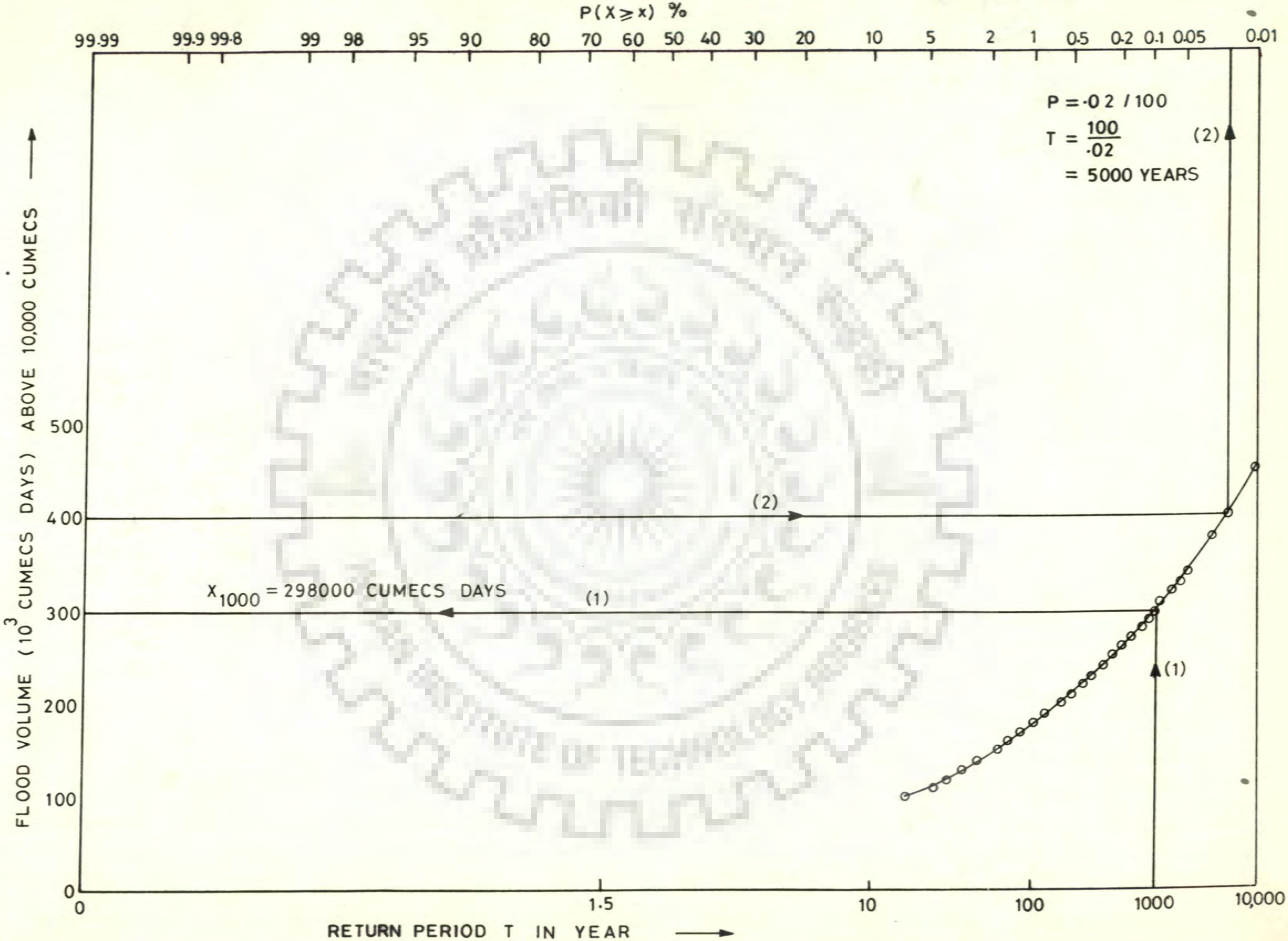

Fig. 6.5 - Plot of flood volumes above 10,000 cumecs and (P X ≥ x)

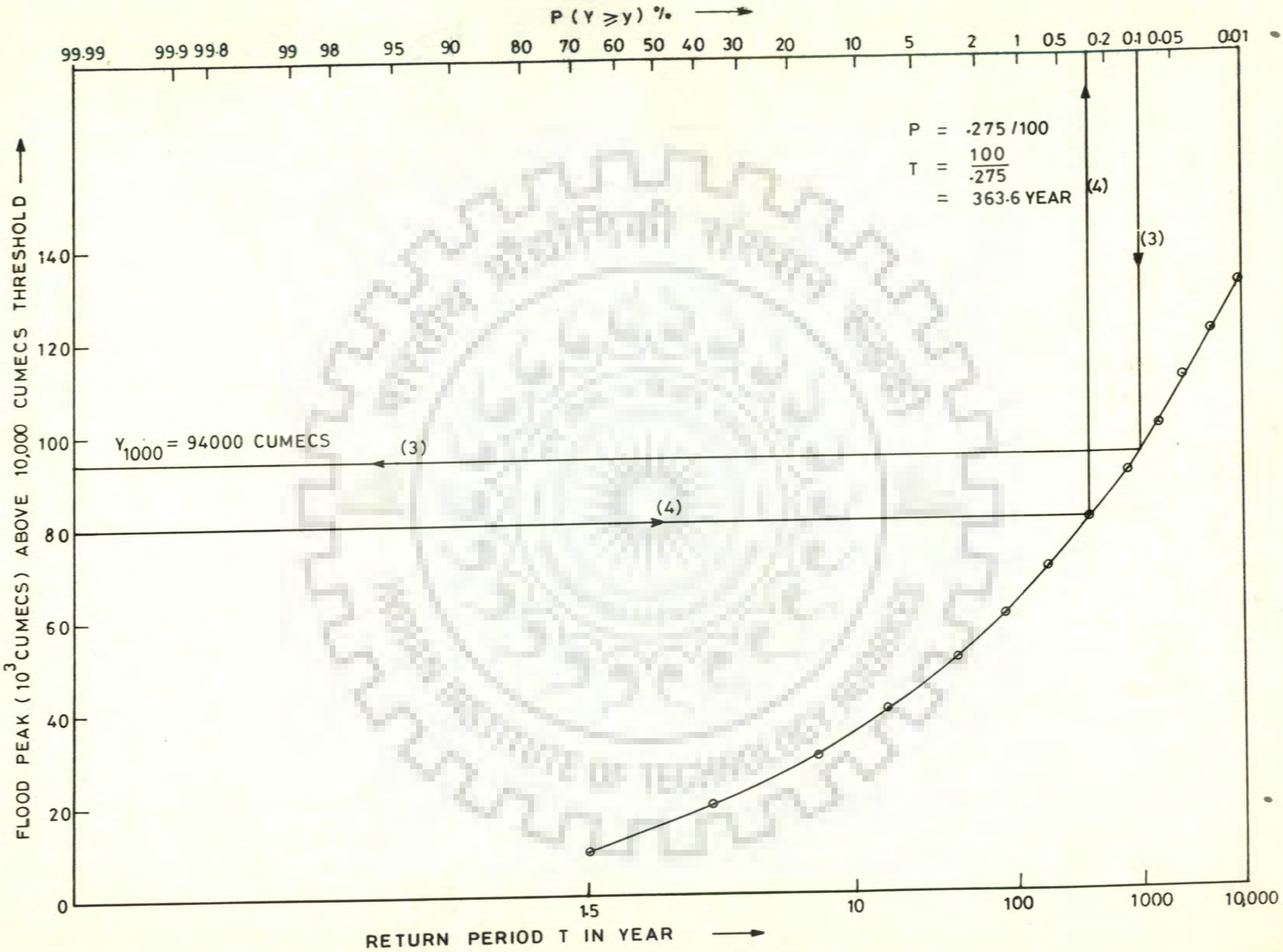

Fig. 6·6-Plot of flood peaks above 10,000 cumecs and P (Y ≥ y)

Table 6.4

Return periods for some selected events

S. No.	Flood volume (cumecdays)	Flood peak (cumecs)	Return period of event X≥x or Y ≥y	X ≥x and Y ≥y	Y≥y|X=x
1	200000	40000	15.0	162.7	1.1
2	200000	60000	71.2	204.9	1.2
3	200000	80000	146.9	466.5	3.6
4	200000	90000	158.0	829.8	10.6
5	200000	100000	160.8	1548.1	39.4
6	300000	40000	15.0	1040.3	1.1
7	300000	60000	78.4	1057.0	1.1
8	300000	80000	337.0	1311.2	1.2
9	300000	90000	579.3	1703.6	1.5
10	300000	100000	810.8	2458.5	2.3
11	400000	40000	15.0	5066.3	1.1
12	400000	60000	78.5	5049.6	1.1
13	400000	80000	360.5	5224.7	1.1
14	400000	90000	729.1	5636.1	1.1
15	400000	100000	1383.4	6532.2	1.2
16	500000	40000	15.0	21053.8	1.1
17	500000	60000	78.5	20481.9	1.1
18	500000	80000	361.4	20481.9	1.1
19	500000	90000	739.9	20802.5	1.1
20	500000	100000	1469.3	22742.7	1.1

Note: X is flood volume and Y is flood peak.

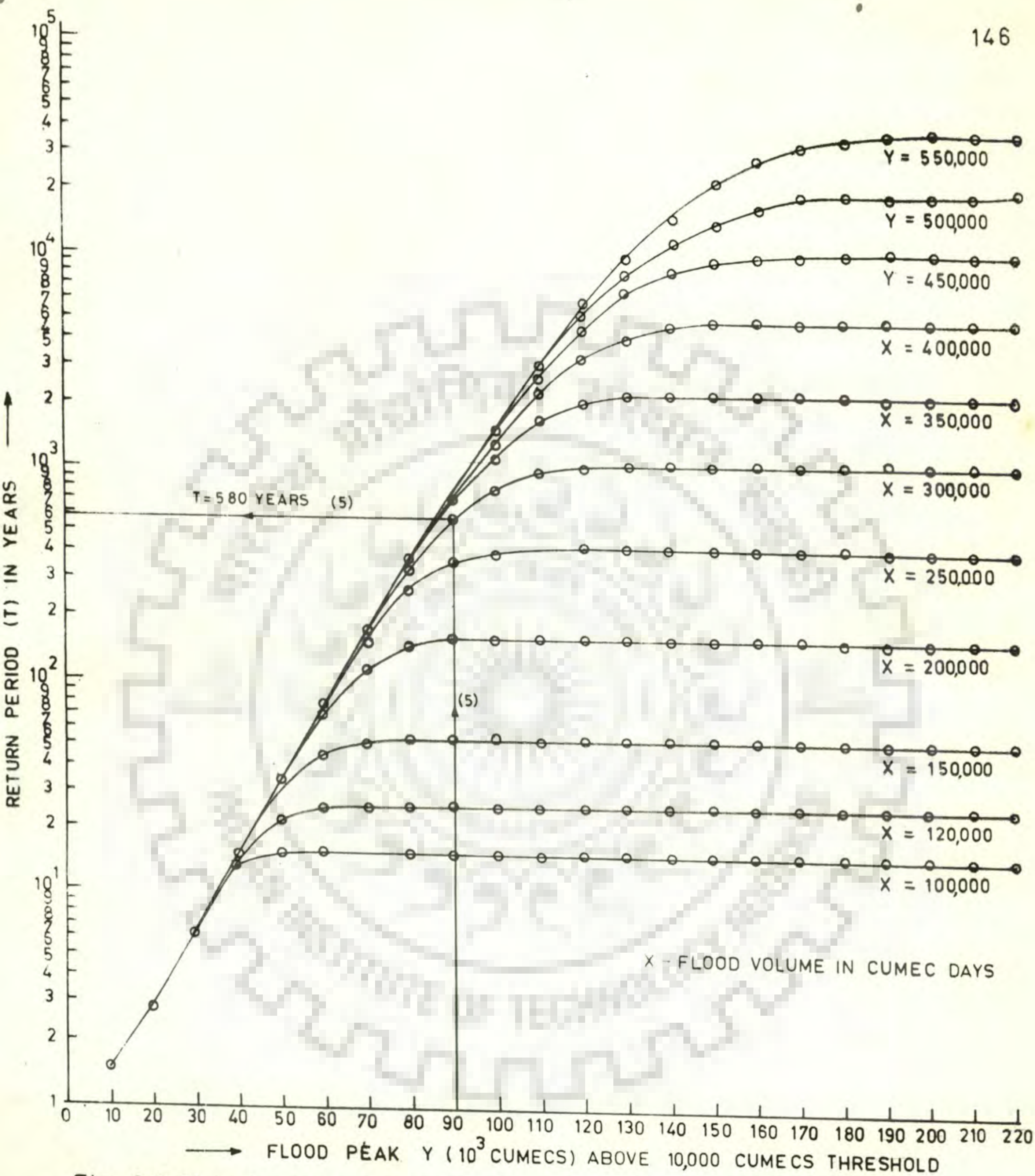

Fig. 6·7-Plot of return period T associated with the event ($X \geq x$ or $Y \geq y$) with several flood peaks (Y) and flood volumes (X)

