REGIONAL FLOOD FREQUENCY ESTIMATION IN INDIA

A THESIS

Submitted in partial fulfilment of the requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

HYDROLOGY

by

RAKESH KUMAR

DEPARTMENT OF HYDROLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE - 247 667 (INDIA)

JULY, 2009

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **REGIONAL FLOOD FREQUENCY ESTIMATION IN INDIA** in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy and submitted in the Department of Hydrology, Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from July, 2003 to July, 2009 under the supervision of Dr. N.K. Goel, Professor, Department of Hydrology, Indian Institute of Technology Roorkee and Dr. K.K.S. Bhatia, Former Scientist F, National Institute of Hydrology, Roorkee and Director, Modinagar Institute of Technology, Modinagar.

The matter in the thesis has not been submitted by me for the awards of any other degree of this or any other Institute.

Rakeshkr

(RAKESH KUMAR)

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

(K. K. S. Bhatia) Supervisor	(N. K. Goel) Supervisor

Date: July 21, 2009

The Ph.D. viva-voce Examination of **Mr. Rakesh Kumar**, Research Scholar, has been held on Nov. 30, 2009.

Signature of Supervisors	Signature of External Examiner

ABSTRACT

Estimation of magnitudes of likely occurrence of floods is of great importance for design of various types of hydraulic structures. Floods of different return periods are also required for taking up some of the non-structural measures of flood management. As per the Bureau of Indian Standards hydrological design criteria, frequency based floods find their applications in estimation of design floods for almost all the types of hydraulic structures viz. small size dams, barrages, weirs, road and railway bridges, cross drainage structures, flood control structures etc., excluding large and intermediate size dams. For design of large and intermediate size dams probable maximum flood (PMF) and standard project flood (SPF) are adopted, respectively. However, in these two cases also flood frequency analysis is invariably performed for assessing the return periods of PMF and SPF. Whenever, rainfall or river flow records are not available at or near the site of interest, it is difficult for hydrologists or engineers to derive reliable design flood estimates directly. In such a situation, regional flood frequency relationships developed for the region are one of the alternative methods, which may be adopted for estimation of design floods especially for small catchments.

As the studies on flood frequency estimation in India are limited, scattered and mostly based on the conventional techniques; hence, there is an urgent need for making systematic efforts for developing a reliable and convenient regional flood frequency estimation procedure based on the state of art technique for gauged and ungauged catchments. Further, the soft computing techniques offer real advantages over conventional modeling, including the ability to handle large amounts of noisy data from dynamic and nonlinear systems, especially when the underlying

hydrological relationships are not fully understood. These techniques viz. Artificial Neural Networks (ANN) and Fuzzy Logic (FL) have been applied for solving some of the hydrological problems such as development of stage-discharge relationship, flood forecasting, rainfall-runoff modeling, estimation of precipitation and evaporation, ground water modeling, water quality modeling etc. However, applications of ANNs in regional flood frequency estimation are limited and use of Fuzzy Logic in regional flood frequency estimation remains to be investigated. Whereas, some of the recent studies show that the fuzzy modeling is more versatile and improved alternative to ANNs.

In this study, regional flood frequency relationships have been developed for 17 hydrometeorologically homogeneous categorized Subzones of India using the L-moments approach. The applicability of soft computing techniques viz. Artificial Neural Networks (ANN) and Fuzzy Inference System (FIS) in regional flood frequency estimation has also been investigated. The L-moments form basis of an elegant mathematical theory and can be used to facilitate the estimation process in regional frequency analysis. The L-moment based methods are demonstrably superior to those that have been used previously, and are now being adopted by many organizations worldwide. For carrying out the regional flood frequency estimation study, screening of the annual maximum peak flood data has been carried out for assessing the suitability of the data for regional flood frequency analysis by the L-moments based Discordancy (D_i) statistic test. The regional homogeneity of the 17 Subzones has been tested employing the L-moments based heterogeneity measure (H) by carrying out 500 simulations using the four parameter Kappa distribution. For carrying out regional flood frequency analysis studies based on the L-moments approach twelve frequency distributions viz. Extreme Value (EV1), General Extreme

Value (GEV), Logistic (LOS), Generalized Logistic (GLO), Normal (NOR), Generalized Normal (GNO), Exponential (EXP), Uniform (UNF), Generalized Pareto (GPA), Pearson Type-III (PE3), Kappa (KAP) and five parameter Wakeby (WAK) have been used. Based on the L-moment ratio diagram as well as Z^{dist} –statistic criteria robust frequency distributions have been identified for the 17 Subzones of India.

The 17 Subzones cover total 25,89,342 km^2 area, which constitutes about 79% of the geographical area of India. The annual maximum peak flood data and catchment areas of 261 streamflow gauging sites of the 17 Subzones of India were collected for carrying out the study. Out of these, the data of 196 streamflow gauging sites and their catchment areas have been used for regional flood frequency estimation. The data of remaining 65 streamflow gauging sites have been excluded as per the data screening and regional homogeneity testing procedures. The record length for these streamflow gauging sites varies from 5 to 38 years. The catchment areas of the streamflow gauging sites range from 6 km^2 to 2,297 km^2 and their mean annual peak floods vary from 12.8 m^3/s to 1687.3 m^3/s.

Out of the 17 Subzones, PE3 has been identified as the robust distribution for 7 Subzones, GNO for 3 Subzones, GEV for 3 Subzones, GPA for 3 Subzones and GLO for 1 Subzone of India. The regional flood frequency relationships have been developed based on the respective robust identified frequency distributions for estimation of floods of various return periods for gauged catchments for the 17 Subzones.

For estimation of floods of various return periods for ungauged catchments, the regional relationships have been developed between mean annual peak floods and catchments areas of the gauged catchments of the 17 Subzones using the Levenberg-

Marquardt (LM) iteration procedure. The performance of this technique has been evaluated based on the statistical performance indices viz. Efficiency (EFF), Correlation Coefficient (CORR), Root Mean Square Error (RMSE) and Mean Average Error (MAE). The regional relationships developed between mean annual peak floods and catchments areas for the 17 Subzones have been coupled with the respective L-moments based robust identified regional flood frequency relationships developed for gauged catchments for each of the Subzones.

The regional flood frequency relationships have also been developed for estimation of floods of various return periods for gauged and ungauged catchments for 4 Subzones out of the 17 Subzones using ANN and FIS techniques. Performances of ANN, FIS and L-moments in regional flood frequency estimation have been compared based on the statistical performance criteria viz. EFF, CORR, RMSE and MAE.

The regional flood frequency relationships developed in the present study based on L-moments provide a convenient method for estimation of floods of various return periods for gauged and ungauged catchments of the 17 Subzones of India for the practitioners. The applicability of ANN and FIS in regional flood frequency estimation is explored and comparison of ANN, FIS and L-moments establishes the potential of FIS in regional flood frequency estimation.

ACKNOWLEDGEMENTS

It is my proud privilege to express my heartfelt gratitude and sincere thanks to my supervisors Dr. N.K. Goel, Professor, Department of Hydrology, Indian Institute of Technology Roorkee and Dr. K.K.S. Bhatia, former Scientist 'F', National Institute of Hydrology, Roorkee and Director, Modinagar Institute of Technology, Modinagar for providing guidance and encouragement throughout this research work.

I am grateful to Prof. B.S. Mathur, Prof. D.K. Srivastava, Prof. D.C. Singhal, Prof. Himanshu Joshi, Dr. M. Perumal and Dr. D.S. Arya faculty members in the Department of Hydrology for their constant encouragement, guidance and cooperation.

I express my sincere thanks to Mr. R.D. Singh, Director, National Institute of Hydrology, Roorkee and other former Directors of NIH namely Dr. Satish Chandra, Dr. S.M. Seth, Dr. K.S. Ramasastri and Dr. K.D. Shrama for extending help, support and facilities during the course of this research work. The help and support received from Dr. J.R.M. Hosking are thankfully acknowledged. I am also grateful to Dr. C. Chatterjee, Dr. J.V. Tyagi, Dr. A.K. Lohani, Dr. R.P. Pandey, Dr. Sanjay Kumar, Dr. P.K. Bhunya, Dr. N. Panigrahy, Dr. Manohar Arora, Mr. T.R. Sapra and my other colleagues in NIH who were helpful in many ways.

I am short of words to express my feelings for all my family members, who stood besides me during the present work at various stages. My father, mother and brothers and teachers were always the source of inspiration for completion of this work. I am extremely grateful to my wife Mukta, and my children Akanksha, Prerana and Utkarsh for their persuasive inspiration, love, affection and patience throughout the study.

Lastly, the greatest of my gratitude is towards the Almighty for giving me power to fulfill the promises he entrusted to me as a human being. All the efforts that could be made in this study have been possible by the grace of God who is the Author and Finisher of every good work.

(Rakesh Kumar)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF NOTATIONS

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

ANN	Artificial Neural Network
FIS	Fuzzy Inference System
PWMs	Probability Weighted Moments
FL	Fuzzy Logic
CV	Coefficient of Variation
LP-3	Log Pearson Type - 3
MLE	Maximum-likelihood Estimators
GML	Generalized Maximum Likelihood
L_i	Input Layer
L_H	Hidden Layer
L_o	Output Layer
MAPE	Mean Absolute Percentage Error
ANGIS	Adaptive Neuro-GA Integrated System
GA	Generic Algorithm
CDF	Cumulative Distribution Function
EV1	Extreme Value Type-I Distribution
GEV	General Extreme Value Distribution
LOS	Logistic Distribution
GLO	Generalized Logistic Distribution
GPA	Generalized Pareto Distribution
GNO	Generalized Normal Distribution
PE3	Pearson Type-III Distribution
KAP	Kappa Distribution
WAK	Wakeby Distribution
NOR	Normal Distribution
EXP	Exponential Distribution
UNF	Uniform Distribution
μ	Mean
σ	Standard Deviation

γ	Skewness
S	Slope
D	Drainage Density
R	Annual Normal Rainfall
$\underline{I}$	Identity Matrix
MSE	Mean Square Error
f ( )	Neuron Transfer Function
j	Neuron
H_{oj}	Real-Value Output
D_i	Discordancy Statistic
H	Heterogeneity Statistic
τ_2	L- Coefficient of Variation
τ_3	L-skewness
τ_4	L-kurtosis
t_4^R	Regional Average Kurtosis
$\overline{Q}$	Mean Annual Peak Flood
Q_T	Flood Estimate for T Year Return Period
$Q_T/\overline{Q}$	Growth Factor for T Year Return Period
LM	Levenberg-Marquardt
EFF	Nash-Sutcliffe Coefficient
CORR	Correlation Coefficient
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
A	Catchment Area in km^2
C_T	Regional Coefficient for estimation of Q_T for ungauged catchments
Qp	Annual Maximum Peak Flood
P	Probability of Non-Exceedance
MF	Membership Function
a	Regional Coefficient for Regional Relationship between $\overline{Q}$ and A
b	Regional Coefficient for Regional Relationship between $\overline{Q}$ and A
LMA	Levenberg-Marquardt Algorithm
GNA	Gauss-Newton Algorithm

CHAPTER 1

INTRODUCTION

1.1 GENERAL

Since the beginning of scientific hydrology in the seventeenth century, one of the problems facing the engineers and hydrologists has been estimation of design floods for basins for which the data required for hydrological analysis are not available or the available data are inadequate. Pilgrim and Cordery (1992) mention that estimation of peak flows on small to medium-sized rural drainage basins is probably the most common application of flood estimation as well as being of greatest overall economic importance. In almost all cases, no observed data are available at the design site, and little time can be spent on the estimate, precluding use of other data in the region. The authors further state that hundreds of different methods have been used for estimating floods on small drainage basins, most involving arbitrary formulas. The three most widely used types of methods are the rational method, the U.S. Soil Conservation Service method and regional flood frequency methods. Considering the importance of prediction in ungauged basins, the International Association of Hydrological Sciences (IAHS) has launched Predictions in Ungauged Basins (PUB) as an IAHS initiative for the decade of 2003-2012, aimed at uncertainty reduction in hydrological practice (Sivapalan et al., 2003). Further, due to rapidly increasing population and speedy economic growth of India, there is a need for judicious and optimal planning, development and management of water resources projects including construction of more and more hydraulic structures for generation of hydropower, interlinking of rivers for enhancing the availability of water for various uses, construction of road and railway bridges as well as taking up structural and non-structural measures of flood management etc. Realizing the importance and requirement of Hydrological Design Aids (HDA); it has been proposed to develop the hydrological design aids under one of the components of the currently ongoing World Bank funded Hydrology Project Phase-II (HP-II). Eight Central government agencies and thirteen States of India are participating in the HP-II. Thus, there is an urgent need for making systematic efforts for developing a reliable and convenient regional flood frequency estimation procedure for gauged and ungauged catchments based on the state of art technique of regional flood frequency estimation for the practicing engineers, academicians and researchers. Also, there is a need for investigating the applicability of the soft computing techniques in regional flood frequency estimation.

Frequency analysis is performed to determine the frequency of the likely occurrence of hydrologic events. Singh (1994) mentions that the information on flood magnitudes and their frequencies is needed for design of various types of water resources projects/ hydraulic structures such as dams, reservoirs, spillways, bridges, road and railway bridges, culverts, levees, urban drainage systems, airfield drainage, irrigation systems, stream control works, water supply systems and hydroelectric power plants. Estimation floods of various return periods is also required for taking up various types of non-structural measures of flood management such as flood hazard modelling, flood risk zoning, flood plain zoning (e.g. Forster et al., 2005; Goyal and Arora, 2007; Forster et al., 2008; Chatterjee et al., 2008) for industrial, residential and recreational use, setting of flood insurance premiums, economic evaluation of flood protection projects, drought mitigation programmes etc. As per the Bureau of Indian Standards (BIS) hydrological design criteria, frequency based floods find their applications in estimation of design floods for almost all the types of hydraulic structures excluding large and intermediate size dams. For design of large and intermediate size dams Probable Maximum Flood (PMF) and Standard Project Flood (SPF) are adopted, respectively (National Institute of Hydrology, Roorkee, 1992). However, for these two cases also flood frequency analysis is generally performed for ascertaining the return periods of PMF and SPF.

The L-moments form basis of an elegant mathematical theory for carrying out regional frequency analysis and are being used by many organizations the worldwide. Hosking (1990) introduced the L-moments approach for estimation of parameters as well as for screening of data, testing the regional homogeneity and identifying the best fit distributions. The L-moments are capable of characterising a wider range of distributions, compared to the conventional moments. Zafirakou-Koulouris et al. (1998) mention that the L-moments offer significant advantages over ordinary product moments, especially for environmental data sets, because of the following:

i. L-moment ratio estimators of location, scale and shape are nearly unbiased, regardless of the probability distribution from which the observations arise (Hosking, 1990).
ii. L-moment ratio estimators such as L-coefficient of variation, L-skewness, and L-kurtosis can exhibit lower bias than conventional product moment ratios, especially for highly skewed samples.
iii. The L-moment ratio estimators of L- coefficient of variation and L-skewness do not have bounds which depend on sample size as do the ordinary product moment ratio estimators of coefficient of variation and skewness.
iv. L-moment estimators are linear combinations of the observations and thus are less sensitive to the largest observations in a sample than product moment estimators, which square or cube the observations.

v. L-moment ratio diagrams are particularly good at identifying the distributional properties of highly skewed data, whereas ordinary product moment diagrams are almost useless for this task (Vogel and Fennessey, 1993).

Robson and Reed (1999) presented the statistical procedures for flood estimation in the Flood Estimation Handbook. In the Handbook L-moments approach has been used for estimation of the parameters of the flood growth curves. The authors mention that L-moments are preferred for flood frequency estimation because of their robust properties in the presence of unusually small or large values (outliers). Griffs and Stedinger (2007 a, b) presented evolution of flood frequency analysis with *Bulletin 17* of USA. The authors mention that the fields of hydrology and flood frequency analysis have substantially evolved since *Bulletin 17* was first published and new techniques are now available which should become part of these standard procedures. A comparison is provided which demonstrates how the standard and weighted *Bulletin 17B* quantile estimators perform relative to alternative Log Pearson Type-III (LP3) quantile estimators that also make use of regional information.

Presently, the soft computing techniques are being used for solving various types of hydrologic problems (e.g. ASCE Task Committee, 2000 a, b; Coulibaly et al., 2000; Xiong and Shamseldin, 2001; Chang et al., 2005; Wu et al., 2005, Raghuwanshi et al., 2006; Nayak and Sudheer, 2007; Nayak et al., 2007). The soft computing techniques such as Artificial Neural Network (ANN) and Fuzzy Inference system (FIS) offer real advantages over conventional modeling, including ability to handle large amounts of noisy data from dynamic and nonlinear systems, especially when underlying hydrological relationships are not fully understood.

1.2 GAPS IN PRESENT PRACITICE OF REGIONAL FLOOD FREQUENCY ESTIMATION

In India studies have been carried out for regional flood frequency estimation by various organizations. Prominent among these include the studies carried out jointly by Central Water Commission (CWC), Research Designs and Standards Organization (RDSO) and India Meteorological Department (IMD) using the method based on synthetic unit hydrograph and design rainfall considering physiographic and meteorological characteristics for estimation of design floods (e.g. CWC, 1982; CWC, 1985) and regional flood frequency analysis studies carried out by RDSO using the USGS and pooled curve methods (e.g. RDSO, 1991) for various hydrometeorological Subzones of India. Besides these, regional flood frequency analysis studies have also been carried out at some of the academic and research Institutions (e.g. Chander et al., 1978; Perumal and Seth, 1985). In most of the regional flood frequency studies the conventional methods such as U.S.G.S. method, regression based methods and Chow's method etc. have been used. Some attempts have been made by Singh (1989), Sankarasubramanian (1995), Upadhyay and Kumar (1999), Kumar et al. (1999), Kumar et al. (2003 a, b), Kumar and Chatterjee (2005) and others to apply the recent approaches of regional flood frequency estimation.

Recently, the soft computing techniques such as Artificial Neural Networks (ANN) and Fuzzy Logic (FL) have been applied for solving some of the hydrological problems such as development of stage-discharge relationship, flood forecasting, rainfall-runoff modeling, estimation of precipitation and evaporation, ground water modeling, water quality modeling etc. (ASCE Task Committee, 2000 a, b; Jain et al., 2004; Raghuwanshi et al., 2006; Kumar et al., 2009). However, applications of ANNs in regional flood frequency estimation are limited and use of Fuzzy Logic in regional flood frequency estimation remains to be investigated. Whereas, recent studies show that the fuzzy modeling is more versatile and improved alternative to ANNs (Aqil et al., 2007; Lohani, 2007).

Thus the studies carried out for regional flood frequency estimation in India are limited to a few regions, scattered as well as they are mostly based on the various types of conventional techniques. As a result, the gap between research and practice in the area of regional flood frequency estimation is increasing. To overcome the problems of prediction of floods of various return periods for gauged, sparsely gauged and ungauged catchments, a robust and convenient method of regional flood frequency estimation is required to be developed for the practitioners in India. Also, there is a need for exploring the applicability of the soft computing techniques in regional flood frequency estimation.

1.3 BROAD OBJECTIVES OF THE STUDY

With a view to bridge the gaps in the procedure of regional flood frequency estimation in India as well as to explore the potential of the soft computing techniques in regional flood frequency estimation the present study has been carried out with the following objectives:

(i) To develop regional flood frequency relationships based on L-moments approach for gauged and ungauged catchments of the 17 Subzones of India, and

(ii) To investigate applicability of the soft computing techniques viz. Artificial Neural Network (ANN) and Fuzzy Inference System (FIS) in regional flood frequency estimation.

The study has been carried out for 17 hydrometeorologically homogeneous

Subzones of India covering about 79% of its geographical area. For this purpose, the annual maximum peak flood data of 196 streamflow gauging sites and their catchment areas have been used.

1.4 LAYOUT OF THESIS

The subject matter of this thesis has been laid out in six chapters. The first chapter gives a brief introduction about regional flood frequency estimation and the broad objectives of the study. The second chapter provides general description of regional flood frequency estimation and reviews the research works in the area of regional flood frequency analysis and soft computing techniques viz. ANN and FIS. The chapter three presents description of the study area and data used in the study. Chapter four provides the details of the methodology of L-moments for regional flood frequency estimation and applications of soft computing techniques viz. ANN and FIS in regional flood frequency estimation. The results are presented in chapter five along with the discussions. Chapter six concludes the findings of the study and provides suggestions for further research work.

CHAPTER 2

REVIEW OF LITERATURE

2.1 GENERAL

Estimation of design flood for various types of hydraulic structures has been engaging the attention of engineers, since long time. Flood frequency analysis has been a very active area of investigation in hydrology. Frequency based floods find their applications in design of various types of hydraulic structures as well as for taking of some of the measures of flood management. Chow (1964) mentions that the frequency analysis of streamflow data is believed to have been first applied to flood studies by Herschel and Freeman in 1880 to 1890 by means of a graphical procedure of using flow-duration curves. The author further quotes that according to Fuller (1914), the use of probability methods in runoff studies had been suggested to him in 1896 by George W. Rafter. Owing to the dearth of long-period records on American rivers at that time, the use of probability methods for flood frequency analysis was apparently hindered until later years. Fuller (1914) gave a full account of the first really comprehensive study of statistical methods applied to floods in the United States. However, Hazen (1914) soon discovered that if the logarithms representing the annual floods are used instead of the number themselves, the agreement with the normal law of errors is closer. This is true because the frequency distributions of annual floods are usually skewed or asymmetrical and the distribution can be suitably represented by such frequency distribution laws as the Galton, or lognormal-probability, law.

Hazen (1914) proposed the use of lognormal-probability paper and developed a procedure of analysis (Hazen, 1921). Hazen's method requires a table of factors for computing theoretical frequency curves by means of the coefficients of variation and skewness. The table was originally obtained by empirical methods and hence has been found to be inaccurate. A corresponding table of exact factors based on a mathematical procedure was later prepared by Chow (1954). Other laws of frequency distribution and methods of frequency analysis of floods were also proposed by many hydrologists. Type 1 and Type 3 of Karl Pearson's curves of frequency distribution were put in a form convenient for use in flood studies by Foster (1924). Gumbel (1941) published the first of a great number of papers (e.g. Gumbel, 1941; Gumbel, 1949) on the application of the Fisher-Tippett theory of extreme values to flood frequency analysis. The use of extreme-value theory has been further extended by other hydrologists. The Type III external distribution was first proposed by Gumbel (1954) for drought frequency analysis.

Jenkinson (1955) proposed the General Extreme Value (GEV) distribution. Its theory and practical applications are reviewed in the Flood Studies Report (Natural Environmental Research Council, 1975). The index flood method developed by the U.S. Geological Survey (Dalrymple, 1960 a, b; Benson, 1962) was also widely used to perform regional flood frequency analysis. The Flood Studies Report of Natural Environmental Research Council (1975) deals with the British flood frequency analysis procedures. Greenwood et al. (1979) introduced the concept of the probability weighted moments (PWMs) and Landwehr et al. (1979 a, b) compared the PWMs with the traditional techniques and carried out studies using the PWMs. Hosking (1990) introduced the theory of L-moments.

The main aspects of flood frequency analysis and its applications have been described by investigators such as Chow (1964), Nash and Shaw (1965), Bell (1968), Thomas and Benson (1970), Larson and Reich (1972), Yevjevich (1972), Filliben

(1975), Kendall (1975), Kite (1977), Kuczera (1982), Interagency Advisory Committee on Water Data (1982), U.S.W.R (1982), Gries and Wood (1983), Stedinger (1983), Lettenmaier and Potter (1985), Hebson and Cunnane (1986), Cunnane (1988), National Research Council (1988), Cunnane (1989), Tasker and Stedinger (1989), Bobee and Ashkar (1991), Lu and Stedinger (1992), Maidment (1992), Mc Cuen (1993), Stedinger et al. (1992), Stedinger et al. (1993), Cong et al. (1993); Barnett and Lewis (1994), Zrinji and Burn (1994), Karim and Chowdhury (1995), Hosking and Wallis (1997), Rao and Hamed (2000), Anderson et al. (2000), Kavvas (2003), Griffis and Stedinger (2007 a), Bhunya et al. (2007), and Bhunya et al. (2008) etc.

Recently the soft computing techniques have also drawn considerable attention for their effective applications in hydrology and water resources. The soft computing techniques such as ANN and FIS offer significant advantages over conventional modeling, including the ability to handle large amounts of noisy data from dynamic and nonlinear systems, especially when the underlying hydrological relationships are not fully understood. The applications of soft computing techniques in hydrology and water resources have been discussed by many investigators (Zimmermann 1991; Baldwin, 1996; ASCE Task Committee 2000 a, b; Coulibaly et al., 2000; Xiong and Shamseldin, 2001; Kumar et al., 2002; Chang et al., 2005; Wu et al.2005, Raghuwanshi et al., 2006; Nayak and Sudheer, 2007; Nayak et al., 2007; Kumar et al., 2009). The various aspects of flood frequency analysis as well as applications of the soft computing techniques in hydrology and water resources have been reviewed as follows.

2.2 FLOOD FRQUENCY ANALYSIS

Flood frequency analysis refers to estimation of floods of various return periods. The primary objective of frequency analysis is to relate the magnitude of extreme events to their frequency of occurrence through the use of probability distributions (Chow et al., 1988). Rao and Hamed (2000) mention that the data observed over an extended period of time in a river system or hydrometeorologically homogeneous region are analyzed in frequency analysis. The data are assumed to be independent and identically distributed. The flood data are considered to be stochastic and are space and time independent. Further, it is assumed that the floods have not been affected by natural or manmade changes in the hydrological regime in the system. The authors further mention that in practice, the true probability distribution of the data at a site or a region is unknown. The assumption that data in a given system arise from a single-parent distribution may be questionable when data from large watersheds are analyzed. In such cases, more than one type of rainfall or flow may contribute to extreme events in a region. However, for the analysis to be of practical use, simpler distributions are often used to characterize the relation between flood magnitudes and their frequencies. The performance of distributions is evaluated by using different statistical tests. Quite often, many assumptions made in flood frequency analysis may be invalid. At any rate these assumptions have been questioned and discussed extensively (Klemes, 1987 a, b: Yevjevich, 1968).

Hosking and Wallis (1997) have presented the L-Moments based regional frequency analysis approach. The authors mention that regional flood frequency analysis resolves the problem of short data records or unavailability of data by "trading space for time"; as the data from several sites are used in estimating flood frequencies at any site. Robson and Reed (1999) state that gauged records are rarely long enough to allow direct estimation of the average interval between major floods at a site, other than very approximately. This average interval defines the return period at which flooding occurs. The authors further mention that the return periods of interest in UK flood design are often as long as 50 or 100 years. For many catchments, streamflow data are not available or the data are inadequate at the site of interest. In such cases the methods of frequency analysis using data from a single site have limited value because of large sampling errors, and as a result, regional flood frequency analysis is performed. By defining a region that is hydrologically similar in terms of the parameters or variables to be studied, data from several gauging sites within this homogeneous region are pooled together into a single regional frequency analysis. Several methods are available to perform a regional analysis. One of the first steps in a regional frequency analysis is to define the region itself. The definition of a region depends on the quantities to be estimated. Many methods are available to define a region that is homogeneous. Regional boundaries can be defined in terms of similarity of flood-frequency curves in a region which can be considered homogeneous (Singh, 1994).

A number of methods have been used for carrying out regional flood frequency analysis. The index flood method developed by the U.S. Geological Survey (Dalrymple, 1960 a, b; Benson, 1962) was widely used to perform regional flood frequency analysis. A uniform approach for determining flood frequencies was recommended for use by U.S. federal agencies in 1967, which consisted of fitting Log Pearson type - 3 (LP-3) distribution to describe the flood data. This procedure was extended in 1976 to fitting LP-3 distribution with a regional estimator of the log-space skew coefficient and this was released as Bulletin 17 by US Water Resources Council (USWRC). Bulletins 17A and 17B were released subsequently, in 1977 and 1981, respectively. These procedures of the USWRC were widely followed in USA and a few other countries USWRC (1981). Cunnane (1988) describes twelve different regional flood frequency analysis methods.

Greis and Wood (1983) presented an initial evaluation of the index-flood approach, which did not reflect the uncertainties in flood quantile estimators, resulting from scaling the regional flood frequency estimates by the at-site means. Some of the prominent flood frequency analysis studies include Potter and Walker (1981), Wallis and Wood (1985), Lettenmaier et al. (1987), Boes et al. (1989), Jin and Stedinger (1989), Potter and Lettenmaier (1990), Farquharson (1992), Burn and Goel (2000) etc. Cunnane (1989) mentions that a procedure for estimating flood magnitudes for return period of T years Q_T is robust if it yields estimates of Q_T which are good (low bias and high efficiency) even if the procedure is based on an assumption which is not true. Farquharson (1992) states that GEV distribution was selected for use in the Flood Studies Report (Natural Environmental Research Council, 1975) and has been found in other studies to be flexible and generally applicable. Hosking and Wallis (1997) mention that the method recommended in the U.K. Flood Studies Report (Natural Environmental Research Council, 1975) has a strong regional component. It divides the British Isles into eleven regions with region boundaries largely following those of major catchments. The frequency distribution of annual maximum stream flow is assumed to be the same at each gauging site in a region after the streamflow values have been divided by the site mean annual maximum streamflow. Some of the recent flood frequency analysis studies have been reviewed in Section 2.8.

Based on data availability and record length of the data the following three types of approaches may be adopted for developing the flood frequency relationships: (a) at-site flood frequency analysis, (b) at-site and regional flood frequency analysis, and (c) regional flood frequency analysis. The steps involved in carrying out flood frequency analysis based on the above approaches are mentioned below.

2.2.1 At-Site Flood Frequency Analysis

(i) Fit various frequency distributions to the annual maximum peak flood data of a stream flow gauging site.

(ii) Select the best fit distribution based on the goodness of fit criteria.

(iii) Use the best fit distribution for estimation of T-year flood.

2.2.2 At-Site and Regional Flood Frequency Analysis

(i) Identify a hydrometeorologically homogeneous region.

(ii) Screen the observed annual maximum peak flood data of the streamflow gauging sites of the homogeneous region and test the regional homogeneity.

(iii) Develop regional flood frequency relationships for the region considering various frequency distributions.

(iv) Select the best fit distribution based on the goodness of fit criteria.

(v) Estimate the at-site mean annual peak flood.

(vi) Use the best fit regional flood frequency relationship for estimation of T-year flood for gauged catchment.

2.2.3 Regional Flood Frequency Analysis

(i) Identify a hydrometeorologically homogeneous region.

(ii) Screen the observed annual maximum peak flood data of the streamflow gauging sites of the homogeneous region and test the regional homogeneity.

(iii) Develop regional flood frequency relationship for the region considering various frequency distributions.

(iv) Select the best fit distribution based on the goodness of fit criteria.

(v) Develop a regional relationship between mean annual peak flood and physiographic and climatic characteristics of the gauged catchments for the region.

(vi) Estimate the mean annual peak flood using the developed regional relationship.

(vii) Use the best fit regional flood frequency relationship for estimation of T-year flood for ungauged catchments.

2.3 ASSUMPTIONS AND DATA REQUIREMENT

The assumptions and data requirement for frequency analysis are described below.

2.3.1 Assumptions in Frequency Analysis

The three assumptions are implicit in frequency analysis.

(i) The data to be analyzed describe random events.

(ii) The natural process of the variable is stationary with respect to time and

(iii) The population parameters can be estimated from the sample data.

2.3.2 Assumptions in Index-Flood Procedure

This index-flood procedure makes the following assumptions (Hosking and Wallis, 1997).

(i) Observations at any given site are identically distributed.

(ii) Observations at any given site are serially independent.

(iii) Observations at different sites are independent.

(iv) Frequency distributions at different sites are identical apart from a scale factor.

(v) The mathematical form of the regional growth curve is correctly specified.

2.3.3 Data Requirement for Frequency Analysis

For flood frequency analysis either annual flood series or partial duration flood series may be used. The requirements with regard to data are that:

(i) It should be relevant.

(ii) It should be adequate and

(iii) It should be accurate.

The term relevant means that data must deal with problem. For example, if the problem is of duration of flooding then data series should represent the duration of flows in excess of some critical value. If the problem is of interior drainage of an area then data series must consist of the volume of water above a particular threshold. The term adequate primarily refers to length of data. The length of data primarily depends upon variability of data and hence there is no guide line for the length of data to be used for frequency analysis. The term accurate also refers to the homogeneity of data and accuracy of the discharge values. The data used for analysis should not have any effect of man made changes. Changes in the stage-discharge relationship may render stage records non-homogeneous and unsuitable for frequency analysis. It is therefore preferable to work with discharge values and if stage frequencies are required then most recent rating curve is used.

2.4 ADEQUACY OF RECORD LENGTH FOR FLOOD FREQUENCY ANALYSIS

Subramanya (1990) mentions that the flood frequency studies are most reliable in climates that are uniform from year to year. In such cases even a relatively short record gives a reliable picture of the frequency distribution. With increasing lengths of flood records, it affords a viable alternative method of flood-flow estimation in most cases. The author further states that the minimum number of years of record required to obtain satisfactory estimates depends upon the variability of data and hence on the physical and climatological characteristics of the basin.

Robson and Reed (1999) states that single site analysis is used when there is a reliable and long record at the site of interest and when the target return period T is not too long. Single-site analysis is not usually appropriate if the record length is shorter than T. If the record is between T and 2T years in length, it is recommended that both a single site analysis and a pooled analysis are carried out. If the record length is more than 2T years long, then a single-site analysis is usually sufficient, but comparison with a pooled analysis is recommended as a precaution. The number of stations included in the pooling-group is determined by a rule of thumb: the *5t rule*. This specifies that the pooled stations should collectively supply five times as many years of record as the target return period, T. Thus, the pooling-group is sized to provide at least 5T *station-years* of flood data.

2.5 PARAMETERS ESTIMATION

Several approaches have been used for estimating the parameters of frequency distributions. Some of the commonly used parameter estimation approaches for most of the frequency distributions include:

(i) Method of least squares

(ii) Method of moments

(iii) Method of mixed moments

(iv) Method of maximum likelihood

(v) Method of probability weighted moments

(vi) Method of maximum entropy

(vii) Method of L-moments

The method of least squares is based on the principal of least squares for the sum of squares of residuals. The method of moments has been one of the simplest and conventional parameter estimation techniques used in statistical literature. In this method, while fitting a probability distribution to a sample of data, the parameters are estimated by equating the sample moments to those of the theoretical moments of the frequency distribution. Even though this method is conceptually simple, and computations are straight-forward, it is found that numerical values of the sample moments can be very different from those of the population from which the sample has been drawn, especially when sample size is small and/or the skewness of the sample is considerable. Further, estimated parameters of distributions fitted by method of moments, are not very accurate. Stedinger et al. (1992) mention that the method that has strong statistical motivation is the method of maximum likelihood. Maximum likelihood estimators (MLEs) have very good statistical properties in large samples, and experience has shown that they generally do well with records available in hydrology. However, often MLEs cannot be reduced to simple formulas, so estimates must be calculated using numerical methods. Cunnane (1989) described statistical distributions for flood frequency analysis. Hosking (1990) introduced the L-moments approach for estimation of parameters as well as for screening of data, testing the regional homogeneity and identifying the best fit distributions.

A number of attempts have been made literature to develop unbiased estimates of skewness for various distributions. However, these attempts do not yield exactly unbiased estimates. Further, a notable drawback with conventional moment ratios such as skewness and coefficient of variation is that, for finite samples, they are bounded, and are not able to attain the full range of values available to population moment ratios (Kirby, 1974). Wallis et al. (1974) have shown that the sample estimates of conventional moments are highly biased for small samples. The L-moments are capable of characterising a wider range of distributions, compared to the conventional moments. A distribution may be specified by its L-moments, even if some of its conventional moments do not exist (Hosking, 1990). Further, L-moments are more robust to outliers in data than conventional moments (Vogel and Fennessey, 1993) and enable more reliable inferences to be made from small samples about an underlying probability distribution.

Stedinger et al. (1992) mention that fitting a distribution to data sets provides a compact and smoothened representation of the frequency distribution revealed by the available data, and leads to a systematic procedure for extrapolation to frequencies beyond the range of the data set. When flood flows, low flows, rainfall, or water-quality variables are well-described by some family of distributions, a task for the hydrologist is to estimate the parameters of that distribution so that required quantiles and expectations can be calculated with the "fitted" model. Appropriate choices for distribution functions can be based on examination of the data using probability plots and moment ratios, the physical origins of the data, previous experience, and administrative guidelines. Stedinger et al. (1992) have also described the theoretical properties of the various distributions commonly used in hydrology, and have summarised the relationships between the parameters and the L-moments. The expressions to compute the biased and the unbiased sample estimates of L-moments and their relevance with respect to hydrologic application have also been presented by the authors.

Hosking (1990) also introduced L-moment ratio diagrams, which are quite useful in selecting appropriate regional frequency distributions of hydrologic and meteorologic data. The advantages offered by L-moment ratio diagrams over conventional moment ratio diagrams are well elucidated by Vogel and Fennessey (1993). The advantages offered by L-moments over conventional moments in hypothesis testing, boundedness of moment ratios and identification of distributions have also been discussed in detail by Hosking and Wallis (1997). Recently a number of regional flood frequency analysis studies have been carried out based on the L-moments approach. The L-moment methods are demonstrably superior to those that have been used previously, and are now being adopted by many organizations worldwide (Hosking and Wallis, 1997).

2.6 GOODNESS-OF-FIT TESTS

Rigorous statistical tests have been used and are useful for assessing whether or not a given set of observations might have been drawn from a particular family of distributions. The goodness of fit tests provide evaluation criteria for identifying the robust frequency distribution based on the comparison of different frequency distributions. The various goodness-of-fit tests include (i) Kolmogorov-Smirnov test, (ii) Chi-square test, (iii) D-index test, (iv) Descriptive ability criteria, (v) Predictive ability criteria, (vi) L-moments ratio diagram and (vii) L-moments based heterogeneity statistic (H) criteria described by Hosking and Wallis (1997). The Kolmogorov-Smirnov test provides bounds within which every observation on a probability plot should lie if the sample is actually drawn from the assumed distribution. It is useful for evaluating visually the adequacy of a fitted distribution (Stephens, 1974). In the Chi-square test, data are first divided into k class intervals.

The statistic Chi square is distributed asymptotically as Chi square with k-1 degrees of freedom. The observed number of events in the class interval is the number of events that would be expected from the theoretical distribution and k is an arbitrary number of classes to which the observed data are divided and the Chi square value is computed (Rao and Hamed, 2000).

The D-index for comparison of the fit of various distributions in upper tail is given as:

$$D\ index = (1/\bar{x})\sum_{i=1}^{6} Abs(x_i - \hat{x}_i) \tag{2.1}$$

where x_i and $\hat{x}_i$ are the ith highest observed and computed values for the distribution.

As per this test the distribution giving the least D-index is considered to be the best fit distribution. The descriptive ability criteria relate to ability of a chosen model to describe/reproduce chosen aspects of observed flood peaks. The descriptive ability criteria which have been used in flood frequency analysis are: (i) average of relative deviations between computed and observed values of annual maximum peak discharge (ADF), efficiency (EFF) and standard error (SE). The predictive ability criteria relate to statistical ability of procedure to achieve its assigned task with minimum bias, and maximum efficiency and robustness and various predictive ability criteria used in flood frequency analysis are: (i) Bias (BIAS), (ii) Root mean square error (RMSE) and (iii) Coefficient of variation (CV). The details of these criteria are discussed elsewhere (Cunnane, 1989; National Institute of Hydrology, 1994-95). The L-moments based goodness of fit test defined by Hosking and Wallis (1997) are L-moment ratio diagram and $|Z_i^{dist}|$ –statistic criteria. These tests of goodness of fit are the most powerful, out of all the available tests. The details of these tests are presented in Chapter 4.

2.6.1 Identification of Homogeneous Region

Hosking and Wallis (1997) mention that of all the stages in regional frequency analysis involving many sites, the identification of homogeneous regions is usually most difficult and requires the greatest amount of subjective judgement. The aim is to form groups of sites that approximately satisfy the homogeneity condition, that the sites' frequency distributions are identical apart from a site-specific scaling factor. Several authors have proposed methods for forming groups of similar sites for use in regional frequency analysis. The authors have categorized the procedures as geographical convenience, subjective partitioning, objective partitioning, cluster analysis and other multivariate analysis methods. A summary of these procedures and some of the examples of their applications in regional frequency analysis, described by the authors is given below.

Under the procedure of geographical convenience the regions are often chosen to be sets of contiguous sites based on administrative areas (Natural Environmental Research Council, 1975), or major physical groupings of sites (Matalas et al., 1975). Cervantes et al. (1983) presented a cluster model for flood analysis. It is sometimes possible, particularly in small scale studies, to define regions subjectively by inspection of the site characteristics. In objective partitioning methods, regions are formed by assigning sites to one of the two groups depending on whether a chosen site characteristic does or does not exceed some threshold value. The threshold is chosen to minimize a within-group heterogeneity criterion, such as a likelihood-ratio statistic within-group variation of the sample coefficient of variation (Wiltshire, 1986 a, b). The groups are then further divided in an iterative process until a final set of acceptably homogeneous regions is obtained.

Acreman and Sinclair (1986) analysed annual maximum streamflow data for 168 gauging sites in Scotland and formed five regions, four of which they judged as homogeneous. Burn (1989) used cluster analysis to derive regions for flood frequency analysis, though his cluster variables include at-site statistics.

Schaefer (1990) analyzed the annual maximum peak flood data for sites in Washington state and formed regions by grouping together sites with similar values of mean annual precipitation.

Pilon and Adamowski (1992) carried out a Monte-Carlo simulation study to show the value of information added to flood frequency analysis, by adopting a GEV regional shape parameter model over the at-site models using the observed data collected from the province of Nova Scotia (Canada). However, authors assumed the at-site mean in all sites considered as 100.0 and they have generated the flood data directly from a GEV distribution (after selecting through L-Moment ratio diagram), whose parameters have been computed from the regional moments. This simulation does not correspond to the true regional Monto-Carlo simulation of the region considered, even though it shows that additional information value is added by regional models. Further, their simulation does not incorporate the degree of heterogeneity present in the region.

Hosking and Wallis (1997) mention that for regional frequency analysis with an index-flood procedure there is little advantage in using very large regions. The authors further mention that little gain in the accuracy of quantile estimates is obtained by using more than about 20 sites in a region. Thus, there is no compelling reason to amalgamate large regions whose estimated regional frequency distributions are similar.