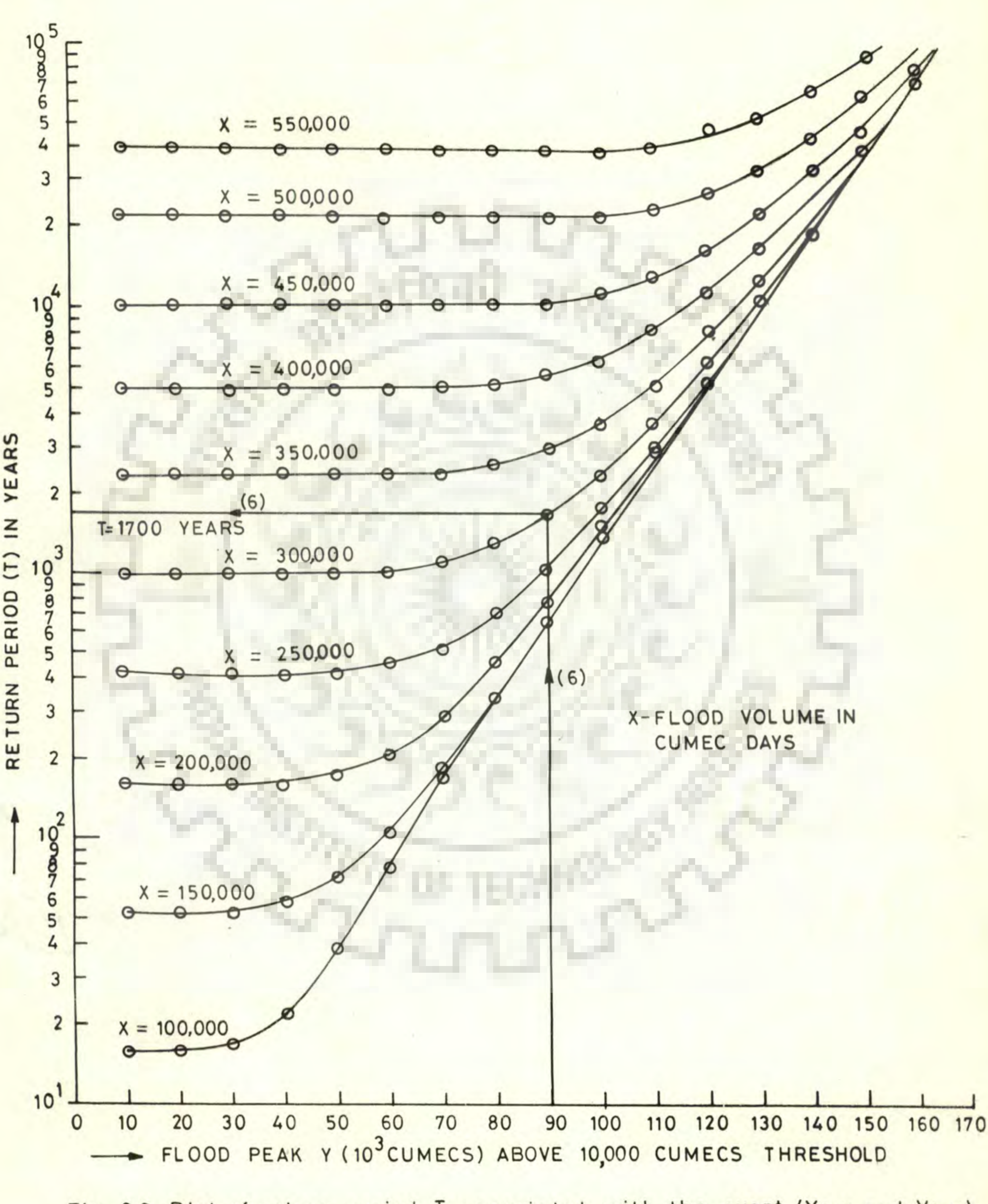

Fig. 6.8 - Plot of return period T associated with the event ($X \geq x$ and $Y \geq y$) with several flood peaks (Y) and flood volumes (X)

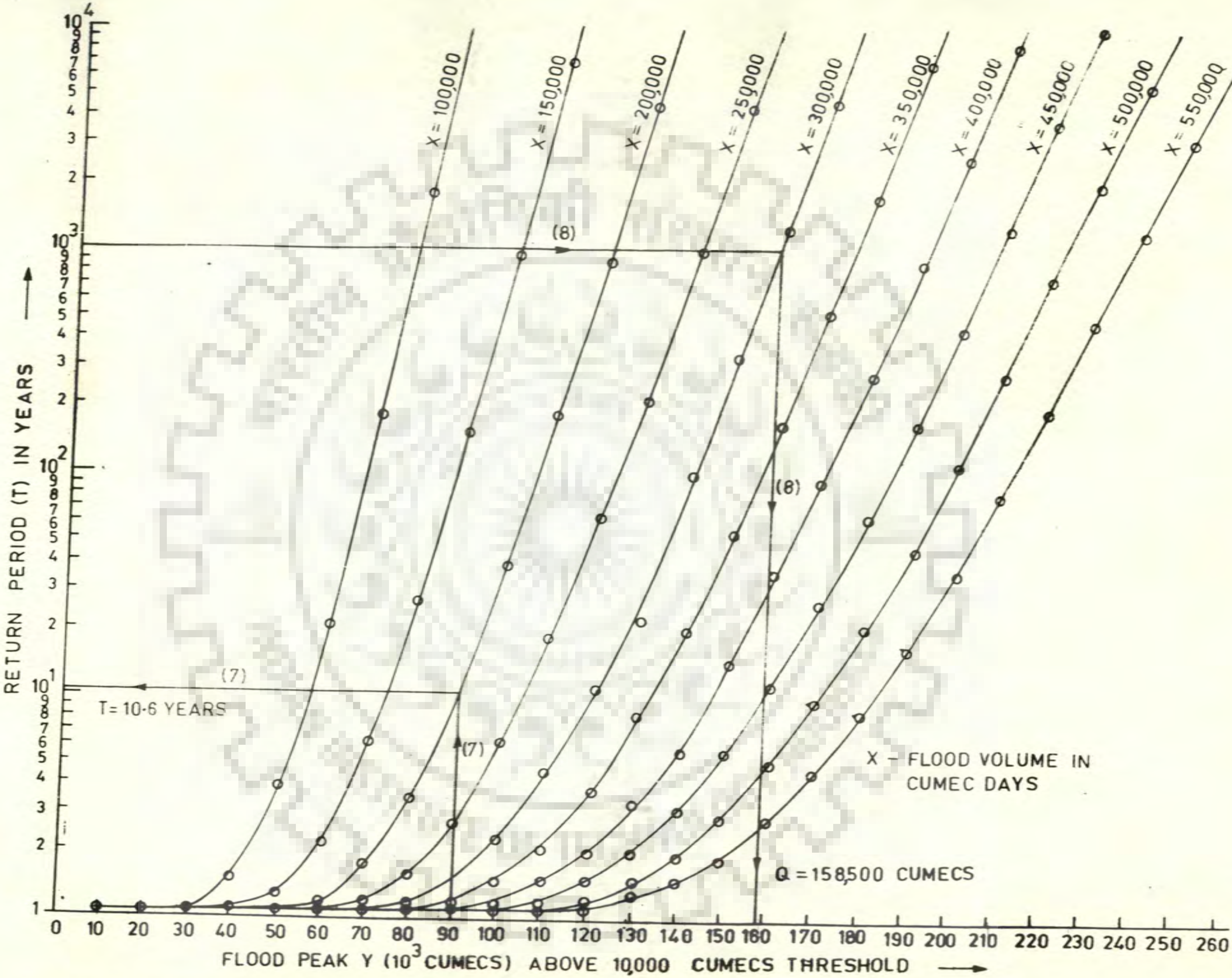

Fig. 6.9 - Plot of return period T associated with the event (Y ≥ y | X = x) with several flood peaks (Y) and flood volumes (X)

In Fig. 6.5 the prob. of exceedance corresponding to 400,000 cumecdays flood volume is 0.0002 giving a return period of 5000 years.

3. 1000 years return period flood peak.

In Fig. 6.6 1000 years return period flood peak will be corresponding to 0.001 prob. of exceedance as 94,000 cumecs.

4. The return period associated with flood peak ≥ 80,000 cumecs.

In Fig. 6.6 the prob. of exceedance corresponding to 80,000 cumecs flood peak is 0.00275 giving a return period of 363.6 years.

5. The return associated with flood peak ≥ 90,000 cumecs or flood volume ≥ 300,000 cumecdays,

Fig. 6.7 gives the return period of event (X ≥ 300,000 or Y ≥ 90,000) as 580 years.

6. The return associated with flood peak ≥ 90,000 cumecs and flood volume ≥ 300,000 cumecdays,

Fig. 6.8 gives the return period of event (X ≥ 300,000 and Y ≥ 90,000) as 1700 years.

7. The return period associated with flood peak ≥ 90,000 cumecs, given a flood volume of 200,000 cumecdays.

Fig. 6.9 gives the return period of (Y≥ 90,000|X=200,000) as 10.6 years.

8. 1000 years return period flood peak given a flood volume of 300,000 cumecdays.

Fig. 6.9 gives the flood peak as 158500 cumecs.

Similar type of answers can be obtained for other flood peaks and flood volumes also with the help of the developed methodology.

The methodology developed can be applied to other river basins also on the similar lines. However, for small basins smaller interval discharge data should be used to obtain various flood features.

6.5 Inferences

The details of validation and application of the methodology of multivariate stochastic modelling of flood flows have been presented in this chapter. On the basis of the analysis carried out in the chapter, the following inferences can be drawn.

(i) The flood peak magnitudes for various return periods given by the methodology and by fitting Wakeby- 4 parameters distribution to AFS, are close enough. This indicates the suitability of two step power transformation (TSPT) method and partial duration series approach in flood frequency analysis.

(ii)The developed methodology is capable of doing univariate as well as bivariate modelling of flood flows. The methodology offers various possibilities of probability interpretation and considers flood event as a whole, which is an improvement over conventional flood frequency methods.

CHAPTER VII

CONCLUSIONS

In the present study a systematic methodology for stochastic modelling of flood flows has been developed. The methodology is based on partial duration series approach. It uses bivariate normal distribution as the parent bivariate distribution function for the two dependent variables such as flood peaks and flood volumes. The normalization of the variables required for the application of bivariate normal distribution is done using two step power transformation.

The inter-comparison of two daily flow models namely linear autoregressive and shot noise models has also been attempted in the study to ascertain their suitability for Indian rivers, where most of the flow is concentrated in the five months of monsoon season. The validation of the methodology has then been done using synthetically generated daily flows by modified shot noise model after verifying its performance in reproducing flood characteristics in generated data. The daily flow data from 1949-79 for river Narmada at Garudeshwar have been used in the study.

Various conclusions drawn from review of literature formed general basis for the development of the methodology. These have been given in section 2.5 of Chapter II. A number of SUBROUTINES were developed during different stages of the study. Based on the study and analysis of flood characteristics using daily discharge data of river Narmada at Garudeshwar, the following conclusions were drawn:

1. For flood frequency analysis using annual flood series Wakeby/PWM performs better than GEV/PWM, EV1/PWM and LPIII/MOM distributions.

2. The number of flood events in a year follow Poisson distribution and once the applicability of this is established for a particular threshold, it remains valid for higher threshold levels also.

3. The exponential distribution fits well to the flood peaks, flood volumes and flood durations for most of the threshold levels. However, the general applicability of this distribution for all the threshold levels is not confirmed.

4. The distribution functions of largest flood peaks, largest flood volumes and largest flood durations as given by Eq. (4.15) to (4.17) fit well the data for most of the threshold levels.

5. Among various cases and distributions considered for linear autoregressive model for daily flow generation, the Case II (i.e. logarithmic transformation in the beginning and exponentiation in the end) with normally distributed independent stochastic component performs better. It closely reproduces the overall statistical parameters of daily flows.

6. In synthetic daily flow generation, the modified shot noise model (MSNM) gives better reproduction of statistical parameters of: (i) number of flood events, (ii) annual flood series, (iii)

annual flood duration series, (iv) annual flood volume series, (v) flood durations, and (vi) flood volumes above a particular threshold, as compared to linear autoregressive model. The reproduction of overall statistical parameters of daily flows, and flood peaks above a particular threshold is also satisfactory in MSNM.

The conclusions given above are site specific. Besides these, the following broad conclusions were also drawn in different stages of the study. These conclusions are expected to be valid for other sites also.

1. The use of bivariate normal distribution is quite adequate for modelling two components among peak, volume and duration of flood events, at a time, as the third variable can be estimated with reasonable accuracy by assuming triangular relationship between these.

2. The two step power transformation is an improvement over original Box-Cox transformation for normalization of a given series, since it preserves C_s and C_k of normal distribution in the transformed series.

3. The use of Wilson- Hilferty transformation for generating PTIII distributed random numbers holds good only when the value of C_s is around 2.0. The performance of this transformation deteriorates as C_s deviates from 2.0.

4. The modified shot noise model (MSNM) requires only 18 parameters as compared to 36 parameters of shot noise model. This model preserves flood related characteristics in the generated

data. The model is suitable for daily flow generation for rivers having floods in monsoon season only.

5. The use of daily mean flow data for obtaining flood related characteristics is applicable only for large basins (wherein the time of concentration is at least more than one day). For small basins shorter interval data should be used for obtaining flood characteristics. The use of bivariate normal distribution and TSPT would remain applicable for this also.

6. The methodology developed in the study is capable of univariate as well as bivariate modelling of flood characteristics. This offers various possibilities of probability interpretation which conventional flood frequency methods do not. The methodology has been validated using long term synthetic data.

Though the study has focussed on flood flows, the methodology can be extended to other dependent hydrological variables such as drought related characteristics, sediment yield and runoff, and many others. Modelling of flood flows by considering more than two variables, at a time, is further possible direction of work in this area. Stochastic analysis of floods using principle of maximum entropy is another area, in which further work is required to make it mathematically simpler.

REFERENCES

1. Acreman, M.C. and Sinclair, C.D. (1986), 'Classification of drainage basins according to their physical characteristics; An application for flood frequency analysis in Scotland', Journal of Hydrology, Vol. 84, pp.365-380.

2. Adamowski, K. (1985), 'Nonparametric kernel estimation of flood frequencies', Water Resources Research, Vol. 21/11, pp.1585-1590.

3. Ahmad, M.I., Sinclair, C.D. and Werritty, A. (1988), 'Log - logistic flood frequency analysis', Journal of Hydrology, Vol. 98, pp.205-224.

4. Amorocho, J. and Espildora, B. (1973), 'Entropy in assessment of uncertainty in hydrologic systems and models', Water Resources Research, Vol. 9/6, pp.1511-1522.

5. Anderson, R.L. (1941), ' Distribution of serial correlation coefficients', Annals of Math. Statist., Vol.8/1, pp.1-13.

6. Arnell, N. and Gabrielle, S. (1985), 'Regional flood frequency analysis with the two-component extreme value distribution: an assessment using computer simulation experiments', Workshop on Combined Efficiency of Direct and Indirect Estimations for Point and Regional Flood Prediction, Perugia, Italy.

7. Ashkar, F. and Rousselle, J. (1982), 'A multivariate statistical analysis of flood magnitude, duration and volume', in V.P. Singh (Editor), 'Statistical Analysis of Rainfall and Runoff'.

8. Ashkar, F. and Rousselle, J. (1983a), ' Some remarks on the truncation used in partial duration flood series models', Water Resources Research, Vol. 19/2, pp.477-480.

9. Ashkar, F. and Rousselle, J. (1983b), 'The effect of certain restrictions imposed on the inter-arrival times of floods', Water Resources Research, Vol. 19/2, pp.481-485.

10. Baker, V.R. (1985), 'Palaeoflood hydrology of extraordinary flood events', Paper pres. at US-China Bilateral Symposium on the Analysis of Extraordinary Flood Events, Nanjing.

11. Benson, M.A. (1950), 'Use of historical data in flood frequency analysis', Trans. Amer. Geophys. Union, Vol. 31/3, pp.419-424.

12. Benson, M.A. (1968), 'Uniform flood frequency estimating methods for federal agencies', Water Resources Research, Vol. 4/5, pp.891-908.

13. Beran, M.A. and Nozdryn-Plotnicki, M.J. (1977), 'Estimation of low return period floods', Hydrological Sciences Bull., Vol. 22, pp.275-281.

14. Beran, M.A., Hosking, J.R.M. and Arnell, N. (1986), 'Comment on 'Two component extreme value distribution for flood frequency analysis' by Rossi et al. (1984)', Water Resources Research, Vol. 22/2, pp.263-266.

15. Bergman, L.E. (1961), 'The frequency distribution of near extremes', J. of Geophysics Research, Vol. 66/10, pp.3295-3307.

16. Berman, S.M. (1962), 'Limiting distribution of the maximum term in sequences of dependent random variables', Ann. of Math. Statist., Vol. 33, pp.894-904.

17. Bernier, J. (1967a), 'Sur le theorie du renouvellement et son application en hydrologie', Electricite de France, HYD, Vol. 67/10.

18. Bernier, J. (1967b), 'Methodes bayesiennes en hydrologie statistique', Proc. Intl. Hydrol. Symp., Fort Collins, Colorado, USA, pp.459-470.

19. Bethlahmy, N. (1977), 'Flood analysis by SMEMAX transformation', Journal of the Hydraulics Division, ASCE, Vol. 103/HY1, pp.69-80.

20. Bobee, B. (1973), 'Sample error of T Year events computed by fitting a Pearson Type 3 distribution', Water Resources Research, Vol. 9/5, pp.1264-1270.

21. Borgman, L.E. (1963), 'Risk criteria', Jour. Waterways and Harb. Div., ASCE, Vol. 89, pp.1-35.

22. Box, G.E.P. and Cox, D.R. (1964), 'An analysis of transformation', Jour. of the Royal Stat. Society, Vol. 26, pp.211-252.

23. Box, G.E.P. and Tiao, G.C. (1973), 'Bayesian inference in statistical analysis', Addison Wesley publishing Co., New York, N.Y., pp.156-160.

24. Buishand, T.A. (1989), 'The partial duration series method with a fixed number of peaks', Journal of Hydrology, Vol. 109, pp.1-9.

25. Burges, S.J., Lettenmaier, D.P. and Bates, C.L. (1975), 'Properties of the three-parameter lognormal distribution', Water Resources Research, Vol. 11/2, pp.229-235.

26. Cervantes, J.E., Kavvas, M.L. and Delleur, J.W. (1983), 'Cluster model for flood analysis', Water Resources Research, Vol. 19/1, pp.209-224.

27. Chander, S., Spolia, S.K. and Kumar, A. (1978), 'Flood frequency analysis by power transformation', Journal of the Hydraulics Division, ASCE, Vol. 104, pp.1495-1504.

28. Chen Jia-Qi, Ye Yong-yi and Ten Wei (1975), 'The important role of historical flood data in the estimation of spillway design floods', Scientia Sinica, Vol. 18/5, pp.669-680.

29. Chow, V.T. (1950), 'Discussion of "Annual floods and the partial duration series" by W.B. Langbein', Trans. Am. Geophys. Union, Vol. 31/6, pp.939-941.

30. Chow, V.T. (1951), 'A general formula for hydrologic frequency analysis', Trans. Amer. Geophys. Union, Vol. 32, pp.231-237.

31. Chow. V.T. (1953), 'Frequency analysis of hydrologic data with special application to rainfall intensities', Univ. Illinois Eng. Expt. St., Bull., Vol. 414.

32. Cole, G. (1966), 'An Application of the Regional Analysis of Flood Flows', Institn. Civil Engineers, Symposium on River Flood Hydrology, pp.39-57.

33. Condie, R. (1977), 'The Log Pearson type 3 distribution: The T-year event and its asymptotic standard error by maximum likelihood theory', Water Resources Research, Vol. 13/6, pp.987-991.

34. Condie, R. and Lee, K.A. (1983), 'Flood frequency analysis with historic information', Journal of Hydrology, Vol. 58, pp.47-62.

35. Correia, F.N. (1983), 'Metodos de analise e determinacao de candais de cheia', Laboratorio Nacional de Engenharia Civil, Lisboa.

36. Correia, F.N. (1987), ' Multivariate partial duration series in flood risk analysis' in V.P. Singh (Editor), ' Hydrologic Frequency Modelling', Reidel, Dordrecht, pp.541-554.

37. Cunnane, C. (1973),' A particular comparison of annual maximum and partial duration series methods of flood frequency estimation', Journal of Hydrology, Vol. 18, pp.257-273.

38. Cunnane, C. (1973), 'A particular comparison of annual maximum and partial duration series methods of flood frequency estimation', Journal of Hydrology, Vol. 18, pp.257-273, with ERRATA Vol. 19, pp.377.

39. Cunnane, C. and Nash, J.E. (1974), 'Bayesian estimation of frequency of hydrologic events', IAHS Sci. Publ. Vol. 100, pp.47-55.

40. Cunnane, C. (1979), 'A note on the Poisson assumption in partial duration series models', Water Resources Research., Vol. 15/2, pp.489-494.

41. Cunnane, C. (1984), 'Condition of separation of skewness of random samples from the general extreme value distribution', Report, Deptt. of Engineering hydrology, University College Galway, pp.7.

42. Cunnane, C. (1987), 'Review of statistical model for flood frequency estimation' in V.P. Singh (Editor), 'Hydrologic Frequency Modelling', Reidel, Dordrecht, pp.49-95.

43. Cunnane, C. (1988), 'Methods and merits of regional flood frequency analysis', Journal of Hydrology, Vol. 100, pp.269-290.

44. CWC (1972), 'Estimation of design flood - recommended procedure' publication of Central Water Commission, India.

45. Dalrymple, T. (1960), 'Flood frequency Methods', USGS water supply paper no. 1543A, pp.11-51, Washington.

46. Davis, D.R., Kisiel, C.C. and Duckstein, L. (1972), 'Bayesian decision theory applied to design in hydrology', Water Resources Research, pp 33-41.