2.6.2 Regional Homogeneity Tests

In carrying out regional flood frequency analysis once a set of physically plausible regions has been identified, it is desirable to assess whether the region is meaningful and may be accepted as homogeneous. Various types of homogeneity tests are reported in literature e.g. Dalrymple's (1960a, b) homogeneity test (U.S.G.S. test), the tests proposed by Acreman and Sinclair (1986), Wiltshire (1986 a, b), Choudhury et al. (1991) etc. Most of these tests involve a statistical value which measures some aspect of frequency distribution which is uniform/constant in a homogeneous region. This statistic may be a 10 year value scaled by mean, coefficient of variation, coefficient of skewness or L-moment ratio of a combination thereof. The test statistic H, termed as heterogeneity measure has been described by Hosking and Wallis (1997). It is a very effective regional homogeneity test and it is being very widely used in carrying out regional flood frequency analysis. It compares inter-site variations in sample L-moments for the group of sites with what would be expected of a homogeneous region. This heterogeneity measure has been discussed in detail in Chapter 4.

2.7 FLOOD FREQUENCY ANALYSIS STUDIES CARRIED OUT IN INDIA

A number of studies have been carried out in the area of regional flood frequency analysis in India. Some of these include Goswami (1972), Thiruvengadachari et al. (1975), Varshney (1979), Jhakade et al. (1984), National Institute of Hydrology (1984-85), Venkataraman and Gupta (1986), Venkataraman et al. (1986), Thirumalai and Sinha (1986), Mehta and Sharma (1986), Huq et al. (1986), Kaur (1988), Upadhyay et al. (1990), Research, Designs and Standards Organization

(1991), National Institute of Hydrology (1990-91), National Institute of Hydrology (1997-98), Kurothe et al. (1997), Kurothe et al. (2001), Ali and Singh (2001), Bhatt (2003), Sikka and Selvi (2005), Goyal and Arora (2007), Bhadra et al. (2008). In most of the regional flood frequency studies the conventional methods such as U.S.G.S. method, regression based methods and Chow's method have been used. Some attempts have been made by Chander et al. (1978), Perumal and Seth (1985), Singh and Seth (1985), Seth and Singh (1987), Singh (1989) and others to study the applications of new approaches of regional flood frequency analysis for some of the typical regions of India for which the conventional methods had been already applied. Some of the recent studies on regional flood frequency estimation are reviewed as follows.

Sankarasubramanian (1995) investigated the sampling properties of L-moments for both unbiased and biased estimators for five of the commonly used distributions. Based on the simulation results, regression equations have been fitted for the bias and the variance in L-skewness for the five distributions. The sampling properties of L-moments have been compared with those of conventional moments and the results of the comparison have been presented for both the biased and unbiased estimators. The performance of evaluation in terms of relative RMSE in third moment ratio reveals that conventional moments are preferable at lower skewness, while L-moments are preferable at higher skewness.

Kumar and Singh (1996) carried out a comparative study for the seven hydrometeorological Subzones of Zone-3 of India using the EV1 distribution by fitting the probability weighted moment (PWM) as well as following the modified U.S.G.S. method. In the study General Extreme Value (GEV) and Wakeby distribution based on PWMs have also been used and performances of the various methods have been evaluated based on the descriptive ability and predictive ability criteria.

Upadhyay and Kumar (1999) applied L-moments approach for regional flood frequency analysis. The study concluded that at gauged sites, regional flood frequency estimates were found to be more accurate than at-site estimates as is clear from root mean square error and standard error of regional estimates as compared to at-site estimates. The authors recommended that alongside the discharge data collection at gauging sites, emphasis should be given collection of data about the physiographic and hydrological characteristics of the catchment. The authors recommended that it would improve reliability and accuracy of regional flood estimates not only at ungauged sites, but also at gauged sites having short record lengths and facilitate reliable and economically viable design of the hydraulic structures.

Parida and Moharram (1999) compared quantile estimates computed using some of the commonly used statistical models and found that based on ranking of mean absolute deviation of the estimates the Generalized Pareto (GPA) distribution, in general, performed well for the study area.

Parmeswaran et al. (1999) developed a flood estimating model for individual catchment and for the region as a whole using the data of fifteen gauging sites of Upper Godavari Basins of Maharashtra. Seven probability distributions were used in the study. Based on the goodness of fit tests log normal distribution is reported to be the best fit distribution. A regional relationship between mean annual peak flood and catchment area has been developed for estimation of mean annual peak flood for ungauged catchments and regional relationship for maximum discharge of a known recurrence interval for the ungauged catchments.

Kumar and Chatterjee (2005) carried out regional flood frequency analysis for North Brahmaputra region of India. In the study, data of 13 stream flow gauging sites were screened using the discordancy measure (D_i) and homogeneity of the region is then tested employing the L-moment based heterogeneity measure (H). Based on this test, it was observed that the data of 10 out of 13 gauging sites constituted a homogeneous region. Comparative regional flood frequency analysis studies were conducted employing the L-moments based commonly used frequency distributions. Based on the L-moment ratio diagram and $|Z_i^{dist}|$ -statistic criteria, Generalized Extreme Value (GEV) distribution was identified as the robust distribution for the study area. Regional flood frequency relationships were developed for estimation of floods of various return periods for gauged and ungauged catchments using the L-moment based GEV distribution and a regional relationship developed use the method of least squares between mean annual peak flood and catchment area. Flood frequency estimates of gauged and ungauged catchments were compared; when, without satisfying the criteria of regional homogeneity, data of all the 13 gauging sites were used instead of data of only 10 gauging sites constituting the homogeneous region.

Some of the recent flood frequency analysis studies carried out abroad have been reviewed as follows.

2.8 RECENT FLOOD FREQUENCY ANALYSIS STUDIES CARRIED OUT ABROAD

Wang (1996) mentioned that the estimation of floods of large return periods from lower bound censored samples may often be advantageous because interpolation and extrapolation are made by exploring the trend of larger floods in each of the records. The method of partial probability weighted moments (partial PWMs) is a useful technique for fitting distributions to censored samples. The author redefined partial PWMs. The expression for partial PWMs is derived for the extreme values type I distribution. Combined with those for the extreme value II and III distributions, an unified expression for partial PWMs is presented for the GEV distribution. The equations for solving the distribution parameters are provided. Monte Carlo simulation shows that lower bound censoring at a moderate level does not unduly reduce the efficiency of high-quantile estimation even if the samples have come from a true GEV distribution.

Zafirakou–Koulouris et al. (1998) introduced L-moments diagrams for the evaluation of goodness of fit for censored data (data containing values above or below the analytical threshold of measuring equipments). The authors also summarized the advantages of the L-moments approach.

Iacobellis and Fiorentino (2000) presented a new rationale, which incorporates the climatic control for deriving the probability distribution of floods which based on the assumption that the peak direct streamflow is a product of two random variates, namely, the average runoff per unit area and the peak contributing area. The probability density function of peak direct streamflow was found as the integral over total basin area, of that peak contributing area times the density function of average runoff per unit area. The model was applied to the annual flood series of eight gauged basins in Basilicata (Southern Italy) with catchment area ranging from 40 to 1600 km^2. The results showed that the parameter tended to assume values in good agreement with geomorphologic knowledge and suggest a new key to understand the climatic control of the probability distribution of floods.

Martins and Stedinger (2000) mention that the three-parameter extreme-value (GEV) distribution has found wide application for describing annual floods, rainfall, wind speeds, wave heights, snow depths and other maxima. Previous studies show that small-sample maximum-likelihood estimators (MLE) of parameters are unstable and recommend L-moment estimators. Examination of the behaviour of MLEs in small samples demonstrates that absurd values the GEV-shape parameter k can be generated. The authors state that use of a Bayesian prior distribution to restrict k values to a statistically/physically reasonable range in a generalized maximum likelihood (GML) analysis eliminates this problem.

Durrans et al. (2003) mention that in some applications, it is desirable to perform joint (i.e., simultaneous) flood frequency analyses on seasonal as well as annual bases. However, a problem one encounters in seasonal flood frequency analysis is that the consistency or interrelationship that must exist between the annual maximum and individual seasonal flood frequency distributions may not be preserved. The most important cause of inconsistencies is that one cannot arbitrarily specify the parametric forms of the annual and all of the seasonal distributions. A correct theoretical analysis of the joint frequency problem would require the use of a rather unusual and complicated distributional model. The authors mention that their study presents two approximate but useful methods for joint frequency analysis using the log Pearson Type 3 distribution. The authors show via examples that the two methods can be applied to reasonably model annual and five seasonal flood distributions in the Tennessee Valley.

Jingyl and Hall (2004) applied the geographical approach (Residual method), Wards' cluster method, the Fuzzy c-means method and a Kohonen neural network to 86 sites in the Gan-Ming river basin of China to delineate homogeneous regions based on site characteristics. The authors state that since the Kohonen neural network can be employed to identify the number of sub-regions as well as the allocation of the sites to sub-regions, this method is preferred over Ward's method and the Fuzzy c-means approach. The regional L-moment algorithm has been used to take advantage of both identifying an appropriate underlying frequency distribution and to construct sub-regional growth curves.

Chokmani and Quarda (2004) proposed a physiographical space-based kriging method for regional flood frequency estimation. The methodology relies on the construction of a continuous physiographical space using physiographical and meteorological characteristics of gauging stations and the use of multivariate analysis techniques. Two multivariate analysis methods were tested: canonical correlation analysis and principal component analysis. Ordinary kriging, a geostatistical technique, was then used to interpolate flow quantiles through the physiographical space. Data from 151 gauging stations across the southern part of the province of Quebec, Canada, were used to illustrate this approach. Results of the proposed method were compared to those produced by a traditional regional estimation method using the canonical correlation analysis. The proposed method estimated the 10 year return period specific flow with a coefficient of determination of 0.78. However, this performance decreases with the increase in quantile return period. The authors also observed that the proposed method works better when the physiographical space is defined using canonical correlation analysis.

Merz and Bloschl (2005) examined the predictive performance of various regionalization methods for the ungauged catchment case, based on a jack-knifing comparison of locally estimated and regionalized flood quantiles of 575 Austrian catchments. It is observed that spatial proximity is a significantly better predictor of regional flood frequencies than are catchment attributes. A method that combines spatial proximity and catchment attributes yields the best predictive performance. The method is based on kriging and takes differences in the length of the flood records into account. It is shown that short flood records contain valuable information which can be exploited by the method proposed by the authors. A method that used only spatial proximity performs second best. The methods that only use catchment attributes perform significantly poorer than those based on spatial proximity. These are a variant of the Region of Influence (ROI) approach, applied in an automatic model and multiple regressions. The authors suggest that better predictive variables and similarity measures need to be found to make these methods more useful.

Cunderlik and Burn (2006) developed a new pooling approach that takes into consideration the sampling variability of flood seasonality measures used as pooling variables. A nonparametric resampling technique is used to estimate the sampling variability for the target site, as well as for every site that is a potential member of the pooling group for the target site. The variability is quantified by Mahalanobis distance ellipses. The similarity between the target site and potential site is then assessed by finding the minimum confidence interval at which their Mahalanobis ellipses intersect. The confidence intervals can be related to regional homogeneity, which allows the target degree of regional homogeneity to be set in advance. The approach is applied to a large set of catchments from Great Britain, and its performance is compared with the performance of a previously used pooling technique based on Euclidean distance. The results demonstrated that the proposed approach outperforms the previously used approach in terms of the overall homogeneity of delineated pooling groups in the study area.

Kjeldsen and Jones (2006) mention that the standard for conducting flood frequency analysis in the UK, as set out in the Flood Estimation Handbook, is based on the index flood method, using the median of the annual maximum flood as the index flood. The authors used a region-of-influence approach is used, involving the creation of a collection of hydrologically similar catchments. The authors also examined the sampling uncertainty of quantile estimates on the basis of pooling groups and using the median as the index flood for both gauged and ungauged sites. Analytical approximations for the variance of the quantile estimates were derived, on the basis of asymptotic theory, and were used to calculate approximate confidence intervals for flood frequency curves obtained using both single-site and pooled analysis at gauged and ungauged sites. The authors showed that the pooled analysis yields narrower confidence intervals than the single-site analysis and that the presence of intersite correlations increases the sampling uncertainty. The method was extended to encompass estimation at ungauged sites in the UK on the basis of a regression model for the index flood, which significantly increases the prediction uncertainty compared with using an estimate of the index flood derived from observations at the target site.

Zhang and Singh (2006) derived bivariate distributions of flood peak and volume, and flood volume and duration using the copula method. The authors state that major advantage of this method is that marginal distributions of individual variables (i.e. flood peak, volume, and duration) can be of any form and the variables can be correlated. The copula method was applied to obtain the conditional return periods that are needed for hydrologic design. The derived distributions were tested using flood data from Amite River at Denham Springs, La., and the Ashuapmushuan River at Saguenay, Quebec, Canada. The derived distributions were also compared with the Gumbel mixed and the bivariate Box-Cox transformed normal distributions. The copula-based distributions were found to be in better agreement with plotting position-based frequency estimates than were other distributions.

Chebana and Quarda (2007) presented a multivariate L-moments homogeneity test with the aim to extend the statistical homogeneity test of Hosking and Wallis (1997) to the multivariate case. The usefulness of the methodology is illustrated on flood events. Monte-Carlo simulations are also performed for a bivariate Gumbel logistic model with Gumbel marginal distributions. Results illustrate the power of the proposed multivariate L-moment homogeneity test to detect heterogeneity on the whole structure of the model and on the marginal distributions. In a bivariate flood setting, a comparison is carried out with the classical homogeneity test of Hosking and Wallis based on several types of regions.

Griffs and Stedinger (2007a) presented evolution of flood frequency analysis with *Bulletin 17*. The authors mention that the current methodology recommended for flood-frequency analyses by U.S. Federal agencies is presented in *Bulletin 17B*. *Bulletin 17* was first published in 1976, minor corrections were made in 1977 resulting in *Bulletin 17A*, which was later succeeded by *Bulletin 17B* published in 1982. The authors further mention that the fields of hydrology and flood frequency analysis have substantially evolved since *Bulletin 17* was first published. New techniques are now available which should become part of these standard procedures. The authors provide a comparison which demonstrates how the standard and weighted *Bulletin 17B* quantile estimators perform relative to alternative Log Pearson Type-III (LP3) quantile estimators that also make use of regional information.

Griffis and Stedinger (2007b) state that since the adoption of the log-Pearson Type 3 (LP3) distribution by U.S. federal agencies, it has been widely used in hydrology, but its properties are not well understood. The authors explore the characteristics of the LP3 distribution in both real space and log space, and their relationship and comparisons with U.S. flood data summaries reveal that the LP3 distribution provides a reasonable model of the distribution of annual flood series from unregulated watersheds for log space skews $|\gamma_x| \leq 1.414$ (through $|\gamma_x| \leq 1$ is more realistic), and for $\gamma_x = 0$ with standard deviations in the range 0.1 to 1.0 with base-e natural logarithms (0.04 to 0.43 with base-10 common logarithms). L-moment ratio relationships for the LP3 distribution are also developed by the authors so they can be compared to summary statistics for a region, and to several other distributions frequently recommended for modeling hydrometeorological extremes.

Genest et al. (2007) introduced metaelliptical copulas as a flexible tool for modeling multivariate data in hydrology. The author reviewed the properties of the broad class of dependence functions, along with associated rank-based procedures for copula parameter estimation and goodness-of-fit testing. A new graphical diagnostic tool is also proposed for selecting an appropriate metaelliptical copula. The author use peak, volume, and duration of the annual spring flood for the Romaine River (Quebec, Canada) for illustration purposes.

Zhang and Singh (2007) derived volume, and duration, and then obtained conditional return periods using the Gumbel-Hougaard copula, trivariate distributions of flood peak. The derived distributions were tested using flood data from the Amite River Basin in Louisiana. The authors mention that a major advantage of the copula method is that marginal distributions of individual variables can be of any form and the variables can be correlated.

Strupczewski et al. (2009) mention that the main objections to the use of a pure statistical approach in the analysis of hydrological extremes are small sample size and unknown distribution function. The maximum likelihood (ML) estimates of large quantiles are highly sensitive to the distributional choice, while the power of discrimination procedures is unacceptably low for hydrological sample sizes. The L-moments method seems to be the best for this purpose. Application of heavy-tailed distributions for extremes modeling is discussed by the authors. Moreover two-shape parameter distributions, while some of them are heavy-tailed, are proposed. Keeping in mind that the largest sample element is a low quality data, the effect of its omission on the L-moments accuracy of upper quantiles of two-parameter heavy-tailed distribution is examined. The authors further mention that recent developments in the statistics of extremes are primarily related to the maximum likelihood estimation in the presence of covariates. Its present and prospective hydrological applications are discussed with emphasis on non-stationary flood frequency analysis. As and alternative a two level estimation technique is proposed by the authors for estimation of non-stationary parameters of the distribution.

It is often necessary to interpret information about flood frequency in terms of the risk of exceedance, i.e. the probability of a flood exceeding a threshold value. There are simple relationships between risk and return periods. Some aspects of risk analysis related to flood frequency analysis are reviewed as follows.

2.9 RISK ANALYSIS

Kite (1977) mentions that for any hydraulic structure there is a total risk of failure which can be broken down into the risk of failure of each project component i.e. hydrologic, hydraulic and structural. The risk within any component can then be broken down into true risk and uncertainty. Yen and Ang (1971) have used the terms objective risk and subjective risk. For the hydrologic component, risk is the calculable probability of failure e.g. occurrence of a certain flood, occurrence of a drought, etc. The calculation of risk is based on the assumption that the underlying event distribution is known. As an example, if it is known that flood magnitudes in a particular river valley location follow the lognormal distribution and that the time-distribution of the floods follow a Poisson distribution then the risk that the flood of a certain magnitude will occur in the next five years can be computed exactly.

Uncertainty occurs because the basic data available contain random measurement and computation errors, systematic errors, non-homogeneity in time, loss of information in changing from a continuous record to a discrete data set and so on. These imperfect data are then used to estimate the parameters of the assumed population distribution. Uncertainty generally increases as the variance of the sample data increases and decreases as the sample length increases. Prasad (1971) describes Risks in hydrologic design of engineering projects. Thomas (1971) has evaluated the errors in streamflow estimates made from a continuous stage record while Moss (1969) has related the standard error of discharge estimates to the number of streamflow measurements made per year and the associated costs of maintaining the station. The effect of uncertainty on the parameters of the population distribution can be included in an analysis by computing the standard error of estimate of the particular distribution at the required probability level. Confidence limits around the expected event magnitude can then be calculated. To summarize this concept, hydrologic risk is made up of basic risk and uncertainty both of which can be evaluated. What cannot be evaluated is the error caused by selecting the wrong distribution to fit the sample data.

Yen (1971) has tabulated values of T, the required design return period, for various expected project lives, n, and permissible risks of failure, p. Table 2.1 is reproduced from Yen (1971).

Table 2.1 Design return period for various project lives and risks of failure (Yen, 1971)

Permissible risk of failure	Expected Project Life, n, in years							
	1	2	5	10	20	25	50	100
0.99	1.01	1.11	1.66	2.71	4.86	5.95	11.4	22.2
0.95	1.05	1.29	2.22	3.86	7.16	8.85	17.2	33.9
0.90	1.11	1.46	2.71	4.86	9.19	11.4	22.2	43.9
0.75	1.33	2.00	4.13	7.73	14.9	18.6	36.6	72.6
0.50	2.00	3.41	7.73	14.9	29.4	36.6	72.6	145.0
0.33	3.00	5.45	12.9	25.2	49.9	62.1	124.0	247.0
0.25	4.00	7.46	17.9	35.3	70.0	87.3	174.0	348.0
0.20	5.00	9.47	22.9	45.3	90.1	113.0	225.0	449.0
0.10	10.0	19.5	48.0	95.4	190.0	238.0	475.0	950.0
0.05	20.0	39.5	98.0	195.0	390.0	488.0	975.0	1950.0
0.02	50.0	99.0	248.0	495.0	990.0	1238.0	2476.0	4951.0
0.01	100.0	199.5	498.0	995.0	1990.0	2488.0	4977.0	9953.0

2.10 SOFT COMPUTING TECHNIQUES

2.10.1 General

Recently soft computing techniques such as Artificial Neural Networks (ANN) and Fuzzy Inference system (FIS) are being applied for solving various types of hydrologic and water resources problems such as flood forecasting, development of rating curves, estimation of evaporation, estimation of sediment yield, approximating the three dimensional flow and transport processes in coastal aquifers, studying soil water retention, etc. Mohan (2007) mentions that the history of the ANNs stems from the 1940s, the decade of the first electronic computer. However, the first significant step took place in 1957 when Rosenblatt introduced the first concrete neural model, the perceptron. In 1959, Bernard Widrow and Marcian Hoff of Stanford developed models they called ADALINE and MADALINE. These models were named for their use of Multiple ADAptive LINear Elements. MADALINE was the first neural network to be applied to a real world problem. In 1974, Werbos introduced a so-called backpropagation algorithm for the three-layered perceptron network. Hopfield brought out his idea of a neural network in 1982.

Fuzzy logic is another area of artificial intelligence that has been applied successfully in different engineering fields. Fuzzy logic concepts were introduced by Lotfi A. Zadeh in 1965 (ASCE, 2000 a, b). He was a professor of computer science at the University of California in Berkeley. Fuzzy logic is a superset of conventional Boolean logic that has been extended to handle imprecise data and the concept of partial truth. In fuzzy logic, variables are "fuzzified" through the use of membership functions that define the membership degree to fuzzy sets. These variables are called linguistic variables. Fuzzy algorithms are formed by the union use of the fuzzy OR operator of individual fuzzy rules. The way in which the fuzzy operators IF, THEN, AND, OR are implemented can have a significant impact on model performance. Fuzzy systems are defined by a number of fuzzy rules, a number of membership functions, and mechanisms to apply logical operators. There are numerous successful applications of fuzzy systems in control and modeling. They are suitable for situations where an exact model of a process is either impractical or very costly to build, but an imprecise model based on existing human expertise can do the job. In such situations, fuzzy systems are considered the best alternative. Fuzzy sets are an aid in providing information in a more human comprehensible or natural form and can handle uncertainties at various levels. The new smart gadgets and fuzzy control systems appeared in mass in Japan and Korea in the 1990s. The soft computing techniques viz. ANN and FIS are described as follows.

2.10.2 Artificial Neural Network (ANN)

Pal and Mitra (1999) mention that Artificial Neural Networks are relatively crude electronic models based on the neural structure of the brain. The brain basically learns from experience. It is natural proof that some problems that are beyond the scope of current computers are indeed solvable by small energy efficient packages. This brain modeling also promises a less technical way to develop machine solutions. These biologically inspired methods of computing are thought to be the next major advancement in the computing industry. Even simple animal brains are capable of functions that are currently impossible for computers. The computers have trouble recognizing even simple patterns much less generalizing those patterns of the past into actions of the future. Now, advances in biological research promise an initial understanding of the natural thinking mechanism. This research shows that brains store information as patterns. Some of these patterns are very complicated and allow us the ability to recognize individual faces from many different angles. This process of storing information as patterns, utilizing those patterns, and then solving problems encompasses a new field in computing. This field does not utilize traditional programming but involves the creation of massively parallel networks and the training of those networks to solve specific problems. This field also utilizes words very different from traditional computing, words like behave, react, self-organize, learn, generalize, and forget.

The ANN methods are capable of adopting the non-linear relationship among the various hydrological variables, e.g. between rainfall and runoff as compared to the conventional techniques, which assume a linear relationship between rainfall and runoff. The ANNs have strong generalisation ability, which means that once they have been properly trained, they are able to provide accurate results even for cases they have never seen before (Haykin, 1994). The neural-network approach, also referred to as connectionism or paralleled distributed processing, adopts a "Brain metaphor" of information processing. Information processing in a neural network occurs through interactions involving large number of simulated neurons. Artificial neural-networks (ANNs) are massively parallel systems composed of many processing elements connected by links of variables weights. The network consists of layers of neurons, with each layer being fully connected to the proceeding layer by inter connection strengths or weights (W). Fig. 2.1, illustrates a three-layer neural network consisting of input layer (L_i), hidden layer (L_H) and the output layer (L_o) with the inter-connection weights W_{ih} and W_{ho} between layers of neurons. Some of the recent applications of ANNs in hydrology include comparison of ANNs and empirical approaches for predicting watershed runoff (Anmala et al., 2000); comparative analysis of event based rainfall runoff modelling techniques-deterministic, statistical and artificial neural networks (Jain and Indurthy, 2003). Wu et al. (2005) demonstrated an application of ANNs for watershed-runoff and stream-flow forecasts. Bhattacharjya et al. (2007) developed a simulation methodology using a trained ANN model to approximate the three-dimensional density dependent flow and transport processes in a coastal aquifer. Some of the studies dealing with the applications of ANN in hydrology are reviewed along with applications of FIS after Section 2.10.3.

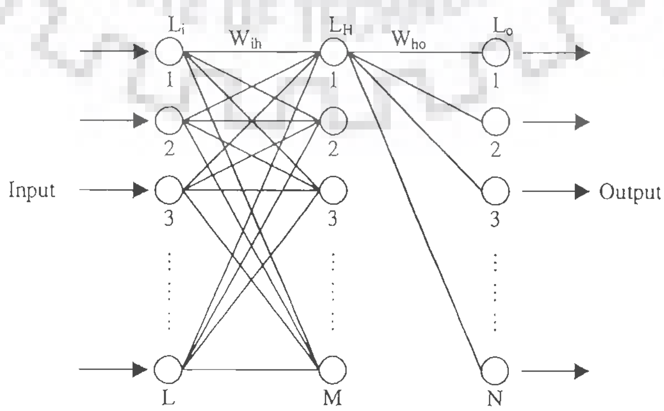


Fig. 2.1 Configuration of three-layer neural network

2.10.3 Fuzzy Inference System (FIS)

In crisp logic, the true value acquired by propositions or predicates are 2-valued, namely True, False, which may be treated numerically equivalent to (0, 1). However, in fuzzy logic, true values are multi-valued such as absolutely true, partly true, absolutely false, very true, and so on and are numerically equivalent to (0-1). Thus, in fuzzy logic, the event may take a range of values between 0 and 1. The fuzzy set theory is an effective tool to handle the problems of uncertainty.

Rajasekaran and Pai (2004) mention that fuzzy set theory is an excellent mathematical tool to handle the uncertainty arising due to vagueness. Fuzziness means 'vagueness'. The fuzzy systems approximate functions. They are universal approximators if they use enough fuzzy rules. In this sense fuzzy systems can model any continuous function or system. Those systems can just as well come from physics or sociology as from control theory or signal processing. The quality of the fuzzy approximation depends on the quality of the rules. In practice experts guess at the fuzzy rules. Or neural schemes learn the rules from data and tune the rules with new data. The result always approximates some unknown nonlinear function that can change in time. Better brains and better neural networks give better function approximations. This is not the standard view of fuzzy systems but it is the view that is generally taken in fuzzy engineering: function approximation with fuzzy systems. The standard view is that fuzzy systems theory or "fuzzy logic" is a linguistic theory that models how we reason with vague rules of thumb and common sense. Fuzzy sets and systems serve as means to this linguistic end. It tends to hold in practice when the number of inputs and outputs in a problem is small enough and when the time scale is slow enough for a human to find some solution paths as when we focus a camera lens or back up a car or grill a steak. It reflects the kinds of issues the first fuzzy engineers addressed and shows the kinds of tools they often used in their work and the language they used to defend it.

The basic structure of a FIS consists of three conceptual components: A rule base, which contains a selection of fuzzy rules; a database which defines the membership function (MF) used in the fuzzy rules; and a reasoning mechanism, which performs the inference procedure upon rules and a given condition to derive a reasonable output conclusion. A FIS implements a nonlinear mapping from its input space to an output space. A FIS can utilize human expertise by storing its essential components in a rule base and database, and perform fuzzy reasoning to infer the overall output value. Derivation of if-then rules and corresponding membership functions depends heavily on a priori knowledge about system under consideration.

Fuzzy logic modeling technique can be classified into three categories, namely the linguistic or Mamdani type (Zadeh, 1973; Mamdani, 1977), the relational equation (Yi and Chung, 1993) and the Takagi, Sugeno (TS) fuzzy model (Takagi and Sugeno, 1985). Fuzzy algorithms are formed by the union use of the fuzzy OR operator of individual fuzzy rules (Brown & Harris, 1994). Fuzzy rule based modeling has been attempted in water resources management, reservoir operation by some of the investigators. Applications of fuzzy set theory in hydrology and water resources are illustrated by Panigrahi and Mujumdar (2000), Cheng et al. (2002), Nayak et al. (2007) and Nayak and Sudheer (2007) etc.

Nauck and Kruse (1997) mention that Neuro-fuzzy systems have recently gained a lot of interest in research and application. Neuro-fuzzy models are fuzzy systems that use local learning strategies to learn fuzzy sets and fuzzy rules. Neuro-fuzzy techniques have been developed to support the development of e.g. fuzzy controllers and fuzzy classifiers. The authors discuss a learning method for fuzzy classification rules. The learning algorithm in a simple heuristics that is able to derive fuzzy rules from a set of training data very quickly, and tunes them by modifying parameters of membership functions. The authors' approach is based on NEFCLASS, a neuro-fuzzy model for pattern classification. The authors also discuss some results obtained by the software implementation of NEFCLASS, which is freely available on the Internet. Applications of some of the soft computing techniques in hydrology and water resources are reviewed as follow.

Whitley and Hromadka (1999) presented approximate confidence intervals for design floods for a single site using a neural network. The authors mention that a basic problem in hydrology is the computation of confidence levels for the value of the T-year flood when it is obtained from a log Pearson 3 distribution using the estimated mean, standard deviation and skewness. The authors gave a practical method for finding approximate one-sided or two-sided confidence intervals for the 100-year flood based on data from a single site. The confidence intervals are generally accurate to within a percent or two, as tested by simulations, and are obtained by use of neural network.

Shi and Mizumoto (2001) improved a neuro-fuzzy learning algorithm based on the fuzzy clustering method. In this approach, before learning fuzzy rules typical data were extracted from training data by using fuzzy c-means clustering algorithm, in order to remove redundant data and resolve conflicts in data, and make them as practical training data. By these typical data, fuzzy rules can be tuned by using the neuro-fuzzy learning algorithm. Therefore, the learning time can be expected to be reduced and the fuzzy rules generated by the improved approach are reasonable and suitable for the identified system model. Moreover, the efficiency of the improved method is also shown by identifying nonlinear functions by the authors.

Jain et al. (2004) presented analysis of soil water retention data using artificial neural networks. The authors mention that many studies of water flow and solute transport in the vadose zone require estimates of the unsaturated soil hydraulic properties, including the soil water retention curve (WRC) describing the relationship between soil suction and water content. An ANN approach was developed to describe the WRC using observed data from several soils. The ANN approach was found to produce equally or more accurate descriptions of the retention data as compared to several analytical retention functions popularly used in the vadose zone hydrology literature. The authors mention that given sufficient input data, the ANN approach was also found to closely describe the hysteretic behavior of a soil, including observed scanning wetting and drying curves.

Keskin and Ozlem (2006) proposed ANN models as an alternative approach of evaporation estimation for Lake Egirdir. The study was carried out to develop ANN models to estimate daily pan evaporation from measured meteorological data; to compare the ANN models to the Penman model; and to evaluate the potential of ANN models. Meteorological data from Lake Egirdir consisting of 490 daily records from 2001 to 2002 were used to develop the model for daily pan evaporation estimation. The measured meteorological variables included daily observations of air and water temperature, sunshine hours, solar radiation, air pressure, relative humidity, and wind speed. The results of the Penman method and ANN models were compared to pan evaporation values. The comparison showed that there is better agreement between the ANN estimations and measurements of daily pan evaporation than for other model.

Raghuwanshi et al. (2006) mention that accurate estimation of both runoff and sediment yield is required for proper watershed management. The ANN models were developed, to predict both runoff and sediment yield on a daily and weekly basis, for a small agricultural watershed. A total of five models were developed for predicting runoff and sediment yield, of which three models were based on a daily interval and the other two were based on a weekly interval. All five models were developed both with one and two hidden layers. Each model was developed with five different network architectures by selecting a different number of hidden neurons. Training was conducted using the Levenberg-Marquardt backpropagation where the input and output were presented to the neural network as a series of learning sets. Simulated surface runoff and sediment yield were compared with observed values and the minimum root-mean-square error and Nash Sutcliff efficiency (coefficient of efficiency) criteria were used for selecting the best performing model. Regression models for predicting daily and weekly runoff and sediment yield were also developed using the above training datasets, whereas these models were tested using the testing datasets. In all cases, the ANN models performed better than the linear regression based models. The ANN models with a double hidden layer were observed to be better than those with single hidden layer. Further, the ANN model prediction performance improved with increased number of hidden neurons and input variables. As a result, models considering both rainfall and temperature as input performed better than those considering rainfall alone as input. Training and testing results revealed that the models were predicting the daily and weekly runoff and sediment yield satisfactorily.

Garbrecht (2006) investigated the performance of three ANN designs that account for the effects of seasonal rainfall and runoff variations for monthly rainfall-runoff simulation on an 815 km^2 watershed in central Oklahoma. The ANN design that accounted explicitly for seasonal variations of rainfall and runoff performed best by all performance measures. Explicit representation of seasonal variations was achieved by use of a separate ANN for each calendar month. For the three ANN designs tested, a regression of simulated versus measured runoff displayed a slope slightly under 1 and positive intercept, pointing to a tendency of the ANN to underpredict high and overpredict low runoff values.

The data required for initially training the ANN model is generated by using a numerical simulation model. The simulated data consisting of corresponding sets of input and output patterns are used to train a multilayer perceptron using the back-propagation algorithm. The trained ANN predicts the concentration at specified observation locations at different times. The performance of the ANN as a simulator of the density dependent saltwater intrusion process in a coastal aquifer is evaluated using an illustrative study area. The authors mention that the evaluation results show that the ANN technique can be successfully used for approximating the three-dimensional flow and transport processes in coastal aquifers.

Kisi (2007) mentions that forecast of future events are required in many activities associated with planning and operation of the components of a water resources system. For the hydrologic components, there is a need for both short term and long term forecasts of streamflow events in order to optimize the system or to plan for future expansion or reduction. The author presents a comparison of different ANNs algorithms for short term daily streamflow forecasting. Four different ANN algorithms, namely, backpropagation, conjugate gradient, cascade correlation and Levenberg-Marquardt are applied to continuous streamflow data of the North Platte river in the United States. The modules are verified with untrained data. The results from the different algorithms are compared with each other. The correlation analysis was used in the study and found to be useful to determine appropriate input vectors to the ANNs.

Aqil et al. (2007) mention that traditionally, the multiple linear regression technique has been one of the most widely used models in simulating hydrological time series. However, when the nonlinear phenomenon is significant, the multiple linear will fail to develop an appropriate predictive model. Recently, neuro-fuzzy systems have gained much popularity for calibrating the nonlinear relationships. The authors evaluated the potential of a neuro-fuzzy system as an alternative to the traditional statistical regression technique for the purpose of predicting flow from a local source in a river basin. The effectiveness of the proposed identification technique was demonstrated through a simulation study of the river flow time series of the Citarum River in Indonesia. Furthermore, in order to provide the uncertainty associated with the estimation of river flow, a Monte Carlo simulation was performed. As a comparison, a multiple linear regression analysis that was being used by the Citarum River Authority was also examined using various statistical indices. The simulation results using 95% confidence intervals indicated that the neuro-fuzzy model consistently underestimated the magnitude of high flow while the low and medium flow magnitudes were estimated closer to the observed data. The comparison of the prediction accuracy of the neuro-fuzzy and linear regression methods indicated that the neuro-fuzzy approach was more accurate in predicting river flow dynamics. The neuro-fuzzy model was able to improve the root mean square error (RMSE) and mean absolute percentage error (MAPE) values of the multiple linear regression forecasts by about 13.52% and 10.73%, respectively. Considering its simplicity and efficiency, the neuro-fuzzy model is recommended as an alternative tool for modeling of flow dynamics in the study area.

Nayak and Sudheer (2007) explored the potential of integrating two different artificial intelligence techniques, namely neural network and fuzzy logic, effectively to model the rainfall-runoff process from rainfall and runoff information. The integration is achieved through representing fuzzy system computations in a generic artificial neural network (ANN) architecture, which is functionally equivalent to a fuzzy inference system. The model is initialized by a hyperellipsoidal fuzzy clustering (HEC) procedure, which identifies suitable numbers of fuzzy if-then rules through proper partition of the input space. The parameters of the membership functions are optimized using a nonlinear optimization procedure. The consequent functions are chosen to be linear in their parameters, and a standard least squares error method is employed for parameter estimation. The proposed model is tested on two case studies: Narmada basin in India and Kentucky basin in the United States. The results are highly encouraging as the model is able to explain more than 92% of the variance. The performance of the proposed model is found to be comparable to that of an adaptive neural based fuzzy inference system (ANFIS) developed for both the basins. The number of parameters in the proposed model is fewer compared to ANFIS, and the former can be trained in lesser time. It is also observed that the proposed model simulates the peak flow better than ANFIS. Overall, the study suggests that the proposed model can potentially be a viable alternative to ANFIS for use as an operational tool for rainfall and runoff modeling purposes.

Mukerji et al. (2009) carried out flood forecasting studies for Jamtara gauging site of the Ajay river basin in Jharkhand, India using an ANN model, an ANFIS model, and an adaptive neuro-GA integrated system (ANGIS) model. Relative performances of these models are also compared. Initially the ANN model is developed and is then integrated with fuzzy logic to develop an ANFIS model.

Further, the ANN weights are optimized by generic algorithm (GA) to develop an ANGIS model. For development of these models, 20 rainfall-runoff events are selected, of which 15 are used for model training and five are used for validation. Various performance measures are used to evaluate and compare the performances of different models. The authors mention that for the same input data set ANGIS model predicts better than the ANN model in most of the cases.

The review of literature reveals that studies carried out in India on regional flood frequency estimation are limited, scattered and based on the conventional techniques and in general do not meet the requirements of the practitioners. Also, there is a need for making systematic efforts for development of reliable and convenient regional flood frequency relationships for gauged and ungauged catchments of India based on the state of art technique of regional flood frequency estimation. Further as the applications of ANNs in regional flood frequency estimation are limited and the applicability of fuzzy techniques in regional flood frequency estimation broadly remains to be investigated; hence, there is also a need for investigating applicability of artificial neural networks and fuzzy techniques in regional flood frequency estimation.

2.11 PROBLEM DEFINITION

Keeping in view the gaps in the existing literature of regional flood frequency estimation, the present study has been taken up. The steps involved in carrying out the study are mentioned below.

(i) To develop regional flood frequency relationships for estimation of floods of various return periods for gauged catchments using the L-moments approach for 17 Subzones of India.

(ii) To develop regional relationships between mean annual peak floods and catchment areas of the gauged catchments for estimation of mean annual peak floods for ungauged catchments for 17 Subzones of India.

(iii) To develop regional flood frequency relationships for estimation of floods of various return periods for ungauged catchments using the L-moments approach for 17 Subzones of India.

(iv) To investigate applicability of the soft computing techniques viz. Artificial Neural Network (ANN) and Fuzzy Inference System (FIS) in development of regional flood frequency relationships.

(v) To compare the performances L-moments, ANN and FIS in regional flood frequency estimation.

(vi) To evolve a robust procedure for regional flood frequency estimation for the 17 Subzones of India.

CHAPTER 3

DESCRIPTION OF STUDY AREA AND DATA USED

3.1 GENERAL

This chapter gives details of study area and data used in carrying out the study. For carrying out the study annual maximum peak flood data and catchment areas of 261 stream flow gauging sites of the 17 Subzones of India were collected. The 17 Subzones for which regional flood frequency analysis has been carried out cover total 25,89,342 km^2 area of India, which forms about 79% of the geographical area of India. The description of the study area and data used in the study is given as follow.

3.2 STUDY AREA

India has been divided into 7 major zones, which are further sub-divided into 26 hydrometeorologically homogeneous Subzones (CWC, 1982). In this study regional flood frequency relationships have been developed for 17 Subzones out of the 26 Subzones of India. As the data for remaining 9 Subzones are not available; hence, the study could be carried out only for 17 Subzones. The names of the 17 Subzones for which study has been carried out are mentioned below and their location map is shown in Fig. 3.1.

(i) Chambal Subzone 1 (b)

(ii) Sone Subzone 1 (d)

(iii) Upper Indo-Ganga Plains Subzone 1 (e)

(iv) Middle Ganga Plains Subzone 1 (f)

(v) Lower Ganga Plains Subzone 1 (g)

Fig 3.1 Location map of 17 Subzones of India

(vi) North Brahmaputra Subzone 2 (a)

(vii) South Brahmaputra Subzone 2 (b)

(viii) Mahi and Sabarmati Subzone 3 (a)

(ix) Lower Narmada and Tapi Subzone 3 (b)

(x) Upper Narmada and Tapi Subzone 3 (c)

(xi) Mahanadi Subzone 3 (d)

(xii) Upper Godavari Subzone 3 (e)

(xiii) Lower Godavari Subzone 3 (f)

(xiv) Krishna and Pennar Subzone 3 (h)

(xv) Kaveri Basin Subzone 3 (i)

(xvi) East Coast Subzone 4 (b)

(xvii) Sub-Himalayan Region Zone-7

The descriptions of the 17 Subzones are given in the flood estimation reports which were jointly prepared by the Central Water Commission, India Meteorological Department and Research Designs and Standards Organization (e.g. CWC, 1982; CWC, 1985). Brief descriptions of these subzones based on the information available in the flood estimation reports are presented in Appendix 3.1. The details of data used in the study are described as follows.

3.3 DATA USED

The annual maximum peak flood data and catchment areas of 261 streamflow gauging sites of the 17 Subzones of India were collected for carrying out the study. Out of the collected data of the 261 stream flow gauging sites the annual maximum peak flood data of 196 stream flow gauging sites and their catchment areas have been used after carrying out the data screening and testing the regional homogeneity, as discussed in Chapter 5. Table 3.1 summarizes the status of data availability and salient features of the data for 17 Subzones of India. Table 3.2 provides the range of catchment areas and mean annual peak floods for 17 Subzones of India. The record lengths for these streamflow gauging sites vary from 5 to 38 years. The catchment areas of the streamflow gauging sites range from 6 km^2 to 2,297 km^2 and their mean annual peak floods vary from 12.8 m^3/s to 1687.3 m^3/s. The station-year record length varies from 165 to 393 and the average station-year record length for the 17 Subzones is about 262 years.