47. Fiorentino, M. and Gabriele, S. (1984), 'A correction for the bias of maximum likelihood estimators of Gumbel parameters', Journal of Hydrology, Vol. 73, pp.39-50.

48. Foster, H.A. (1924), 'Theoretical frequency curves and their application to engineering problems', Trans. ASCE, Vol. 87, pp.142-173.

49. Fuller, W.E. (1914), 'Flood Flows', Trans. Am. Soc. Civl. Engrs., Vol. 77, pp.564-617.

50. Goel, N.K., Seth, S.M. and Chandra S. (1987),' Design flood estimation for Narmada Sagar project using partial duration series - a case study', in V.P.Singh (Editor), 'Application of Frequency and Risk in Water Resources', Reidel, Dordrecht, pp.113-122.

51. Goel, N.K. and Seth, S.M. (1988), ' Comparative study of different parameter estimation techniques for EV-1 distribution', Proceedings of International Seminar on ' Hydrology of Extremes', Roorkee, India, pp.53-63.

52. Goel, R.S. (1989), 'Objective comparison of annual maximum series, peaks over threshold series and time series methods of flood frequency prediction', M.Sc., thesis, Deptt. of Engineering Hydrology, University College, Galway, pp.104.

53. Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979), 'Probability weighted moments: definition and relation to parameters of distributions expressible in inverse form', Water Resources Research, Vol. 15/5, pp.1049-1054.

54. Greis, N.P. and Wood, E.F. (1981), 'Regional flood frequency estimation and network design', Water Resources Research, Vol. 17/4, pp.1167-1177.

55. Gumbel, E.J. (1941), 'The return period of flood flows', Ann. Math. Statist., Vol. 12/2, pp.163-190.

56. Gumbel, E.J. (1958), 'Statistics of extremes', Columbia Univ. Press., 375 pp.

57. Gupta, D.K., Asthana, B.N. and Bhargava, A.N. (1989),' Flood frequency analysis by two step power transformation', Journal of Institution of Engineers (India), Vol. 70, pp.127-131.

58. Gupta, V.K., Duckstein, L. and Peebles, R.W. (1976), 'On the joint distribution of the largest flood and its time of occurrence', Water Resources Research, Vol. 12/2, pp.295-304.

59. Hall, M.J. and O'Connel, P.E. (1972), 'Time series analysis of mean daily river flows', Water and Water Engg., Vol. 76, pp.125-133.

60. Hastings, C. (1955), ' Approximations for digital computers', Princeton University Press, Princeton, N.J.

61. Hazen, A. (1932), 'Flood flows', John Wiley, New York, 199 pp.

62. Hoshi, K. and Burges, S.J. (1981), 'Sampling properties of parameter estimates for the log Pearson Type 3 distribution using moments in real space, Journal of Hydrology, Vol. 53, pp.305-316.

63. Hosking, J.R.M. and Wallis, J.R. (1984), 'Palaeoflood hydrology and flood frequency analysis', EOS, Vol. 65/45, pp.890.

64. Hosking, J.R.M., Wallis, J.R., and Wood, E.F. (1985a), 'An appraisal of the regional flood frequency procedure in the UK flood studies report', Hydrol. Sci. J., Vol. 30/1, pp.85-109.

65. Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985b), 'Estimation of the generalized extreme value distribution by the method of probability weighted moments', Technometrics, Vol. 27/3, pp.251-261.

66. Houghton, J.C. (1978a), 'Birth of a parent: The Wakeby distribution for modelling flood flows', Water Resources Research, Vol. 14/6, pp.1105-1109.

67. Houghton, J.C. (1978b), 'The incomplete means estimation procedure applied to flood frequency analysis', Water Resources Research, Vol. 14/6, pp.1111.

68. Hua, Shi-Qian (1985), 'A general survey of flood frequency analysis in China', Paper pres. at US-China Bilateral Symposium on the Analysis of Extraordinary Flood Events, Nanjing.

69. Jaynes, E.T. (1957), 'Information theory and statistical mechanics - Part I', Physical Review, Vol. 106, pp.620 - 630, Part II, Vol. 108, pp.171 - 191.

70. Jenkinson, A.F. (1955), 'The frequency distribution of the annual maximum (or minimum) values of meteorological elements', Quart. J. Roy. Meteor. Soc., Vol. 81, pp.158-171.

71. Jenkinson, A.F. (1969), 'Statistics of extremes. In: Estimation of maximum floods', WMO Tech. Note no. 98, pp.183-228.

72. Jing, D., Song, D., Yang, R. and Hon, Y. (1989), 'Expressions relating probability weighted moments to parameters of several distributions inexpressible in inverse form', Journal of Hydrology, Vol. 110, pp.259-270.

73. John, J.A. and Draper, N.R. (1980),' Alternative family of transformations', Journal of Applied Statistics, Vol. 29, pp.190-197.

74. Kaczmarek, Z. (1957), 'Efficiency of the estimation of floods with a given return period', Proc. Toronto Sympos., IAHS Public. No. 45, pp.145-159.

75. Kavvas, M.L. and Delleur, J.W. (1975), 'The stochastic and chronological structure of rainfall sequences : application to Indiana', Tech. Report 57, Water Resources Research Center, Purdue University, West Lafayette, Indiana.

76. Kavvas, M.L. (1982), 'Stochastic trigger model for flood peaks : 1. Development of the model, 2. Application of the model', Water Resources Research, Vol. 18/2, pp.383-411.

77. Kavvas, M.L., Cervantes, J.E. and Delleur, J.W. (1983), 'A cluster model for flood analysis', Water Resources Research, Vol. 19/1, pp.209-224.

78. Kavvas, M. L. and Delleur, J.W. (1984), ' A statistical analysis of the daily streamflow hydrograph', Journal of Hydrology, Vol. 71, pp.253-275.

79. Kelman, J. (1980), ' A stochastic model for daily streamflow', Journal of Hydrology, Vol. 47, pp.235-249.

80. Kendall, G.R. (1959), 'Statistical analysis of extreme events', Nat. Res. Council, Canada, Ass. Comm. Geod. Geophysics, Proc. Hydrol. Symp., pp.54-78, Ottawa, Canada.

81. Kimball, B.F. (1946), 'Sufficient statistical estimation functions for the parameters of the distribution of maximum values', Ann. Math. Statist., Vol. 17, pp.299- .

82. Kite, G.W. (1977), 'Frequency and risk analysis in hydrology', Water Resources Publications, Fort Collins, Co 80522, USA, pp.2

83. Kottegoda, N.T. (1980), ' Stochastic water resources technology', Macmillan, New York.

84. Kottegoda, N.T. and Horder, M.A. (1980), ' Daily flow model based on rainfall occurrences using pulses and a transfer function', Journal of Hydrology, Vol. 47, pp.215-234.

85. Krstanovic, P.F. and Singh, V.P. (1987), 'A multivariate stochastic flood analysis using entropy', in V.P. Singh (Editor), 'Hydrologic Frequency Modelling', Reidel, Dordrecht, pp.515-539.

86. Kuczera, G. (1982a), 'Combining site-specific and regional information: An empirical Bayes approach', Water Resources Research, Vol. 18/2, pp.306-314.

87. Kuczera, G. (1982b), 'Robust flood frequency models', Water Resources Research, Vol. 18/2, pp.315-324.

88. Kuczera, G. (1983), 'A Bayesian surrogate for regional skew in flood frequency analysis', Water Resources Research, Vol. 19/3, pp.821-832.

89. Kumar, A. (1982), ' Hydrologic time series modelling -an overview', State of Art Report, SA-1, National Institute Hydrology, Roorkee.

90. Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1978), 'Some comparisons of Flood Statistics in Real and Log Space', Water Resources Research, Vol. 14/5, pp.902-920.

91. Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979a), 'Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles', Water Resources Research, Vol. 15/5, pp.1055-1064.

92. Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979b), 'Estimation of parameters and quantiles of Wakeby distribution', Water Resources Research, Vol. 15/6, pp.1361-1379.

93. Langbein, W.B. and others (1947), 'Topographic characteristics of drainage basins', U.S. Geol. Surv. Prof. Paper 968-C, pp.125-157.

94. Langbein, W.B. (1949), 'Annual floods and the partial duration flood series', Trans. Am. Geophys. Union, Vol. 30, pp.879-881.

95. Leese, M.N. (1973), 'Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series', Water Resources Research, Vol. 9/6, pp.1534-1542.

96. Lettenmaier, D.P. and Potter, K.W. (1985), 'Testing flood frequency estimation methods using a regional flood generating model', Water Resources Research, Vol. 21/12, pp.1903-1914.

97. Lettenmaier, D.P., Wallis, J.R. and Wood, E.F. (1985), 'Note on the comparative robustness of estimates of extreme flood quantiles', Presented at American Geophysical Union, Spring Meeting, Baltimore.

98. Lowery, M.D. and Nash, J.E. (1970), 'A comparison of methods of fitting the double exponential distribution', Journal of Hydrology, Vol. 10, pp.259-275.

99. Luo, Cheng-Zheng (1985), 'A survey of historical flood and its regionalisation in China', Paper pres. at US-China Bilateral Symposium on the Analysis of Extraordinary Flood Events, Nanjing.

100. Matalas, N.C. and Wallis, J.R. (1972), 'An approach to formulating strategies for flood frequency analysis', Proc. Int. Symp. on Uncertainties in Hydrologic and Water Resource Systems, Tucson, Arizona, pp.940-961.

101. Matalas, N.C. and Wallis, J.R. (1973), 'Eureka! It fits a Pearson type 3 distribution', Water Resources Research, Vol. 9/2, pp.281-289.

102. Matalas, N.C., Wallis, J.R. and Slack, J.R. (1975), 'Regional skew in search of a parent', Water Resources Research, Vol. 11/6, pp.815-826.

103. McMahon, T.A. and Srikanthan, R. (1981), 'Log Pearson type 3 distribution - is it applicable to flood frequency analysis of Australian streams?' Journal of Hydrology, Vol. 52, pp.139-147.

104. Mosley, M.P. (1981), 'Delimitation of New Zealand hydrological regions', Journal of Hydrology, Vol. 49, pp.173-192.

105. Nash, J.E. and Amorocho, J. (1966), 'The accuracy of the prediction of floods of high return periods', Water Resources Research, Vol. 2/2, pp.191-198.

106. Nash, J.E. and Shaw, B.L. (1966), 'Flood frequency as a function of catchment characteristics', Proc. Sympos. on River Flood Hydrology, Inst. Civ. Engnrs., London, pp.115-136.

107. NERC (1975), 'Flood studies report', Natur. Environ. Res. Council, London, Vols. 1-5, 1100pp.

108. NIH (1985), ' Design flood estimation for Narmada Sagar and Sardar Sarovar projects', Final Report, National Institute of Hydrology, Roorkee, India.

109. North, M. (1980), 'Time dependent stochastic model of floods', Journal of Hydraulics Division, ASCE, Vol. 106/HY5, pp.649-665.

110. Nozdryn-Plotnicki, M.J. and Watt, W.E. (1979), 'Assessment of fitting techniques for the log Pearson type 3 distribution using Monte Carlo simulation', Water Resources Research, Vol. 15/3, pp.714-718.

111. O'Connell, P.E. (1974), 'Shot noise model in synthetic hydrology', Proceedings of the workshop on ' Mathematical models for Surface Water Hydrology' Pisa, Italy.

112. O'Connell, P.E. and Jones, D.A. (1979), 'Some experience with the development of models for stochastic simulation daily flows', in Mc Bean et al. (editor), ' Inputs for Risk Analysis in Water Systems'.

113. O'Donnell, T., Hall, M.J. and O'Connell, P.E. (1972), 'Some applications of stochastic hydrological models', International Symposium on Modelling Techniques in Water Resources Systems, (Editor AK Biswas) Vol. 1, pp.227-239, Publ. by Environment, Canada, Ottawa.

114. Pang, K.D. (1985), 'Extraordinary floods in early Chinese history and their absolute dates', Paper pres. at US-China Bilateral Symposium on the Analysis of Extraordinary Flood Events, Nanjing.

115. Parzen, E. (1962), ' Stochastic processes', Holden Day, Sanfrancisco.

116. Payne, K. W., Newman, R. and Kerri, D. (1969), ' Daily streamflow simulation', Journal of Hydraulics Division, ASCE, Vol.95/HY4, pp.1163-1180.

117. Phien, H.N. and Lung-Cheng, Hsu (1984), 'Variance of the T-year event in the log Pearson Type 3 Distribution', Journal of Hydrology, Vol. 77, pp.141-158.

118. Quimpo, R.G. (1967), 'Stochastic model of daily flow sequences', Hydrology Paper No. 18, Colorado State Univ.

119. Rao, D.V. (1980), 'Log Pearson type III distribution, method of mixed moments', Journal of the Hydraulics Division, ASCE, Vol. 106, pp.999-1019.

120. Riggs, H.C. (1961), 'Frequency of natural events', Proc. ASCE, Journal of the Hydraulic Division, Vol. 87/HY1, paper 2706, pp.15-26.

121. Rosbjerg, D. (1985), 'Estimation in partial duration series with independent and dependent peak values', Journal of Hydrology, Vol. 76, pp.183-195.

122. Rossi, F., Fiorentino, M. and Versace, P. (1984), 'Two component extreme value distribution for flood frequency analysis', Water Resources Research, Vol. 20/7, pp.847-856.

123. Rousselle, J. (1972), 'On some problems of flood analysis', Ph.D. Thesis, C.S.U., Fort Collins, Colorado, 226 p.

124. Sackl, B. and Bergmann, H. (1987), 'A bivariate flood model and its application' in V.P.Singh (Editor), 'Hydrologic Frequency Modelling', Reidel, Dordrecht, pp.571-582.

125. Sangal, B.P. and Biswas, A.K. (1970), 'The three-parameter log normal distribution and its application in hydrology', Water Resources Research, Vol. 6/2, pp.505-515.

126. Sardar Sarovar Narmada Nigam (1989), 'Planning for prosperity : Sardar Sarovar development plan'.

127. Schroder, R. (1969), 'Haufigkeitsanalyse hydrologischer daten', Technischer Bericht Nr. 5 des Institutes fur Hydraulik und Hydrologie der Technischen Hochschule Darmstadt, Darmstadt.

128. Seth, S.M., Perumal, M., Goel, N.K. and Nirupma, P. (1983), ' Development and application of program on power transformation method to flood frequency analysis - a comparative study with other normalization procedures', Technical Report No. 13, National Institute of Hydrology, Roorkee, India.

129. Seth, S.M., Singh, R.D. and Perumal, M. (1986), ' Suitability of power transformation based on Gumbel EV-1 distribution for flood frequency analysis', Technical Report, Tr-2, National Institute of Hydrology, Roorkee, India.

130. Shane, R.M. and Lynn, W.R. (1964), 'Mathematical model for flood risk evaluation', Journal of the Hydraulics Division, ASCE, Vol. 90, pp.1-20.

131. Shannon, C.E. (1948), 'A mathematical theory of communication', Bell System Tech. Journal, Vol. 27, pp.379-423.

132. Singh, K.P. and Sinclair, R.A. (1972), 'Two distribution method for flood frequency analysis', Journal of the Hydraulics Division, ASCE, Vol. 98, pp.24 - 44.

133. Singh, V.P., Singh, K. and Rajagopal, A.K. (1985), 'Application of the principle of maximum entropy to hydrologic frequency analysis', Completion Report, Louisiana Water Resources Research Institute, Louisiana State University, Baton Rouge, Louisiana.

134. Singh, V.P. and Krstanovic, P.F. (1986), 'A multivariate stochastic flood analysis using entropy', Completion Report, Louisiana Water Resources Research Institute, Louisiana State University, Baton Rouge, Louisiana.

135. Singh, V.P. and Krstanovic, P.F. (1987), 'A stochastic model for sediment yield using the principle of maximum entropy', Water Resources Research, Vol. 23/5, pp.781-793.

136. Slack, J.R., Wallis, J.R. and Matalas, N.C. (1975), 'On the value of information to flood frequency analysis', Water Resources Research, Vol. 11/5, pp.629-647.

137. Sonuga, J.O. (1972), 'Principle of maximum entropy in hydrologic frequency analysis', Journal of Hydrology, Vol. 27, pp.177-191.

138. Sonuga, J.O. (1976), 'Entropy principle applied to the rainfall - runoff process', Journal of Hydrology, Vol. 30, pp.81-94.

139. Stedinger, J.R. (1980), 'Fitting log normal distributions to hydrologic data', Water Resources Research, Vol. 16/3, pp.481-490.

140. Stedinger, J.R. and Cohn, T.A. (1985), 'The use of historical flood information in flood frequency analysis', Paper pres. at US-China Bilateral Symposium on the Analysis of Extraordinary Flood Events, Nanjing.

141. Sutcliffe, J.V. (1985), 'The use of historical records in flood frequency analysis', Paper pres. at US-China Bilateral Symposium on the Analysis of Extraordinary Flood Events, Nanjing.

142. Taesombut, V. and Yevjevich, V. (1978), "Use of partial flood series for estimating distribution of maximum annual flood peak', Hydrol. Paper No. 97, Colorado State University, Fort Collins, 71 pp.

143. Takeuchi, K. (1984), 'Annual maximum series and partial duration series - evaluation of Langbein's formula and Chow's discussion', Journal of Hydrology, Vol 68, pp.275-284.

144. Tavares, G.D. and Da Silva, J.E. (1983), 'Partial duration series method revisited', Journal of Hydrology, Vol. 64, pp.1-24.

145. Tiao, G.C. and Lund, D.R.C. (1970), 'The use of OLUMV estimates in inference robustness studies of the location parameter of a class of symmetric distribution', Jour. Amer.Statist. Ass., Vol. 65.

146. Thomas, H.A. (1948), 'Frequency of minor floods', Journ. Boston Soc. Civ. Engrs., Vol. 35, pp.425.

147. Todorovic, P. (1970), 'On some problems involving random number of random variables', Annals Math. Statist., Vol. 41(3), pp.1059-1063.

148. Todorovic, P. and Zelenhasic, E. (1970), 'A stochastic model for flood analysis', Water Resources Research, Vol. 6/6, pp.1641-1648.

149. Todorovic, P. (1971), 'On extreme problems in hydrology', paper presented at Joint Statistics Meeting, American Statist. Ass. and Inst. of Math. Statist., Colorado State University, Fort Collins, Co.

150. Todorovic, P. and Rousselle, J. (1971), 'Some problems of flood analysis', Water Resources Research, Vol. 7/5, pp.1144-1150.

151. Todorovic, P. and Woolhiser, D.A. (1972), 'On the time when the extreme flood occurs', Water Resources Research, Vol. 8/6, pp.1433-1438.

152. Todorovic, P. (1978a), ' Stochastic models of extreme flows and their application to design', Proceedings of International symposium on 'Risk and Reliability in Water Resources', University of Waterloo, Ontario, Canada.

153. Todorovic, P. (1978b), 'Stochastic models of floods', Water Resources Research, Vol. 14/2, pp.345-356.

154. Treiber, B. and Plate, E. J. (1975), ' A stochastic model for the simulation of daily flows,' IAHS workshop on 'Application of Mathematical Models in Hydrology and Water Resources Systems', Brastilova.

155. USDA (1957), ' Engineering handbook- section 4, hydrology', Soil Conservation Service, U.S. Deptt. of Agriculture.

156. Vandewiele, G.L. and Dom, A. (1989), ' A non-gaussian multicomponent model for river flow', Water Resources Research, Vol. 25/3, pp.397-404.

157. Wallis, J.R., Matalas, N.C. and Slack, J.R. (1977), 'Apparent regional skew', Water Resources Research, Vol. 13, pp.159-182.

158. Wallis, J.R. (1980), 'Risk and uncertainties in the evaluation of flood events for the design of hydrologic structures', Keynote address at "Seminar on Extreme Hydrological Events - Floods and Droughts", Erice, Italy, 33pp.

159. Wallis, J.R. and Wood, E.F. (1985), 'Relative accuracy of Log Pearson III procedures', ASCE Journal of Hydraulic Engineering, Vol. 111/7, pp.1043-1056.

160. Wallis, J.R., Matalas, N.C. and Slack, J.R. (1974), 'Just a moment!', Water Resources Research, Vol. 10/2, pp.211-219.

161. Weiss, G. (1973), 'Shot noise models for synthetic generation of multi-site daily streamflow data', IAHS Sympos. on Design of Water Resource Projects with Inadequate Data, Madrid.

162. Weiss, G. (1977),' Shot noise model for the generation of synthetic streamflow data', Water Resources Research, Vol. 13/1, pp.101-108.

163. Wiltshire, S.W. (1985), 'Grouping Basins for Regional Flood Frequency Analysis', Hydrol. Sci. Journ., Vol. 30/1, pp.151-159.

164. Wiltshire, S.W. (1986a), 'Identification of homogeneous regions for flood frequency analysis', Journal of Hydrology Vol. 84, pp.287-302.

165. Wiltshire, S.W. (1986b), 'Regional flood frequency analysis I: homogeneity statistics', Hydrol. Sci. Jour., Vol. 31/3, pp.321-333.

166. Wiltshire, S.W. (1986c), 'Regional flood frequency analysis II: multivariate classification of drainage basins in Britain', Hydrol. Sci. Jour., Vol. 31/3, pp.335-346.

167. Wood, E.F. and Rodriguez-Iturbe, I. (1975), 'Bayesian inference and decision making for extreme hydrologic events', Water Resources Research, Vol. 11/4, pp.533-542.

168. WRC (1966), 'Methods of flow frequency analysis', Subcommittee on Hydrology, Inter Agency Committee on Water Resources, U.S. Government Printing Office, Washington D.C., Bull. 13, 42 p.

169. WRC (1967), 'Guidelines for determining flood flow frequency', Bulletin 15, Hydrology Committee, Water Resources Council, Washington, 1967. (Also revised versions, Bulletin 17A, 1977 and Bulletin 17B, 1981).

170. Wu, B. and Goodridge, J.D. (1976), 'Selection of frequency distributions for hydrologic frequency analysis', Dept. of Water Resources, State of California, Sacramento, 85 p.

171. Xu Runzi and Ye Bangham (1985), 'Estimating rare floods by geomorphological methods', Paper pres. at US-China Bilateral Symposium on the Analysis of Extraordinary Flood Events, Nanjing.

172. Yakowitz, S.J. (1973), ' A stochastic model for daily streamflow in arid region', Water Resources Research, Vol.9, pp.1271-1285.

173. Yevjevich, V. (1972a), ' Probability and statistics in hydrology', Water Resources Publication, Fort Collins, Co.

174. Yevjevich, V. (1972b), ' Structural analysis of hydrologic time series', CSU paper no. 56, Fort Collins, Co.

175. Yevjevich, V. (1976),'Generation of hydrologic and water resources samples', Chapter 4 in 'Stochastic Approaches to Water Resources', Vol. I, Edited by Shen, H.W., USA.

176. Yevjevich, V. and Taesombut, V. (1978), 'Information on flood peaks in daily flow series', Proc. Int. Symp. on Risk and Reliability in Water Resources, Univ. of Waterloo, Ontario, Canada, pp.451-470.

177. Zelenhasic, E. (1970),' Theoretical probability distributions for flood peaks', Hydrology paper no. 42, Colorado State University, Fort Collins.

178. Zelenhasic, E. and Salvai, A. (1987), ' A method of streamflow drought analysis', Water Resources Research, Vol. 23/1, pp.156-168.