Table 3.1 Details of data availability and salient features of the data for 17 Subzones of India

Subzone	Area of Subzone (km^2)	No. of gauging sites for which data are available	No. of gauging sites whose data are used in analysis	Record length (Years)	Station-year record length (Years)
1 (b)	146630	13	12	10-31	231
1 (d)	128900	12	10	13-33	232
1 (e)	226000	21	12	25-34	356
1 (f)	171350	13	8	11-33	231
1 (g)	130280	13	10	10-33	267
2 (a)	121444	24	13	13-27	279
2 (b)	73556	16	11	5-28	217
3 (a)	138400	10	10	14-25	191
3 (b)	77700	19	14	12-28	296
3 (c)	86353	15	13	14-30	301
3 (d)	195256	23	15	11-31	326
3 (e)	88870	12	9	14-32	192
3 (f)	174201	19	17	14-29	393
3 (h)	280881	18	16	14-33	373
3 (i)	96051	12	8	12-33	193
4 (b)	131300	10	8	17-38	219
Zone -7	322170	11	10	13-20	165

Table 3.2 Range of catchment areas and mean annual peak floods for 17 Subzones of India

Subzone	Range of catchment area (km^2)	Range of mean annual peak flood (m^3/s)
1 (b)	26.2-2297.3	18.8-1549.0
1 (d)	34.0-1658.0	130.2-584.4
1 (e)	25.3-2072.0	13.7-780.5
1 (f)	32.9-447.8	24.3-555.2
1 (g)	15.0-569.8	51.2-650.5
2 (a)	21.4-595.7	18.5-852-7
2 (b)	21.4-497.3	22.2-321.1
3 (a)	18.4-1094.0	74.0-448.7
3 (b)	17.2-1017.0	34.9-558.3
3 (c)	53.7-2110.9	209.2-1687.3
3 (d)	19.0-1150.0	25.1-1071.9
3 (e)	31.3-2227.4	60.1-868.9
3 (f)	35.0-824.0	77.8-1212.8
3 (h)	31.7-1689.9	28.3-794.9
3 (i)	30.0-953.0	12.8-309.1
4 (b)	51.2-663.0	43.2-316.0
Zone -7	6.0-2072.0	17.1-1606.8

CHAPTER 4

METHODOLOGY

4.1 GENERAL

This chapter deals with the methodologies of regional flood frequency estimation using L-moments and the soft computing techniques used in study. Section 4.2 describes the methodology of L-moments, data screening, test of regional homogeneity, frequency distributions used and goodness of fit measures employed in the study. Development of regional relationships between mean annual peak floods and catchment areas of the gauged catchments for estimation of mean annual peak floods for ungauged catchments is discussed in Section 4.3. The methodology of regional flood frequency estimation using the soft computing techniques viz. Artificial Neural Network (ANN) and Fuzzy Inference System (FIS) are presented in Section 4.4. The description of various techniques are available in different books, reports and research papers also. In the present thesis, the methods scattered at different places have been brought at one place for completeness, better readability and continuity of the present work.

4.2 L-MOMENTS APPROACH

4.2.1 General

The L-moments were introduced by Hosking (1990) and these are a recent development within statistics. In a wide range of hydrologic applications, L-moments provide simple and reasonably efficient estimators of characteristics of hydrologic data and of a distribution's parameters (Stedinger et al., 1992). Like the ordinary product moments, L-moments summarize the characteristics or shapes of theoretical probability distributions and observed samples. Both moment types offer measures of distributional location (mean), scale (variance), skewness (shape), and kurtosis (peakedness). Recently a number of regional flood frequency analysis studies have been carried out based on the L-moments approach. The L-moment methods are demonstrably superior to those that have been used previously, and are now being adopted by many organizations worldwide (Hosking and Wallis, 1997).

4.2.2 Probability Weighted Moments and L-Moments

The L-moments are an alternative system of describing the shapes of probability distributions (Hosking and Wallis, 1997). They arose as modifications of probability weighted moments (PWMs) of Greenwood et al. (1979). Probability weighted moments is defined as:

$$M_{p,r,s} = E\left(x^p \{F\}^r \{1-F\}^s\right) = \int_0^1 \{x(F)\}^p F^r \{1-F\}^s \, dF \tag{4.1}$$

where, F = F(x) is the cumulative distribution function (CDF) for x, x(F) is the inverse CDF of x evaluated at the probability F, and p, r and s are real numbers. If p is a nonnegative integer, $M_{p,0,0}$ represents the conventional moment of order p about the origin. If p = 1 and s = 0,

$$M_{1,r,0} = \beta_r = \int_0^1 x(F) F^r dF \tag{4.2}$$

For an ordered sample $x_1 \le x_2 \ldots. \le x_N$, N > r, the unbiased sample PWM's are given by

$$\hat{\beta}_r = \frac{1}{N} \frac{\sum_{i=1}^{N} \binom{i-1}{r} x_i}{\binom{N-1}{r}} \tag{4.3}$$

For any distribution the r^{th} L-moment λ_r is related to the r^{th} PWM (Hosking, 1990), through:

$$\lambda_{r+1} = \sum_{k=0}^{r} \beta_k (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \tag{4.4}$$

These L-moments are linear functions of PWMs. For example, the first four L-moments are related to the PWMs using:

$$\begin{aligned} \lambda_1 &= \beta_0 \\ \lambda_2 &= 2\beta_1 - \beta_0 \\ \lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{aligned} \tag{4.5}$$

The L-moments are analogous to their conventional counterparts as they can be directly interpreted as measures of scale and shape of probability distributions and hence, are more convenient than the PWMs. Hosking (1990) defined L-moment ratios which are analogous to conventional moment ratios as:

$$\begin{aligned} &\text{L-coefficient of variation, L-CV: } \tau_2 = \lambda_2 / \lambda_1 \\ &\text{L-coefficient of skewness, L-skew: } \tau_3 = \lambda_3 / \lambda_2 \\ &\text{L-coefficient of kurtosis, L-kurtosis: } \tau_4 = \lambda_4 / \lambda_2 \end{aligned} \tag{4.6}$$

Analogous to the conventional moment ratios, λ_1 is a measure of location, τ_2 is a measure of scale and dispersion, τ_3 is a measure of skewness and τ_4 is a measure of kurtosis. Hosking (1990) showed that for $x \geq 0$, the value of τ_2 lies between 0 and 1, while the absolute values of τ_3 and τ_4 lie between 0 and 1. This restriction in the values of the L-coefficients works out to be an advantage in their interpretation as opposed to the conventional moments which do not have any bounds (Rao and Hamed, 2000).

4.2.3 Screening of Data Using Discordancy Statistic Test

The objective of screening of data is to check that the data are appropriate for performing the regional flood frequency analysis. In this study, screening of the data was performed using the L-moments based Discordancy statistic (D_i). Discordancy is measured in terms of the L-moments of the sites' data and the aim is to identify those sites that are grossly discordant with the group as a whole. The sample L-moment ratios (t_2, t_3 and t_4) of a site are considered as a point in a three-dimensional space. A group of sites form a cluster of such points in the three-dimensional space. A site is considered discordant if it is far from the centre of the cluster.

Hosking and Wallis (1997) defined the Discordancy statistic D_i for a site i in a group of N sites. Let $u_i = [t_2^{(i)} \ t_3^{(i)} \ t_4^{(i)}]^T$ be a vector containing the sample L-moment ratios t_2, t_3 and t_4 values for site i,

$$\bar{u} = N^{-1} \sum_{i=1}^{N} u_i \tag{4.7}$$

analogous to their regional values termed as τ_2, τ_3, and τ_4, expressed in Eq. (4.6). T denotes transposition of a vector or matrix. Let

$$A_m = \sum_{i=1}^{N} (u_i - \bar{u})(u_i - \bar{u})^T \tag{4.8}$$

be the (unweighted) group average. The sample covariance matrix is defined as:

The Discordancy measure for site i is defined as:

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T A_m^{-1} (u_i - \bar{u}) \tag{4.9}$$

The site i is declared to be discordant, if D_i is greater than the critical value of the Discordancy statistic D_i, given in a tabular form by Hosking and Wallis (1997).

4.2.4 Test of Regional Homogeneity

For testing regional homogeneity, a test statistic H, termed as heterogeneity measure has been discussed by Hosking and Wallis (1997). It compares the "inter-site variations in sample L-moments for the group of sites" with "what would be expected of a homogeneous region". The inter-site variations in sample L-moments are evaluated based on any of the three measures of variability V_1 (based on L-CV), V_2 (based on L-CV and L-skewness) and V_3 (based on L- skewness and L-Kurtosis). These measures of variability are computed as follows:

(i) V_1 is the weighted standard deviation of at site L-CV's ($t_2^{(i)}$)

$$V_1 = \left[\sum_{i=1}^{N} n_i \; (t_2^{(i)} - t_2^R)^2 \Big/ \sum_{i=1}^{N} n_i \right]^{1/2} \tag{4.10}$$

where, n_i is the record length at each site and t_2^R is the regional average L-CV weighted proportionally to the sites' record length as given below.

$$t_2^R = \sum_{i=1}^{N} n_i \; t_2^{(i)} \Big/ \sum_{i=1}^{N} n_i \tag{4.11}$$

(ii) V_2 is the weighted average distance from the site to the group weighted mean (based on L-CV and L-skewness) on a graph of t_2 versus t_3

$$V_2 = \sum_{i=1}^{N} n_i \left\{ (t_2^{(i)} - t_2^R)^2 + (t_3^{(i)} - t_3^R)^2 \right\}^{1/2} \Big/ \sum_{i=1}^{N} n_i \tag{4.12}$$

where, t_3^R is the regional average L-Skew weighted proportionally to the sites' record length.

(iii) V_3 is the weighted average distance from the site to the group weighted mean (based on L-skewness and L-kurtosis) on a graph of t_3 versus t_4

$$V_3 = \sum_{i=1}^{N} n_i \left\{ (t_3^{(i)} - t_3^R)^2 + (t_4^{(i)} - t_4^R)^2 \right\}^{1/2} \Big/ \sum_{i=1}^{N} n_i \tag{4.13}$$

where, t_4^R is the regional average L-Kurtosis weighted proportionally to the sites' record length.

To establish "what would be expected of a homogeneous region", firstly simulations are used to generate homogeneous regions with sites having same record lengths as those of observed data. In order to generate the simulated data, a four parameter Kappa distribution is used. The four parameter Kappa distribution is chosen so as not to commit to a particular two or three parameter distribution. Further, the four parameter Kappa distribution includes as special cases the Generalised Logistic (GLO), Generalised Extreme Value (GEV) and Generalised Pareto (GPA) distributions and hence, acts as a good representation of many of the probability distributions occurring in environmental sciences.

The parameters of the Kappa distribution are obtained using the regional average L-moment ratios t_2^R, t_3^R, t_4^R and mean = 1. A large number of data regions are generated (say N_{sim} = 500) based on this Kappa distribution. The simulated regions are homogeneous and have no cross-correlation or serial correlation. Further, the sites have the same record lengths as the observed data. For each generated region, V_j (i.e. any of V_1, V_2 or V_3) is computed using Eqns. 4.10 to 4.13. Subsequently, their mean (μ_v) and standard deviation (σ_v) are computed.

The heterogeneity measure H(j) (i.e. H(1), H(2) or H(3)) is computed as:

$$H(j) = \frac{V_j - \mu_V}{\sigma_V} \tag{4.14}$$

If the heterogeneity measure is sufficiently large, the region is declared to be heterogeneous. Hosking and Wallis (1997) mention the following criteria for assessing heterogeneity of a region:

If H(j) < 1, the region is acceptably homogeneous;

If $1 \leq H(j) < 2$, the region is possibly heterogeneous; and if

$H(j) \geq 2$, the region is definitely heterogeneous.

These boundary values of H(j) being 1 and 2 are determined by performing a series of Monte Carlo experiments in which the accuracy of quantile estimates corresponding to different values of H(j) are computed (Hosking and Wallis, 1997). The authors further mention that for both real world data and artificially simulated regions, H(1) has much better power to discriminate between homogeneous and heterogeneous regions as compared to H(2) and H(3).

4.2.5 Frequency Distributions Used

The following frequency distributions have been used in this study. The details about these distributions and relationships among parameters of these distributions and L-moments are available in literature (e.g. Hosking and Wallis, 1997).

4.2.5.1 Extreme Value Type-I Distribution (EV1)

Extreme Value Type-I distribution (EV1) is a two parameter distribution and it is popularly known as Gumbel distribution. The quantile function or the inverse form of the distribution is expressed as:

$$x(F) = u - \alpha \ln(-\ln F) \tag{4.15}$$

Where, u and α are the location and scale parameters respectively, F is the non-exceedence probability viz. (1-1/T) and T is return period in years.

4.2.5.2 General Extreme Value Distribution (GEV)

General Extreme Value distribution (GEV) is a generalized three parameter extreme value distribution. Its theory and practical applications are reviewed in the

Flood Studies Report (NERC, 1975). The quantile function or the inverse form of the distribution is expressed as:

$$x(F) = u + \alpha \{1 - (-\ln F)^{k}\} / k ; \qquad k \neq 0 \qquad (4.16)$$

$$x(F) = u - \alpha \ln(-\ln F) \qquad k = 0 \qquad (4.17)$$

Where, u, α and k are location, scale and shape parameters of GEV distribution respectively. EV1 distribution is the special case of the GEV distribution, when k = 0.

4.2.5.3 Logistic Distribution (LOS)

Inverse form of the Logistic distribution (LOS) is expressed as:

$$x(F) = u - \alpha \ln \{(1-F)/F\} \qquad (4.18)$$

Where, u and α are location and scale parameters respectively.

4.2.5.4 Generalized Logistic Distribution (GLO)

Inverse form of the Generalized Logistic distribution (GLO) is expressed as:

$$x(F) = u + \alpha [1 - \{(1-F)/F\}^{k}] / k; \qquad k \neq 0 \qquad (4.19)$$

$$x(F) = u - \alpha \ln \{(1-F)/F\}; \qquad k = 0 \qquad (4.20)$$

Where, u, α and k are location, scale and shape parameters respectively. Logistic distribution is the special case of the Generalized Logistic distribution, when k = 0.

4.2.5.5 Generalized Pareto Distribution (GPA)

Inverse form of the Generalized Pareto distribution (GPA) is expressed as:

$$x(F) = u + \alpha \{1 - (1-F)^{k}\} / k; \qquad k \neq 0 \qquad (4.21)$$

$$x(F) = u - \alpha \ln(1-F) \qquad k = 0 \qquad (4.22)$$

where u, α and k are location, scale and shape parameters respectively. Exponential distribution is special case of Generalized Pareto distribution, when k = 0.

4.2.5.6 Generalized Normal Distribution (GNO)

The cumulative density function of the three parameter Generalized Normal distribution (GNO) is given below.

$$F(x)=\phi\left[-k^{-1}\log\{1-k(x-\xi)/\alpha\}\right] \tag{4.23}$$

where, ξ, α and k are its location, scale and shape parameters respectively. When k = 0, it becomes normal distribution with parameters ξ and α. This distribution has no explicit analytical inverse form.

4.2.5.7 Pearson Type-III Distribution (PE3)

The inverse form of the Pearson type-III (PE3) distribution is not explicitly defined. Hosking and Wallis (1997) mention that the PE3 distribution combines Gamma distributions (which have positive skewness), reflected Gamma distributions (which have negative skewness) and the normal distribution (which has zero skewness). The authors parameterize the Pearson type-III distribution by its first three conventional moments viz. mean μ, the standard deviation σ, and the skewness γ. The relationship between these parameters and those of the Gamma distribution is as follows. Let X be a random variable with a Pearson type-III distribution with parameters μ, σ and γ. If $\gamma > 0$, then $X - \mu + 2\sigma/\gamma$ has a Gamma distribution with parameters $\alpha = 4/\gamma^2$, $\beta = \sigma\gamma/2$. If $\gamma = 0$, then X has normal distribution with mean μ and standard deviation σ. If $\gamma < 0$, then $-X + \mu - 2\sigma/\gamma$ has a Gamma distribution with parameters $\alpha = 4/\gamma^2$, $\beta = |\sigma\gamma/2|$.

If $\gamma \neq 0$, let $\alpha = 4/\gamma^2$, $\beta = |\sigma\gamma/2|$, and $\xi = \mu - 2\sigma/\gamma$ and Γ (.) is Gamma function. If $\gamma > 0$, then the range of x is $\xi \leq x < \infty$ and the cumulative distribution function is:

$$F(x) = G\left(\alpha, \frac{x-\xi}{\beta}\right) \Big/ \Gamma(\alpha) \qquad (4.24)$$

If $\gamma < 0$, then the range of x is $-\infty < x \leq \xi$ and the cumulative distribution function is:

$$F(x) = 1 - G\left(\alpha, \frac{\xi - x}{\beta}\right) \Big/ \Gamma(\alpha) \qquad (4.25)$$

4.2.5.8 Kappa Distribution (KAP)

The kappa distribution is a four parameter distribution that includes as special cases the Generalized Logistic (GLO), Generalized Extreme Value (GEV) and Generalized Pareto distribution (GPA).

$$x(F) = \xi + \alpha\left[1 - \left\{(1-F)^h / h\right\}^k\right] / k \qquad (4.26)$$

where, ξ is the location parameter, α is the scale parameter.

When h = -1, it becomes Generalized logistic (GLO) distribution; h = 0 is the Generalized Extreme Value distribution (GEV); and h = 0 is the Generalized Pareto distribution (GPA). It is useful as a general distribution with which to compare the fit of two and three parameter distributions and for use in simulating artificial data in order to assess the accuracy of statistical methods (Hosking and Wallis, 1997).

4.2.5.9 Wakeby Distribution (WAK)

Inverse form of the five parameter Wakeby distribution (WAK) is expressed as:

$$x(F) = \xi + \frac{\alpha}{\beta}\left\{1 - (1-F)^{\beta}\right\} - \frac{\gamma}{\delta}\left\{1 - (1-F) - \delta\right\} \qquad (4.27)$$

where, ξ, α, β, γ, and δ are the parameters of the Wakeby distribution.

4.2.6 Goodness of Fit Measures

In a realistically homogeneous region, all the sites follow the same frequency distribution. But as some heterogeneity is usually present in a region so no single distribution is expected to provide a true fit for all the sites of the region. In regional flood frequency analysis the aim is to identify a distribution which will yield reasonably accurate quantile estimates for each site of the homogeneous region. Assessment of validity of the candidate distribution may be made on the basis of how well the distribution fits the observed data. The goodness of fit measures assess the relative performance of various fitted distributions and help in identifying the robust viz. most appropriate distribution for the region. Recently introduced L-moment ratio diagram and the goodness of fit or behavior analysis measure for frequency distributions given by Z_i^{dist} statistic mentioned by Hosking and Wallis (1997) have been used in the study. A description of these goodness of fit measures is given as follows.

4.2.6.1 L-moment Ratio Diagram

The L-moment statistics of a sample reflect every information about the data and provide a satisfactory approximation to the distribution of sample values. The L-moment ratio diagram can therefore be used to identify the underlying frequency distribution. The average L-moment statistics of the region is plotted on the L-moment ratio diagram and the distribution nearest to the plotted point is identified as the underlying frequency distribution. One big advantage of L-moment ratio diagram is that one can compare fit of several distributions using a single graphical instrument (Vogel and Fennessey, 1993).

4.2.6.2 $|Z^{dist}|$ –Statistic Criteria

The best fit frequency distribution for a homogeneous region is determined by how well the L-skewness and L-kurtosis of the fitted distribution match the regional average L-skewness and L-kurtosis of the observed data (Hosking and Wallis, 1997). This procedure is described below.

Initially, several three parameter distributions are fitted to the regional average L-moments t_2^R, t_3^R and mean =1. Let τ_4^{Dist} be the L-kurtosis of the fitted distribution which may be GEV, GLO, GNO, PE3 etc. Using the N_{sim} number of simulated regions of the Kappa distribution, the regional average L-kurtosis, t_4^m is computed for the m[th] simulated region. The bias of t_4^R is computed as:

$$B_4 = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} \left(t_4^m - t_4^R \right) \tag{4.28}$$

The standard deviation of t_4^R is computed as:

$$\sigma_4 = \left[\left(N_{sim} - 1 \right)^{-1} \left\{ \sum_{m=1}^{N_{sim}} \left(t_4^m - t_4^R \right)^2 - N_{sim} B_4^2 \right\} \right]^{1/2} \tag{4.29}$$

The goodness-of-fit measure for each distribution is computed as (Hosking and Wallis, 1997):

$$Z^{dist} = \left(\tau_4^{dist} - t_4^R + B_4 \right) \Big/ \sigma_4 \tag{4.30}$$

The fit is considered to be adequate if $|Z^{dist}|$-statistic is sufficiently close to zero, a reasonable criterion being $|Z^{dist}|$ –statistic less than 1.64. Hosking and Wallis (1997) state that the $|Z^{dist}|$ –statistic has the form of a normal distribution under suitable assumptions. Thus the criterion $|Z^{dist}|$ –statistic less than 1.64 corresponds to acceptance of the hypothesized distribution at a confidence level of 90%.

4.3 DEVELOPMENT OF REGIONAL RELATIONSHIPS BETWEEN MEAN ANNUAL PEAK FLOODS AND CATCHMENT AREAS

For estimation of T-year return period flood at a site, the estimate for mean annual peak flood is required. For gauged catchments, such estimates can be obtained based on the at-site mean of the annual maximum peak flood data. However, for ungauged catchments at-site mean can not be computed in absence of the flow data. In such a situation, a regional relationship between the mean annual peak flood of gauged catchments in the region and their pertinent physiographic and climatic characteristics is needed for estimation of the mean annual peak flood. For example, the form of this regional relationship may be:

$$\overline{Q} = a\, A^{b} S^{c} D^{d} R^{e} \tag{4.31}$$

Here, ($\overline{Q}$) is the mean annual peak flood, A is the catchment area, S is the slope, D is the drainage density, R is the annual normal rainfall or rainfall for the duration of annual maximum peak flood for the catchment etc., a, b, c, d, and e are the regional coefficients. The regional coefficients are estimated using the mean annual peak floods of the gauged catchments and their pertinent physiographic and climatic characteristics for a region. The physiographic and climatic characteristics which are considered pertinent for generation of annual maximum peak floods from a catchment and can be obtained from the observed records e.g. rainfall for the duration of occurrence of the annual maximum peak floods and derived from the toposheets/maps of the gauged catchments may be considered for development of this relationship.

This form of regional relationship may be developed using an optimization technique such as Levenberg-Marquardt technique.

4.3.1 Levenberg- Marquardt Algorithm

The Levenberg-Marquardt algorithm (LMA) provides a numerical solution to the problem of minimizing a function, generally nonlinear, over a space of parameters of the function (Levenberg, 1944; Marquardt, 1963; Jacoby et al., 1972; Kuester and Mize, 1973; Gill et al., 1981). These minimization problems arise especially in least squares curve fitting and nonlinear programming. The LMA interpolates between the Gauss-Newton algorithm (GNA) and the method of gradient descent. The LMA is more robust than the GNA, which means that in many cases it finds a solution even if it starts very far off the final minimum. On the other hand, for well-behaved functions and reasonable starting parameters, the LMA tends to be a bit slower than the GNA.

The LMA is a very popular curve-fitting algorithm used in many software applications for solving generic curve-fitting problems.

The primary application of the Levenberg-Marquardt algorithm is in the least squares curve fitting problem: given a set of empirical data pairs of independent and dependent variables, (x_i, y_i), optimize the parameters β of the model curve $f(x,\beta)$ so that the sum of the squares of the deviations

$$S(\beta) = \sum_{i=1}^{m} \left[y_i - f(x_i,\beta)\right]^2 \tag{4.32}$$

becomes minimal.

The details of LMA are given in Appendix 4.1.

4.4 REGIONAL FLOOD FRQUENCY ESTIMATION USING ANN AND FIS

The methodology used for regional flood frequency estimation employing the soft computing techniques viz. ANN and FIS is described below.

4.4.1 Artificial Neural Network

The ANNs have shown a good potential to efficiently model complex input-output relationships where there is presence of nonlinearity and inconsistent/noisy data that adversely affects other approaches. The ANNs have gained popularity in a large array of engineering applications where conventional analytical methods show inferior performance. An ANN consists of a number of interconnected computational elements called neurons that are arranged in a number of layers. The connection between each pair of neurons is called a link and is associated with a weight that is a numerical estimate of the connection strength. Every neuron in a layer receives and processes weighted inputs from neurons in the previous layer and transmits its output to neurons in the next layer. The weighted summation of the inputs to a neuron is converted to an output according to a transfer function, typically a sigmoid function.

There are a wide range of ANN architectures, among which the three-layer feed-forward architecture is widely used. This network contains three distinctive modes: training, cross validation, and testing. In the training mode, the training data sets consisting of input-output patterns are presented to the network. The weights are found through an iterative process, in which the back propagation learning algorithm is used to find the weights such that the difference between the given outputs and the outputs computed by the network is sufficiently small. While training, it is a usual practice that the training data sets are further subdivided into two sets training and cross testing sets according to data availability. A training data set is used for training, during which the training data set mean square error (MSE) and the cross-validation data set MSE, which is not used for training are monitored together to find the optimal termination point for training. This check avoids overtraining. After training, the network is tested with the testing data set to determine how accurately the network can simulate the input-output relationship.

ANNs have been proven to provide better solutions when applied to (i) complex systems that may be poorly described or understood; (ii) problems that deal with noise or involve pattern recognition, diagnosis, abstraction, and generalization; and (iii) situations where input is incomplete or ambiguous by nature. An ANN has the ability to extract patterns in phenomena and overcome difficulties due to the selection of a model form such as linear, power, or polynomial. An ANN algorithm is capable of modeling the hydrological process due to its ability to generalize patterns in noisy and ambiguous input data and to synthesize a complex model without prior knowledge or probability distributions. The ANN model is calibrated using automatic calibration techniques. Thus, an ANN model eliminates subjectivity and lengthy calibration cycles.

4.4.1.1 Structure of ANN

Artificial Neural Networks are massively parallel systems composed of many processing elements connected by links of variables weights. The ANN is characterized by its architecture that represents the pattern of connection between the nodes, its method of determining the connection weights and the activation function. Artificial neural network consists of a number of artificial neurons known as *processing elements* or *nodes*. Each node performs a mapping of its inputs to its output in a three step process: firstly, it calculates the sum of the activation of its inputs, and then decides its new activation level based on the derived sum, and finally generates an output signal corresponding to the new level. The neurons in ANN are usually arranged in layers shown in the Fig. 4.1: an input layer, an output layer and one or more intermediate layers known as hidden layers. Each neuron in a specific layer is connected to many other neurons via weighted connections. The weights determine the strength of the connections between interconnected neurons.

4.4.1.2 Input Layer

The first layer, known as input layer, consists of neurons that represent the inputs received from the external environment. It does not perform any transformations upon the inputs but just sends them to the neurons of the second layer (hidden layer). The sole role of the nodes of the input layer is to relay the external inputs to the neurons of the hidden layer. Hence, the number of input nodes corresponds to the number of input variables.

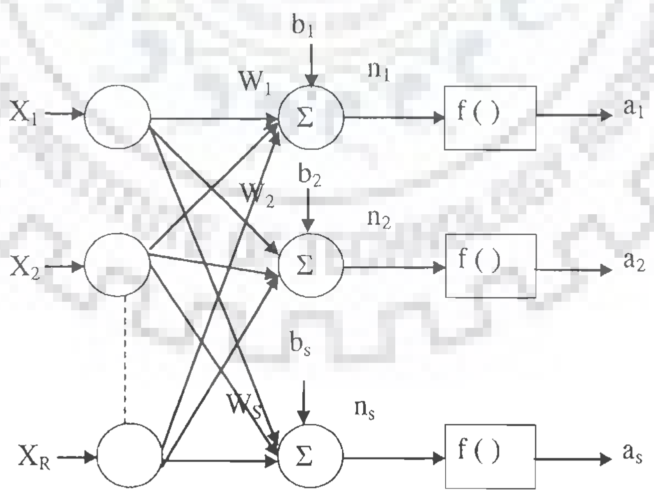


Fig 4.1 Schematic representation of a multilayer perceptron

4.4.1.3 Hidden Layer

The layer between the input and the output layer is known as the hidden layer, the purpose of which is to extract higher order (nonlinear) statistics from the input data. It is the hidden layer nodes that allow the network to detect and capture the relevant pattern(s) in the data and to perform the complex nonlinear mapping between the input and output variables. Hidden layer consists of neurons that typically receive the inputs from the input layer, perform transformation on it, and pass the output to the layer next to it, which can be a second hidden layer or the output layer.

4.4.1.4 Output Layer

The last layer is the output layer consisting of neurons that receive the hidden layer output and send it to the user. Number of neurons in this layer corresponds to the number of network outputs.

4.4.1.5 Transfer Function

The mapping or final activation level of a node is determined by its transfer function. Multilayer networks typically use sigmoid transfer functions in the hidden layers. These functions are often called as *squashing functions*, since they compress an infinite input range into a finite output range. Logistic sigmoid, tangent sigmoid and linear type transfer functions are the most popular transfer functions used for the modeling purposes.

The connection weights W_1, W_2, W_3......W_s reflect the relative importance of each input to the neuron. The sum of the weighted inputs and the bias forms the input to the transfer function f. Neurons may use any differentiable transfer function, f ( ), to generate their output. Output from individual node after summation operation is given by the following expressions:

$$n_1=W_1X+b_1 \tag{4.33}$$

Output from individual node after passing through transfer function is given by:

$$a_1 = f(n_1) = f(\vec{W}_1^T \vec{X} + b_1) \qquad (4.34)$$

Where, f ( ) is the neuron transfer function for limiting the amplitude of the output of a neuron, X is the input to the network; b is the bias of the hidden layer node.

4.4.1.6 Feed Forward Network (FFN)

Feed forward network (FFN) is the most commonly used network in ANN modeling. Multi-layer FNN can have more than one hidden layer. The network ability to learn from examples and to generalize depends on the number of hidden nodes. A too small network (i.e. with very few hidden nodes) will have difficulty in learning the data, while a too complex network tends to overfit the training samples and thus has a poor generalization capability. Finding a parsimonious model for accurate prediction is particularly critical since there are no formal methods for determining the appropriate number of hidden nodes prior to training. Therefore, trial-and-error method is commonly used for network design.

Multi-layer FNN training (supervised type of training) consists of providing input--output examples to the network, and minimizing the objective function (i.e. error function) using either a first-order or a second order optimization method. There are two modes of feeding the data into the network: incremental mode and batch mode. In incremental mode of training, the weights and biases of the network are updated after one set of training data is being applied to the network. In batch mode of training, the weights and biases of the network are updated only after the entire training set of data is applied to the network.

4.4.1.7 Development of Model Architecture

The architecture of a network (model) consists of a description of how many layers a network has, the number of neurons in each layer, transfer function of each layer and how the layers connect to each other. In order to improve network performance, many factors like determination of adequate model inputs, data division and pre-processing, the choice of suitable network architecture, selection of network internal parameters, the stopping criteria and model testing need careful addressing (Maier and Dandy, 2000).

4.4.1.8 Training Algorithms

A major concern in the development of a neural network is determining an appropriate set of weights that make it perform the desired function. There are many ways that this can be done; the most popular class of these algorithms is based on supervised training. Supervised training starts with a network comprising an arbitrary number of hidden neurons, a fixed topology of connections, and randomly selected values for weights. The network is then presented with a set of training patterns, each comprising an example of the problem to be solved (the inputs) and its corresponding solution (the targeted output). Each problem is input into the network in turn, and the resultant output is compared to the targeted solution providing a measure of total error in the network for the set of training patterns. Properly trained back propagation network give reasonable results when presented with new input during validation. In the process of model development several network architectures with different number of input neurons in input layer with varying number of hidden neurons are considered to select the optimal architecture of the network. A trial and error procedure based on the minimum error during validation is used to select the best network architecture.

4.4.1.9 Back Propagation Algorithm

The generalised delta rule, which determines the appropriate weight adjustments necessary to minimise the errors can be explained through Figure 4.2.

The Figure 4.2 shows a neuron (j) and its functions.

The total input H_{ij} to hidden units j is a linear function of outputs x_i of the units that are connected to j and of the weights w_{ij} on these connections i.e.

$$H_{ij} = \sum_i x_i w_{ij} \tag{4.35}$$

Units can be given biases (θ_j) by introducing on extra input to each unit which always has a value of 1.

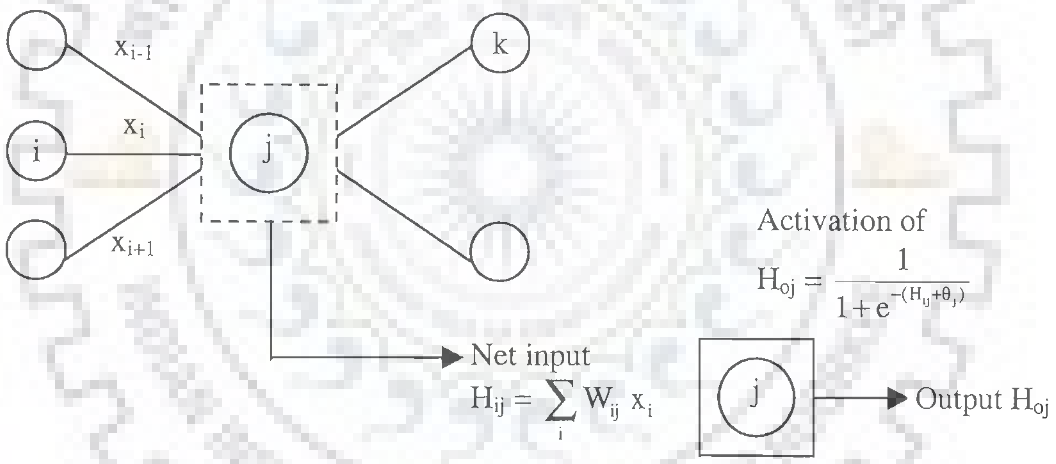


Fig. 4.2 A Neuron and its function

A hidden unit has a real-value output H_{oj}, which is a non-linear function of its total input.

$$H_{oj} = \frac{1}{1+e^{-(H_{ij}+\theta_j)}} \tag{4.36}$$

The use of a linear function for combing the inputs to a unit before applying the non-linearity greatly simplifies the learning procedure.

The aim is to find a set of weights that ensure that for each input vector, the output vector produced by the network is the same as (or sufficiently close to) the desired output vector. If there is a fixed, finite set of input-output cases, the total error in the performance of the network with a particular set of weights can be computed by comparing the actual and desired output vectors for every case. The total error E, is defined as:

$$E = \frac{1}{2}\sum_{C}\sum_{f}(O_{j,c} - T_{j,c})^2 \tag{4.37}$$

where 'c' is the an index over cases (input-output pairs), j is an index over output units, 'O' is the actual state of an output unit and T is its targeted state. To minimise E by gradient descent, it is necessary to compute the partial derivative of E with respect to each weight in the network i.e. $\partial E/\partial W_{ji}$. This can computed successively as follows:

Firstly differentiate Eqn (3.37) for a particular case, c,

$$\frac{\partial E}{\partial O_j} = (O_j - T_j) \tag{4.38}$$

Next $\partial E/\partial x_j$ is computed using chain rule i.e.

$$\frac{\partial E}{\partial xj} = \frac{\partial E}{\partial Oj}\,\frac{\partial Oj}{\partial xj} \tag{4.39}$$

Differentiating Eqn (3.36) to get the value of $\partial Oj/\partial xj$ and substituting in (3.39)

$$\frac{\partial Ej}{\partial xj} = \frac{\partial E}{\partial Oj}\; O_j\,(1 - O_j) \tag{4.40}$$

Eqn (3.39) calculates how the change in the total input 'x' to an output unit, will affect the error E. The total input is just a linear function, of the states of the lower level units and it is also a linear function of the weights on the connections, it is, therefore, easy to compute how the error will be affected by changing these states and weights. For a weight w_{ij}, from i to j the derivative is

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial xj}\frac{\partial x_j}{\partial w_{ji}} = \frac{\partial E}{\partial xj} \cdot O_i \quad (4.41)$$

and for the output of the ith unit the contribution to $\partial E/\partial O_i$ resulting from the effect of i on j is simply.

$$\frac{\partial E}{\partial xj}\frac{\partial xj}{\partial Oj} = \frac{\partial E}{\partial xj} \cdot w_{ji} \quad (4.42)$$

So taking into account all the connections emanating from unit i we have

$$\frac{\partial E}{\partial Oi} = \sum_j \frac{\partial E}{\partial xj} \cdot w_{ij} \quad (4.43)$$

Given $\partial E/\partial O$ for all units j, in the previous layer, the $\partial E/\partial O_I$ in the penultimate layer can be computed using Eqn (3.43). This procedure can therefore be repeated for successively layers.

The simplest version of gradient descent is to change each weight by an amount proportional to the accumulated $\partial E/\partial w$.

$$\Delta w = -\epsilon \ \partial E/\partial w \quad (4.44)$$

The convergence of Eqn (3.44) can be significantly improved, by an acceleration method wherein the incremental weights at t can related to the previous incremental weights given in Eqn (3.45).

$$\Delta w(t) = \epsilon \frac{\partial E}{\partial w(t)} + \alpha \Delta w(t-1) \quad (4.45)$$

where α is an exponential decay factor between '0' and '1' that determines the relative contribution of the current gradient and earlier gradients to the weight change. The term back propagation refers to the process by which derivatives of network error, with respect to network weights and biases, can be computed. In back propagation algorithm, the weights are moved in the direction of the negative gradient, i.e. in the direction in which the performance function decreases most rapidly. Thus, as the training proceeds, the back propagation learning algorithm constantly adjusts the weight towards the minimum.

4.4.2 Fuzzy Inference System (FIS)

Fuzzy logic is another area of artificial intelligence that has been applied successfully in different engineering fields. Fuzzy logic concepts were introduced by Zadeh in 1965 (ASCE, 2000 a,b). Fuzzy logic is a superset of conventional Boolean logic that has been extended to handle imprecise data and the concept of partial truth. In fuzzy logic, variables are "fuzzified" through the use of membership functions that define the membership degree to fuzzy sets. These variables are called linguistic variables.

The basic structure of fuzzy modeling, (Fig 4.3) commonly known as Fuzzy Inference System (FIS), is a rule-based or knowledge-based system consisting of three conceptual components: viz. (i) a rule base that consists of a collection of fuzzy IF–THEN rules; (ii) a database that defines the membership function (MF) used in fuzzy rules; and a reasoning mechanism that combines these rules into a mapping routine from the inputs to the outputs of the system, to derive a reasonable Output conclusion.

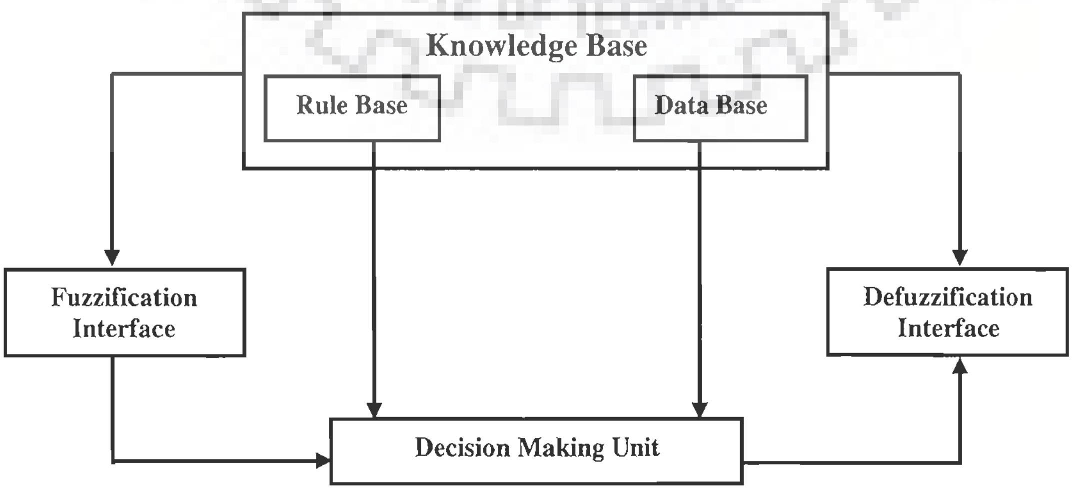


Fig 4.3 Schematic representation of Fuzzy Inference System

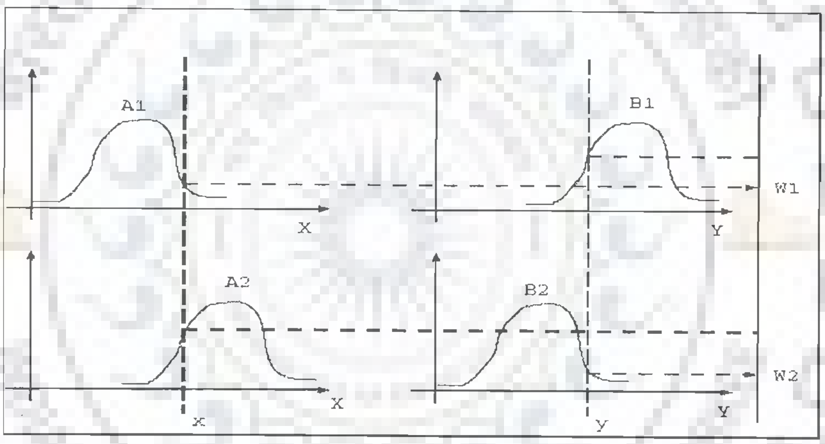
4.4.2.1 Architecture of FIS

Consider that the FIS has two inputs x, y and one output z. Figs. 4.4 and 4.5 illustrate a TSK fuzzy inference system. For a first-order Takagi Sugeno (TSK) model, a common rule set with two fuzzy if-then rules can be written as follows:

Rule 1, if x is A1 and y is B1, then f1 = p1x + q1y + r1, and

Rule 2, if x is A2 and y is B2, then f2 = p2x + q2y + r2, where (4.46)

The "if" statement is the antecedent, the "then" statement is the consequent, x and y are linguistic variables and A1, A2, B1, B2 are corresponding fuzzy sets, and p1, q1, r1 and p2, q2, r2 are linear parameters.



f1 = p1x + q1y + r1 f2 = p2x + q2y + r2

Fig 4.4 FIS membership functions (MFs) and rule generation

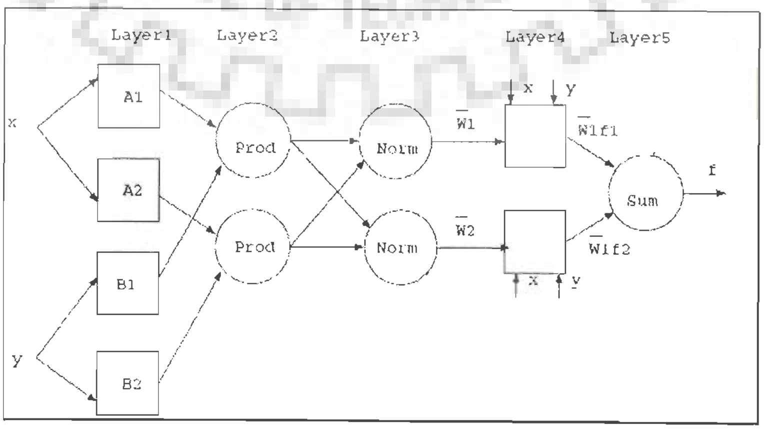


Fig 4.5 FIS network

4.4.2.2 Functionality of Each Layer in FIS

Layer 1

Each node in this layer generates membership grades of an input variable. The node output OP_i is defined by:

$$OP_i^1 = \mu_{Ai}(x) \qquad \text{for } i = 1,2 \tag{4.47}$$

$$OP_i^1 = \mu_{Bi-2}(y) \qquad \text{for } i = 3,4 \tag{4.48}$$

where x (or y) is the input to the node; Ai (or Bi-2) is a fuzzy set associated with this node, characterized by the shape of the MFs (μ) in this node and can be any appropriate functions that are continuous and piecewise differentiable such as Gaussian, generalized bell shaped, trapezoidal shaped and triangular shaped functions. Assuming a Gaussian function as the MF, the output OP_i (1) can be computed as:

$$OP_i^1 = \mu_{Ai}(x) = e^{-\frac{1}{2}\left(\frac{x-c_i}{s_i}\right)^2} \tag{4.49}$$

where $\{c_i, s_i\}$ is the parameter set that changes the shapes of the membership function with maximum equal to 1 and minimum equal to 0. These parameters are called premise parameters or antecedent parameters.