APPENDIX-I

IMPORTANT SUBROUTINES DEVELOPED AND USED IN THE STUDY

Some of the important subroutines developed and used in the study are given in this Appendix. These subroutines were implemented on DEC-2Ø computer system. Instructions for use are given in the subroutines in the form of comment statements.

S. NO.	Subroutine	Purpose
1.	NKG	Multivariate flood analysis,
2.	POWER2	Two step power transformation,
3.	POWER1	Normalizing the series using Box-Cox transformation,
4.	DVPA	Computs number of flood events, flood peaks, volumes and durations for a given threshold. Also computes some information about low flows but not used in the study,
5.	STAT	Computes statistical parameters,
6.	CORRE	Computes correlation coeff.,
7.	SEQ	Arranges the series in ascending order,
8.	FCT	Computes function in Subroutine QG32,
9.	FCTI	Computes function in Subroutine QG32
1Ø.	SORTX	Arranges the series in discending order,
11.	LEAP	Number of days in Feb. in a leap year,
12.	NDTRI	Inverse normal distribution function (Source: SSP),
13.	NDTR	Normal distribution function (Source: SSP),
14.	QG32	Integration of a function by 32 point Gaussian Quadrature formula,
15.	GAUSS	Generation of normally distributed random numbers.

APPENDIX-I

IMPORTANT SUBROUTINES DEVELOPED AND USED IN THE STUDY

```
C*********************************************************************
C       SUBROUTINE NKG
C       PURPOSE:
C       SUBROUTINE FOR MULTIVARIATE ANALYSIS.
C       DESCRIPTION OF PARAMETERS
C       X IS INPUT SERIES OF VARIABLE X (FLOOD VOLUMES)
C       Y IS INPUT SERIES OF VARIABLE Y (FLOOD PEAKS)
C       N IS TOTAL NO. OF VALUES IN X SERIES OR Y SERIES
C         TOTAL NO. FLOOD EVENTS.
C       ALEMDA AVERAGE NO. OF FLOOD EVENTS PER YEAR.
C       XSTART STARTING VALUE OF X
C       NX  END VALUE OF X
C       XINCRE INCREMENT DESIRED IN X
C       YSTART STARTING VALUE OF Y
C       NY END VALUE OF Y
C       YINCRE INCREMENT DESIRED IN Y
C       SUBROUTINES REQUIRED
C       POWER2 FOR TWO STEP POEWER TRANSFORMATION
C       CORRE  FOR CORRELATION COEFF.
C       STAT   FOR STATISTICAL PARAMETERS
C       NDTRI
C       NDTR
C       QG32
        SUBROUTINE NKG(X,Y,N,ALEMDA,XSTART,NX,XINCRE,YSTART
     1,NY,YINCRE)
C       SUBROUTINE
        DIMENSION X(1),Y(1),R1(1Ø)
        COMMON/BR1/A,B,XBAR,YBAR,SX,SY,SXY
        COMMON/RD4/C1,X1T,Y1T
       COMMON/BR/QSUM,QAVE,QMAX,QMIN,QSTD,QCV,QSKEW,QKURT,R1
C       COMPUTE CORRELATION BETWEEN X AND Y IN ORIGINAL DOMAIN.
        CALL CORRE(X,Y,N,R)
        S=SQRT(1.-R*R)*SY
        A1=A
        B1=B
C       NORMALIZE THE SERIES
        WRITE(4,1)
1       FORMAT(2ØX'ANALYSIS OF THE POWER TRANSFORMED SERIES')
        CALL POWER2(X,N,ALX,AGX)
        CALL POWER2(Y,N,ALY,AGY)
C       NORMALIZE THE X AND Y SERIES
        DO 51 I=1,N
        X(I)=(X(I)*ALX-1.)/ALX
51      Y(I)=(Y(I)**ALY-1.)/ALY
        CALL STAT(X,N,Ø,1)
        X1BAR=QAVE
        CALL STAT(Y,N,Ø,1)
        Y1BAR=QAVE
        DO 82 I=1,N
        XT1=X(I)-X1BAR
```

```
      YT1=Y(I)-Y1BAR
      X(I)=(ABS(XT1)**AGX)*XT1/ABS(XT1)
      Y(I)=(ABS(YT1)**AGY)*YT1/ABS(YT1)
82    CONTINUE
C     COMPUTE CORRELATION BETWEEN NORMALIZED VARIABLES
      CALL CORRE(X,Y,N,R)
      WRITE(4,5)A,B,XBAR,YBAR,SX,SY,R,ALX,ALY,AGX,AGY
5     FORMAT(' A,B,XBAR,YBAR,SX,SY,R,ALX,ALY,AGX,AGY='/11F11.4)
C     COMPUTE 5Ø, 1ØØ, 5ØØ, 1ØØØ, AND 1ØØØØ YEARS X AND Y QUANTILES.
      F1=(ALOG(1.-.Ø2)+ALEMDA)/ALEMDA
      F2=(ALOG(1.-.Ø1)+ALEMDA)/ALEMDA
      F3=(ALOG(1.-.ØØ2)+ALEMDA)/ALEMDA
      F4=(ALOG(1.-.ØØ1)+ALEMDA)/ALEMDA
      F5=(ALOG(1.-.ØØØ2)+ALEMDA)/ALEMDA
      CALL NDTRI(F1,AX1,C,IER)
      CALL NDTRI(F2,AX2,C,IER)
      CALL NDTRI(F3,AX3,C,IER)
      CALL NDTRI(F4,AX4,C,IER)
      CALL NDTRI(F5,AX5,C,IER)
      X5Ø=XBAR+AX1*SX
      X5Ø=(ABS(X5Ø)**(1./AGX))X5Ø/ABS(X5Ø)
      X5Ø=((X5Ø+X1BAR)*ALX+1.)**(1./ALX)
      X1ØØ=XBAR+AX2*SX
      X1ØØ=(ABS(X1ØØ)**(1./AGX))*X1ØØ/ABS(X1ØØ)
      X1ØØ=((X1ØØ+X1BAR)*ALX+1.)**(1./ALX)
      X5ØØ=XBAR+AX3*SX
      X5ØØ=(ABS(X5ØØ)**(1./AGX))*X5ØØ/ABS(X5ØØ)
      X5ØØ=((X5ØØ+X1BAR)*ALX+1.)**(1./ALX)
      X1ØØØ=XBAR+AX4*SX
      X1ØØØ=(ABS(X1ØØØ)**(1./AGX))*X1ØØØ/ABS(X1ØØØ)
      X1ØØØ=((X1ØØØ+X1BAR)*ALX+1.)**(1./ALX)
      X5ØØØ=XBAR+AX5*SX
      X5ØØØ=(ABS(X5ØØØ)**(1./AGX))*X5ØØØ/ABS(X5ØØØ)
      X5ØØØ=((X5ØØØ+X1BAR)*ALX+1.)**(1./ALX)
      Y5Ø=YBAR+AX1*SY
      Y5Ø=(ABS(Y5Ø)**(1./AGY))*Y5Ø/ABS(Y5Ø)
      Y5Ø=((Y5Ø+Y1BAR)*ALY+1.)**(1./ALY)
      Y1ØØ=YBAR+AX2*SY
      Y1ØØ=(ABS(Y1ØØ)**(1./AGY))*Y1ØØ/ABS(Y1ØØ)
      Y1ØØ=((Y1ØØ+Y1BAR)*ALY+1.)**(1./ALY)
      Y5ØØ=YBAR+AX3*SY
      Y5ØØ=(ABS(Y5ØØ)**(1./AGY))*Y5ØØ/ABS(Y5ØØ)
      Y5ØØ=((Y5ØØ+Y1BAR)*ALY+1.)**(1./ALY)
      Y1ØØØ=YBAR+AX4*SY
      Y1ØØØ=(ABS(Y1ØØØ)**(1./AGY))*Y1ØØØ/ABS(Y1ØØØ)
      Y1ØØØ=((Y1ØØØ+Y1BAR)*ALY+1.)**(1./ALY)
      Y5ØØØ=YBAR+AX5*SY
      Y5ØØØ=(ABS(Y5ØØØ)**(1./AGY))*Y5ØØØ/ABS(Y5ØØØ)
      Y5ØØØ=((Y5ØØØ+Y1BAR)*ALY+1.)**(1./ALY)
      WRITE(4,6)
6     FORMAT(7X'X5Ø'6X'X1ØØ'6X'X5ØØ'5X'X1ØØØ'5X'X5ØØØ'7X'Y5Ø'
     16X'Y1ØØ'6X'Y5ØØ'5X'Y1ØØØ'5X'Y5ØØØ')
      WRITE(4,7)X5Ø,X1ØØ,X5ØØ,X1ØØØ,X5ØØØ,Y5Ø,Y1ØØ,Y5ØØ,Y1ØØØ,Y5ØØØ
7     FORMAT(1ØF1Ø.1)
71    CONTINUE
```

```
C       PERFORMS MULTIVARIATE ANALYSIS
        X11=XSTART
        DO 53 I=1,NX
        Y11=YSTART
        X1T=((X11**ALX-1.)/ALX)-X1BAR
        SIGN=X1T/ABS(X1T)
        X1T=(ABS(X1T)**AGX)*SIGN
        ZX=(X1T-XBAR)/SX
        CALL NDTR(ZX,HX,DEN)
        FX=EXP(ALEMDA*HX-ALEMDA)
        TX=1./(1.-FX)
        WRITE(4,8)X11,FX,TX
8       FORMAT(1ØX'X='F1Ø.1,'FX='F7.4,'TX='F1Ø.1)
        WRITE(4,2)
        DO 54 J=1,NY
        Y1T=((Y11**ALY-1.)/ALY)-Y1BAR
       SIGN=Y1T/ABS(Y1T)
        Y1T=(ABS(Y1T)**AGY)*SIGN
        ZY=(Y1T-YBAR)/SY
        CALL NDTR(ZY,HY,DEN)
        FY=EXP(ALEMDA*HY-ALEMDA)
        TY=1./(1.-FY)
        C1=SX/SQRT(2.*(SX*SX*SY*SY-SXY*SXY))
        ZYIX=SQRT(2.)*C1*((Y1T-YBAR)-(SXY/(SX*SX))*(X1T-XBAR))
        CALL NDTR(ZYIX,HYIX,DEN)
        FYIX=EXP(ALEMDA*HYIX-ALEMDA)
        TYIX=1./(1.-FYIX)
        XL=-5.*SX+XBAR
        CALL QG32(XL,X1T,FCT,HXYD)
        HXYI=HX*HY
        FXYD=EXP(ALEMDA*HXYD-ALEMDA)
        FXYI=EXP(ALEMDA*HXYI-ALEMDA)
        HXYIX=HX*HYIX
        FXYIX=EXP(ALEMDA*HXYIX-ALEMDA)
        TXORYD=1./(1.-FXYD)
        TXORI=1./(1.-FXYI)
        TXORYX=1./(1.-FXYIX)
        TXANYD=1./(1.-FX-FY+FXYD)
        TXANYI=1./(1.-FX-FY+FXYI)
        TXANYX=1./(1.-FX-FYIX+FXYIX)
         WRITE(4,3)Y11,FY,TY,FYIX,TYIX,FXYD,FXYI,
     1FXYIX,TXORYD,TXORYI,TXORYX,TXANYD,TXANYI
3       FORMAT(F1Ø.1,2(F7.4,F1Ø.1),3F7.4,6F1Ø.1)
2       FORMAT(9X'Y'5X'FY'8X'TY'3X'FYIX'6X'TYIX'3X'FXYD'3X'FXYI'
     12X'FXYIX'4X'TXORYD'4X'TXORYI'4X'TXORYX'4X'TXANYD'4X'TXANYI
     2')
        Y11=Y11+YINCRE
54      CONTINUE
        X11=X11+XINCRE
53      CONTINUE
        RETURN
        END
```

```
C*****************************************************************************
C       SUBROUTINE FOR TWO STEP POWER TRANSFORMATION
        SUBROUTINE POWER2(X,N,AL,AG)
C       METHOD ADOPTED FROM 'FLOOD FREQUENCY ANALYSIS BY TWO
C       STEP POWER TRANSFORMATION'BY GUPTA,D.K.,ASTHANA B.N.
C       AND BHARGAVA A.N.,JOUNAL OF INSTITUTION OF ENGINEERS
C       VOL. 7Ø/4,NOV.1989.
C       X=INPUT SERIES
C       N=NO. OF DATA POINTS.
C       AL=LEMDA, AND AG= GAMMA
        DIMENSION X(1),R(1Ø),T(12ØØØ)
        COMMON/BR/QSUM,QAVE,QMAX,QMIN,QSTD,QCV,QSKEW,QKURT,R
        AG=1.Ø
        CALL POWER1(X,N,AL)
        YBAR=QAVE
        CS=QSKEW