Layer 2

Every node in this layer multiplies the incoming signals, denoted as Π, and the output $OP_i{}^2$ that represents the firing strength of a rule is computed as,

$$OP_i^2 = \mu_{Ai}(x)\mu_{Bi}(y) = w_i \qquad \text{for } i = 1,2 \tag{4.50}$$

Layer 3

The ith node of this layer, labeled as N, computes the normalized firing strengths as

$$OP_i^3 = \frac{w_i}{w_1 + w_2} = \overline{w}_i \qquad \text{for } i = 1,2 \tag{4.51}$$

Layer 4

Node i in this layer computes the contribution of the ith rule toward the model output, with the following node function:

$$OP_i^4 = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i) \qquad (4.52)$$

Where, w is the output of layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set.

The clustering algorithms are used extensively not only to organize and categorize data, but are also useful for data compression and model construction. Clustering partitions a data set into several groups such that the similarity within a group is larger than that among groups. The clustering techniques are validated on the basis of the two assumptions viz. (i) similar inputs to the target system to be modeled should produce similar outputs, and (ii) These similar input-output pairs are bundled into clusters in the training data set.

The subtractive clustering method assumes that each data point is a potential cluster center and calculates a measure of the likelihood that each data point would define the cluster center, based on the density of surrounding data points. Using a fuzzy clustering algorithm, membership functions can be determined according to two possible methods. In the first method, the clusters are projected orthogonally onto the axes of the antecedent variables, and the membership functions are fitted to these projections. The second method uses multi-dimensional antecedent membership functions, i.e. the fuzzy clusters are projected onto the input space. The subtractive clustering algorithm performs the following tasks (i) selects the data point with the highest potential to be the first cluster centre, (ii) removes all data points in the vicinity of the first cluster centre (as determined by radii), in order to determine the next data cluster and its centre location and (iii) iterates on this process until all of the data is within radii of a cluster centre.

In the present study the MATLAB software (Math Works, 1994; The Math Works MATLAB Digest. 2(5)) has been used for regional flood frequency estimation by ANN and FIS.

4.5 STATISTICAL PERFORMANCE INDICES

The statistical performance indices for evaluation of the performances of various approaches used in this study are described below.

4.5.1 Correlation Coefficient (CORR)

The Correlation Coefficient (CORR) is expressed as:

$$r = \frac{\sum_{i=1}^{n}(O_i - \overline{O}_i)(P_i - \overline{P}_i)}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O}_i)^2 \sum_{i=1}^{n}(P_i - \overline{P}_i)^2}} \qquad (4.53)$$

Where, n = total number of data sets; O_i = Observed peak floods for i^{th} data set; P_i = predicted peak flood for i^{th} data set and $\overline{O_i}$ = mean of observed peak flood for i^{th} dataset. Correlation coefficient is a measure of how well the variation in the output is explained by the targets. 'r' value equal to one implies a perfect fit between the outputs and the targets.

4.5.2 Nash-Sutcliffe Efficiency (EFF)

The Nash-Sutcliffe efficiency (EFF) is expressed as:

$$E = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O}_i)^2} \qquad (4.54)$$

The value of Nash-Sutcliffe efficiency EFF varies between $-\infty$ to 1. It may also be expressed in terms of percentage. The closer the value to 1 or 100%, the better is the model performance.

4.5.3 Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) is expressed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \qquad (4.55)$$

RMSE indicates the discrepancy between the observed and predicted values. A RMSE value close to zero indicates better performance of the model. The best fit between observed and predicted values, which is unlikely to occur, would have RMSE as 0.

4.5.4 Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) is expressed as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i| \qquad (4.56)$$

In the statistical performance index taking the absolute value of the error term rather than its square removes the bias towards outlying points in the data set. The best fit between observed and predicted values would have MAE close too.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 GENERAL

The methodology discussed in Chapter 4 has been applied for development of regional flood frequency relationships for gauged and ungauged catchments using the L-moments, ANN and FIS techniques. The annual maximum peak flood data of 196 stream flow gauging sites of the 17 hydrometeorologically homogeneous categorized Subzones of India, described in the Chapter 3 have been used for the development of the regional flood frequency relationships. Regional relationships have been developed between mean annual peak floods and catchment areas of the 17 Subzones using the Levenberg-Marquardt (LM) iteration on the data of the mean annual peak floods and catchment areas for the 17 Subzones. For this regional relationship the statistical performance indices viz. efficiency (EFF), correlation coefficient (CORR), root mean square error (RMSE) and mean average error (MAE) have been computed. For estimation of floods of various return periods for ungauged catchments the regional relationships developed between mean annual peak floods and catchments areas for the 17 Subzones have been coupled with the regional flood frequency relationships developed based on the robust identified frequency distributions for the respective Subzones.

For investigating the applicability of the soft computing techniques in regional flood frequency estimation the ANN and FIS techniques have been applied for the data of 4 Subzones out of the 17 Subzones viz. Subzone 3(c), Subzone 3(d), Subzone 3(f) and Zone-7. These four Subzones have been identified for carrying the detailed regional flood frequency estimation studies based on their lower values of the heterogeneity measure (H), more number of streamflow gauging sites in the Subzones and better representation of the regional relationships between mean annual peak floods and catchment areas as compared to the other Subzones. The regional flood frequency relationships computed by the L-moments, ANN and FIS have also been compared using the aforementioned statistical performance indices. The following aspects of analysis and discussion of results are presented in this chapter:

(i) Screening of the data using the discordancy measure, D_i.

(ii) Testing of homogeneity of the region using the heterogeneity measure, H.

(iii) Identification of the robust frequency distributions for 17 Subzones based on the goodness of fit measures viz. the L-moment ratio diagram and Z^{dist} statistic criteria.

(iv) Development of regional flood frequency relationships using the L-moments approach for gauged catchments.

(v) Development of regional relationships between mean annual peak floods and catchment areas using the Levenberg-Marquardt technique.

(vi) Development of regional flood frequency relationships using the L-moments approach for ungauged catchments.

(vii) Regional flood frequency estimation for gauged catchments using ANN and FIS techniques.

(viii) Comparison of performances of L-moments, ANN and FIS techniques for regional flood frequency estimation.

(ix) Development of regional flood frequency relationships for ungauged catchments using the better identified soft computing technique.

5.2 SCREENING OF DATA USING DISCORDANCY STATISTIC TEST

The objective of the discordancy statistic (D_i) test is to identify those streamflow gauging sites from a group of given sites that are grossly discordant with the group as a whole. Values of D_i have been computed in terms of the L-moments for all the gauging sites of each of the Subzones. The computed values of D_i for each of the sites are compared with the critical value of D_i. The critical values of D_i corresponding to the number of stream flow gauging sites whose data have been used in the analysis are given by Hosking and Wallis (1997) and the same are reproduced in Table 5.1.

Table 5.1 Critical values for the Discordancy statistic, D_i (Hosking and Wallis, 1997)

Number of sites in region	Critical value of D_i
5	1.333
6	1.648
7	1.917
8	2.140
9	2.329
10	2.491
11	2.632
12	2.757
13	2.869
14	2.971
≥15	3.000

If for a Subzone the computed value of D_i for some site is more than the critical value of D_i then that site is discarded from the analysis and D_i values for the remaining sites are again computed. These re-computed D_i values are again compared with the critical D_i value. When the computed D_i values of all the sites are less than the critical value of D_i for a Subzone; then such a group of the streamflow gauging sites is considered for further analysis as described below.

5.3 TESTING OF REGIONAL HOMOGENEITY

The test based on the heterogeneity measure 'H' takes into consideration that in a homogeneous region, all sites have same population L-moment ratios. But their sample L-moment ratios may differ at each site due to sampling variability. The intersite variation of L-moment ratio is measured as the standard deviation of the at-site LCV's weighted proportionally to the record length at each site. To establish what would be the expected inter-site variation of L-Moment ratios for a homogeneous region, 500 simulations were carried out using the Kappa distribution for computing the heterogeneity measure (H) (Hosking and Wallis, 1997).

The heterogeneity measure (H) has been computed for each Subzone using the data of the streamflow gauging sites which are found suitable for regional flood frequency analysis as per the screening of data using the D_i statistic described in Section 5.2. The values of the heterogeneity measures H(1), H(2) and H(3) were computed utilizing the data of all the sites passing the D_i test by generating 500 regions using the fitted Kappa distribution. Hosking and Wallis (1997) suggested the following criteria for assessing heterogeneity of a region: if $H(j) < 1$, the region is acceptably homogeneous; if $1 \leq H(j) < 2$, the region is possibly heterogeneous; and if $H(j) \geq 2$, the region is definitely heterogeneous. Hence, if by following the above procedure the values heterogeneity measure H(1), H(2) and H(3) for a Subzone are obtained more than 1; then the site exhibiting the maximum D_i as per the D_i statistic test is excluded from the analysis and the heterogeneity measure H(1), H(2) and H(3) values are again computed. This procedure is repeated until the heterogeneity measure (H) is obtained within the range suggested by Hosking and Wallis (1997), as mentioned above. In the case of analysis of the data of 17 Subzones of India considered in the present study the efforts to reduce the values of H(1), H(2) and H(3) to 1 led to elimination of data of a large number of the stream flow gauging sites resulting in the significant loss of data and hence, this was considered appropriate to accept the H(1), H(2) and H(3) value close to 2 for formation of a homogeneous region. Similar procedure has also been adopted in some of the studies carried out earlier also (Kumar and Chatterjee, 2005; Kumar et al., 2003a, b).

The values of number of streamflow gauging sites, the range of discordancy statistic (D_i) as well as the heterogeneity measures(H1, H2, H3) computed by carrying out 500 simulations using the Kappa distribution for the 17 Subzones, which have been considered hydrometeorologically homogeneous for carrying out the regional flood frequency analysis, are given in Table 5.2. The catchment area, sample statistics, sample size and discordancy statistic for each of the streamflow gauging sites the 17 Subzones are given in Tables 5.3.1 to Table 5.3.17.

Table 5.2 Discordancy statistic D_i and Heterogeneity measures for 17 Subzones

S.N.	Subzone	No. of gauging sites	Range of discordancy statistic	Heterogeneity measures		
			Di	H1	H2	H3
1	1 (b)	12	0.09-2.12	1.65	0.91	0.08
2	1 (d)	10	0.35-2.12	0.97	0.93	0.17
3	1 (e)	12	0.00-2.16	2.12	2.06	1.09
4	1 (f)	8	0.08-2.08	0.71	0.89	1.76
5	1 (g)	10	0.08-2.27	1.32	1.06	0.53
6	2 (a)	13	0.25-2.51	1.66	0.93	0.39
7	2 (b)	11	0.23-1.73	2.26	2.16	0.79
8	3 (a)	10	0.36-1.89	0.46	0.74	0.56
9	3 (b)	14	0.08-2.49	1.84	0.91	0.14
10	3 (c)	13	0.17-2.56	1.79	1.84	0.54
11	3 (d)	15	0.08-2.11	1.68	0.71	1.98
12	3 (e)	9	0.36-2.10	1.21	0.34	0.39
13	3 (f)	17	0.34-2.03	1.13	1.64	0.61
14	3 (h)	16	0.09-2.51	2.26	0.75	0.52
15	3 (i)	8	0.36-2.01	2.18	1.26	0.90
16	4 (b)	8	0.19-2.00	1.28	0.87	0.00
17	Zone 7	10	0.10-1.95	0.47	0.43	0.81

Table 5.3.1 Catchment area, sample statistic, sample size and discordancy statistic for Chambal Subzone 1(b)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
94	2297.330	1549.000	20	0.3816	0.2044	0.1582	0.64
72	662.800	597.520	23	0.5447	0.2907	0.1185	0.57
118	41.000	70.800	10	0.2574	0.1193	0.3110	2.12
1116/3	361.050	339.560	16	0.5308	0.4779	0.3076	1.63
1	44.750	100.078	13	0.5197	0.3285	0.2394	0.95
437	237.140	197.130	10	0.5253	0.2520	-0.0186	1.43
77	26.180	18.820	11	0.6632	0.4965	0.2180	1.56
306	43.770	76.770	31	0.4650	0.2911	0.1820	0.09
35	39.520	184.654	26	0.3348	0.0108	0.0340	1.67
44	109.000	202.680	22	0.4598	0.2517	0.0875	0.64
406	48.090	92.210	24	0.4068	0.1454	0.1263	0.39
519	1500.020	1551.600	25	0.5212	0.2804	0.0749	0.31

Table 5.3.2 Catchment area, sample statistic, sample size and statistic statistic for Sone Subzone 1(d)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
1198/1	341	224.258	31	0.4145	0.2782	0.1320	0.37
1136/1	158	166.850	20	0.2439	0.2951	0.2983	2.12
611	440	201.483	29	0.5276	0.4881	0.3127	1.09
171	373	203.970	33	0.4597	0.4699	0.2851	1.87
462	517	130.217	23	0.4206	0.2360	0.0668	1.17
184	249	337.125	24	0.3643	0.3281	0.2643	0.54
155	181	235.375	24	0.4881	0.4550	0.3100	0.81
187	1658	404.889	18	0.3341	0.2213	0.1473	0.35
108 K	279	269.056	18	0.3297	0.1377	0.0629	1.12
31	812	584.417	12	0.4602	0.4899	0.3496	0.55

Table 5.3.3 Catchment area, sample statistic, sample size and discordancy statistic for Upper Indo -Ganga Plains Subzone 1(e)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
89	810.94	64.256	32	0.5799	0.3717	0.1527	0.22
1227	41.15	33.575	28	0.6813	0.5115	0.2527	1.03
181	352.00	38.611	32	0.6828	0.4773	0.1728	1.11
99	90.65	187.893	28	0.5533	0.4012	0.1743	0.62
1307	322.19	190.707	27	0.5194	0.3237	0.2049	0.27
1231	49.47	56.497	30	0.4985	0.3240	0.1076	1.72
93	264.18	411.029	34	0.5307	0.3505	0.1819	0.02
146	194.26	9.152	25	0.3712	0.1259	0.1194	1.03
20	2425.53	112.774	31	0.5128	0.2606	0.1700	1.06
325	252.60	350.897	29	0.3944	0.1337	0.1102	0.89
50	38.85	13.702	30	0.4760	0.2505	0.1165	0.21
61	278.94	97.113	30	0.6472	0.4846	0.2024	0.60

Table 5.3.4 Catchment area, sample statistic, sample size and discordancy statistic for Middle Ganga Plains Subzone 1(f)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
59	54.39	97.485	33	0.3107	-0.0475	0.0216	1.63
30	447.76	490.500	30	0.3268	0.1201	0.0013	1.24
160	150.40	70.313	32	0.2958	0.1394	0.1792	0.28
3	32.89	24.290	31	0.3867	0.2226	0.1335	0.89
60	130.00	140.556	27	0.2205	0.3578	0.3932	2.08
24	69.75	59.308	26	0.3289	0.1231	0.1188	0.08
141	59.83	79.391	23	0.3354	0.1958	0.0918	0.55
104	234.19	555.207	29	0.3647	0.2429	0.2497	1.26

Table 5.3.5 Catchment area, sample statistic, sample size and discordancy statistic for Lower Ganga Plains Subzone 1(g)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
181	212.90	260.423	23	0.4363	0.3478	0.1857	0.25
94	336.70	356.062	18	0.4105	0.2673	0.1120	0.21
286	136.83	137.120	29	0.3244	0.1607	0.0184	1.08
49	393.68	650.450	36	0.5482	0.6362	0.5179	2.16
462	516.30	163.200	10	0.2540	0.1351	0.2107	1.40
656	79.50	99.918	33	0.5400	0.4261	0.2784	0.89
676	92.39	185.539	33	0.2718	-0.0183	0.1660	2.27
101	244.24	223.987	23	0.3782	0.2385	0.1500	0.08
167	569.80	342.441	32	0.4525	0.2266	0.1209	0.65
27	15.01	51.215	30	0.4425	0.1961	-0.0048	1.01

Table 5.3.6 Catchment area, sample statistic, sample size and discordancy statistic for North Brahmaputra Subzone 2(a)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
450	233.10	208.108	17	0.3417	0.0677	0.0492	0.74
242	230.00	698.710	24	0.3812	0.1872	0.0848	0.25
285	92.45	149.421	22	0.3377	0.1857	0.1395	0.48
566	46.26	18.542	23	0.2815	0.1499	0.1403	0.60
70	21.42	22.665	13	0.3164	0.2739	0.1654	0.92
91	132.35	852.743	22	0.3133	0.0565	0.0321	0.84
139	22.17	50.940	26	0.4745	0.3511	0.1668	2.51
215	135.66	33.510	25	0.3959	0.1918	0.0822	0.41
24	42.10	49.782	23	0.3864	0.3218	0.1492	0.85
363	326.00	380.289	27	0.2366	0.1054	0.1611	2.25
12	230.40	581.713	21	0.3221	0.2269	0.1264	0.74
196	85.47	79.715	19	0.3809	0.4082	0.2303	1.34
373	595.70	526.603	17	0.3441	0.0586	0.0134	1.08

Table 5.3.7 Catchment area, sample statistic, sample size and discordancy statistic for South Brahmaputra Subzone 2(b)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
8	259.83	75.265	21	0.5954	0.5218	0.3374	0.76
526	303.68	158.606	5	0.2120	0.0906	0.1910	1.29
215	135.66	31.233	27	0.4318	0.2028	0.0738	0.91
160	497.28	74.262	18	0.3652	0.2567	0.2735	0.82
141	59.80	76.000	21	0.3498	0.2542	0.1313	1.59
414	338.72	321.148	16	0.2986	0.1084	0.0034	1.28
269	76.93	78.197	24	0.5576	0.5199	0.3732	1.00
404	20.98	28.699	26	0.2307	-0.0276	0.0198	1.57
566	46.62	18.319	24	0.2811	0.1536	0.1501	0.41
130	46.44	61.014	14	0.5654	0.4114	0.2348	0.71
170	32.37	30.877	21	0.5616	0.4109	0.1968	0.65

Table 5.3.8 Catchment area, sample statistic, sample size and discordancy statistic for Mahi and Sabarmati Subzone 3(a)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
8	30.14	74.000	25	0.4728	0.4372	0.2557	0.36
192/253	48.43	189.684	19	0.3545	0.2002	0.1432	1.20
281/334	18.44	75.588	17	0.4854	0.4657	0.3758	1.89
5	230.00	352.722	18	0.5839	0.4512	0.1824	1.72
46	580.00	352.955	22	0.4807	0.2842	0.0626	0.53
99	144.50	258.143	21	0.3826	0.2427	0.0846	0.53
945	231.11	212.071	14	0.4698	0.2561	0.1035	0.47
26	1094.00	448.650	20	0.404	0.2748	0.0666	0.90
11	98.16	164.667	18	0.4183	0.4671	0.2888	1.46
141	73.19	108.941	17	0.4280	0.1648	0.0403	0.93

Table 5.3.9 Catchment area, sample statistic, sample size and discordancy statistic for Lower Narmada and Tapi Subzone 3(b)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
105	59.59	223.821	28	0.4580	0.4467	0.3591	1.83
110	18.90	116.654	26	0.4092	0.1943	0.0935	0.65
502/3	105.07	234.154	26	0.3489	0.2670	0.1626	1.37
200	27.18	34.952	21	0.4471	0.3240	0.2037	0.08
162	17.22	69.273	22	0.3838	0.2424	0.1425	0.29
21(DEV)	378.04	492.526	19	0.5757	0.4946	0.3178	1.32
701	28.23	239.000	18	0.5760	0.4647	0.2355	0.58
374/1	225.84	316.095	21	0.5630	0.3981	0.1367	1.17
497/1	53.09	77.652	23	0.3984	0.1045	0.0318	2.49
50	193.73	352.053	19	0.4721	0.3686	0.2603	0.61
411/1	261.59	558.286	21	0.4741	0.419	0.2288	0.88
485/4	284.90	248.333	21	0.4570	0.3129	0.1278	0.53
361/2	828.00	244.053	19	0.3072	0.1712	0.1306	1.09
53	103.26	274.917	12	0.5972	0.4514	0.1841	1.09

Table 5.3.10 Catchment area, sample statistic, sample size and discordancy statistic for Upper Narmada and Tapi Subzone 3(c)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
731/6	115.90	252.867	30	0.2922	0.1900	0.0962	2.05
294	518.67	919.600	30	0.3470	0.1613	0.1106	0.17
897/1	314.88	856.462	26	0.4119	0.3103	0.1610	0.18
634/2	348.92	380.103	29	0.3434	0.2221	0.2018	1.07
831/1	53.68	209.174	23	0.2729	-0.0222	0.0548	1.56
505	67.37	211.792	24	0.3104	0.1045	0.0571	0.55
863/1	2110.85	1687.273	22	0.4515	0.3783	0.1546	0.95
253	114.22	216.900	20	0.3600	0.1247	0.0923	0.86
584/1	139.08	248.783	23	0.4298	0.3415	0.1700	0.35
512/3	142.97	219.955	22	0.3848	0.2383	0.0869	0.72
776/1	179.90	572.778	18	0.2791	0.1842	0.1622	1.38
644/1	989.89	546.250	20	0.4498	0.3764	0.1903	0.59
787/2	321.16	811.786	14	0.4457	0.5553	0.3520	2.56

Table 5.3.11 Catchment area, sample statistic, sample size and discordancy statistic for Mahanadi Subzone 3(d)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
48	109	103.900	30	0.4020	0.2950	0.1658	0.46
93K	74	153.071	28	0.2740	0.1235	0.1974	1.44
59KGP	30	72.897	29	0.4079	0.2770	0.1780	0.74
308	19	41.222	27	0.3461	0.2339	0.0882	0.87
332NGP	225	188.591	22	0.2899	0.2117	0.2020	1.23
59BSP	136	196.227	22	0.4068	0.3471	0.2283	1.48
698	113	247.000	25	0.4240	0.3210	0.1356	1.09
121	1150	1003.857	21	0.2690	0.1622	0.0787	1.19
332KGP	175	71.833	24	0.3102	0.1569	0.1647	0.51
40K	115	260.667	21	0.3469	0.2328	0.1784	0.14
42	49	53.500	20	0.2260	0.0488	0.0530	1.92
69	173	238.895	19	0.3457	0.2392	0.1455	0.08
90	190	130.727	11	0.3570	0.1566	0.1335	2.11
195	615	963.769	13	0.2394	0.1305	0.1614	1.10
235	312	176.143	14	0.3128	0.2205	0.1130	0.63

Table 5.3.12 Catchment area, sample statistic, sample size and discordancy statistic for Upper Godavari Subzone 3(e)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
139	93.60	163.344	32	0.3907	0.2321	0.1497	0.36
234	2227.39	868.875	24	0.4171	0.2080	0.0425	0.52
79	35.22	60.130	23	0.4562	0.2001	0.0089	0.93
346	64.88	203.696	23	0.3615	0.1103	0.0867	0.36
295	77.70	90.864	22	0.2933	0.1775	0.1395	2.08
368	136.75	206.286	21	0.3896	0.0993	0.0681	0.96
76	1197.76	695.333	18	0.4804	0.3040	0.0544	0.86
44	152.33	214.643	14	0.5022	0.4267	0.2009	2.10
289	458.00	263.800	15	0.3062	0.0452	0.0885	0.81

Table 5.3.13 Catchment area, sample statistic, sample size and discordancy statistic for Lower Godavari Subzone 3(f)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
184	364	344.483	29	0.3879	0.2106	0.1462	0.34
57	163	189.393	28	0.2567	0.1229	0.1154	1.02
973/1	362	505.036	28	0.3414	0.0600	0.0323	0.60
912/1	137	404.862	29	0.4042	0.2779	0.1095	0.94
20	60	204.714	28	0.3335	0.0219	0.0529	1.02
4	50	237.966	29	0.2834	0.1236	0.0878	0.75
214	35	77.750	24	0.2813	0.2460	0.2389	1.25
51	87	206.680	25	0.2802	0.0747	0.1428	1.35
807/1	824	1212.826	23	0.3730	0.1823	0.0663	0.63
228	483	1075.273	22	0.3827	0.2806	0.1172	0.93
15	459	854.913	23	0.3767	0.1968	0.1170	0.11
881/1	158	307.783	23	0.2855	0.0763	0.0990	0.59
875/1	751	778.095	21	0.4119	0.0773	0.0030	1.56
161	53	93.882	17	0.2992	0.3648	0.2329	2.03
36	139	170.800	15	0.4150	0.3185	0.2357	1.43
224	750	687.357	14	0.4067	0.3365	0.2431	1.18
65	731	725.133	15	0.4147	0.4224	0.2557	1.27

Table 5.3.14 Catchment area, sample statistic, sample size and discordancy measure for Krishna and Pennar Subzone 3(h)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
642	326.08	283.469	32	0.3890	0.2883	0.1590	0.58
123	64.75	111.485	33	0.3394	0.0994	0.1331	1.31
16	270.60	65.679	28	0.4413	0.1889	0.0215	0.96
53(i)	102.45	78.517	29	0.4668	0.1517	-0.0435	2.51
378/3	79.00	89.773	22	0.4071	0.1837	0.0608	0.47
53(ii)	1689.92	794.885	26	0.4752	0.3098	0.1966	0.74
215	167.32	44.308	26	0.4827	0.3099	0.1560	0.25
215(GTL)	139.08	88.040	25	0.4072	0.2559	0.1567	0.13
18	131.52	117.760	25	0.3674	0.1761	0.1851	0.99
322	31.72	50.920	25	0.3013	0.2454	0.1507	2.18
480/3	118.23	92.235	17	0.5419	0.3959	0.1710	1.02
63	1357.15	403.368	19	0.3687	0.1162	0.1504	1.55
601	398.60	280.235	17	0.4775	0.2811	0.1189	0.09
313	220.45	443.167	18	0.3966	0.3250	0.1547	1.17
66	70.84	28.294	17	0.6073	0.3750	0.0998	1.63
98	348.40	125.357	14	0.5162	0.3361	0.1198	0.40

Table 5.3.15 Catchment area, sample statistic, sample size and discordancy statistic for Kaveri Basin Subzone 3(i)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
37	294.00	51.630	33	0.7054	0.5443	0.2805	0.52
26	74.70	27.305	30	0.7008	0.5307	0.2633	0.36
244	30.00	60.593	29	0.5090	0.1712	-0.0441	0.58
583	146.00	54.196	27	0.7496	0.5780	0.2904	0.78
28	953.00	309.133	18	0.6050	0.5091	0.2691	2.01
683	287.00	64.333	12	0.4643	0.1275	-0.0484	1.23
683	287.50	58.536	14	0.4870	0.1950	-0.0206	0.68
845	31.23	12.803	30	0.6719	0.4242	0.1149	1.84

Table 5.3.16 Catchment area, sample statistic, sample size and discordancy statistic for East Coast Subzones 4(b)

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
765	663.00	104.635	20	0.5922	0.3866	0.1973	1.88
583	146.20	53.990	28	0.7406	0.5673	0.2875	0.19
172	51.18	48.063	38	0.6703	0.4184	0.1103	1.57
313	258.22	43.204	27	0.8054	0.6505	0.3467	0.88
346	266.76	160.017	29	0.7525	0.6040	0.3369	0.35
252	401.32	315.983	29	0.5803	0.4960	0.3535	2.00
60	96.60	116.677	31	0.6117	0.4221	0.1913	0.62
152	626.78	167.941	17	0.7096	0.5536	0.3206	0.51

Table 5.3.17 Catchment area, sample statistic, sample size and discordancy statistic for Sub-Himalayan region Zone - 7

Stream Gauging Site	Catchment Area (km^2)	Mean Annual Peak Flood (m^3/s)	Sample Size (Years)	L-CV (τ_2)	L-skew (τ_3)	L-kurtosis (τ_4)	Discordancy Statistic (D_i)
104	2072	855.000	20	0.4008	0.2660	0.1555	0.47
232	710	1606.769	13	0.2363	-0.0031	0.1784	0.72
104	234	677.750	20	0.3025	0.2978	0.3595	1.54
65	190	296.000	20	0.2732	0.1599	0.1964	1.13
3	178	145.846	13	0.4139	0.2869	0.1923	0.67
629	104	530.846	13	0.2551	0.0776	0.2621	0.83
154	43	264.462	13	0.2337	-0.0453	0.1391	1.05
48	27	68.400	20	0.3999	0.1508	0.0137	1.53
50	25	17.100	20	0.3518	0.2118	0.1891	0.10
278	6	20.462	13	0.4250	0.1059	0.0676	1.95

5.4 IDENTIFICATION OF ROBUST REGIONAL FREQUENCY DISTRIBUTIONS

The choice of an appropriate frequency distribution for a homogeneous region is made by comparing the L-moments of the distributions to the average L-moments statistics from regional data. The aim of goodness-of-fit measure is to identify a distribution that fits the observed data acceptably closely. The goodness of fit is judged by how well the L-Skewness and L-Kurtosis of the fitted distribution match the regional average L-Skewness and L-Kurtosis of the observed data. The L-moment ratio diagram and $|Z_i^{dist}|$ –statistic are used as the best fit criteria for identifying the robust distribution for the study area. For each of the Subzones the regional average values of L-skewness i.e. τ_3 and L-kurtosis i.e. τ_4 are plotted on the L-moment ratio diagram and the frequency distribution lying closest to the point defined by the regional average value of L-skewness i.e. τ_3 and L-kurtosis i.e. τ_4 on the L-moment ratio diagram is considered as the suitable frequency distribution as per this test.

For identification of the robust frequency distribution for a Subzone, $|Z_i^{dist}|$ –statistic of the various distributions having its value lower than 1.64 are also compared and the frequency distribution exhibiting the lowest value of and $|Z_i^{dist}|$ is considered appropriate distribution as per this test (Hosking and Wallis, 1997). The comparison of various frequency distributions based on these two tests viz. the L-moment ratio diagram and the $|Z_i^{dist}|$ –statistic provides the robust distribution for each of the Subzones. The L-moment ratio diagrams for the 17 Subzones are shown in Figs. 5.1.1 to 5.1.17 and the Z_i^{dist} values for various frequency distributions for the 17 Subzones are given in Tables 5.4.1 to 5.4.17. Based on the L-moment ratio diagram and $|Z_i^{dist}|$ –statistic criteria, robust distributions are identified for each of the 17 Subzones. The names of the robust identified distributions and values of $|Z_i^{dist}|$ –

statistic of the robust identified frequency distributions for each of the 17 Subzones are given in Table 5.5.

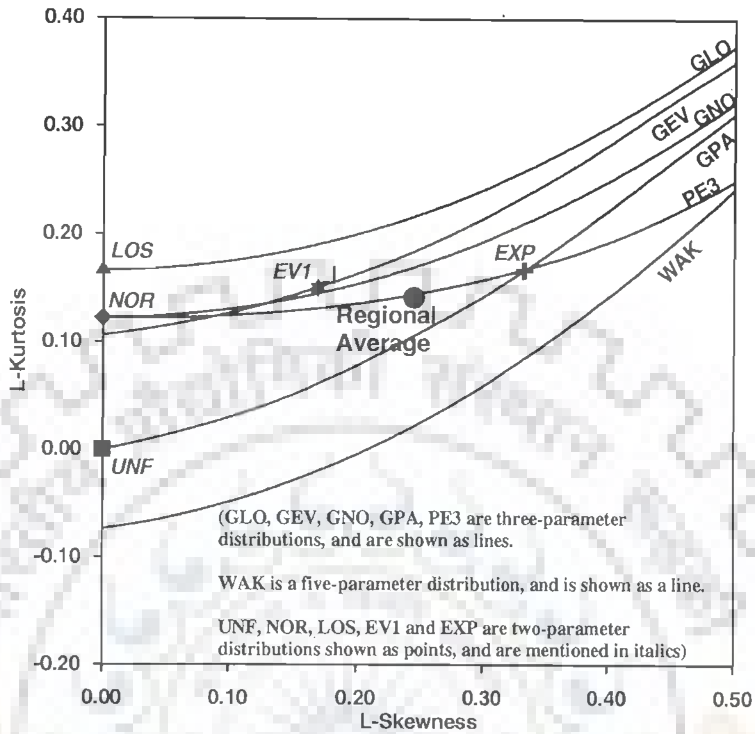


Fig 5.1.1 L-moments ratio diagram for Chambal Subzone 1(b)

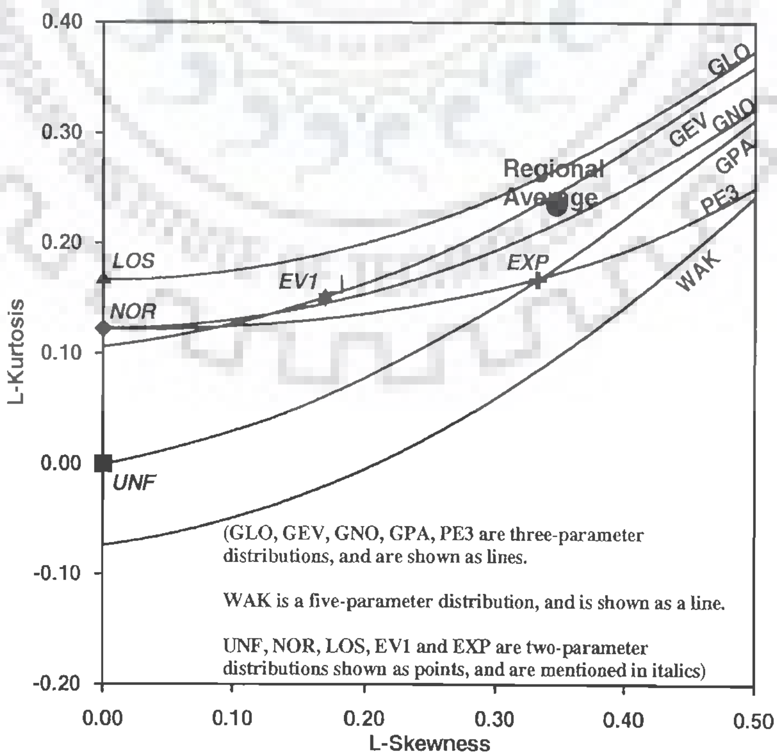


Fig 5.1.2 L-moments ratio diagram for Sone Subzone 1(d)

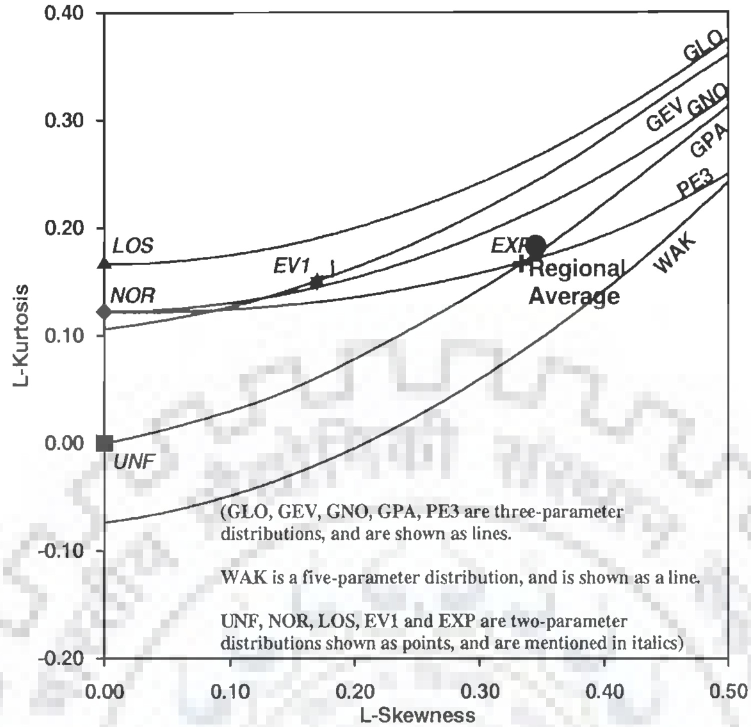


Fig 5.1.3 L-moments ratio diagram for Upper Indo-Ganga Plains Subzone 1(e)

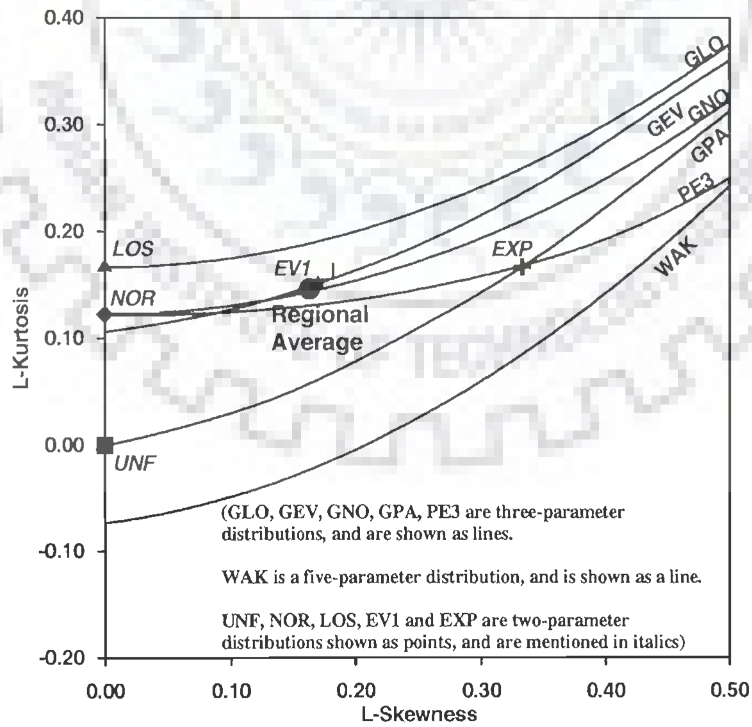


Fig 5.1.4 L-moments ratio diagram for Middle Ganga Plains Subzone 1(f)

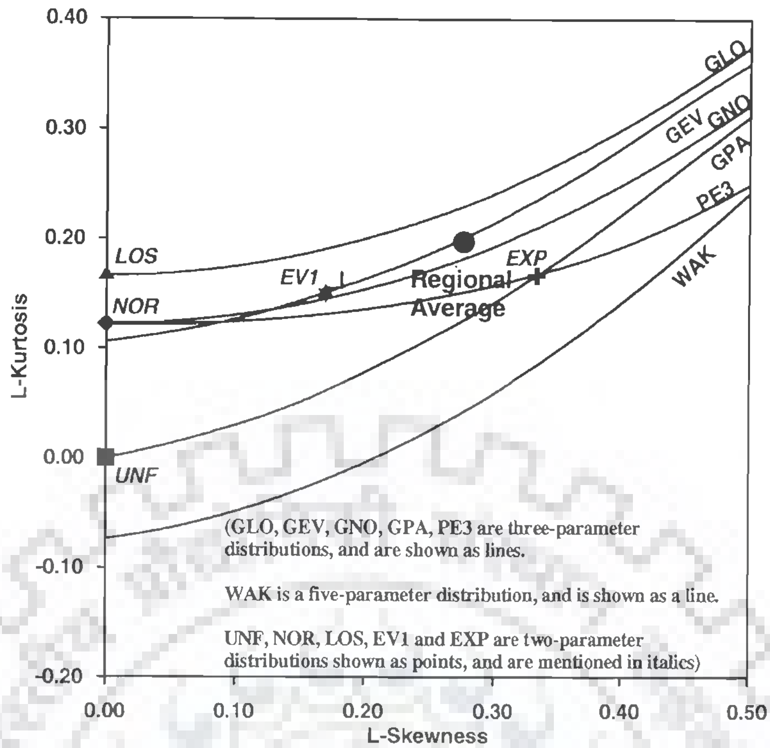


Fig 5.1.5 L-moments ratio diagram for Lower Ganga Plains Subzone 1(g)

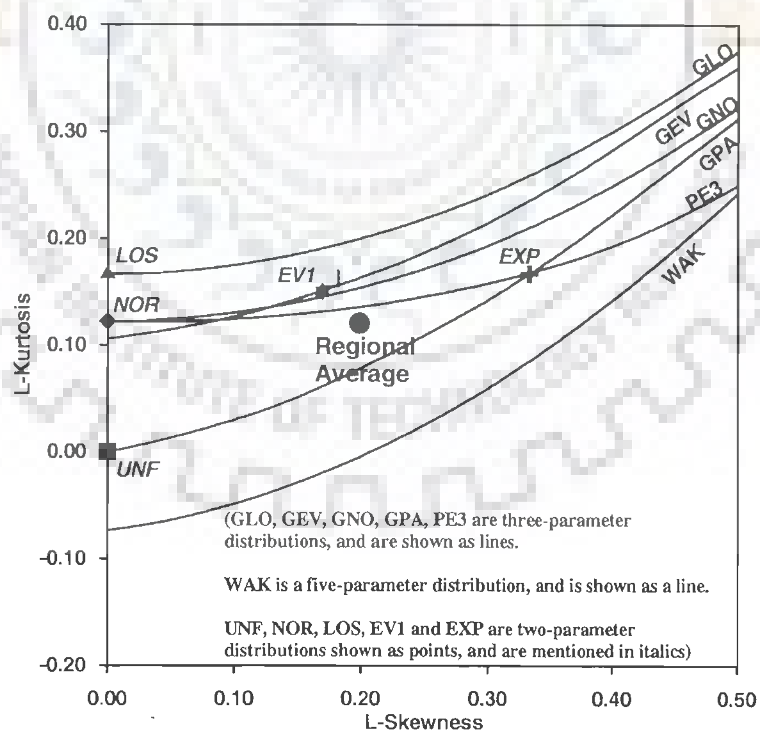


Fig 5.1.6 L-moments ratio diagram for North Brahmaputra Subzone 2(a)

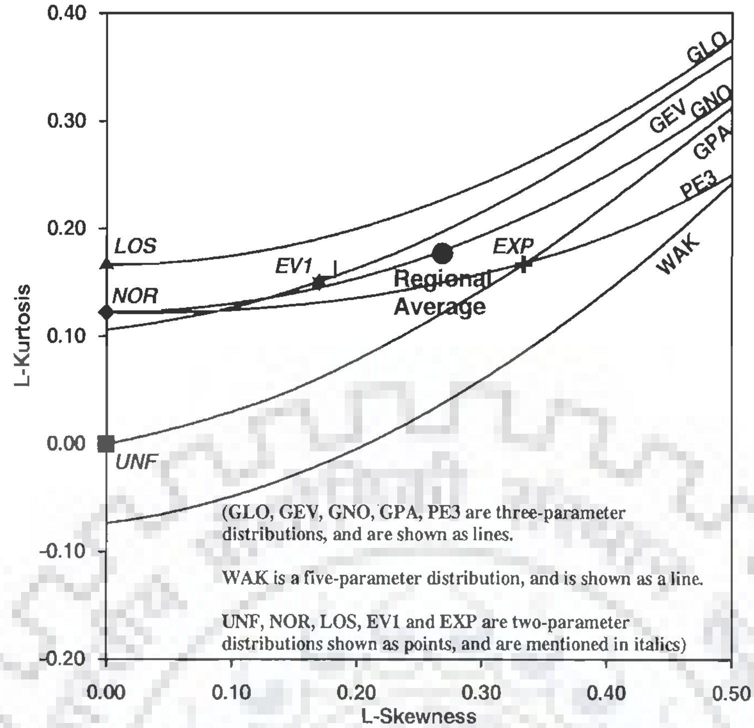


Fig 5.1.7 L-moments ratio diagram for South Brahmaputra Subzone 2(b)

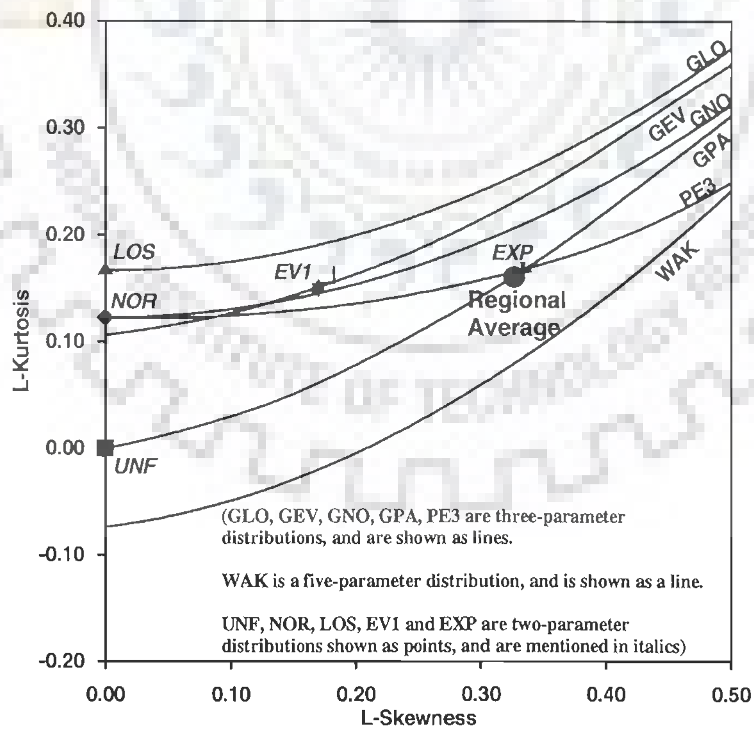


Fig 5.1.8 L-moments ratio diagram for Mahi and Sabarmati Subzone 3(a)

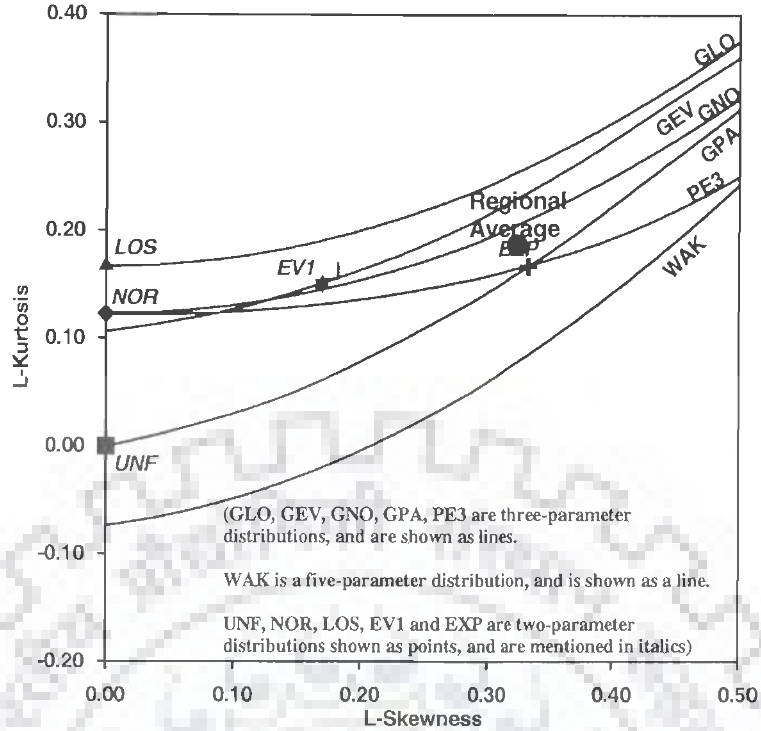


Fig 5.1.9 L-moments ratio diagram for Lower Narmada and Tapi Subzone 3(b)

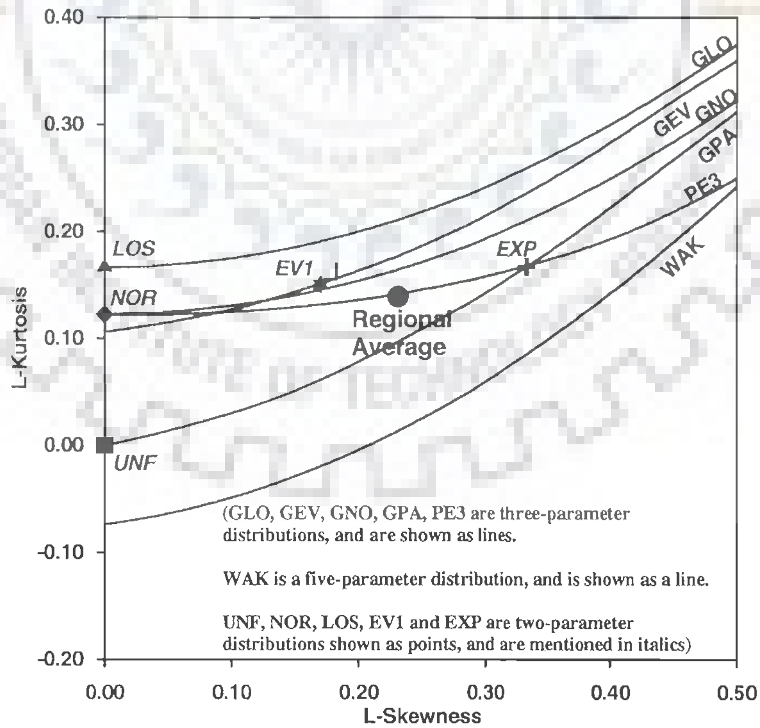


Fig 5.1.10 L-moments ratio diagram for Upper Narmada and Tapi Subzone 3(c)

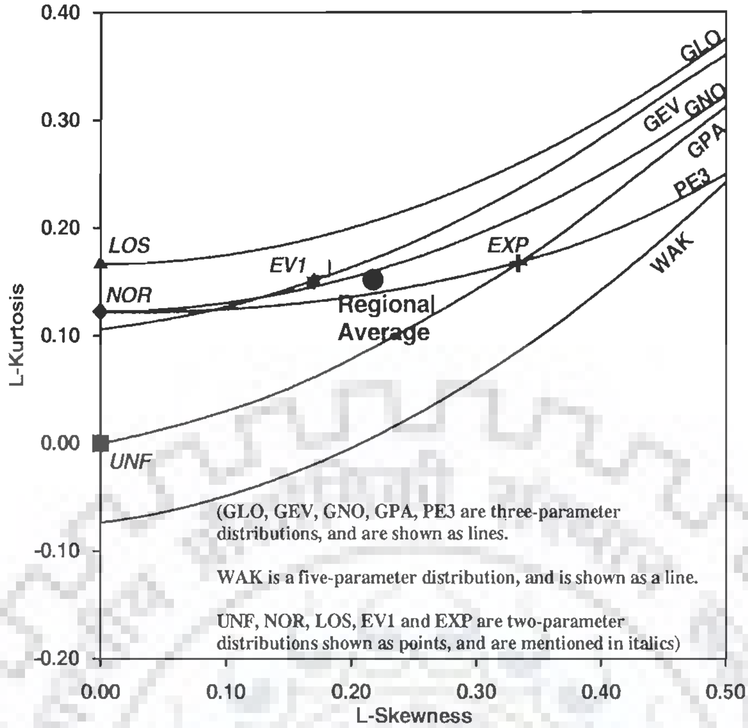


Fig 5.1.11 L- moments ratio diagram for Mahanadi Subzone 3(d)

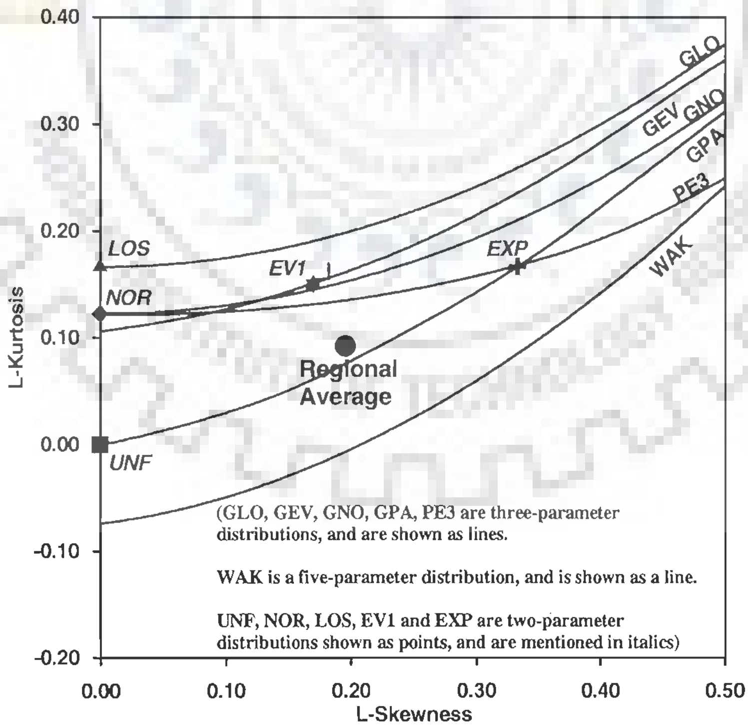


Fig 5.1.12 L-moments ratio diagram for Upper Godavari Subzone 3(e)

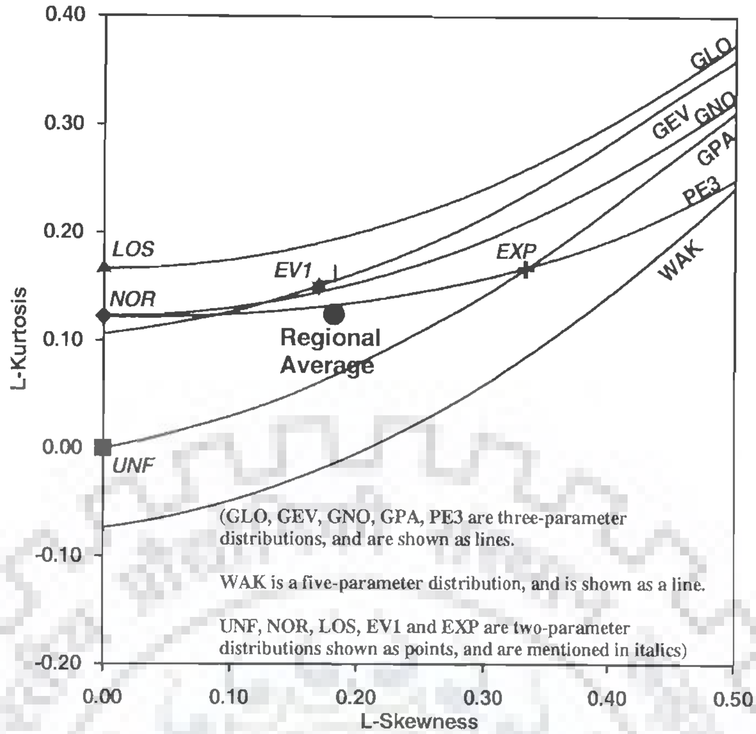


Fig 5.1.13 L-moments ratio diagram for Lower Godavari Subzone 3(f)

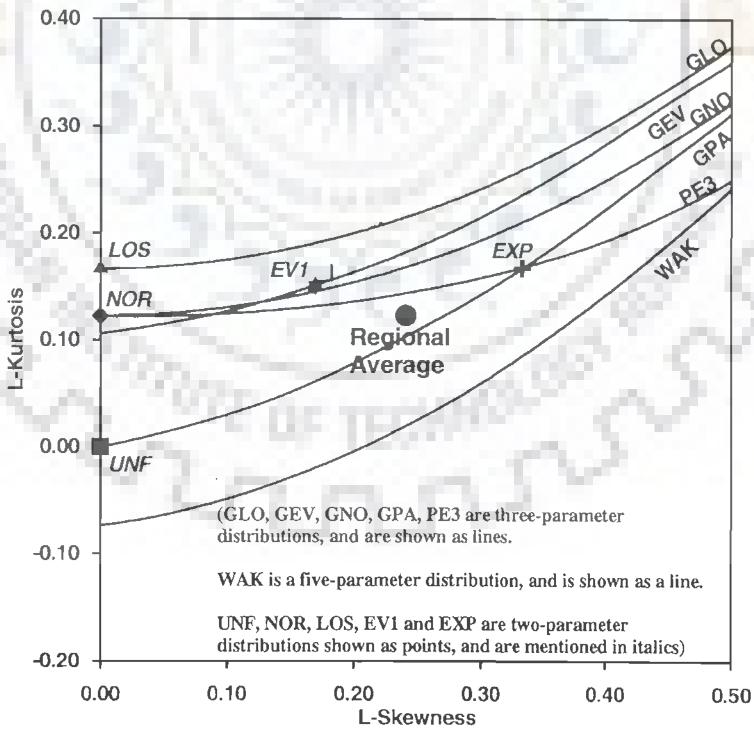


Fig 5.1.14 L-moments ratio diagram for Krishna and Pennar Subzone 3(h)

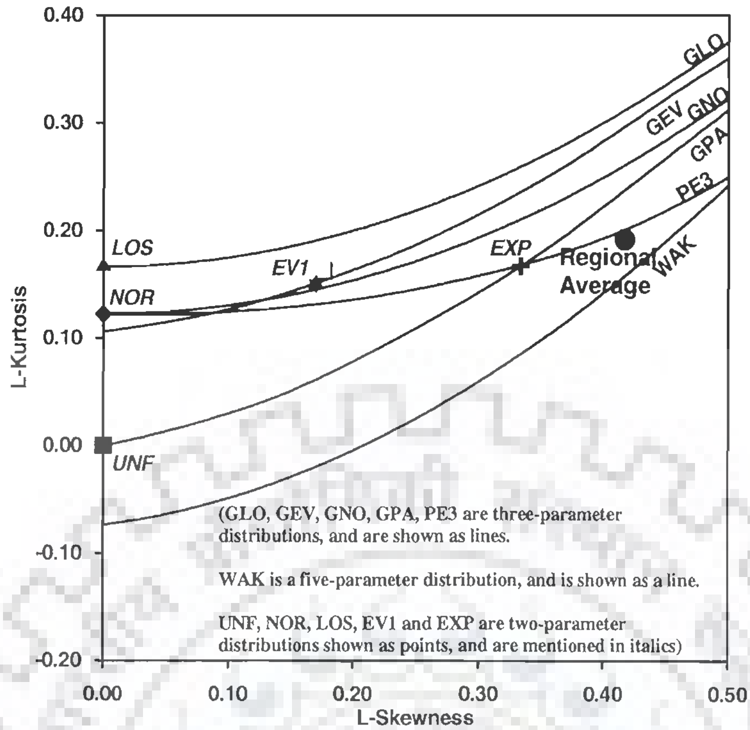


Fig 5.1.15 L-moments ratio diagram for Kaveri Basin Subzone 3(i)

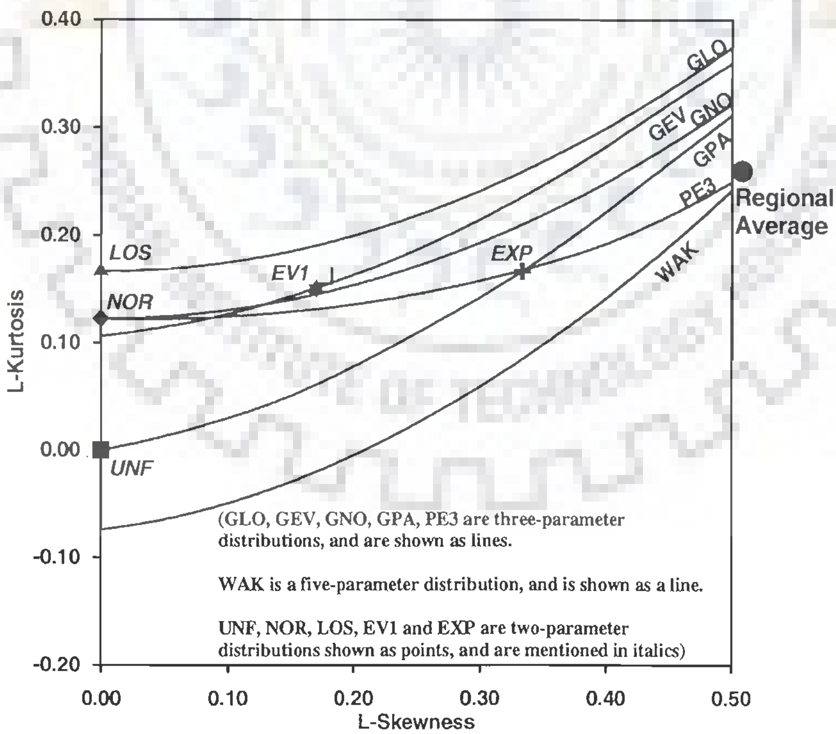


Fig 5.1.16 L-moments ratio diagram for East Coast Subzone 4(b)

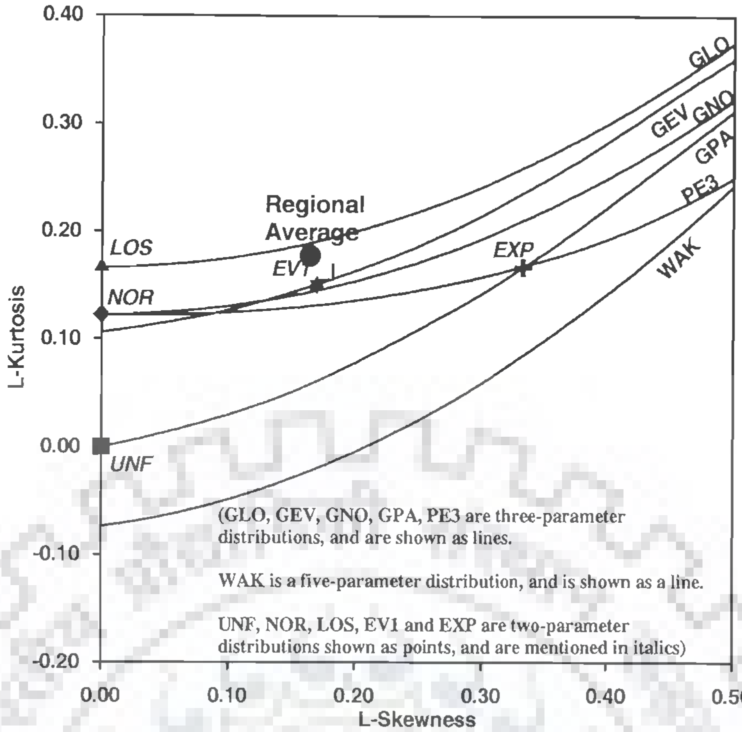


Fig 5.1.17 L-moments ratio diagram for Sub-Himalayan region Zone-7

Table 5.4.1 Z_i^{dist} statistic for various distributions for Chambal Subzone 1 (b)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Pearson Type III (PE3)	0.01
2.	Generalized Normal (GNO)	0.88
3.	Generalized Pareto (GPA)	-1.32
4.	Generalized Extreme Value (GEV)	1.37
5.	Generalized logistic (GLO)	2.46

Table 5.4.2 Z_i^{dist} statistic for various distributions for Sone Subzone 1 (d)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Generalized Extreme Value (GEV)	0.13
2.	Generalized Normal (GNO)	-0.57
3.	Generalized logistic (GLO)	0.70
4.	Generalized Pareto (GPA)	-1.59
5.	Pearson Type III (PE3)	-1.76

Table 5.4.3 Z_i^{dist} statistic for various distributions for Upper Indo-Ganga Plains Subzone 1 (e)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Generalized Pareto (GPA)	-0.30
2.	Pearson Type III (PE3)	-0.73
3.	Generalized Normal (GNO)	0.90
4.	Generalized Extreme Value (GEV)	1.85
5.	Generalized logistic (GLO)	2.54

Table 5.4.4 Z_i^{dist} statistic for various distributions for Middle Ganga Plains Subzone 1 (f)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Generalized Extreme Value (GEV)	0.01
2.	Generalized Normal (GNO)	-0.14
3.	Pearson Type III (PE3)	-0.62
4.	Generalized logistic (GLO)	1.58
5.	Generalized Pareto (GPA)	-3.40

Table 5.4.5 Z_i^{dist} statistic for various distributions for Lower Ganga Plains Subzone 1 (g)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Generalized Extreme Value (GEV)	0.27
2.	Generalized Normal (GNO)	-0.32
3.	Generalized logistic (GLO)	1.22
4.	Pearson Type III (PE3)	-1.36
5.	Generalized Pareto (GPA)	-2.19

Table 5.4.6 Z_i^{dist} statistic for various distributions for North Brahmaputra Subzone 2(a)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Pearson Type III (PE3)	0.68
2.	Generalized Normal (GNO)	1.44
3.	Generalized Pareto (GPA)	-1.80
4.	Generalized Extreme Value (GEV)	1.80
5.	Generalized logistic (GLO)	3.38

Table 5.4.7 Z_i^{dist} statistic for various distributions for South Brahmaputra Subzone 2(b)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Generalized Normal (GNO)	-0.11
2.	Generalized Extreme Value (GEV)	0.40
3.	Pearson Type III (PE3)	-1.00
4.	Generalized logistic (GLO)	1.28
5.	Generalized Pareto (GPA)	-1.84

Table 5.4.8 Z_i^{dist} statistic for various distributions for Mahi and Sabarmati Subzone 3 (a)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Pearson Type III (PE3)	-0.06
2.	Generalized Pareto (GPA)	-0.14
3.	Generalized Normal (GNO)	1.13
4.	Generalized Extreme Value (GEV)	1.82
5.	Generalized logistic (GLO)	2.51

Table 5.4.9 Z_i^{dist} statistic for various distributions for Lower Narmada and Tapi Subzone 3 (b)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Generalized Normal (GNO)	0.32
2.	Pearson Type III (PE3)	-1.02
3.	Generalized Extreme Value (GEV)	1.10
4.	Generalized Pareto (GPA)	-1.14
5.	Generalized logistic (GLO)	1.88

Table 5.4.10 Z_i^{dist} statistic for various distributions for Upper Narmada and Tapi Subzone 3(c)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Pearson Type III (PE3)	0.05
2.	Generalized Normal (GNO)	1.02
3.	Generalized Extreme Value (GEV)	1.54
4.	Generalized Pareto (GPA)	-1.84
5.	Generalized logistic (GLO)	2.95

Table 5.4.11 Z_i^{dist} statistic for various distributions for Mahanadi Subzone 3 (d)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Generalized Normal (GNO)	0.22
2.	Pearson Type III (PE3)	-0.62
3.	Generalized Extreme Value (GEV)	0.66
4.	Generalized logistic (GLO)	2.08
5.	Generalized Pareto (GPA)	-2.68

Table 5.4.12 Z_i^{dist} statistic for various distributions for Upper Godavari Subzone 3 (e)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Generalized Pareto (GPA)	-0.54
2.	Pearson Type III (PE3)	1.72
3.	Generalized Normal (GNO)	2.38
4.	Generalized Extreme Value (GEV)	2.68
5.	Generalized logistic (GLO)	4.10

Table 5.4.13 Z_i^{dist} statistic for various distributions for Lower Godavari Subzone 3(f)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Pearson Type III (PE3)	0.49
2.	Generalized Normal (GNO)	1.28
3.	Generalized Extreme Value (GEV)	1.60
4.	Generalized Pareto (GPA)	-2.94
5.	Generalized logistic (GLO)	3.64

Table 5.4.14 Z_i^{dist} statistic for various distributions for Krishna and Pennar Subzone 3(h)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Generalized Pareto (GPA)	-0.95
2.	Pearson Type III (PE3)	0.99
3.	Generalized Normal (GNO)	2.17
4.	Generalized Extreme Value (GEV)	2.82
5.	Generalized logistic (GLO)	4.37

Table 5.4.15 Z_i^{dist} statistic for various distributions for Kaveri Basin Subzone 3(i)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Pearson Type III (PE3)	1.13
2.	Generalized Pareto (GPA)	2.13
3.	Generalized Normal (GNO)	2.74
4.	Generalized Extreme Value (GEV)	3.69
5.	Generalized logistic (GLO)	4.11

Table 5.4.16 Z_i^{dist} statistic for various distributions for East Coast Subzones 4 (b)

S. No.	Distribution	Z_i^{dist} –statistic
1.	Pearson Type III (PE3)	-0.30
2.	Generalized Pareto (GPA)	1.30
3.	Generalized Normal (GNO)	1.36
4.	Generalized Extreme Value (GEV)	2.34
5.	Generalized logistic (GLO)	2.53

Table 5.4.17 Z_i^{dist} statistic for various distributions for Sub-Himalayan region Zone 7

S. No.	Distribution	Z_i^{dist} –statistic
1.	Generalized logistic (GLO)	0.19
2.	Generalized Extreme Value (GEV)	-0.91
3.	Generalized Normal (GNO)	-1.02
4.	Pearson Type III (PE3)	-1.35
5.	Generalized Pareto (GPA)	-3.29

Table 5.5 Robust identified distributions for 17 Subzones and their Z_i^{dist} statistic

S. No.	Subzone	Distribution	Z_i^{dist} –statistic
1	1 (b)	PE3	0.01
2	1 (d)	GEV	0.13
3	1 (e)	GPA	-0.30
4	1 (f)	GEV	0.01
5	1 (g)	GEV	0.27
6	2 (a)	PE3	0.68
7	2 (b)	GNO	0.08
8	3 (a)	PE3	-0.06
9	3 (b)	GNO	0.32
10	3 (c)	PE3	0.05
11	3 (d)	GNO	0.22
12	3 (e)	GPA	-0.54
13	3 (f)	PE3	0.49
14	3 (h)	GPA	0.95
15	3 (i)	PE3	1.13
16	4 (b)	PE3	-0.30
17	Zone 7	GLO	0.19

5.5 DEVELOPMENT OF REGIONAL FLOOD FREQUENCY RELATIONSHIPS USING L-MOMENTS FOR GAUGED CATCHMENTS

For estimation of floods of various return periods for gauged catchments regional flood frequency relationships have been developed for the 17 Subzones based on the robust identified distribution. The values of regional parameters for the various distributions which have and $|Z_i^{dist}|$ –statistic values less than 1.64 as well as the five parameter Wakeby distribution are given in Table 5.6.1 to 5.6.17. The frequency distribution exhibiting lowest value of $|Z_i^{dist}|$ –statistic among the various distributions is given in the first row and the second lowest in the second row and so on. The regional parameters of the Wakeby distribution have also been included in Table 5.6.1 to 5.6.17, because the Wakeby distribution has five parameters, more than most of the common distributions and it can attain a wider range of distributional shapes than can the common distributions. This makes the Wakeby distribution particularly useful for simulating artificial data for use in studying the robustness, under changes in distributional form of methods of data analysis. It is preferred to use Wakeby distribution for heterogeneous regions (Hosking and Wallis, 1997).

For the commonly used return periods the values of the growth factors ($Q_T/\overline{Q}$) estimated by various distributions having $|Z_i^{dist}|$ –statistic less than 1.64 for each of the 17 Subzones are given in Tables 5.7.1 to 5.5.17. Here, (Q_T) is the value of flood for T-year return period and ($\overline{Q}$) is the mean annual peak flood of the gauged catchment. In the Tables 5.7.1 to 5.7.17 the growth factor values are given in first row for the best fit distribution i.e. distribution having lowest value of $|Z_i^{dist}|$ statistics as compared to other distributions. In the second row for the second best identified distribution and so on. The values of growth factors for the robust identified distributions for the 17 Subzones are summarized in Table 5.8.

For estimation of flood of desired return period (Q_T) for a gauged catchment of a Subzone, the growth factor (Q_T / $\overline{Q}$) value of the corresponding return period of the respective Subzone is to be multiplied by the mean annual peak flood ($\overline{Q}$) of the gauged catchment. Floods of various return periods may also be computed by substituting the values of the parameters of the robust identified distribution of each of the Subzones into the equation of the robust distribution for the respective Subzone and multiplying it by the mean annual peak flood of a gauged catchment.

Table 5.6.1 Regional parameters for various distributions for Chambal Subzone 1 (b)

Distribution	Parameters of the Distribution				
PE3	μ = 1.000	σ = 0.875	γ = 1.477		
GNO	ξ = 0.801	α = 0.734	k = -0.509		
GPA	ξ = -0.021	α = 1.237	k = 0.212		
GEV	ξ = 0.584	α = 0.592	k = -0.114		
WAK	ξ = -0.074	α = 0.990	β = 1.631	γ = 0.663	δ = 0.050

Table 5.6.2 Regional parameters for various distributions for Sone Subzone 1 (d)

Distribution	Parameters of the Distribution				
GEV	$\xi = 0.597$	$\alpha = 0.439$	$k = -0.260$		
GNO	$\xi = 0.754$	$\alpha = 0.584$	$k = -0.734$		
GLO	$\mu = 0.777$	$\alpha = 0.335$	$k = -0.348$		
GPA	$\xi = 0.188$	$\alpha = 0.786$	$k = -0.033$		
WAK	$\xi = 0.082$	$\alpha = 1.516$	$\beta = 7.077$	$\gamma = 0.628$	$\delta = 0.139$

Table 5.6.3 Regional parameters for various distributions for Upper Indo-Ganga Plains Subzone 1 (e)

Distribution	Parameters of the Distribution				
GPA	$\xi = -0.034$	$\alpha = 0.968$	$k = -0.064$		
PE3	$\mu = 1.000$	$\sigma = 1.091$	$\gamma = 2.176$		
GNO	$\xi = 0.670$	$\alpha = 0.741$	$k = -0.766$		
WAK	$\xi = -0.034$	$\alpha = 0.000$	$\beta = 0.000$	$\gamma = 0.968$	$\delta = 0.064$

Table 5.6.4 Regional parameters for various distributions for Middle Ganga Plains Subzone 1(f)

Distribution	Parameters of the Distribution				
GEV	$\xi = 0.734$	$\alpha = 0.468$	$k = 0.010$		
GNO	$\xi = 0.906$	$\alpha = 0.544$	$k = -0.337$		
PE3	$\mu = 1.000$	$\sigma = 0.588$	$\gamma = 0.994$		
GLO	$\mu = 0.915$	$\alpha = 0.308$	$k = -0.164$		
WAK	$\xi = 0.109$	$\alpha = 1.708$	$\beta = 2.525$	$\gamma = 0.362$	$\delta = 0.108$

Table 5.6.5 Regional parameters for various distributions for Lower Ganga Plains Subzone 1(g)

Distribution	Parameters of the Distribution				
GEV	$\xi = 0.610$	$\alpha = 0.512$	$k = -0.159$		
GNO	$\xi = 0.797$	$\alpha = 0.649$	$k = -0.576$		
GLO	$\mu = 0.816$	$\alpha = 0.370$	$k = -0.276$		
PE3	$\mu = 1.000$	$\sigma = 0.812$	$\gamma = 1.661$		
WAK	$\xi = 0.028$	$\alpha = 1.173$	$\beta = 1.650$	$\gamma = 0.407$	$\delta = 0.232$

Table 5.6.6 Regional parameters for various distributions for North Brahmaputra Subzone 2(a)

Distribution	Parameters of the Distribution				
PE3	$\mu = 1.000$	$\sigma = 0.646$	$\gamma = 1.207$		
GNO	$\xi = 0.876$	$\alpha = 0.575$	$k = -0.412$		
WAK	$\xi = 0.099$	$\alpha = 1.312$	$\beta = 5.562$	$\gamma = 0.824$	$\delta = -0.175$

Table 5.6.7 Regional parameters for various distributions for South Brahmaputra Subzone 2(b)

Distribution	Parameters of the Distribution				
GNO	$\xi = 0.805$	$\alpha = 0.643$	k = -0.561		
GEV	$\xi = 0.619$	$\alpha = 0.510$	k = -0.149		
PE3	$\mu = 1.000$	$\sigma = 0.795$	$\gamma = 1.618$		
GLO	$\mu = 0.823$	$\alpha = 0.366$	k = -0.269		
WAK	$\xi = 0.002$	$\alpha = 1.461$	$\beta = 4.417$	$\gamma = 0.698$	$\delta = 0.041$

Table 5.6.8 Regional parameters for various distributions for Mahi and Sabarmati Subzone 3 (a)

Distribution	Parameters of the Distribution				
PE3	$\mu = 1.000$	$\sigma = 0.890$	$\gamma = 1.961$		
GPA	$\xi = 0.099$	$\alpha = 0.914$	k = 0.015		
GNO	$\xi = 0.748$	$\alpha = 0.651$	k = -0.686		
WAK	$\xi = 0.099$	$\alpha = 0.914$	$\beta = 0.015$	$\gamma = 0.000$	$\delta = 0.000$

Table 5.6.9 Regional parameters for various distributions for Lower Narmada and Tapi Subzone 3 (b)

Distribution	Parameters of the Distribution				
GNO	$\xi = 0.746$	$\alpha = 0.661$	k = -0.683		
PE3	$\mu = 1.000$	$\sigma = 0.902$	$\gamma = 1.950$		
GEV	$\xi = 0.564$	$\alpha = 0.504$	k = -0.228		
GPA	$\xi = 0.085$	$\alpha = 0.932$	k = 0.019		
WAK	$\xi = 0.027$	$\alpha = 0.783$	$\beta = 5.303$	$\gamma = 0.803$	$\delta = 0.054$

Table 5.6.10 Regional parameters for various distributions for Upper Narmada and Tapi Subzone 3 (c)

Distribution	Parameters of the Distribution				
PE3	$\mu = 1.000$	$\sigma = 0.684$	$\gamma = 1.394$		
GNO	$\xi = 0.852$	$\alpha = 0.585$	k = -0.479		
GEV	$\xi = 0.677$	$\alpha = 0.477$	k = -0.093		
WAK	$\xi = 0.040$	$\alpha = 2.773$	$\beta = 11.709$	$\gamma = 0.844$	$\delta = -0.137$

Table 5.6.11 Regional parameters for various distributions for Mahanadi Subzone 3 (d)

Distribution	Parameters of the Distribution				
GNO	$\xi = 0.870$	$\alpha = 0.548$	k = -0.451		
PE3	$\mu = 1.000$	$\sigma = 0.629$	$\gamma = 1.316$		
GEV	$\xi = 0.704$	$\alpha = 0.452$	k = -0.073		
WAK	$\xi = 0.100$	$\alpha = 1.985$	$\beta = 6.486$	$\gamma = 0.684$	$\delta = -0.078$

Table 5.6.12 Regional parameters for various distributions for Upper Godavari Subzone 3 (e)

Distribution	Parameters of the Distribution				
GPA	$\xi = 0.069$	$\alpha = 1.251$	$k = 0.344$		
WAK	$\xi = 0.048$	$\alpha = 0.656$	$\beta = 1.261$	$\gamma = 0.759$	$\delta = -0.148$

Table 5.6.13 Regional parameters for various distributions for Lower Godavari Subzone 3 (f)

Distribution	Parameters of the Distribution				
PE3	$\mu = 1.000$	$\sigma = 0.633$	$\gamma = 1.104$		
GNO	$\xi = 0.888$	$\alpha = 0.575$	$k = -0.376$		
GEV	$\xi = 0.709$	$\alpha = 0.487$	$k = -0.019$		
WAK	$\xi = 0.102$	$\alpha = 1.258$	$\beta = 2.720$	$\gamma = 0.591$	$\delta = -0.056$

Table 5.6.14 Regional parameters for various distributions for Krishna and Pennar Subzone 3 (h)

Distribution	Parameters of the Distribution				
GPA	$\xi = 0.050$	$\alpha = 1.161$	$k = 0.223$		
PE3	$\mu = 1.000$	$\sigma = 0.808$	$\gamma = 1.453$		
WAK	$\xi = 0.026$	$\alpha = 0.786$	$\beta = 0.940$	$\gamma = 0.549$	$\delta = 0.034$

Table 5.6.15 Regional parameters for various distributions for Kaveri Basin Subzone 3(i)

Distribution	Parameters of the Distribution				
PE3	$\mu = 1.000$	$\sigma = 1.356$	$\gamma = 2.522$		
WAK	$\xi = -0.158$	$\alpha = 0.000$	$\beta = 0.000$	$\gamma = 0.952$	$\delta = 0.178$

Table 5.6.16 Regional parameters for various distributions for East Coast Subzone 4 (b)

Distribution	Parameters of the Distribution				
PE3	$\mu = 1.000$	$\sigma = 1.584$	$\gamma = 3.145$		
GPA	$\xi = -0.127$	$\alpha = 0.733$	$k = -0.349$		
GNO	$\xi = 0.445$	$\alpha = 0.717$	$k = -1.115$		
WAK	$\xi = -0.127$	$\alpha = 0.000$	$\beta = 0.000$	$\gamma = 0.733$	$\delta = 0.349$

Table 5.6.17 Regional parameters for various distributions for Sub-Himalayan region Zone 7

Distribution	Parameters of the Distribution				
GLO	$\mu = 0.911$	$\alpha = 0.318$	$k = -0.165$		
GEV	$\xi = 0.725$	$\alpha = 0.483$	$k = 0.008$		
GNO	$\xi = 0.902$	$\alpha = 0.562$	$k = -0.340$		
PE3	$\mu = 1.000$	$\sigma = 0.608$	$\gamma = 1.002$		
WAK	$\xi = -0.028$	$\alpha = 3.230$	$\beta = 5.406$	$\gamma = 0.514$	$\delta = 0.020$

Table 5.7.1 Values of growth factors ($Q_T/\overline{Q}$) for Chambal Subzone 1 (b)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
PE3	0.793	2.167	2.873	3.392	3.901	4.403	5.059	5.550
GNO	0.801	2.127	2.874	3.460	4.070	4.709	5.599	6.310
GPA	0.777	2.233	2.865	3.267	3.615	3.915	4.249	4.463
GEV	0.805	2.102	2.869	3.493	4.164	4.888	5.935	6.802
WAK	0.805	2.136	2.844	3.395	3.963	4.551	5.360	5.996

Table 5.7.2 Values of growth factors ($Q_T/\overline{Q}$) for Sone Subzone 1(d)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
GEV	0.766	1.939	2.786	3.563	4.489	5.594	7.393	9.068
GNO	0.754	1.997	2.835	3.552	4.348	5.230	6.540	7.648
GLO	0.777	1.884	2.727	3.548	4.584	5.896	8.190	10.479
GPA	0.739	2.067	2.855	3.467	4.093	4.734	5.603	6.279
WAK	0.752	2.002	2.848	3.564	4.352	5.220	6.503	7.589

Table 5.7.3 Values of growth factors ($Q_T/\overline{Q}$) for Upper Indo-Ganga Plains Subzone 1 (e)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
GPA	0.652	2.367	3.426	4.269	5.150	6.071	7.353	8.374
PE3	0.643	2.403	3.440	4.232	5.028	5.828	6.889	7.693
GNO	0.670	2.284	3.400	4.366	5.449	6.659	8.472	10.019
WAK	0.652	2.367	3.426	4.269	5.150	6.071	7.353	8.374

Table 5.7.4 Values of growth factors ($Q_T/\overline{Q}$) for Middle Ganga Plains Subzone 1(f)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
GEV	0.906	1.776	2.209	2.527	2.840	3.151	3.557	3.862
GNO	0.906	1.777	2.203	2.516	2.826	3.136	3.549	3.864
PE3	0.904	1.788	2.200	2.493	2.775	3.048	3.400	3.659
GLO	0.915	1.728	2.197	2.589	3.023	3.505	4.231	4.857
WAK	0.929	1.731	2.180	2.549	2.947	3.375	3.993	4.503