        CKEX=QKURT-3.
        QSKEW1=QSKEW
        QKURT1=QKURT
 7      FORMAT(' GAMMA LOOP'4F1Ø.4)
 56     IF(ABS(CKEX).LT.Ø.Ø5)GO TO 55
        IF(CKEX.GT.Ø.Ø)GO TO 51
        AG=AG+Ø.Ø1
        GO TO 52
 51     AG=AG-Ø.Ø1
 52     CONTINUE
 67     DO 53 I=1,N
        T(I)=((X(I)**AL-1.)/AL)-YBAR
        SIGN=T(I)/ABS(T(I))
 53     T(I)=(ABS(T(I))**AG)*SIGN
        CALL STAT(T,N,Ø,1)
        CS=QSKEW
        CKEX=QKURT-3.
        QKURT1=QKURT
        QSKEW1=QSKEW
C       TYPE 7,AL,AG,CS,CKEX
        GO TO 56
 55     IF(ABS(CS).LT.Ø.Ø2)GO TO 66
        IF(CS.LT.Ø.Ø)GO TO 54
        AL=AL-Ø.ØØØ1
        GO TO 57
 54     AL=AL+Ø.ØØØ1
 57     DO 58 I=1,N
 58     T(I)=(X(I)**AL-1.)/AL
        CALL STAT(T,N,Ø,1)
        YBAR=QAVE
        GO TO 67
 8      FORMAT(' LEMDA LOOP'4F9.4)
 66     TYPE 1
        WRITE(4,1)
        WRITE(4,2)QMAX,QMIN,QAVE,QSTD,QSKEW1,QKURT1,R(1),AL,AG
        TYPE 2,QMAX,QMIN,QAVE,QSTD,QSKEW1,QKURT1,R(1),AL,AG
 1      FORMAT(5X'MAXIMUM'5X'MINIMUM'8X'MEAN'8X'STD.'8X'SKEW'
     14X'KURTOSIS'2X'CORR. COEFF.'5X'LEMDA'5X'GAMMA')
 2      FORMAT(6F12.3,2X,5F1Ø.4)
        RETURN
        END
```

```
C*********************************************************************
C       SUBROUTINE FOR NORMALISING THE SERIES USING POWER TRAN.
        SUBROUTINE POWER1(X,N,AL)
C       X IS INPUT SERIES.
C       AL IS VALUE OF LEMDA.
        DIMENSION X(1),Y(1000),R(10)
        COMMON/BR/QSUM,QAVE,QMAX,QMIN,QSTD,QCV,QSKEW,QKURT,R
        AL1=1.0
        DO 51 I=1,N
51      Y(I)=(X(I)**AL1-1.)/AL1
        CALL STAT1(Y,N,0,1)
        SK1=QSKEW
        AL2=AL1+0.2
54      DO 52 I=1,N
        Y(I)=X(I)**AL2-1.)/AL2
52      CONTINUE
        CALL STAT1(Y,N,0,1)
        SK2=QSKEW
        DL=-SK2*(AL2-AL1)/(SK2-SK1)
        IF(ABS(DL).LE.0.00001)GO TO 53
        SK1=SK2
        AL1=AL2
        AL2=AL2+DL
        GO TO 54
53      AL=AL2
        DO 55 I=1,N
55      Y(I)=(X(I)**AL-1.)/AL
        CALL STAT1(Y,N,0,1)
        WRITE(4,1)
        TYPE 1
        WRITE(4,2)QMAX,QMIN,QAVE,QSTD,QSKEW,QKURT,R(1),AL
        TYPE 2,QMAX,QMIN,QAVE,QSTD,QSKEW,QKURT,R(1),AL
1       FORMAT(5X'MAXIMUM'5X'MINIMUM'8X'MEAN'8X'STD.'8X'SKEW'
     14X'KURTOSIS'2X'CORR. COEFF.'5X'LEMDA')
2       FORMAT(6F12.3,2X,5F10.4)
        RETURN
        END
C*********************************************************************
C       SUBROUTINE DVPA
C       PURPOSE:
C       SUBROUTINE COMPUTES NO. OF FLOOD EVENTS, FLOOD PEAKS,
C       VOLUMES AND DURATIONS, AND THE LOCATION OF FLOOD PEAKS.
C       DROUGHT VOLUMES DURATION, PEAK ETC. ARE ALSO COMPUTED BUT NOT
C       USED IN THE ANALYSIS.
C       DESCRIPTION OF PARAMETERS
C       Q IS THE SERIES OF DAILY FLOWS.
C       N IS TOTAL NO. OF DAILY FLOW VALUES.N= 365
C       YEAR IS THE YEAR FOR WHICH FLOWS ARE BEING ANALYSED
C       THRESH IS THE THRESHOLD FOR WHICH THE FLOOD FEATURES ARE REQUIRED.
C       SUBROUTINE FOR PEAK VOLUME DURATION ANALYSIS
        SUBROUTINE DVPA(Q,THRESH,N,YEAR)
        DIMENSION Q(1),LZERO(100),V(100),VD(100),VF(100)
     1,DD(100),DF(100),PF(100),PD(100),T(365),LOCF(100)
        DIMENSION LOCD(100)
```

```
      INTEGER YEAR
      COMMON/BL1/VF,DF,PF,LOCF,IK
      DO 51 I=1,N
51    Q(I)=Q(I)-THRESH
      K=1
      DO 52 I=2,N
      IF(Q(I-1).LE.Ø.Ø.AND.Q(I).GT.Ø.Ø)GO TO 53
      IF(Q(I-1).GE.Ø.Ø.AND.Q(I).LT.Ø.Ø)GO TO 53
      GO TO 52
53    LZERO(K)=I-1
      K=K+1
52    CONTINUE
      IF (K.EQ.1)G TO 6Ø
      K=K-1
      KK=1
      KKK=1
      DO 54 J=1,K-1
      V(J)=Ø.Ø
      DUR=Ø.Ø
      N1=LZERO(J)
      N2=LZERO(J+1)
      QMAX=Ø.Ø
      LOC=N1
      QN1=Q(N1)
      QN2=Q(N2+1)
      Q(N1)=Ø.Ø
      Q(N2+1)=Ø.Ø
      DO 55 =N1,N2
      T(L)=1.Ø
      T(N1)=Q(N1+1)/(Q(N1+1)-QN1)
      T(N2)=Q(N2)/(Q(N2)-QN2)
      IF(QMAX-ABS(Q(L)))58,59,59
58    QMAX=ABS(Q(L))
      LOC=L
59    V(J)=V(J)+T(L)*(Q(L)+Q(L+1))/2.
      DUR=DUR+T(L)
55    CONTINUE
      Q(N1)=QN1
      Q(N2+1)=QN2
      IF(V(J).LT.Ø.Ø)GO TO 56
      IF(V(J).GE.Ø.Ø)GO TO 57
56    VD(KK)=-V(J)
      DD(KK)=DUR
      PD(KK)=THRESH-QMAX
      LOCD(KK)=LOC
      IF(VD(KK).LT.5ØØ)GO TO 54
      KK=KK+1
      GO TO 54
57    VF(KKK)=V(J)
      DF(KKK)=DUR
      PF(KKK)=QMAX
      LOCF(KKK)=LOC
      IF(VF(KKK).LT.5ØØ)GO TO 54
      KKK=KKK+1
```

```
54        CONTINUE
          KK=KK-1
          KKK=KKK-1
          IK=KKK
C         WRITE(2,7)KK,KKK
7         FORMAT(5X,'TOTAL TROUGHS AND PEAKS ARE',I5,'AND',I5,'
     1RESPECTIVELY')
          GO TO 61
60        WRITE(2,8)YEAR
8         FORMAT(I6)
          WRITE(2,9)
9         FORMAT(1X,'NO PEAKS AND TROUGHS DETECTED. ALL THE
     1VALUES ARE BELOW THRESHOLD')
          IK=0
61        CONTINUE
          RETURN
          END
C*************************************************************************
          SUBROUTINE STAT(X,KK,OPT,K)
          DIMENSION X(1),Y(12000),R(10)
          COMMON/BR/QSUM,QAVE,QMAX,QMIN,QSTD,QCV,QSKEW,QKURT,R
          REAL N1
          INTEGER OPT
          N1=KK
          SUM1=0.0
          SUM2=0.0
          SUM3=0.0
          SUM4=0.0
          QMAX=X(1)
          QMIN=X(1)
          DO 51 I=1,KK
          IF(X(I).NE.-999)GO TO 57
          IF(X(I).EQ.-999)N1=N1-1
          GO TO 51
57        SUM1=SUM1+X(I)
          SUM2=SUM2+X(I)*X(I)
          SUM3=SUM3+X(I)*X(I)*X(I)
          SUM4=SUM4+X(I)*X(I)*X(I)*X(I)
          IF(X(I).GT.QMAX)QMAX=X(I)
          IF(X(I).LT.QMIN)QMIN=X(I)
51        CONTINUE
          QSUM=SUM1
          QAVE=SUM1/N1
          QSTD=SQRT((SUM2-N1*QAVE*QAVE)/(N1-1))
          QCV=QSTD/QAVE
          QSKEW=((N1*N1)/((N1-1)*(N1-2)*QSTD*QSTD*QSTD))*
     1((SUM3/N1)-(3.*QAVE*SUM2/N1)+2.*QAVE*QAVE*QAVE)
          QKURT=((N1*N1*N1)/((N1-1)*(N1-2)*(N1-3)*(QSTD**4)))*
     1((SUM4/N1)-4.*QAVE*(1./N1)*SUM3+6.*QAVE*QAVE*(1./N1)*SUM2
     1-3.*(QAVE**4))
          DO 56 J=1,K
          DO 55 I=1,KK-J
55        Y(I)=X(I+J)
          CALL CORRE(X,Y,KK-J,R1)
          R(J)=R1
```

```
56      CONTINUE
        IF(OPT.EQ.Ø)GO TO 54
        WRITE(2,1)
1       FORMAT(5X,'MAXIMUM',5X,'MINIMUM',8X,'MEAN',8X,'STD.',
     18X,'SKEW',4X,'KURTOSIS',2X,'CORRELATION COEFFICIENTS')
        WRITE(2,2)QMAX,QMIN,QAVE,QSTD,QSKEW,QKURT,(R(J),J=1,K)
        IF(OPT.EQ.3)TYPE 1
        IF(OPT.EQ.3)TYPE 2,QMAX,QMIN,QAVE,QSTD,QSKEW,QKURT,(R(J),J=1,K)
2       FORMAT(6F12.3,2X,5F8.3)
54      RETURN
        END
C*************************************************************
        SUBROUTINE CORRE(X,Y,N,R)
        DIMENSION X(1),Y(1)
        COMMON /BR1/A,B,XBAR,YBAR,SX,SY,SXY
        SUMX=Ø.Ø
        SUMY=Ø.Ø
        DO 51 I=1,N
        SUMX=SUMX+X(I)
51      SUMY=SUMY+Y(I)
        XBAR=SUMX/N
        YBAR=SUMY/N
        VARX=Ø.Ø
        VARY=Ø.Ø
        VARXY=Ø.Ø
        DO 52 I=1,N
        VARX=VARX+(X(I)-XBAR)**2.
        VARY=VARY+(Y(I)-YBAR)**2.
        VARXY=VARXY+(X(I)-XBAR)*(Y(I)-YBAR)
52      CONTINUE
        R=VARXY/((SQRT(VARX))*(SQRT(VARY)))
        B=VARXY/VARX
        A=YBAR-B*XBAR
        SX=SQRT(VARX/(N-1.))
        SY=SQRT(VARY/(N-1.))
        SXY=VARXY/(N-1.)
        RETURN
        END
C******************************************************************
C       SUBROUTINE FOR ARRANGING THE SERIES IN ASCENDING ORDER
C       INPUT DATA ARE AS FOLLOWS:
C               N=NUMBER OF VALUES TO BE ARRANGED
C               IYEAR=YEAR IN CHRONOLOGICAL ORDER
C               X=DATA SERIES IN CHRONOLOGICAL ORDER
C      OUTPUT RESULTS ARE AS FOLLOWS:
C               X=DATA SERIES IN ASCENDING ORDER
C               IYEAR=YEAR CORRESPONDING TO X
        SUBROUTINE SEQ(N,IYEAR,X)
        DIMENSION X(1),IYEAR(1)
        IF(N.LE.1)GO TO 3Ø
        J=Ø
2Ø      J=J+1
        DO 1Ø I=J,N
        IF (X(I).GT.X(J)) GO TO 1Ø
        XT=X(J)
```

```
      XTY=IYEAR(J)
      X(J)=X(I)
      IYEAR(J)=IYEAR(I)
      X(I)=XT
      IYEAR(I)=XTY
10    CONTINUE
      IF(J.NE.(N-1)) GO TO 2Ø
3Ø    RETURN
      END
C***************************************************************************
      SUBROUTINE FCT(X,Y)
      COMMON /BR1/A,B,XBAR,YBAR,SX,SY,SXY
      COMMON/RD4/C1,X1T,Y1T
      Z1=SQRT(2.)*C1*((Y1T-YBAR)-(SXY/(SX*SX))*(X-XBAR))
      CALL NDTR(Z1,F1,DEN)
      PAI=3.141592654
      Y=(1./(SQRT(2.*PAI)*SX))*EXP(-Ø.5*((X-XBAR)/SX)**2.)
     1*F1
      RETURN
      END
C***************************************************************************
      SUBROUTINE FCTI(D,Y)
      COMMON /RD/Y1,S,XØ,BX
      COMMON /BR1/A,B,XBAR,YBAR,SX,SY,SXY
      Z=(Y1-A-B*D)/S
      CALL NDTR(Z,FPD,DEN)
      Y=EXP(-(D-XØ)/BX)*FPD/BX
      RETURN
      END
C***************************************************************************
      SUBROUTINE SORTX (N,X)
C  SORTS IN DECREASING ORDER, X(1)=LARGEST
      DIMENSION X(1)
      K=N-1
      DO 2 L=1,K
      M=N-L
      DO 2 J=1,M
      IF (X(J)-X(J+1)) 1,1,2
1     XT=X(J)
      X(J)=X(J+1)
      X(J+1)=XT
2     CONTINUE
      RETURN
      END
C***********************************************************************
C     SUBROUTINE FOR DAYS IN A LEAP YEAR
      SUBROUTINE LEAP(YEAR,NDAY)
      INTEGER YEAR
      YEAR1=YEAR/4.Ø
      NYEAR=YEAR1
      BYEAR=NYEAR
      IF(YEAR1-BYEAR)1Ø,11,1Ø
1Ø    NDAY=28
      GO TO 12
```

```
11      NDAY=29
12      RETURN
        END
C*********************************************************************
C       SUBROUTINE NDTRI
C       PURPOSE
C         COMPUTES X=P**(-1)(Y), THE ARGUMENT X SUCH THAT Y=P(X)
C         =THE PROBABILITY THAT THE RANDOM VARIABLE U,DISTRIBUTED
C         NORMALLY(Ø,1), IS LESS THAN OR EQUAL TO X. F(X), THE
C         ORDINATE OF THE NORMAL DENSITY, AT X, IS ALSO COMPUTED.
C       USAGE
C         CALL NDTRI(P,X,C,IER)
C       DESCRIPTION OF PARAMETERS
C         P   -INPUT PROBABILITY
C         X   -OUTPUT ARGUMENT SUCH THAT P=Y=THE PROBABILITY THAT
C              THE RANDOM VARIABLE IS LESS THAN OR EQUAL TO X
C         C   -OUTPUT DENSITY,F(X)
C         IER -OUTPUT ERROR CODE
C              =-1  IF P IS NOT IN THE INTERVAL (Ø,1),INCLUSIVE
C             X=C=.99999E+37 IN THIS CASE
C              =C  IF THERE IS NO ERROR
C                     SEE REMARKS BELOW
C       REMARKS
C         MAXIMUM ERROR IS Ø.ØØØ45
C         IF P=Ø,X IS SET TO -(1Ø)**74.D IS SET TO C
C         IF P=1,X IS SET TO  (1Ø)**74.D IS SET TO C
C       SUBROUTINES AND SUBPROGRAMS REQUIRED
C         NONE
C       METHOD
C          BASED ON APPROXIMATIONS IN C.HASTINGS,``APPROXIMATIONS
C          FOR DIGITAL COMPUTERS'',PRINCETON UNIV.PRESS,PRINCETON,
C          N.J.,1955.SEE EQUATION 26.2.23,HAND BOOK OF MATHEMAICAL
C          FUNCTIONS,ABRAMOWITZ AD STEGUN,DOVER PUBLICATIONS,INC.,
C          NEW YORK.
        SUBROUTINE NDTRI(P,X,D,IE)
        IE=C
        X=.99999E+37
        D=X
        IF(P)1,4,2
1       IE=-1
        GO TO 12
2       IF(P-1.Ø)7,5,1
4       X=-Ø.999999E+37
5       D=Ø.Ø
        GO TO 12
7       D=P
       IF(D-Ø.5)9,9,8
8       D=1.Ø-D
9       T2=ALOG(1.Ø/(D*D))
        T=SQRT(T2)
        X=T-(2.515517+Ø.8Ø2853*T+Ø.Ø1Ø328*T2)/(1.Ø+1.432788*T+
     1Ø.189269*T2+Ø.ØØ13Ø8*T*T2)
        IF(P-Ø.5)1Ø,1Ø,11
1Ø      X=-X
11      D=Ø.3989423*EXP(-X*X/2.Ø)
12      RETURN
        END
```

```
C*******************************************************************
C       SUBROUTINE NDTR
C       PURPOSE:
C       COMPUTES PROBABILITY THAT THE RANDOM VARIABLE X
C       DISTRIBUTED NORMALLY (Ø,1) IS LESS THAN OR EQUAL TO X.
C       THE ORDINATE OF THE NORMAL DENSITY AT X IS ALSO COMPUTED
C       USAGE:
C       CALL NDTR(X,P,D)
C       DESCRIPTION OF PARAMETERS
C       X:INPUT SCALER FOR WHICH P(X) IS COMPUTED.
C       P:OUTPUT PROBABILITY.
C       D:OUTPUT DENSITY.
C       REMARKS:
C       MAXIMUM ERROR IS Ø.ØØØØØØ7.
C       SUBROUTINES AND SUBPROGRAMMES REQUIRED NONE.
C       METHOD:
C       BASED ON APPROXIMATIONS IN C. HASTINGS,'APPROXIMATIONS
C       FOR DIGITAL COMPUTERS',PRINCETON UNI. PRESS ,PRINCETON,
C       N.J.1955,SEE EQUATION 26.2.17,IN HANDBOOK BOOK OF
C       MATHEMATICAL FUNCTIONS BY ABRAMOWITZ AND STEGUN OOVER
C       PUBLICATIONS INC. NEW YORK
C...................................................................
        SUBROUTINE NDTR(X,P,D)
        AX=ABS(X)
        T=1.Ø/(1.Ø+Ø.2316419*AX)
        D=Ø.3989423*EXP(-X*X/2.)
        P=1.Ø-D*T*((((1.33Ø2*T-1.821256)*T+1.781478)*T-Ø.3565638
     1)*T+Ø.3193815)
        IF(X)1,2,2
1       P=1.-P
2       RETURN
        END
C*******************************************************************
C       SUBROUTINE QG32
C       PURPOSE:
C       TO COMPUTE INTEGRAL  SUMMED OVER X FROM XL TO XU.
C       USAGE:
C       CALL QG32(XL,XU,FCT1,Y)
C       DESCRIPTION OF PARAMETERS:
C       XL:THE LOWER BOUND OF THE INTERVAL
C       XU:THE UPPER BOUND OF THE INTRVAL
C       FCT1:THE NAME OF THE EXTERNAL FUNCTION SUBPRO. OR SUBROUTINE  USED
C       Y:THE RESULTING INTEGRAL VALUE
C       REMARKS: NONE
C       SUBROUTINE AND FUNCTION SUBPROGRAMMES REQUIRED.
C       THE SUBROUTINE FCT MUST BE FURNISHED BY THE USER.
C       METHOD:EVALUATION IS DONE BY EANS OF 32 POINT GAUSS
C       QUARDATURE FORMULA WHICH INTEGRATES POLYNOMIALS UPTO
C       DEGREE 63 EXACTLY.
        SUBROUTINE QG32(XL,XU,FCT1,Y)
        A=Ø.5*(XU+XL)
        B=XU-XL
        C=.4986319*B
        CALL FCT(A+C,FCT1)
        CALL FCT(A-C,FCT2)
```

```
Y=.ØØ35Ø93Ø*(FCT1+FCT2)
C=.4928Ø575*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.ØØ813719*(FCT1+FCT2)
C=.48238112*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FT2)
Y=Y+.Ø12696Ø3*(FCT1+FCT2)
C=.467453Ø3*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.Ø1713693*(FCT1+FCT2)
C=.44816Ø57*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.Ø2141794*(FCT1+FCT2)
C=.4246838Ø*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.Ø25499Ø2*(FCT1+FCT2)
C=.39724189*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.Ø29342Ø4*(FCT1+FCT2)
C=.366Ø91Ø5*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.Ø3291111*(FCT1+FCT2)
C=.33152213*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.Ø3617289*(FCT1+FCT2)
C=.29385787*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.Ø39Ø9694*(FCT1+FCT2)
C=.25344995*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.Ø4165596*(FCT1+FCT2)
C=.21Ø67563*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.Ø43826Ø4*(FCT1+FCT2)
C=.1659343Ø*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.Ø4558693*(FCT1+FCT2)
C=.11964368*B
CALL FCT(A+C,FCT1)
CALL FCT(A-C,FCT2)
Y=Y+.Ø4692219*(FCT1+FCT2)
C=.Ø7223598*B
```

```
      CALL FCT(A+C,FCT1)
      CALL FCT(A-C,FCT2)
      Y=Y+.Ø4781936*(FCT1+FCT2)
      C=.Ø2415383*B
      CALL FCT(A+C,FCT1)
      CALL FCT(A-C,FCT2)
      Y=B*(Y+.Ø4827ØØ4*(FCT1+FCT2))
      RETURN
      END
C***********************************************************************
      SUBROUTINE GAUSS(IX,S,AM,V)
      A=Ø.Ø
      DO 5Ø I=1,12
C
      Y=RAN(IX)
C
   5Ø A=A+Y
      V=(A-6.Ø)*S+AM
      RETURN
      END
```

***