Table 5.7.5 Values of growth factors ($Q_T/\overline{Q}$) for Lower Ganga Plains Subzone 1 (g)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
GEV	0.803	1.995	2.745	3.379	4.083	4.867	6.042	7.052
GNO	0.797	2.028	2.760	3.349	3.975	4.641	5.587	6.356
GLO	0.816	1.934	2.699	3.401	4.243	5.257	6.930	8.506
PE3	0.787	2.076	2.764	3.274	3.778	4.278	4.934	5.427
WAK	0.818	1.961	2.681	3.329	4.087	4.976	6.394	7.685

Table 5.7.6 Values of growth factors ($Q_T/\overline{Q}$) for North Brahmaputra Subzone 2 (a)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
PE3	0.873	1.866	2.350	2.699	3.038	3.370	3.799	4.118
GNO	0.876	1.848	2.353	2.735	3.122	3.517	4.052	4.470
WAK	0.868	1.895	2.361	2.666	2.937	3.176	3.451	3.632

Table 5.7.7 Values of growth factors ($Q_T/\overline{Q}$) for South Brahmaputra Subzone 2(b)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
GNO	0.805	2.011	2.718	3.285	3.884	4.518	5.416	6.143
GEV	0.811	1.981	2.707	3.315	3.985	4.726	5.828	6.767
PE3	0.796	2.055	2.721	3.214	3.701	4.182	4.814	5.288
GLO	0.823	1.920	2.663	3.340	4.148	5.117	6.704	8.192
WAK	0.808	2.019	2.735	3.295	3.871	4.463	5.273	5.905

Table 5.7.8 Values of growth factors ($Q_T/\overline{Q}$) for Mahi and Sabarmati Subzone 3(a)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
PE3	0.731	2.162	2.971	3.581	4.191	4.800	5.604	6.213
GPA	0.730	2.169	2.973	3.574	4.168	4.757	5.526	6.101
GNO	0.748	2.085	2.953	3.682	4.481	5.355	6.637	7.708
WAK	0.730	2.169	2.973	3.574	4.168	4.757	5.526	6.101

Table 5.7.9 Values of growth factors ($Q_T/\overline{Q}$) for Lower Narmada and Tapi Subzone 3 (b)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
GNO	0.746	2.101	2.978	3.713	4.518	5.398	6.687	7.763
PE3	0.729	2.178	2.995	3.612	4.227	4.842	5.653	6.267
GEV	0.757	2.047	2.938	3.736	4.664	5.749	7.470	9.031
GPA	0.727	2.186	2.997	3.601	4.198	4.787	5.553	6.125
WAK	0.738	2.144	2.998	3.674	4.375	5.103	6.109	6.903

Table 5.7.10 Values of growth factors ($Q_T/\overline{Q}$) for Upper Narmada and Tapi Subzone 3 (c)

Distri-bution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
PE3	0.847	1.483	2.454	2.848	3.234	3.614	4.108	4.477
GNO	0.852	1.458	2.455	2.896	3.352	3.825	4.478	4.997
GEV	0.854	1.445	2.454	2.921	3.416	3.942	4.691	5.300
WAK	0.834	1.495	2.473	2.832	3.158	3.455	3.807	4.045

Table 5.7.11 Values of growth factors ($Q_T/\overline{Q}$) for Mahanadi Subzone 3 (d)

Distri-bution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
GNO	0.870	1.821	2.331	2.723	3.125	3.538	4.105	4.552
PE3	0.866	1.843	2.213	2.683	3.028	3.366	3.806	4.134
GEV	0.872	1.809	2.332	2.745	3.175	3.627	4.260	4.767
WAK	0.865	1.848	2.353	2.712	3.052	3.374	3.774	4.058

Table 5.7.12 Values of growth factors ($Q_T/\overline{Q}$) for Upper Godavari Subzone 3(e)

Distri-bution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
GPA	0.841	2.059	2.504	2.759	2.960	3.118	3.277	3.368
WAK	0.852	2.022	2.504	2.820	3.103	3.357	3.654	3.854

Table 5.7.13 Values of growth factors ($Q_T/\overline{Q}$) for Lower Godavari Subzone 3 (f)

Distri-bution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
PE3	0.886	1.849	2.308	2.637	2.955	3.265	3.665	3.962
GNO	0.888	1.834	2.311	2.667	3.024	3.384	3.868	4.242
GEV	0.889	1.830	2.316	2.683	3.051	3.423	3.922	4.304
WAK	0.896	1.840	2.306	2.642	2.965	3.276	3.669	3.953

Table 5.7.14 Values of growth factors ($Q_T/\overline{Q}$) for Krishna & Pennar Subzone 3 (h)

Distri-bution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
GPA	0.796	2.142	2.718	3.082	3.393	3.660	3.956	4.142
PE3	0.812	2.079	2.727	3.203	3.669	4.129	4.728	5.177
WAK	0.812	2.082	2.691	3.140	3.591	4.047	4.663	5.141

Table 5.7.15 Values of growth factors ($Q_T/\overline{Q}$) for Kaveri Basin Subzone 3(i)

Distri-bution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
PE 3	0.509	2.692	4.069	5.140	6.227	7.328	8.796	9.916
WAK	0.544	2.552	3.980	5.228	6.639	8.236	10.674	12.802

Table 5.7.16 Values of growth factors ($Q_T/\overline{Q}$) for East Coast Subzones 4 (b)

Distri-bution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
PE3	0.363	2.833	4.608	6.032	7.503	9.009	11.039	12.598
GPA	0.448	2.466	4.235	6.005	8.259	11.132	16.170	21.209
GNO	0.445	2.488	4.335	6.158	8.416	11.179	15.742	19.995
WAK	0.448	2.466	4.235	6.005	8.259	11.132	16.170	21.209

Table 5.7.17 Values of growth factors ($Q_T/\overline{Q}$) for Sub-Himalayan region Zone 7

Distri-bution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
GLO	0.911	1.753	2.240	2.646	3.097	3.599	4.355	5.006
GEV	0.902	1.803	2.252	2.583	2.909	3.233	3.657	3.975
GNO	0.902	1.804	2.246	2.571	2.894	3.216	3.645	3.974
PE3	0.900	1.816	2.243	2.547	2.840	3.123	3.488	3.758
WAK	0.914	1.779	2.276	2.658	3.046	3.439	3.966	4.371

Table 5.8 Values of growth factors ($Q_T/\overline{Q}$) for robust distributions for 17 Subzones of India

S.No.	**Subzone**	**Return Period (Years)**							
		2	**10**	**25**	**50**	**100**	**200**	**500**	**1000**
1	1 (b)	0.793	2.167	2.873	3.392	3.901	4.403	5.059	5.550
2	1 (d)	0.766	1.939	2.786	3.563	4.489	5.594	7.393	9.068
3	1 (e)	0.652	2.367	3.426	4.269	5.150	6.071	7.353	8.374
4	1 (f)	0.906	1.776	2.209	2.527	2.840	3.151	3.557	3.862
5	1 (g)	0.803	1.995	2.745	3.379	4.083	4.867	6.042	7.052
6	2 (a)	0.873	1.866	2.350	2.699	3.038	3.370	3.799	4.118
7	2 (b)	0.805	2.011	2.718	3.285	3.884	4.518	5.416	6.143
8	3 (a)	0.731	2.162	2.971	3.581	4.191	4.800	5.604	6.213
9	3 (b)	0.746	2.101	2.978	3.713	4.518	5.398	6.687	7.763
10	3 (c)	0.847	1.483	2.454	2.848	3.234	3.614	4.108	4.477
11	3 (d)	0.870	1.821	2.331	2.723	3.125	3.538	4.105	4.552
12	3 (e)	0.841	2.059	2.504	2.759	2.960	3.118	3.277	3.368
13	3 (f)	0.886	1.849	2.308	2.637	2.955	3.265	3.665	3.962
14	3 (h)	0.796	2.142	2.718	3.082	3.393	3.660	3.956	4.142
15	3 (i)	0.509	2.692	4.069	5.140	6.227	7.328	8.796	9.916
16	4 (b)	0.363	2.833	4.608	6.032	7.503	9.009	11.039	12.598
17	Zone 7	0.911	1.753	2.240	2.646	3.097	3.599	4.355	5.006

5.6 DEVELOPMENT OF REGIONAL RELATIONSHIPS BETWEEN MEAN ANNUAL PEAK FLOODS AND CATCHMENT AREAS

For ungauged catchments the value of mean annual peak flood ($\overline{Q}$) i.e. the at-site mean cannot be estimated in absence of the observed streamflow data. Hence, relationships between the mean annual peak floods of gauged catchments in the region and their pertinent physiographic and climatic characteristics are needed for estimation of the mean annual peak floods for ungauged catchments. Therefore, the regional relationships have been developed in the form of a power law using the Levenberg-Marquardt (LM) iteration on the data of the mean annual peak floods and catchment areas in the following form for the 17 Subzones.

$$\overline{Q} = a*A^b \qquad (5.1)$$

Figs. 5.2.1 to 5.2.17 show the variation of mean annual peak floods and catchment areas along with the best fitted regional relationships developed by the Levenberg-Marquardt (LM) iteration technique for the 17 Subzones. The values of regional coefficients i.e. 'a' and 'b' as well as the statistical performance indices viz. EFF, CORR, RMSE and MAE for the developed regional relationships based on the Levenberg-Marquardt procedure for the 17 Subzones are given in Table 5.9.

Table 5.9 Regional coefficients and statistical performance indices for 17 Subzones

Sub-zone	a	b	CORR	EFF	RMSE	MAE
1 (b)	4.939	0.756	0.976	0.952	116.059	79.078
1 (d)	30.768	0.363	0.570	0.325	105.165	81.778
1 (e)	86.231	0.102	0.163	0.025	131.249	115.973
1 (f)	2.111	0.913	0.845	0.711	105.013	72.735
1 (g)	20.333	0.465	0.629	0.394	127.299	83.988
2 (a)	18.709	0.555	0.614	0.370	223.165	178.008
2 (b)	6.863	0.521	0.583	0.339	67.352	45.267
3 (a)	31.851	0.383	0.923	0.851	46.863	39.941
3 (b)	63.140	0.289	0.652	0.420	110.836	87.232
3 (c)	23.449	0.547	0.867	0.751	207.729	167.318
3 (d)	2.519	0.863	0.913	0.834	118.881	88.326
3 (e)	11.741	0.561	0.979	0.959	53.369	43.701
3 (f)	10.313	0.676	0.875	0.765	165.305	131.500
3 (h)	3.652	0.701	0.857	0.735	102.425	72.619
3 (i)	0.060	1.244	0.972	0.927	23.911	17.658
4 (b)	16.039	0.371	0.467	0.217	75.122	57.879
Zone 7	63.597	0.387	0.737	0.534	322.405	245.636

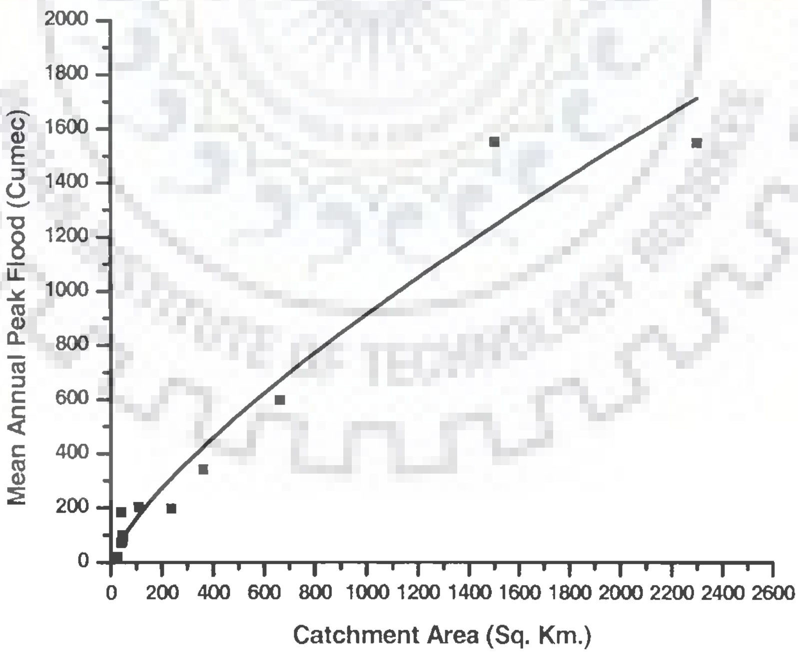


Fig. 5.2.1 Variation of mean annual peak floods with catchment area for various gauging sites of Chambal Subzone 1(b)

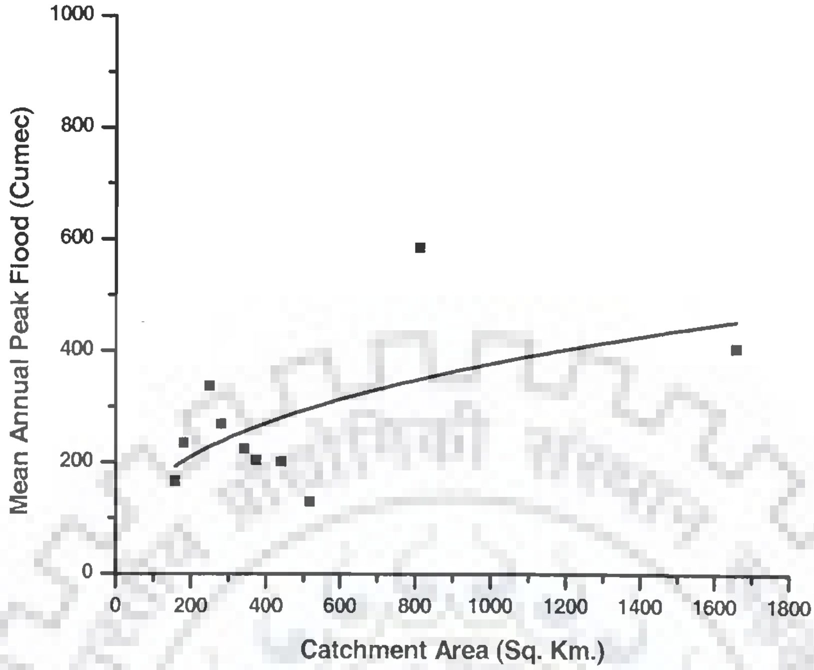


Fig. 5.2.2 Variation of mean annual peak floods with catchment area for various gauging sites of Sone Subzone 1(d)

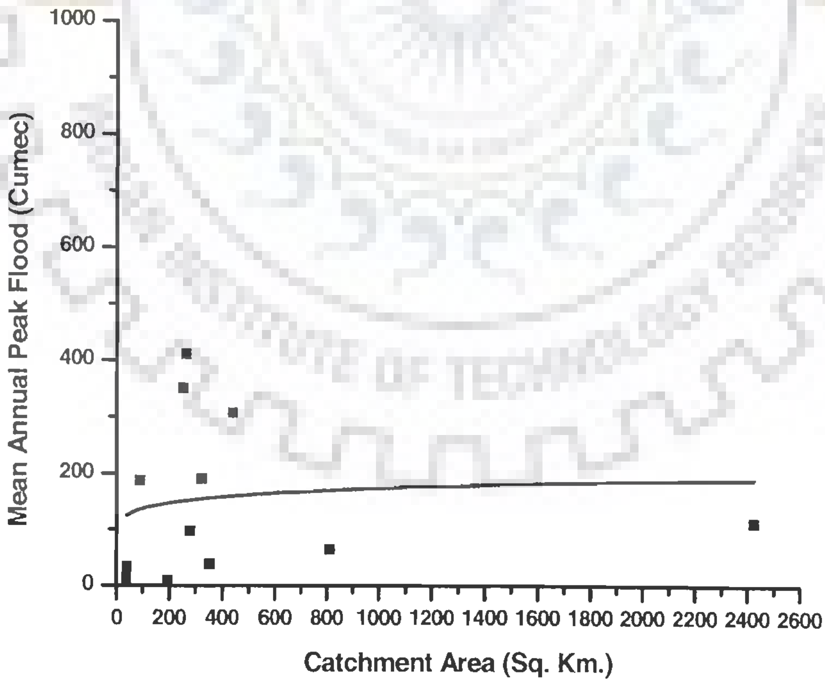


Fig. 5.2.3 Variation of mean annual peak floods with catchment area for various gauging sites of Upper Indo-Ganga Plains Subzone 1(e)

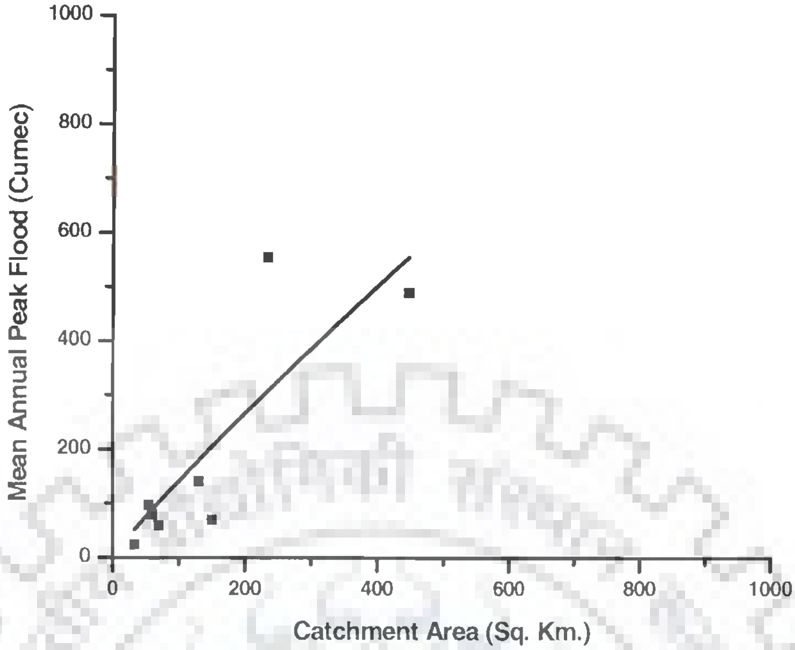


Fig. 5.2.4 Variation of mean annual peak floods with catchment area for various gauging sites of Middle Ganga Plains Subzone 1(f)

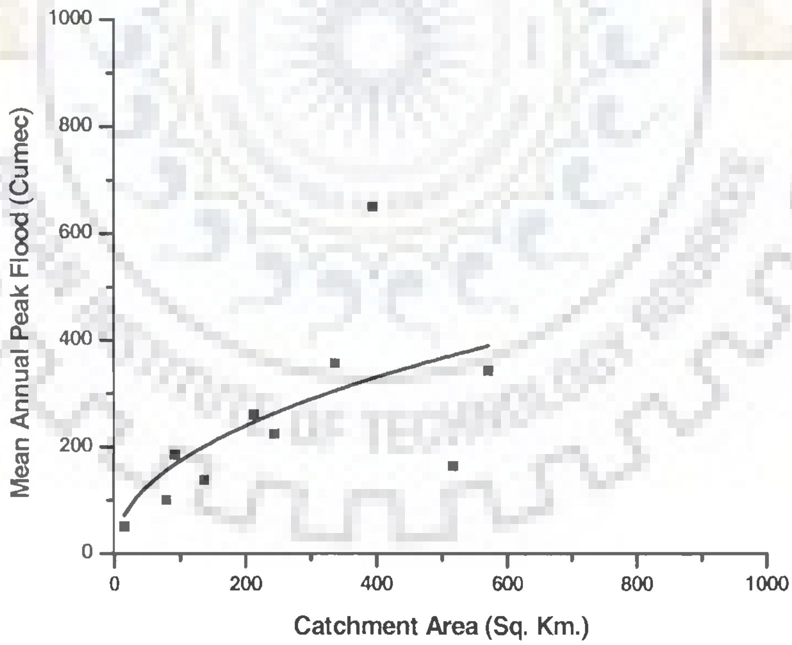


Fig. 5.2.5 Variation of mean annual peak floods with catchment area for various gauging sites of Lower Ganga Plains Subzone 1(g)

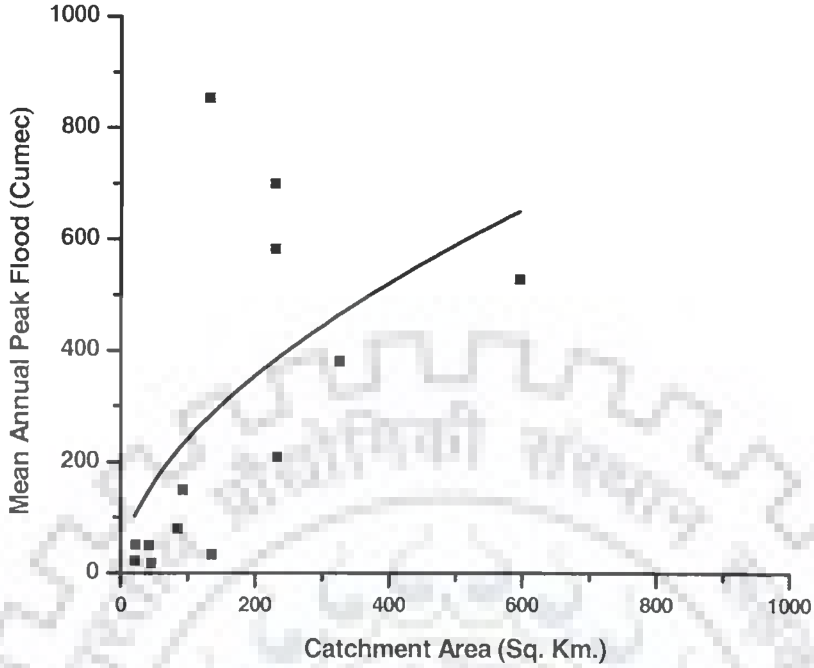


Fig. 5.2.6 Variation of mean annual peak floods with catchment area for various gauging sites of North Brahmaputra Subzone 2(a)

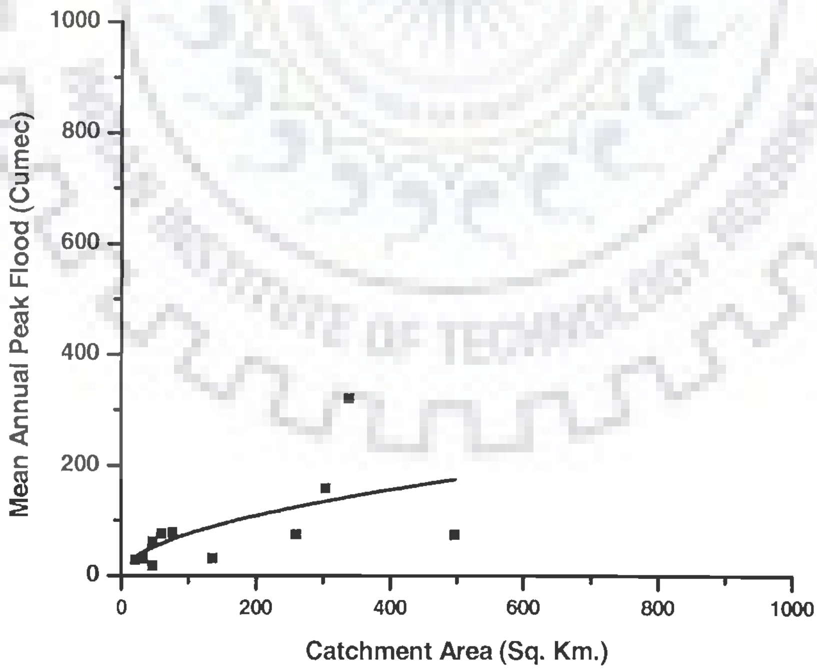


Fig. 5.2.7 Variation of mean annual peak floods with catchment area for various gauging sites of South Brahmaputra Subzone 2(b)

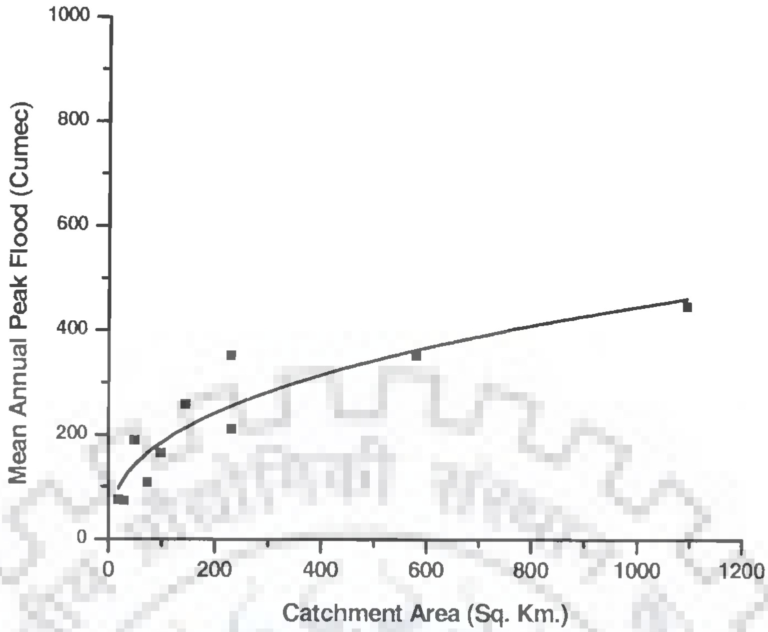


Fig. 5.2.8 Variation of mean annual peak floods with catchment area for various gauging sites of Mahi and Sabarmati Subzone 3(a)

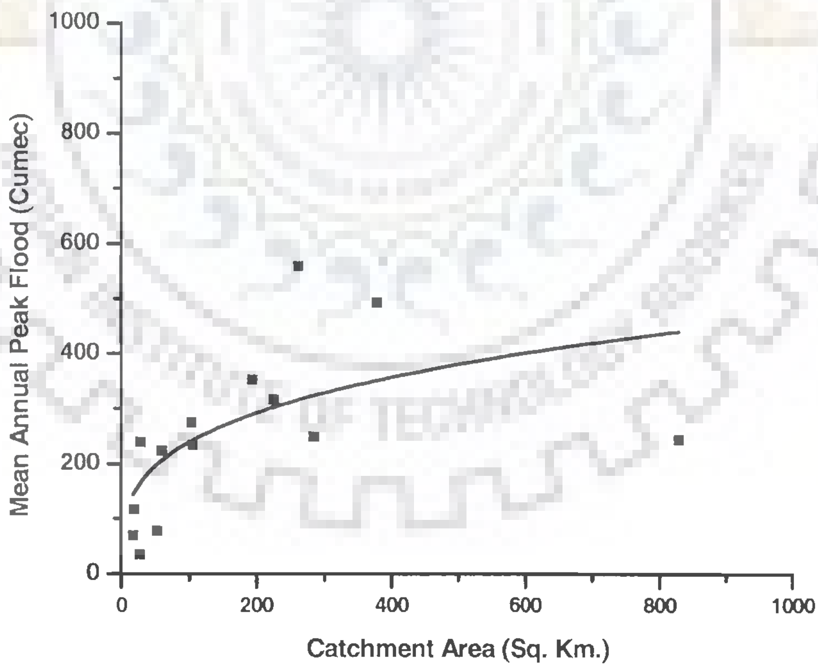


Fig. 5.2.9 Variation of mean annual peak floods with catchment area for various gauging sites of Lower Narmada and Tapi Subzone 3(b)

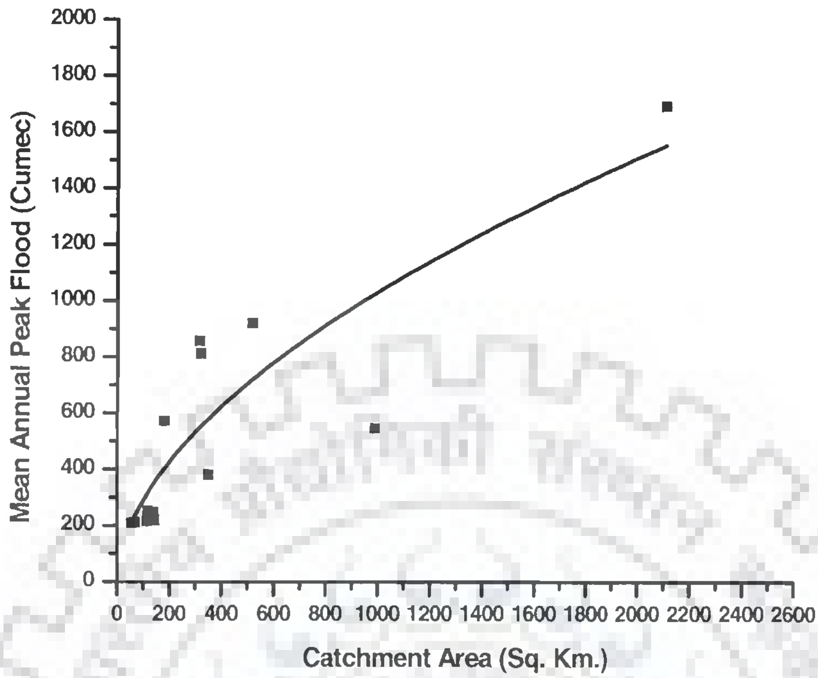


Fig. 5.2.10 Variation of mean annual peak floods with catchment area for various gauging sites of Upper Narmada and Tapi Subzone 3(c)

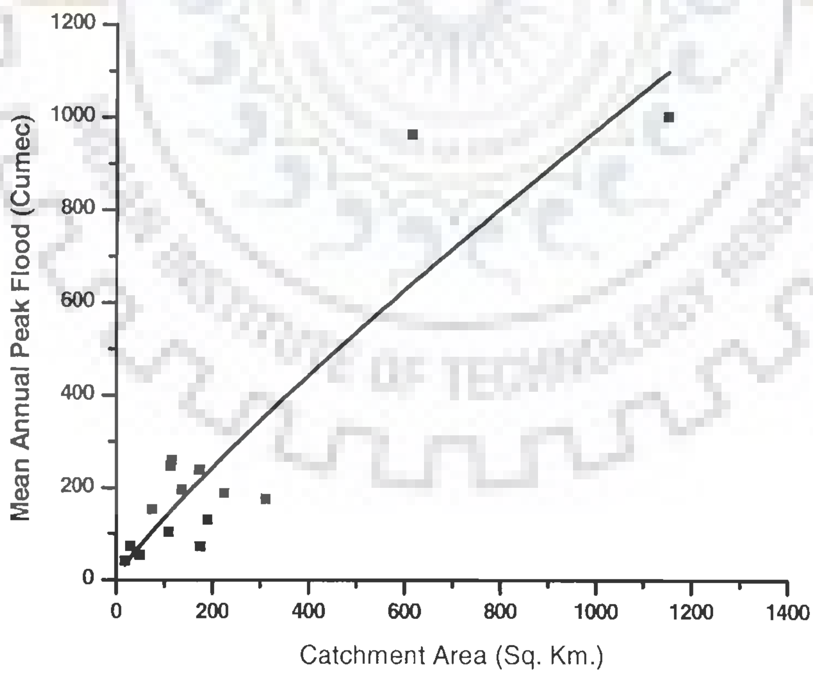


Fig. 5.2.11 Variation of mean annual peak floods with catchment area for various gauging sites of Mahanadi Subzone 3(d)

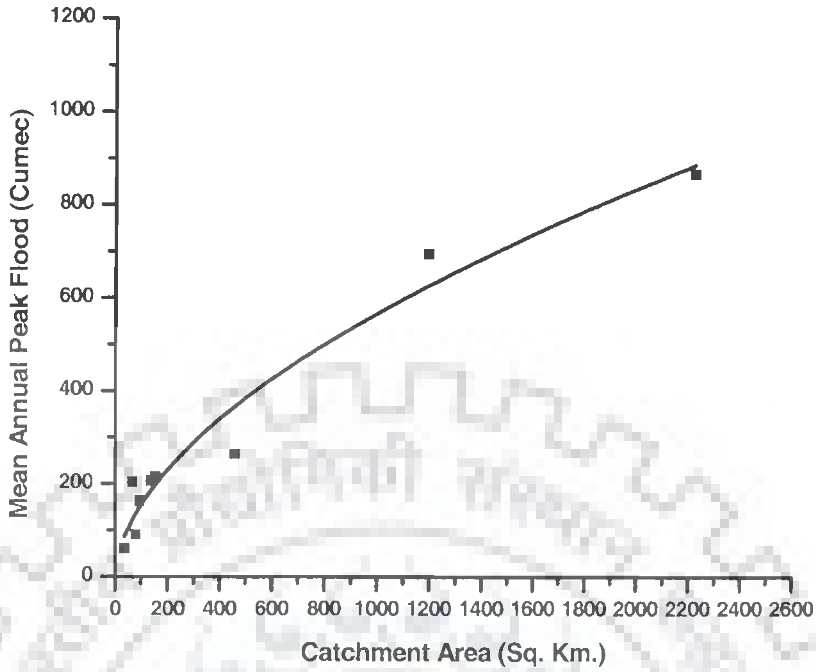


Fig. 5.2.12 Variation of mean annual peak floods with catchment area for various gauging sites of Upper Godavari Subzone 3(e)

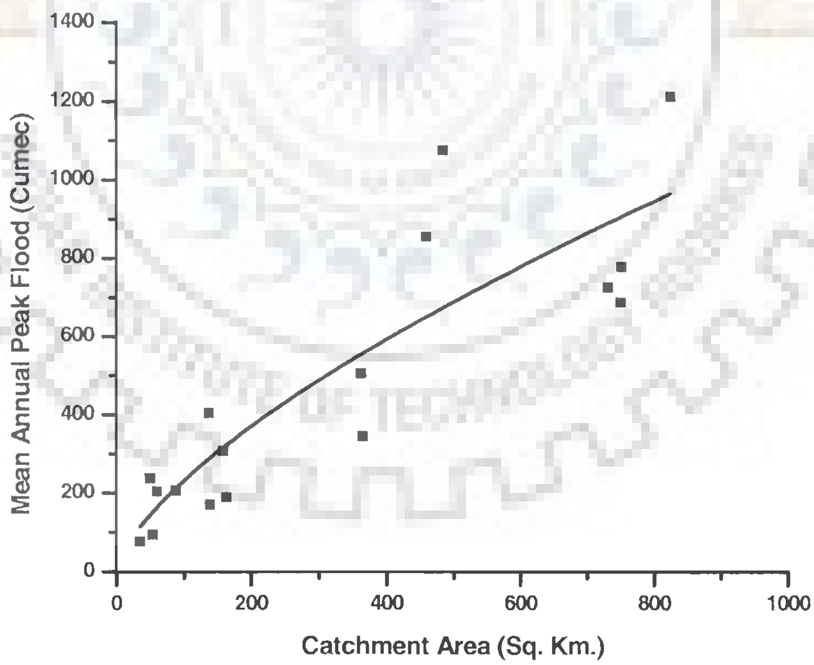


Fig. 5.2.13 Variation of mean annual peak floods with catchment area for various gauging sites of Lower Godavari Subzone 3(f)

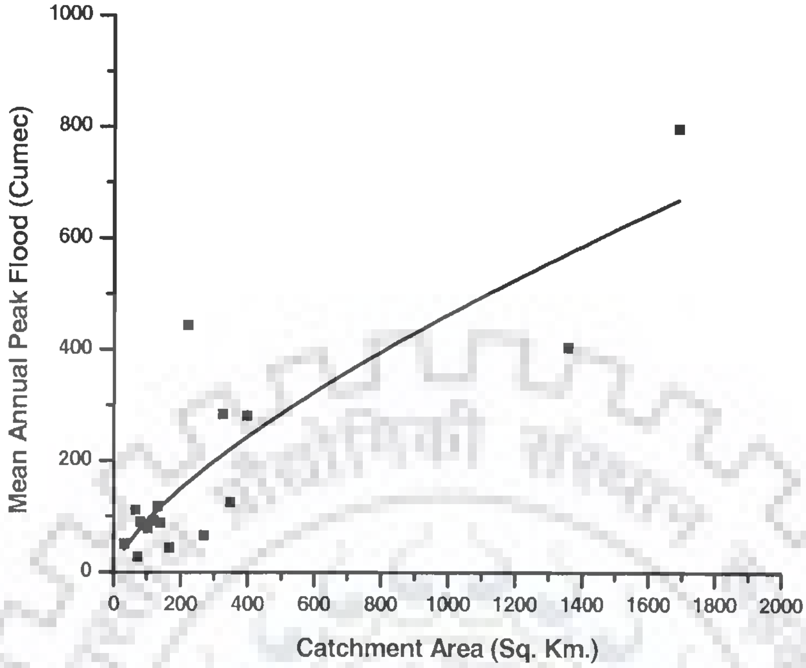


Fig. 5.2.14 Variation of mean annual peak floods with catchment area for various gauging sites of Krishna and Pennar Subzone 3(h)

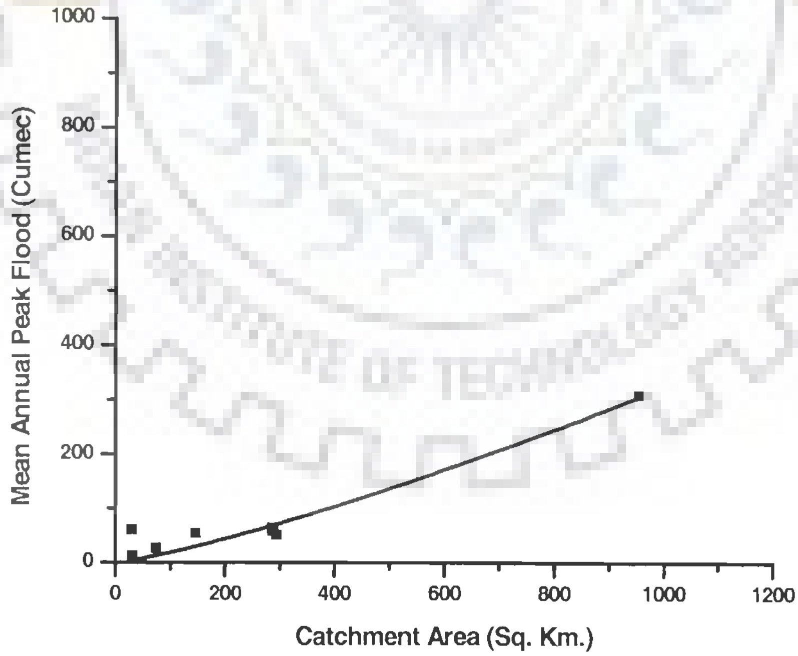


Fig. 5.2.15 Variation of mean annual peak floods with catchment area for various gauging sites of Kaveri Basin Subzone 3(i)

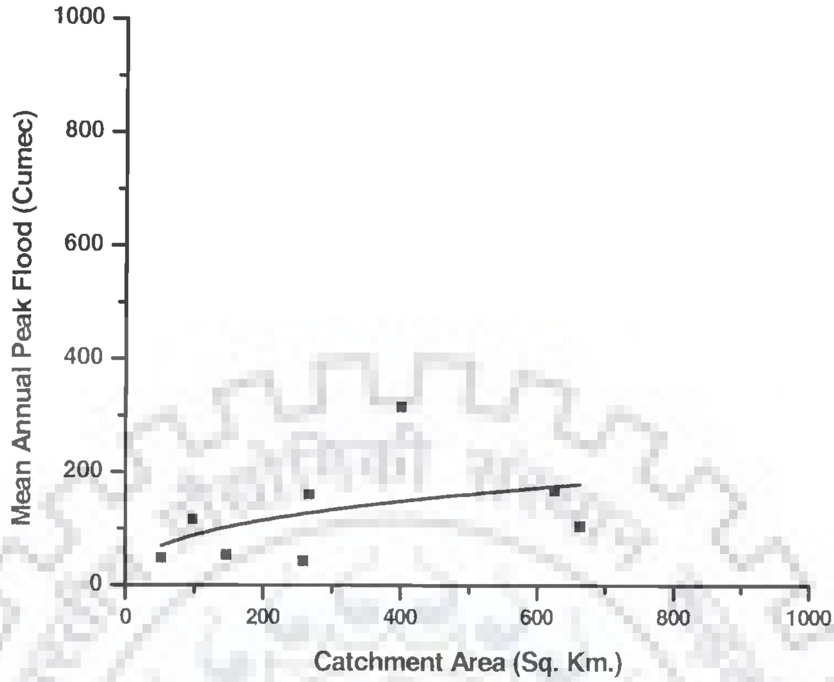


Fig. 5.2.16 Variation of mean annual peak floods with catchment area for various gauging sites of East Coast Subzone 4(b)

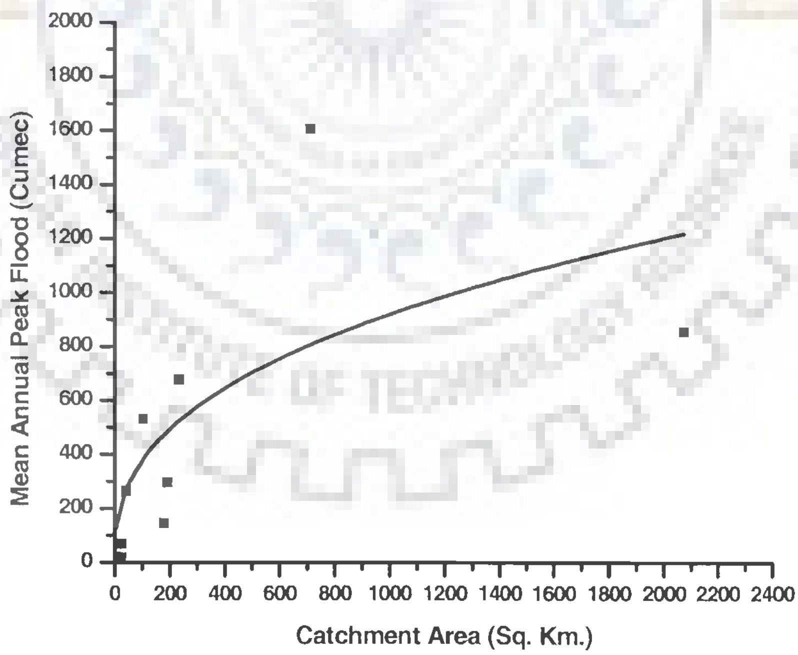


Fig. 5.2.17 Variation of mean annual peak floods with catchment area for various gauging sites of Sub-Himalayan region Zone-7

5.7 DEVELOPMENT OF REGIONAL FLOOD FREQUENCY RELATIONSHIP USING L-MOMENTS APPROACH FOR UNGAUGED CATCHMENTS

For development of regional flood frequency relationships for ungauged catchments, the regional flood frequency relationships developed for gauged catchments (growth factors given in Table 5.8) have been coupled with the regional relationships between mean annual peak floods and catchment areas of the respective Subzones (Table 5.9). In this manner the following form of regional flood frequency relationships have been developed for ungauged catchments of all the 17 Subzones.

$$Q_T = C_T * A^b \quad (5.2)$$

Where, Q_T is the flood estimate for an ungauged catchment in m^3/s for T year return period, A is the catchment area in km^2 and C_T is a regional coefficient. The values of regional coefficients (C_T) for some of the commonly used return periods and 'b' for the 17 Subzones are given in Table 5.10. The tabular and graphical forms of the regional flood frequency relationship developed in equation 5.2 has also been developed for the 17 Subzones and same are given in Tables 5.11.1 to 5.11.17 and Figs. 5.3.1 to 5.3.17.

For estimation of floods of commonly used returns periods for an ungauged catchments for a given catchment area the value of flood estimates may be directly obtained from the Tables 5.11.1 to 5.11.17 for the respective Subzones. The values of flood soft desire return periods for an ungauged catchments for given catchment area may also be obtained from the Figs. 5.3.1 to 5.3.17.

Table 5.10 Values of regional coefficients 'b' and 'C_T' for 17 Subzones of India

S. No	Sub-zone	Coeff. 'b'	Return Period (Years)							
			2	10	25	50	100	200	500	1000
			C_T for various Subzones							
1	1 (b)	0.756	3.917	10.703	14.190	16.753	19.267	21.746	24.986	27.411
2	1 (d)	0.363	23.568	59.659	85.720	109.626	138.118	172.116	227.468	279.004
3	1 (e)	0.102	56.223	204.109	295.427	368.120	444.090	523.508	634.057	722.098
4	1 (f)	0.913	1.913	3.749	4.663	5.334	5.995	6.652	7.509	8.153
5	1 (g)	0.465	16.327	40.564	55.814	68.705	83.020	98.961	122.852	143.388
6	2 (a)	0.555	16.333	34.911	43.966	50.496	56.838	63.049	71.075	77.044
7	2 (b)	0.521	5.525	13.801	18.654	22.545	26.656	31.007	37.170	42.159
8	3 (a)	0.383	23.283	68.862	94.629	114.058	133.488	152.885	178.493	197.890
9	3 (b)	0.289	47.102	132.657	188.031	234.439	285.267	340.830	422.217	490.156
10	3 (c)	0.547	19.861	34.775	57.544	66.783	75.834	84.745	96.328	104.981
11	3 (d)	0.863	2.192	4.587	5.872	6.859	7.872	8.912	10.340	11.466
12	3 (e)	0.561	9.874	24.175	29.399	32.393	34.753	36.608	38.475	39.544
13	3 (f)	0.676	9.137	19.069	23.802	27.195	30.475	33.672	37.797	40.860
14	3 (h)	0.701	2.907	7.823	9.926	11.255	12.391	13.366	14.447	15.127
15	3 (i)	1.244	0.031	0.162	0.244	0.308	0.374	0.440	0.528	0.595
16	4 (b)	0.371	5.822	45.438	73.908	96.747	120.341	144.495	177.055	202.059
17	Zone 7	0.387	57.937	111.486	142.457	168.278	196.960	228.886	276.965	318.367

Table 5.11.1 Variation of floods of various return periods with catchment area based on L-moments for Chambal Subzone 1 (b)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	22	61	81	96	110	124	142	156
20	38	103	137	161	186	209	241	264
50	75	206	273	322	371	419	481	528
100	127	348	461	545	626	707	812	891
200	215	588	779	920	1058	1194	1372	1505
300	292	798	1058	1250	1437	1622	1864	2045
400	363	992	1316	1553	1786	2016	2317	2542
500	430	1175	1557	1839	2115	2387	2742	3009
600	493	1348	1788	2110	2427	2740	3148	3453
700	554	1515	2009	2371	2727	3078	3537	3880
800	613	1676	2222	2623	3017	3405	3912	4292
900	670	1832	2429	2868	3298	3722	4277	4692
1000	726	1984	2630	3105	3571	4031	4631	5081
1100	780	2132	2827	3337	3838	4332	4977	5460
1200	833	2277	3019	3564	4099	4627	5316	5832
1300	885	2419	3207	3787	4355	4915	5647	6196
1400	936	2558	3392	4005	4606	5198	5973	6553
1500	986	2695	3574	4219	4852	5477	6293	6903
1600	1036	2830	3752	4430	5095	5751	6607	7249
1700	1084	2963	3928	4638	5334	6020	6917	7589
1800	1132	3094	4102	4843	5569	6286	7223	7924
1900	1179	3223	4273	5045	5802	6548	7524	8254
2000	1226	3350	4442	5244	6031	6807	7821	8581
2500	1451	3966	5258	6208	7139	8058	9259	10157
3000	1666	4552	6035	7125	8195	9249	10627	11658
3500	1872	5115	6781	8006	9207	10392	11941	13099
4000	2071	5658	7501	8856	10185	11496	13209	14491
4500	2264	6185	8200	9681	11134	12567	14439	15840
5000	2451	6698	8880	10484	12057	13609	15636	17154

Table 5.11.2 Variation of floods of various return periods with catchment area based on L-moments for Sone Subzone 1 (d)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	54	138	198	253	319	397	525	644
20	70	177	254	325	410	511	675	828
50	98	247	355	454	571	712	941	1154
100	125	317	456	583	735	916	1210	1485
200	161	408	587	750	945	1178	1557	1909
300	187	473	680	869	1095	1365	1804	2212
400	207	525	754	965	1216	1515	2002	2456
500	225	569	818	1046	1318	1643	2171	2663
600	240	608	874	1118	1408	1755	2320	2845
700	254	643	924	1182	1489	1856	2453	3009
800	267	675	970	1241	1563	1948	2575	3158
900	278	705	1013	1295	1632	2033	2687	3296
1000	289	732	1052	1346	1695	2113	2792	3425
1100	299	758	1089	1393	1755	2187	2890	3545
1200	309	782	1124	1438	1811	2257	2983	3659
1300	318	805	1157	1480	1865	2324	3071	3767
1400	327	827	1189	1520	1916	2387	3155	3870
1500	335	848	1219	1559	1964	2448	3235	3968
1600	343	869	1248	1596	2011	2506	3311	4062
1700	351	888	1276	1631	2055	2561	3385	4152
1800	358	906	1302	1666	2099	2615	3456	4239
1900	365	924	1328	1699	2140	2667	3525	4323
2000	372	942	1353	1731	2180	2717	3591	4404
2500	403	1021	1467	1877	2364	2946	3894	4776
3000	431	1091	1568	2005	2526	3148	4160	5103
3500	456	1154	1658	2120	2671	3329	4400	5397
4000	479	1211	1740	2226	2804	3494	4618	5665
4500	499	1264	1816	2323	2927	3647	4820	5912
5000	519	1313	1887	2413	3041	3789	5008	6142

Table 5.11.3 Variation of floods of various return periods with catchment area based on L-moments for Upper Indo-Ganga Plains Subzone 1(e)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	71	258	374	466	562	662	802	913
20	76	277	401	500	603	711	861	980
50	84	304	440	549	662	780	945	1076
100	90	326	473	589	710	837	1014	1155
200	97	350	507	632	762	899	1089	1240
300	101	365	529	659	795	937	1134	1292
400	104	376	544	678	818	965	1168	1330
500	106	385	557	694	837	987	1195	1361
600	108	392	567	707	853	1005	1218	1387
700	110	398	576	718	866	1021	1237	1409
800	111	404	584	728	878	1035	1254	1428
900	113	409	591	737	889	1048	1269	1445
1000	114	413	598	745	898	1059	1283	1461
1100	115	417	603	752	907	1069	1295	1475
1200	116	421	609	759	915	1079	1307	1488
1300	117	424	614	765	923	1088	1317	1500
1400	118	427	619	771	930	1096	1327	1512
1500	119	430	623	776	936	1104	1337	1523
1600	119	433	627	781	943	1111	1346	1533
1700	120	436	631	786	948	1118	1354	1542
1800	121	438	635	791	954	1125	1362	1551
1900	121	441	638	795	959	1131	1369	1560
2000	122	443	641	799	964	1137	1377	1568
2500	125	453	656	818	986	1163	1408	1604
3000	127	462	669	833	1005	1185	1435	1634
3500	129	469	679	846	1021	1203	1458	1660
4000	131	476	688	858	1035	1220	1478	1683
4500	133	481	697	868	1047	1235	1495	1703
5000	134	487	704	878	1059	1248	1512	1721

Table 5.11.4 Variation of floods of various return periods with catchment area based on L-moments for Middle Ganga Plains Subzone 1(f)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	16	31	38	44	49	54	61	67
20	29	58	72	82	92	103	116	126
50	68	133	166	190	213	237	267	290
100	128	251	312	357	402	446	503	546
200	241	473	588	673	756	839	947	1028
300	349	685	852	974	1095	1215	1371	1489
400	454	890	1107	1267	1424	1580	1783	1936
500	557	1092	1358	1553	1746	1937	2186	2374
600	658	1289	1604	1834	2062	2287	2582	2804
700	757	1484	1846	2112	2373	2633	2972	3227
800	855	1677	2085	2386	2681	2975	3358	3646
900	952	1867	2322	2656	2985	3312	3739	4060
1000	1048	2055	2557	2925	3287	3647	4117	4470
1100	1144	2242	2789	3190	3586	3978	4491	4876
1200	1238	2428	3020	3454	3882	4307	4862	5279
1300	1332	2612	3248	3716	4176	4634	5231	5679
1400	1425	2795	3476	3976	4469	4958	5597	6077
1500	1518	2976	3702	4235	4759	5280	5961	6472
1600	1610	3157	3927	4492	5048	5601	6323	6865
1700	1702	3337	4150	4747	5335	5920	6683	7256
1800	1793	3515	4372	5002	5621	6237	7040	7644
1900	1884	3693	4594	5255	5906	6552	7397	8031
2000	1974	3870	4814	5507	6189	6867	7751	8416
2500	2420	4745	5902	6751	7587	8418	9503	10318
3000	2858	5604	6970	7974	8962	9943	11224	12187
3500	3290	6451	8024	9179	10316	11446	12920	14028
4000	3717	7288	9064	10369	11653	12930	14596	15847
4500	4139	8115	10093	11546	12976	14397	16253	17646
5000	4557	8934	11112	12712	14287	15851	17894	19428

Table 5.11.5 Variation of floods of various return periods with catchment area based on L-moments for Lower Ganga Plains Subzone 1(g)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	48	118	163	200	242	289	358	418
20	66	163	225	277	334	399	495	577
50	101	250	344	424	512	610	758	884
100	139	345	475	585	707	842	1046	1220
200	192	477	656	807	975	1163	1443	1685
300	232	575	792	975	1178	1404	1743	2034
400	265	658	905	1114	1346	1605	1992	2325
500	294	730	1004	1236	1493	1780	2210	2580
600	320	794	1093	1345	1626	1938	2406	2808
700	343	853	1174	1445	1746	2082	2584	3016
800	365	908	1249	1538	1858	2215	2750	3210
900	386	959	1320	1624	1963	2340	2905	3390
1000	405	1007	1386	1706	2061	2457	3051	3561
1100	424	1053	1449	1783	2155	2569	3189	3722
1200	441	1096	1509	1857	2244	2675	3320	3876
1300	458	1138	1566	1927	2329	2776	3446	4023
1400	474	1178	1621	1995	2411	2874	3567	4164
1500	490	1216	1674	2060	2489	2967	3684	4299
1600	504	1253	1724	2123	2565	3058	3796	4430
1700	519	1289	1774	2183	2638	3145	3904	4557
1800	533	1324	1822	2242	2709	3230	4009	4680
1900	546	1358	1868	2299	2778	3312	4112	4799
2000	560	1390	1913	2355	2846	3392	4211	4915
2500	621	1542	2122	2612	3157	3763	4671	5452
3000	676	1679	2310	2843	3436	4096	5084	5934
3500	726	1804	2482	3055	3691	4400	5462	6375
4000	772	1919	2641	3250	3928	4682	5812	6784
4500	816	2027	2789	3433	4149	4945	6139	7166
5000	857	2129	2929	3606	4357	5194	6448	7526

Table 5.11.6 Variation of floods of various return periods with catchment area based on L-moments for North Brahmaputra Subzone 2(a)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	59	125	158	181	204	226	255	277
20	86	184	232	266	300	332	375	406
50	143	306	386	443	498	553	623	676
100	210	450	566	651	732	812	916	993
200	309	661	832	956	1076	1193	1345	1458
300	387	828	1042	1197	1347	1494	1685	1826
400	454	971	1223	1404	1580	1753	1976	2142
500	514	1099	1384	1589	1789	1984	2237	2425
600	569	1216	1531	1758	1979	2196	2475	2683
700	620	1324	1668	1916	2156	2392	2696	2923
800	667	1426	1796	2063	2322	2576	2904	3147
900	712	1523	1917	2202	2479	2750	3100	3360
1000	755	1614	2033	2335	2628	2915	3286	3562
1100	796	1702	2143	2462	2771	3074	3465	3756
1200	836	1786	2249	2583	2908	3226	3636	3942
1300	874	1867	2352	2701	3040	3372	3802	4121
1400	910	1946	2450	2814	3168	3514	3961	4294
1500	946	2022	2546	2924	3291	3651	4116	4461
1600	980	2095	2639	3031	3411	3784	4266	4624
1700	1014	2167	2729	3134	3528	3914	4412	4782
1800	1047	2237	2817	3235	3642	4040	4554	4937
1900	1078	2305	2903	3334	3753	4163	4693	5087
2000	1110	2372	2987	3430	3861	4283	4828	5234
2500	1256	2684	3381	3883	4370	4848	5465	5924
3000	1390	2970	3741	4296	4836	5364	6047	6555
3500	1514	3235	4075	4680	5268	5843	6587	7140
4000	1630	3484	4388	5040	5673	6293	7094	7689
4500	1740	3720	4684	5380	6056	6718	7573	8209
5000	1845	3944	4967	5704	6421	7122	8029	8703

Table 5.11.7 Variation of floods of various return periods with catchment area based on L-moments for South Brahmaputra Subzone 2(b)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	18	46	62	75	88	103	123	140
20	26	66	89	107	127	148	177	201
50	42	106	143	173	205	238	285	324
100	61	152	205	248	294	342	409	464
200	87	218	295	356	421	490	588	666
300	108	269	364	440	520	605	726	823
400	125	313	423	511	605	703	843	956
500	141	352	475	574	679	790	947	1074
600	155	387	523	632	747	869	1041	1181
700	168	419	566	684	809	941	1128	1280
800	180	449	607	734	868	1009	1210	1372
900	191	478	646	780	922	1073	1286	1459
1000	202	505	682	824	975	1134	1359	1541
1100	212	530	717	866	1024	1191	1428	1620
1200	222	555	750	906	1072	1247	1494	1695
1300	232	578	782	945	1117	1300	1558	1767
1400	241	601	813	982	1161	1351	1619	1837
1500	250	623	842	1018	1204	1400	1679	1904
1600	258	645	871	1053	1245	1448	1736	1969
1700	266	665	899	1087	1285	1495	1792	2032
1800	274	685	926	1120	1324	1540	1846	2094
1900	282	705	953	1152	1362	1584	1899	2153
2000	290	724	979	1183	1398	1627	1950	2212
2500	326	813	1099	1329	1571	1827	2190	2484
3000	358	894	1209	1461	1727	2009	2409	2732
3500	388	969	1310	1583	1872	2177	2610	2960
4000	416	1039	1404	1697	2007	2334	2798	3174
4500	442	1105	1493	1805	2134	2482	2975	3375
5000	467	1167	1577	1906	2254	2622	3143	3565

Table 5.11.8 Variation of floods of various return periods with catchment area based on L-moments for Mahi and Sabarmati Subzone 3(a)

Catchment Area (km²)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m³/s)							
10	56	166	229	276	322	369	431	478
20	73	217	298	359	420	482	562	623
50	104	308	423	510	597	684	799	885
100	136	402	552	665	779	892	1041	1155
200	177	524	720	868	1016	1163	1358	1506
300	207	612	841	1014	1186	1359	1586	1759
400	231	683	939	1132	1324	1517	1771	1963
500	252	744	1023	1233	1443	1652	1929	2139
600	270	798	1097	1322	1547	1772	2068	2293
700	286	847	1163	1402	1641	1879	2194	2433
800	301	891	1224	1476	1727	1978	2309	2560
900	315	932	1281	1544	1807	2069	2416	2679
1000	328	970	1334	1607	1881	2155	2515	2789
1100	340	1007	1383	1667	1951	2235	2609	2893
1200	352	1041	1430	1724	2017	2310	2697	2991
1300	363	1073	1475	1777	2080	2382	2781	3084
1400	373	1104	1517	1828	2140	2451	2861	3172
1500	383	1134	1558	1877	2197	2517	2938	3257
1600	393	1162	1597	1924	2252	2580	3012	3339
1700	402	1189	1634	1970	2305	2640	3082	3417
1800	411	1215	1670	2013	2356	2699	3151	3493
1900	420	1241	1705	2055	2405	2755	3217	3566
2000	428	1266	1739	2096	2453	2810	3280	3637
2500	466	1378	1894	2283	2672	3060	3573	3961
3000	500	1478	2031	2448	2865	3282	3831	4248
3500	530	1568	2155	2597	3040	3481	4064	4506
4000	558	1650	2268	2733	3199	3664	4278	4743
4500	584	1726	2372	2860	3347	3833	4475	4961
5000	608	1798	2470	2977	3485	3991	4659	5166

Table 5.11.9 Variation of floods of various return periods with catchment area based on L-moments for Lower Narmada and Tapi Subzone 3(b)

Catchment Area (km²)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m³/s)							
10	92	258	366	456	555	663	821	954
20	112	315	447	557	678	810	1004	1165
50	146	411	582	726	884	1056	1308	1518
100	178	502	712	887	1080	1290	1598	1855
200	218	613	869	1084	1319	1576	1952	2266
300	245	690	977	1219	1483	1772	2195	2548
400	266	749	1062	1324	1612	1925	2385	2769
500	284	799	1133	1413	1719	2054	2544	2953
600	299	843	1194	1489	1812	2165	2682	3113
700	313	881	1249	1557	1894	2263	2804	3255
800	325	916	1298	1618	1969	2352	2914	3383
900	336	947	1343	1674	2037	2434	3015	3500
1000	347	977	1384	1726	2100	2509	3108	3609
1100	356	1004	1423	1774	2159	2579	3195	3709
1200	366	1029	1459	1819	2214	2645	3277	3804
1300	374	1054	1493	1862	2266	2707	3353	3893
1400	382	1076	1526	1902	2315	2765	3426	3977
1500	390	1098	1556	1941	2361	2821	3495	4057
1600	397	1119	1586	1977	2406	2874	3561	4134
1700	404	1138	1614	2012	2448	2925	3624	4207
1800	411	1157	1641	2046	2489	2974	3684	4277
1900	417	1176	1666	2078	2528	3021	3742	4344
2000	424	1193	1691	2109	2566	3066	3798	4409
2500	452	1273	1804	2249	2737	3270	4051	4703
3000	476	1342	1902	2371	2885	3447	4270	4957
3500	498	1403	1988	2479	3016	3604	4464	5183
4000	518	1458	2066	2576	3135	3746	4640	5387
4500	536	1508	2138	2666	3244	3875	4801	5573
5000	552	1555	2204	2748	3344	3995	4949	5746

Table 5.11.10 Variation of floods of various return periods with catchment area based on L-moments for Upper Narmada and Tapi Subzone 3(c)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	70	123	203	235	267	299	339	370
20	102	179	296	344	390	436	496	540
50	169	296	489	568	644	720	819	892
100	247	432	714	829	942	1052	1196	1303
200	360	631	1044	1212	1376	1537	1747	1904
300	450	788	1303	1512	1717	1919	2181	2377
400	526	922	1525	1770	2010	2246	2553	2783
500	595	1041	1723	2000	2271	2538	2885	3144
600	657	1151	1904	2210	2509	2804	3187	3473
700	715	1252	2071	2404	2730	3051	3468	3779
800	769	1347	2228	2586	2937	3282	3730	4065
900	820	1436	2377	2758	3132	3500	3979	4336
1000	869	1521	2518	2922	3318	3708	4215	4593
1100	915	1603	2652	3078	3495	3906	4440	4839
1200	960	1681	2782	3228	3666	4097	4657	5075
1300	1003	1756	2906	3373	3830	4280	4865	5302
1400	1045	1829	3026	3512	3988	4457	5066	5521
1500	1085	1899	3143	3647	4142	4628	5261	5734
1600	1124	1968	3256	3779	4291	4795	5450	5940
1700	1162	2034	3366	3906	4435	4956	5634	6140
1800	1198	2098	3472	4030	4576	5114	5813	6335
1900	1234	2161	3577	4151	4713	5267	5987	6525
2000	1270	2223	3678	4269	4848	5417	6158	6711
2500	1434	2512	4156	4823	5477	6121	6957	7582
3000	1585	2775	4592	5329	6051	6762	7687	8377
3500	1724	3019	4996	5798	6584	7357	8363	9114
4000	1855	3248	5374	6237	7083	7915	8997	9805
4500	1978	3464	5732	6652	7554	8442	9595	10457
5000	2096	3669	6072	7047	8002	8942	10165	11078

Table 5.11.11 Variation of floods of various return periods with catchment area based on L-moments for Mahanadi Subzone 3(d)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	16	33	43	50	57	65	75	84
20	29	61	78	91	104	118	137	152
50	64	134	172	201	230	261	303	335
100	117	244	312	365	419	474	550	610
200	212	444	568	664	762	863	1001	1110
300	301	630	806	942	1081	1224	1420	1575
400	386	807	1034	1207	1386	1569	1820	2018
500	468	979	1253	1464	1680	1902	2207	2447
600	548	1146	1467	1713	1966	2226	2583	2864
700	625	1309	1675	1957	2246	2543	2950	3271
800	702	1469	1880	2196	2520	2853	3310	3671
900	777	1626	2081	2431	2790	3159	3665	4064
1000	851	1780	2279	2662	3056	3459	4013	4451
1100	924	1933	2475	2891	3317	3756	4358	4832
1200	996	2084	2668	3116	3576	4049	4697	5209
1300	1067	2233	2858	3339	3832	4338	5033	5581
1400	1137	2380	3047	3559	4085	4625	5366	5950
1500	1207	2526	3234	3778	4336	4908	5695	6315
1600	1276	2671	3419	3994	4584	5190	6021	6677
1700	1345	2815	3603	4209	4830	5468	6345	7035
1800	1413	2957	3785	4421	5074	5745	6665	7391
1900	1480	3098	3966	4633	5317	6019	6984	7744
2000	1547	3238	4145	4842	5557	6292	7300	8095
2500	1876	3926	5026	5871	6738	7628	8850	9814
3000	2196	4595	5882	6871	7886	8928	10358	11486
3500	2508	5249	6719	7849	9008	10198	11832	13120
4000	2815	5890	7540	8807	10108	11443	13277	14723
4500	3116	6520	8347	9750	11189	12668	14697	16298
5000	3412	7141	9141	10678	12255	13874	16097	17849

Table 5.11.12 Variation of floods of various return periods with catchment area based on L-moments for Upper Godavari Subzone 3(e)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	36	88	107	118	126	133	140	144
20	53	130	158	174	187	197	207	212
50	89	217	264	291	312	329	345	355
100	131	320	389	429	460	485	510	524
200	193	472	575	633	679	715	752	773
300	242	593	721	795	852	898	944	970
400	285	697	848	934	1002	1055	1109	1140
500	323	790	961	1058	1135	1196	1257	1292
600	357	875	1064	1172	1258	1325	1392	1431
700	390	954	1160	1278	1371	1444	1518	1560
800	420	1028	1251	1377	1478	1557	1636	1682
900	449	1098	1336	1472	1579	1663	1748	1796
1000	476	1165	1417	1561	1675	1764	1854	1906
1100	502	1229	1495	1647	1767	1861	1956	2010
1200	527	1291	1570	1729	1855	1954	2054	2111
1300	551	1350	1642	1809	1940	2044	2148	2208
1400	575	1407	1712	1885	2023	2131	2239	2302
1500	597	1463	1779	1960	2103	2215	2328	2393
1600	619	1517	1845	2032	2180	2297	2414	2481
1700	641	1569	1909	2102	2256	2376	2497	2567
1800	662	1620	1971	2171	2329	2453	2579	2650
1900	682	1670	2032	2238	2401	2529	2658	2732
2000	702	1719	2091	2303	2471	2603	2736	2812
2500	796	1948	2370	2610	2800	2950	3100	3186
3000	881	2158	2625	2891	3102	3268	3434	3530
3500	961	2353	2862	3153	3382	3563	3745	3848
4000	1036	2536	3085	3398	3645	3840	4036	4148
4500	1106	2709	3296	3630	3894	4102	4311	4431
5000	1174	2874	3496	3851	4132	4352	4574	4701

Table 5.11.13 Variation of floods of various return periods with catchment area based on L-moments for Lower Godavari Subzone 3(f)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	43	90	113	129	145	160	179	194
20	69	144	180	206	231	255	286	310
50	129	268	335	383	429	474	532	575
100	205	429	535	612	685	757	850	919
200	328	685	855	977	1095	1210	1358	1468
300	432	901	1125	1285	1440	1592	1786	1931
400	525	1095	1367	1561	1750	1933	2170	2346
500	610	1273	1589	1816	2034	2248	2523	2728
600	690	1440	1797	2054	2301	2543	2854	3086
700	766	1598	1995	2279	2554	2822	3168	3424
800	838	1749	2183	2495	2795	3089	3467	3748
900	908	1894	2364	2701	3027	3345	3754	4059
1000	975	2034	2539	2901	3250	3591	4031	4358
1100	1039	2169	2708	3094	3467	3831	4300	4648
1200	1102	2301	2872	3281	3677	4063	4560	4930
1300	1164	2429	3031	3464	3881	4288	4814	5204
1400	1223	2553	3187	3642	4081	4509	5061	5471
1500	1282	2675	3339	3815	4275	4724	5303	5732
1600	1339	2795	3488	3986	4466	4935	5539	5988
1700	1395	2911	3634	4152	4653	5141	5771	6239
1800	1450	3026	3777	4316	4836	5344	5998	6484
1900	1504	3139	3918	4476	5016	5543	6222	6726
2000	1557	3250	4056	4634	5193	5738	6441	6963
2500	1811	3779	4717	5389	6039	6672	7490	8097
3000	2048	4274	5335	6096	6831	7548	8472	9159
3500	2273	4744	5921	6765	7581	8377	9403	10165
4000	2488	5192	6481	7404	8297	9168	10291	11125
4500	2694	5622	7018	8018	8985	9928	11144	12047
5000	2893	6037	7536	8610	9648	10661	11967	12936

Table 5.11.14 Variation of floods of various return periods with catchment area based on L-moments for Krishna and Pennar Subzone 3(h)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	15	39	50	57	62	67	73	76
20	24	64	81	92	101	109	118	124
50	45	121	154	175	192	207	224	235
100	73	197	250	284	313	337	365	382
200	119	321	407	462	508	548	593	620
300	158	426	541	613	675	729	787	824
400	194	522	662	751	826	891	963	1009
500	227	610	774	878	966	1042	1127	1179
600	258	693	880	997	1098	1184	1280	1340
700	287	772	980	1111	1223	1319	1426	1493
800	315	848	1076	1220	1343	1449	1566	1640
900	342	921	1169	1325	1459	1574	1701	1781
1000	369	992	1258	1427	1571	1694	1831	1917
1100	394	1060	1345	1525	1679	1811	1958	2050
1200	419	1127	1430	1621	1785	1925	2081	2179
1300	443	1192	1512	1715	1888	2036	2201	2305
1400	467	1255	1593	1806	1989	2145	2318	2427
1500	490	1318	1672	1896	2087	2251	2433	2548
1600	512	1379	1749	1983	2184	2355	2546	2666
1700	535	1438	1825	2070	2278	2458	2656	2781
1800	556	1497	1900	2154	2372	2558	2765	2895
1900	578	1555	1973	2237	2463	2657	2872	3007
2000	599	1612	2045	2319	2553	2754	2977	3117
2500	700	1885	2392	2712	2986	3221	3481	3645
3000	796	2142	2718	3082	3393	3660	3956	4142
3500	887	2386	3028	3433	3780	4077	4407	4614
4000	974	2620	3325	3770	4151	4477	4840	5067
4500	1058	2846	3611	4095	4508	4863	5256	5503
5000	1139	3064	3888	4409	4854	5236	5659	5925

Table 5.11.15 Variation of floods of various return periods with catchment area based on L-moments for Kaveri Basin Subzone 3(i)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	1	3	4	5	7	8	9	10
20	1	7	10	13	16	18	22	25
50	4	21	32	40	49	57	69	77
100	10	50	75	95	115	135	162	183
200	23	118	178	224	272	321	385	434
300	37	195	294	372	451	531	637	718
400	53	280	421	532	645	759	911	1027
500	71	369	556	702	852	1002	1203	1355
600	89	463	697	880	1069	1257	1509	1700
700	107	561	845	1066	1295	1523	1828	2060
800	127	662	997	1259	1529	1798	2158	2432
900	147	767	1155	1458	1770	2082	2499	2816
1000	167	874	1316	1662	2018	2374	2849	3210
1100	188	984	1482	1871	2272	2673	3207	3614
1200	210	1097	1652	2085	2531	2978	3574	4027
1300	232	1211	1824	2303	2797	3290	3948	4449
1400	254	1328	2001	2525	3067	3608	4329	4879
1500	277	1447	2180	2752	3341	3931	4717	5316
1600	300	1568	2362	2982	3621	4260	5112	5760
1700	324	1691	2547	3215	3904	4593	5512	6211
1800	347	1816	2735	3452	4192	4932	5918	6669
1900	372	1942	2925	3692	4484	5275	6330	7133
2000	396	2070	3118	3936	4779	5623	6747	7603
2500	523	2732	4116	5195	6308	7421	8906	10036
3000	656	3428	5163	6518	7914	9311	11173	12591
3500	795	4153	6255	7895	9587	11279	13535	15252
4000	938	4903	7385	9322	11320	13317	15981	18009
4500	1086	5677	8550	10793	13106	15419	18502	20850
5000	1238	6472	9748	12305	14941	17578	21094	23770

Table 5.11.16 Variation of floods of various return periods with catchment area based on L-moments for East Coast Subzone 4(b)

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	14	107	174	227	283	340	416	475
20	18	138	225	294	366	439	538	614
50	25	194	316	413	514	617	756	863
100	32	251	408	534	664	798	977	1116
200	42	324	528	691	859	1032	1264	1443
300	48	377	613	803	999	1199	1469	1677
400	54	420	682	893	1111	1334	1635	1866
500	58	456	741	970	1207	1449	1776	2027
600	62	488	793	1038	1292	1551	1900	2169
700	66	516	840	1099	1368	1642	2012	2296
800	70	543	883	1155	1437	1725	2114	2413
900	73	567	922	1207	1501	1802	2209	2521
1000	76	589	959	1255	1561	1874	2297	2621
1100	78	611	993	1300	1617	1942	2379	2715
1200	81	631	1026	1343	1670	2005	2457	2804
1300	83	650	1057	1383	1721	2066	2531	2889
1400	86	668	1086	1422	1769	2124	2602	2970
1500	88	685	1114	1459	1814	2179	2669	3046
1600	90	702	1141	1494	1858	2231	2734	3120
1700	92	718	1167	1528	1901	2282	2796	3191
1800	94	733	1192	1561	1941	2331	2856	3260
1900	96	748	1216	1592	1981	2378	2914	3326
2000	98	762	1240	1623	2019	2424	2970	3390
2500	106	828	1347	1763	2193	2633	3227	3682
3000	114	886	1441	1886	2346	2817	3452	3940
3500	120	938	1526	1997	2485	2983	3656	4172
4000	126	986	1603	2099	2611	3135	3841	4384
4500	132	1030	1675	2193	2727	3275	4013	4579
5000	137	1071	1742	2280	2836	3405	4173	4762

Table 5.11.17 Variation of floods of various return periods with catchment area based on L-moments for Sub-Himalayan region Zone 7

Catchment Area (km^2)	Return periods (Years)							
	2	10	25	50	100	200	500	1000
	Floods of various return periods (m^3/s)							
10	141	269	347	410	480	558	675	776
20	185	351	454	536	628	730	883	1015
50	263	501	647	765	895	1040	1259	1447
100	344	655	847	1000	1171	1360	1646	1892
200	450	857	1107	1308	1531	1779	2152	2474
300	527	1002	1295	1530	1791	2081	2518	2895
400	589	1120	1448	1710	2002	2326	2815	3235
500	642	1221	1578	1864	2182	2536	3069	3527
600	689	1310	1694	2001	2342	2721	3293	3785
700	731	1391	1798	2124	2486	2888	3495	4018
800	770	1465	1893	2236	2617	3042	3681	4231
900	806	1533	1981	2341	2739	3184	3852	4428
1000	839	1597	2064	2438	2854	3316	4013	4612
1100	871	1657	2141	2530	2961	3441	4163	4786
1200	901	1713	2215	2616	3062	3558	4306	4950
1300	929	1767	2284	2699	3158	3670	4441	5105
1400	956	1819	2351	2777	3250	3777	4571	5254
1500	982	1868	2415	2852	3338	3879	4694	5396
1600	1007	1915	2476	2924	3423	3978	4813	5533
1700	1031	1961	2534	2994	3504	4072	4927	5664
1800	1054	2005	2591	3061	3582	4163	5038	5791
1900	1076	2047	2646	3125	3658	4251	5144	5913
2000	1098	2088	2699	3188	3731	4336	5247	6032
2500	1197	2276	2942	3476	4068	4727	5720	6576
3000	1284	2443	3157	3730	4365	5073	6139	7056
3500	1363	2593	3352	3959	4634	5385	6516	7490
4000	1435	2730	3529	4169	4880	5670	6862	7887
4500	1502	2858	3694	4363	5107	5935	7182	8255
5000	1565	2977	3848	4545	5320	6182	7480	8599

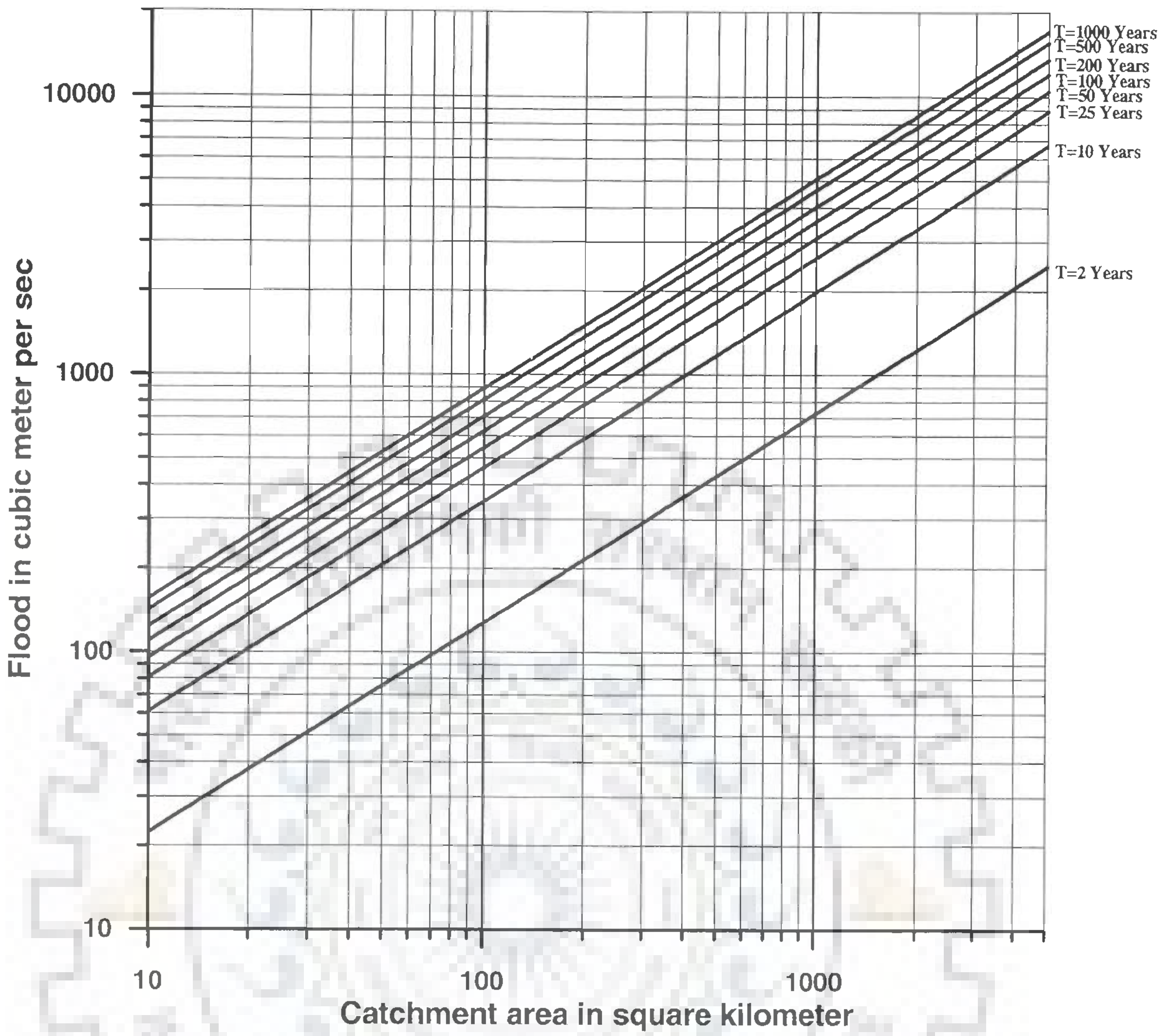


Fig. 5.3.1 Variation of floods of various return periods with catchment area based on L-moments for Chambal Subzone 1 (b)

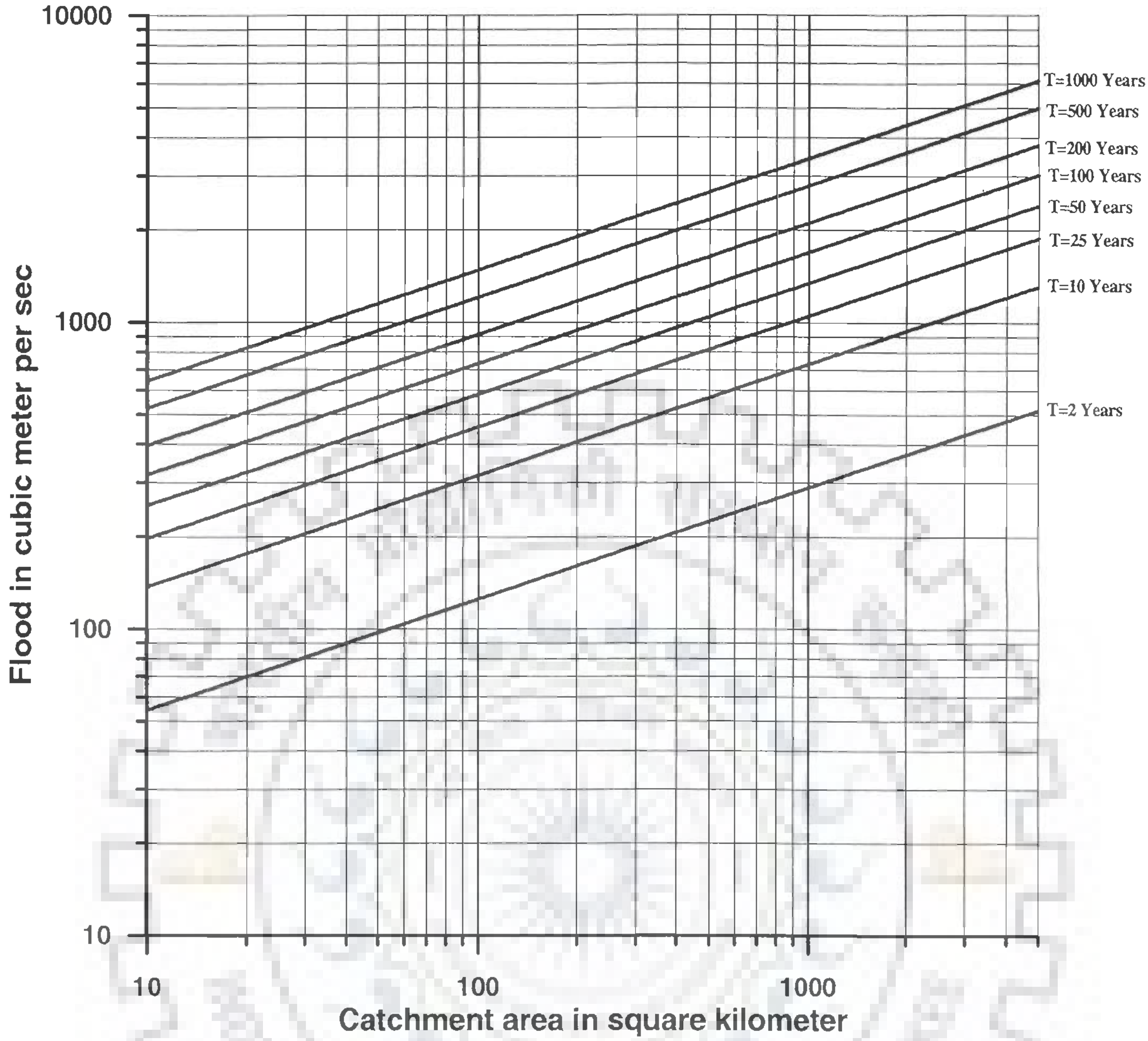


Fig. 5.3.2 Variation of floods of various return periods with catchment area based on L-moments for Sone Subzone 1 (d)

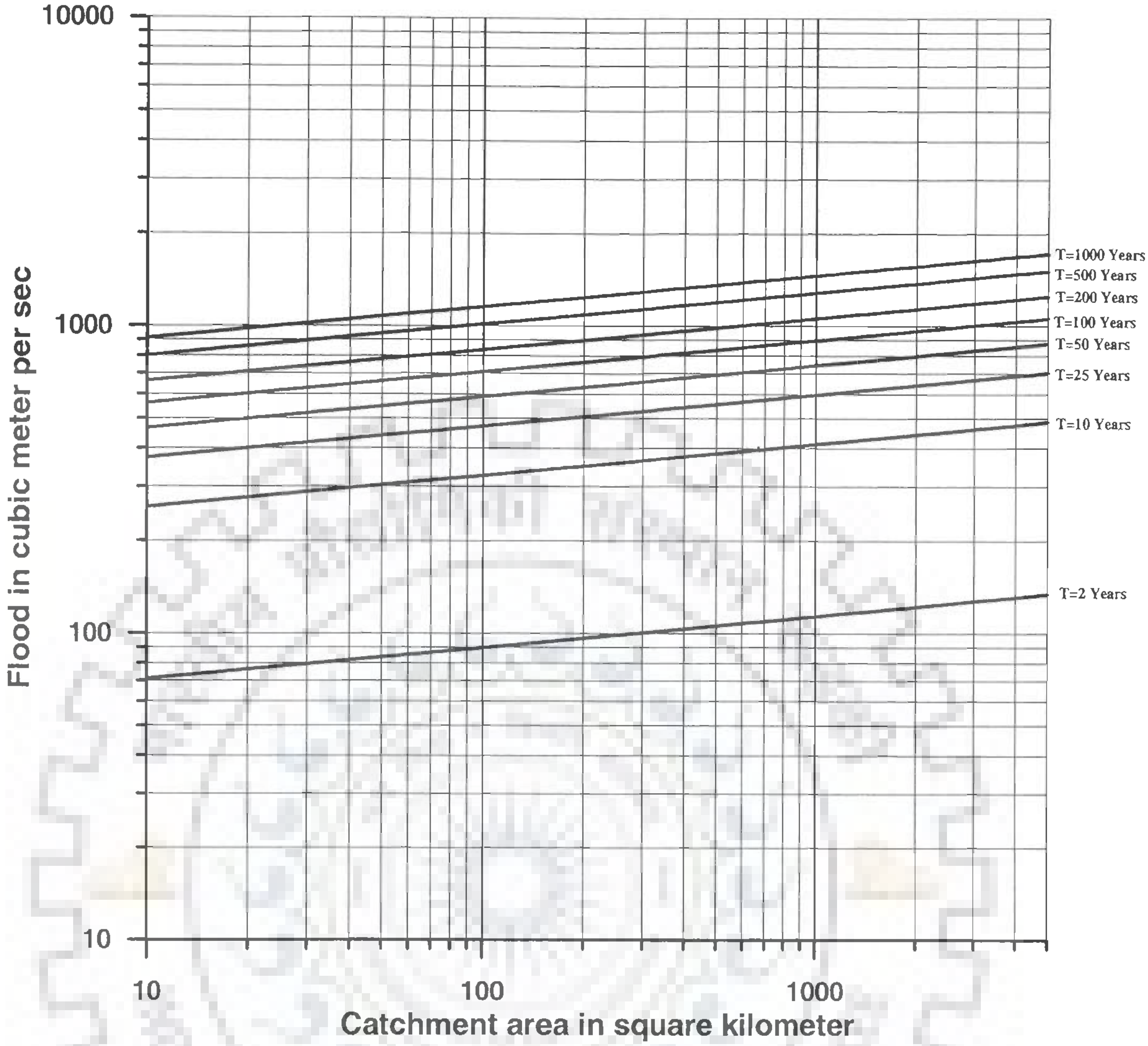


Fig. 5.3.3 Variation of floods of various return periods with catchment area based on L-moments for Upper Indo-Ganga Plains Subzone 1 (e)

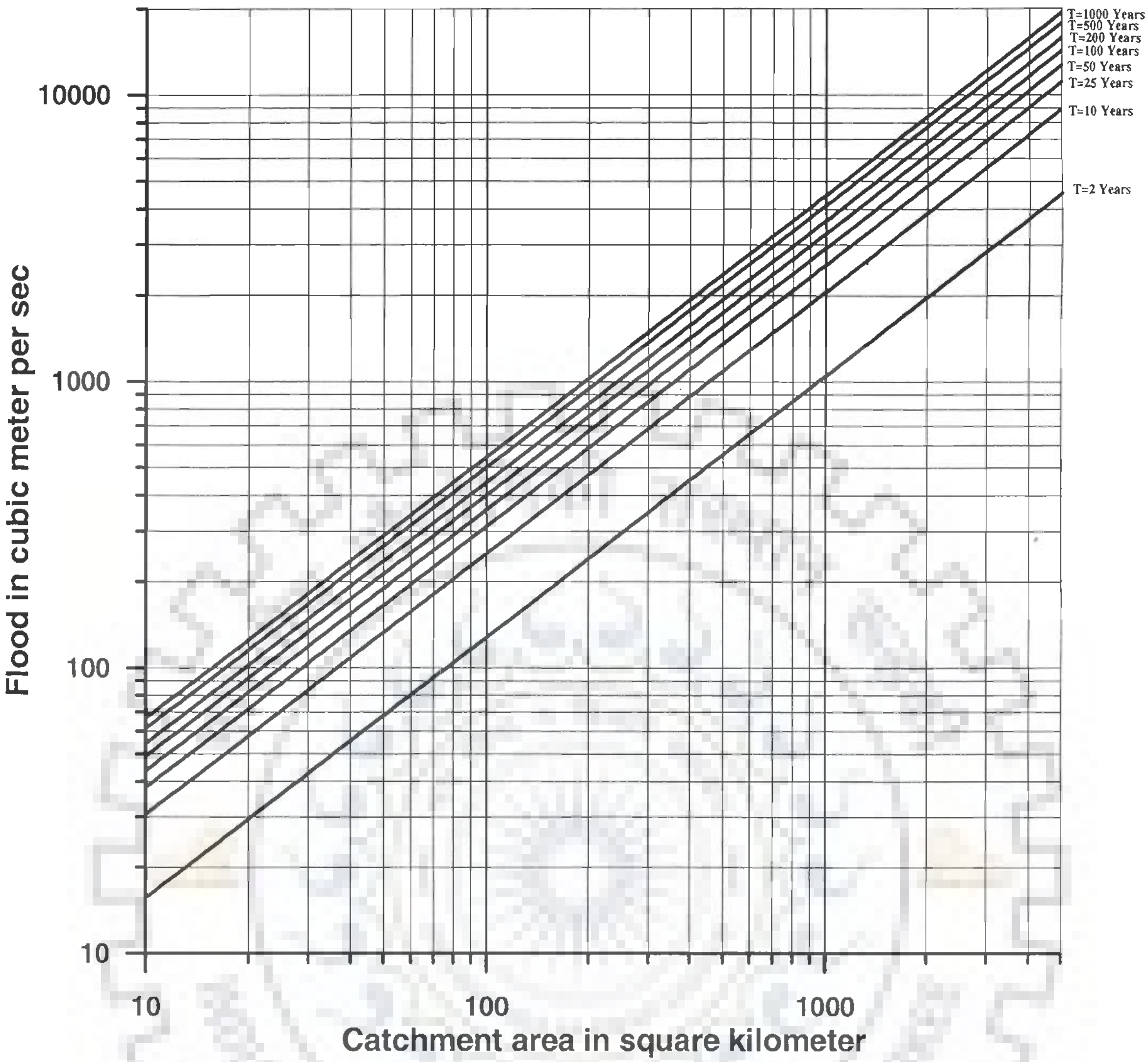


Fig. 5.3.4 Variation of floods of various return periods with catchment area based on L-moments for Middle Ganga Plains Subzone 1 (f)

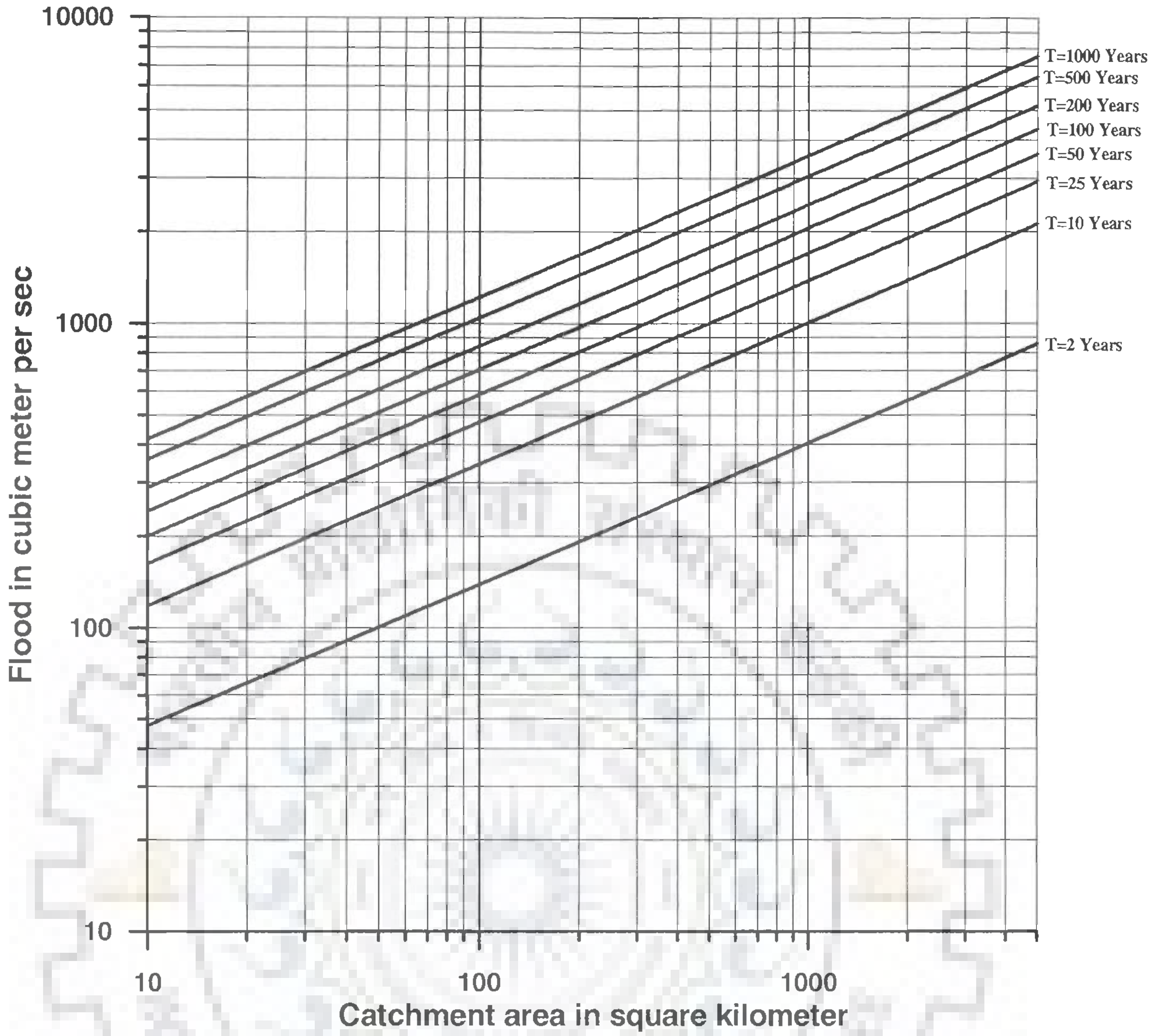


Fig. 5.3.5 Variation of floods of various return periods with catchment area based on L-moments for Lower Ganga Plains Subzone 1 (g)

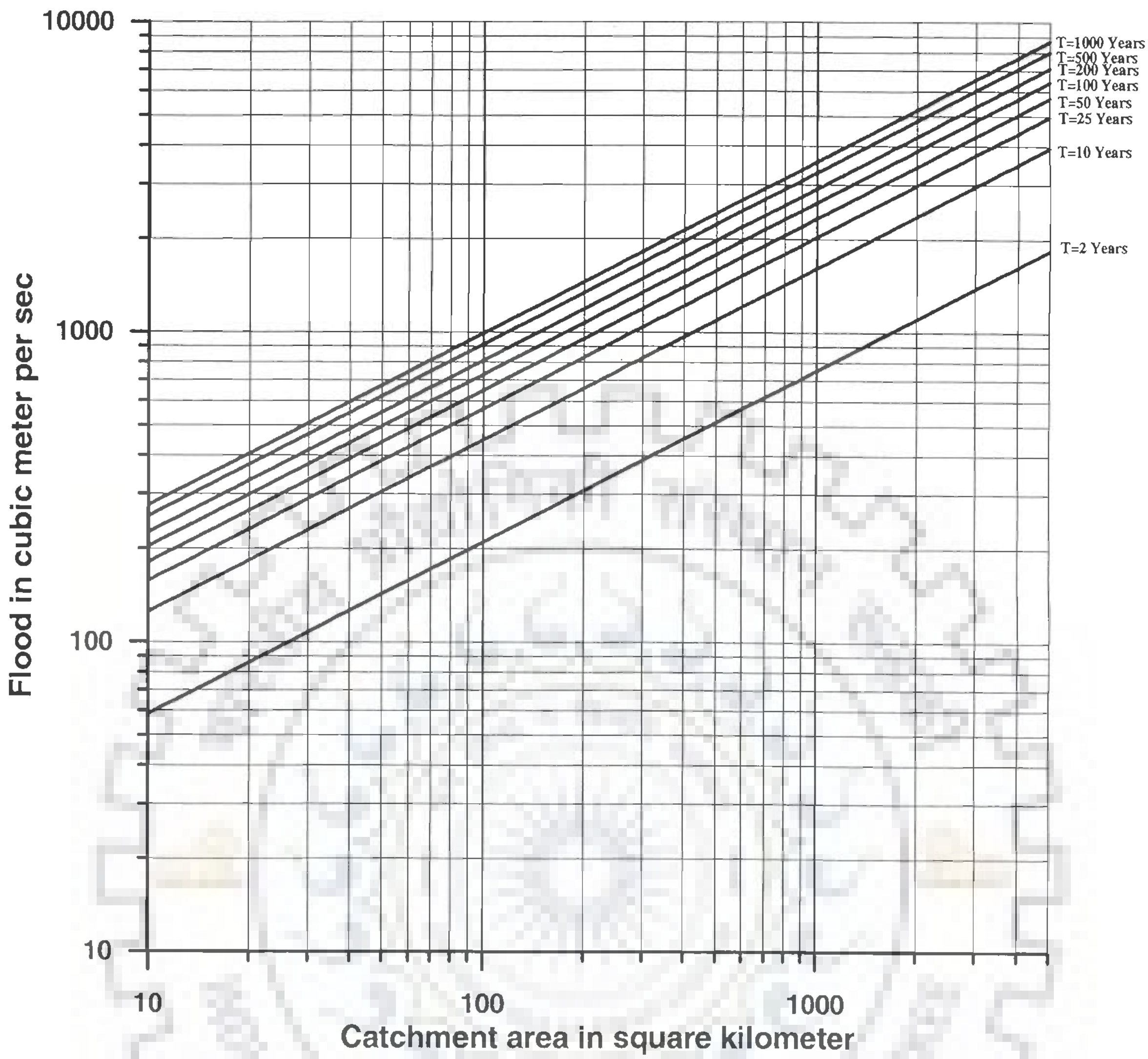


Fig. 5.3.6 Variation of floods of various return periods with catchment area based on L-moments for North Brahmaputra Subzone 2 (a)

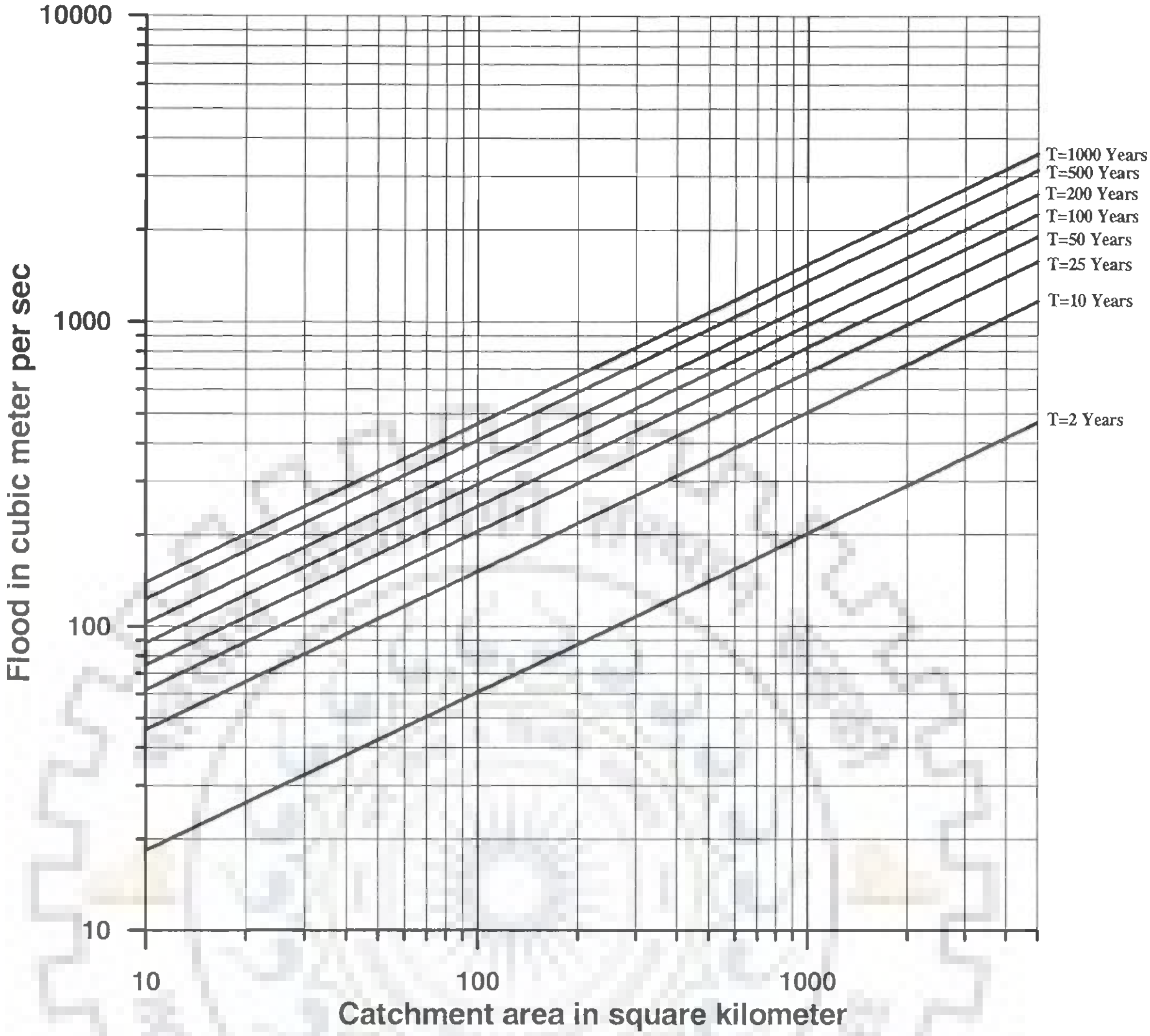


Fig. 5.3.7 Variation of floods of various return periods with catchment area based on L-moments for South Brahmaputra Subzone 2 (b)

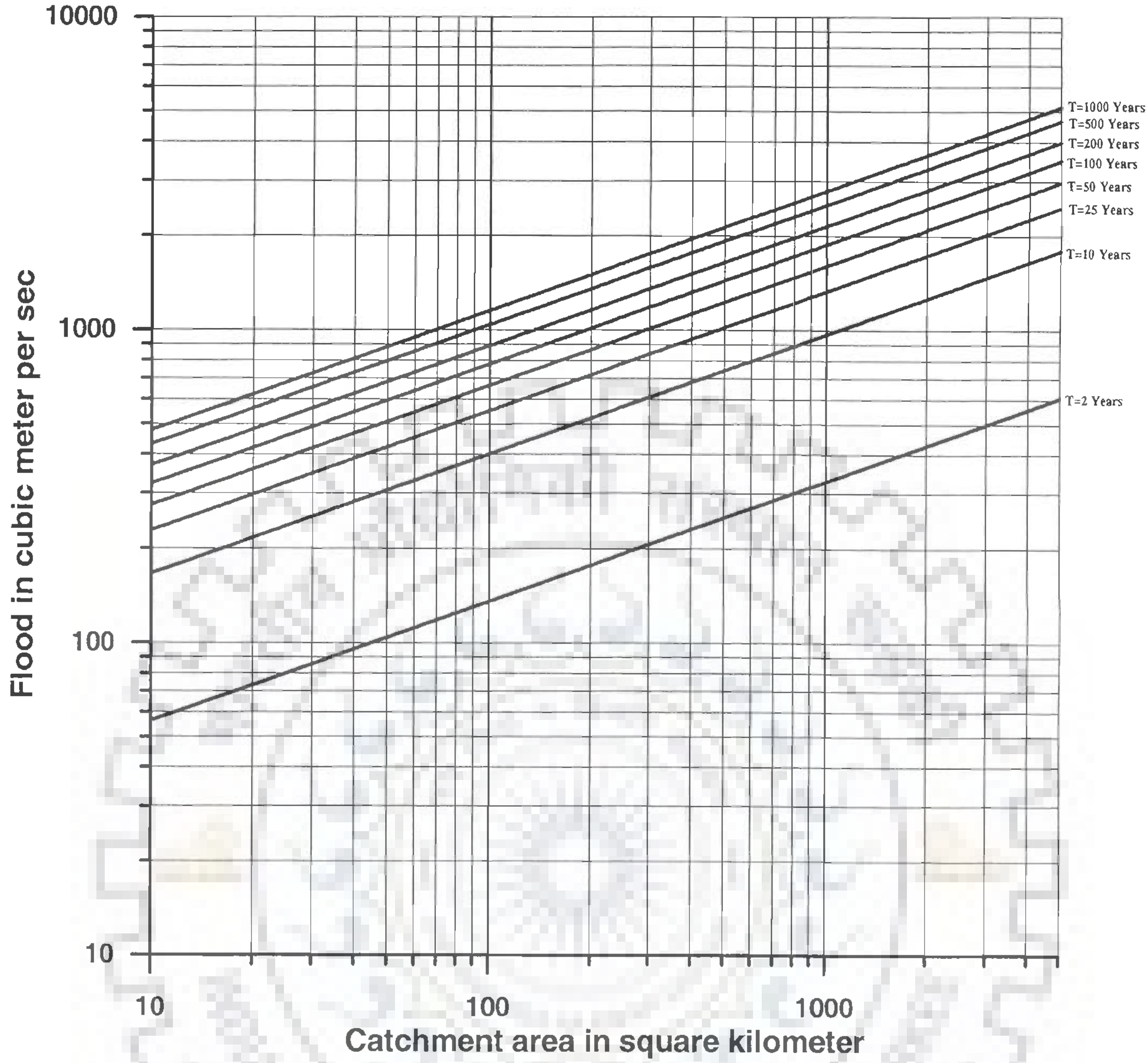


Fig. 5.3.8 Variation of floods of various return periods with catchment area based on L-moments for Mahi and Sabarmati Subzone 3 (a)

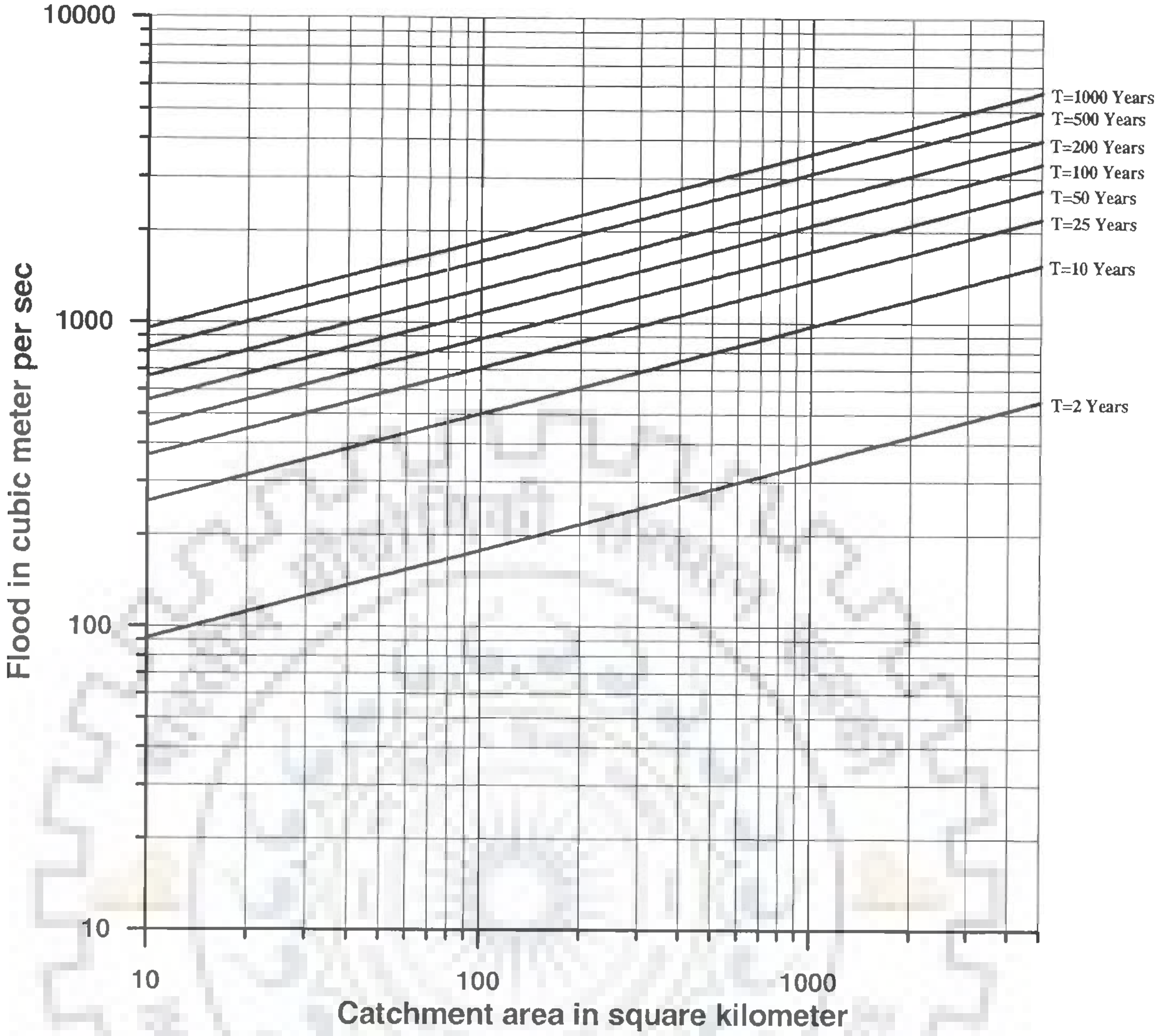


Fig. 5.3.9 Variation of floods of various return periods with catchment area based on L-moments for Lower Narmada and Tapi Subzone 3 (b)

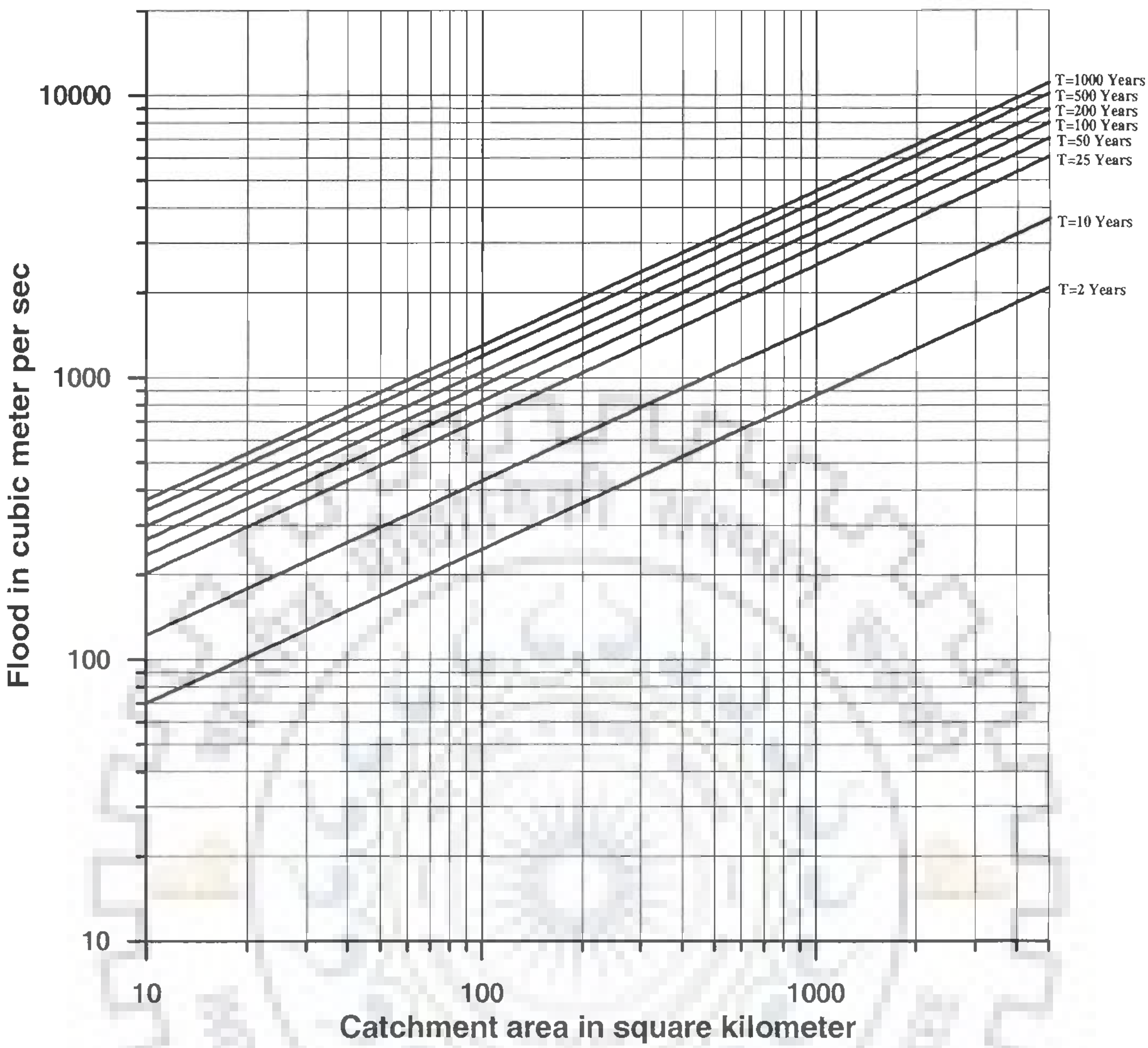


Fig. 5.3.10 Variation of floods of various return periods with catchment area based on L-moments for Upper Narmada and Tapi Subzone 3 (c)

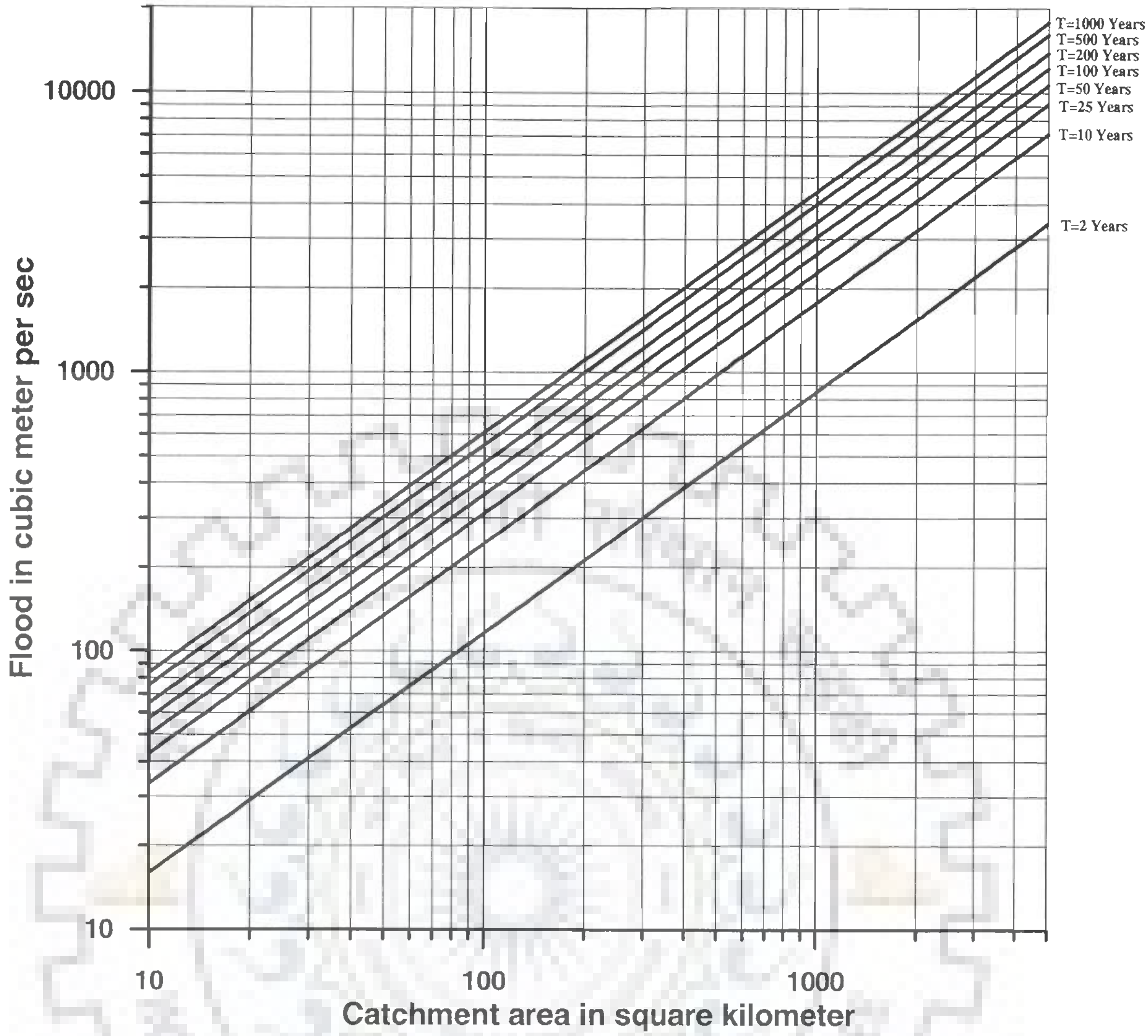


Fig. 5.3.11 Variation of floods of various return periods with catchment area based on L-moments for Mahanadi Subzone 3 (d)

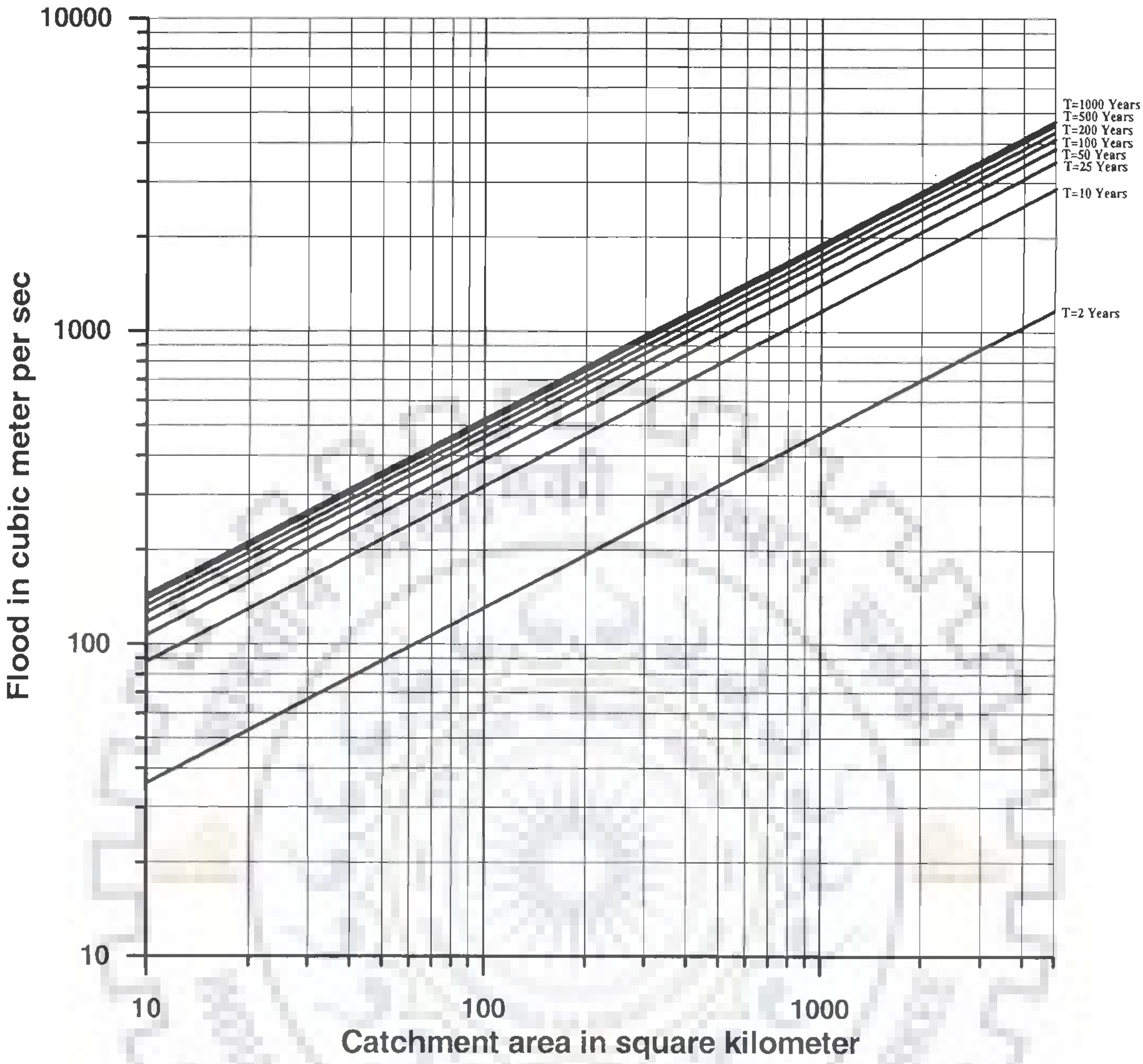


Fig. 5.3.12 Variation of floods of various return periods with catchment area based on L-moments for Upper Godavari Subzone 3 (e)

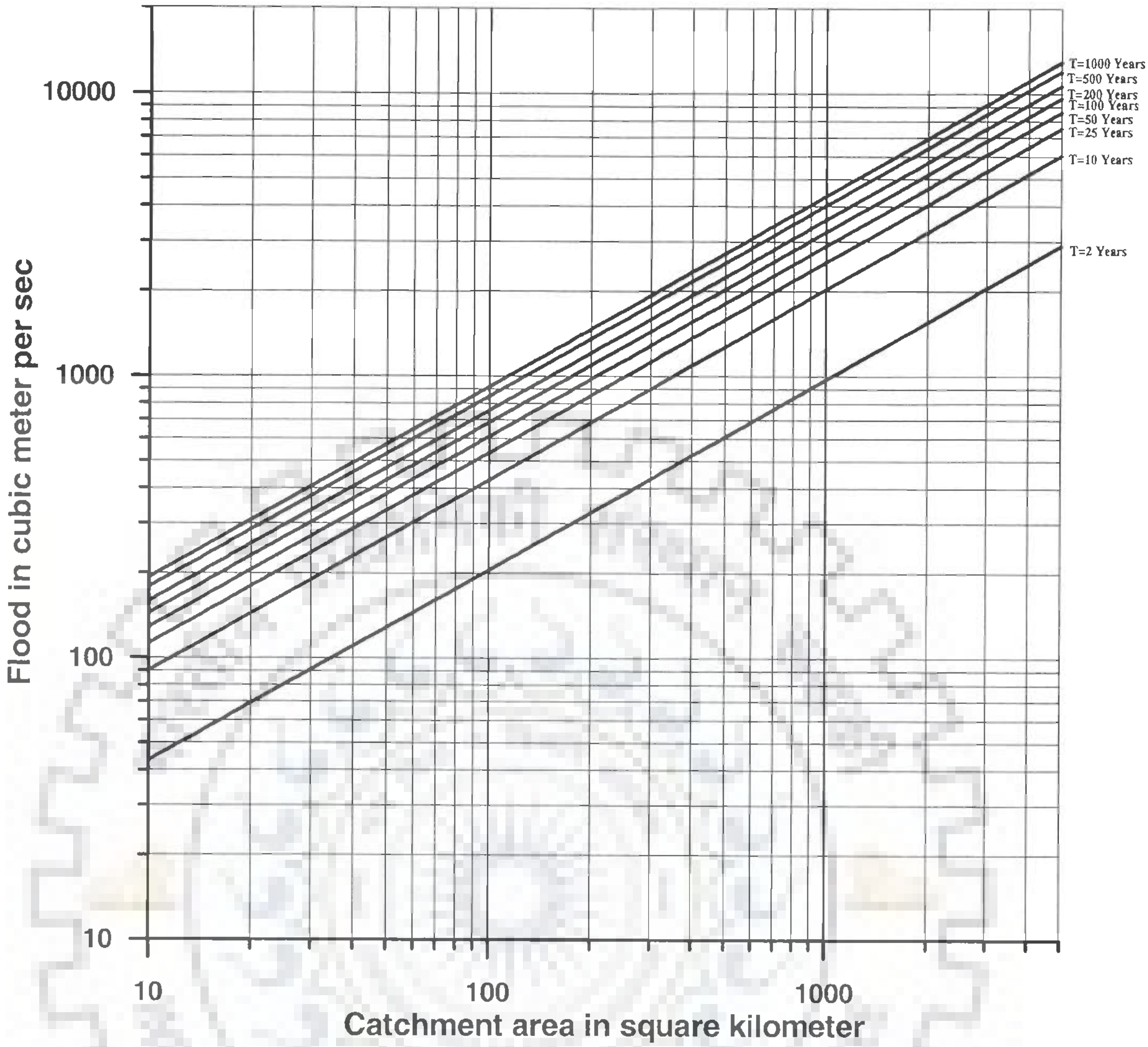


Fig. 5.3.13 Variation of floods of various return periods with catchment area based on L-moments for Lower Godavari Subzone 3 (f)

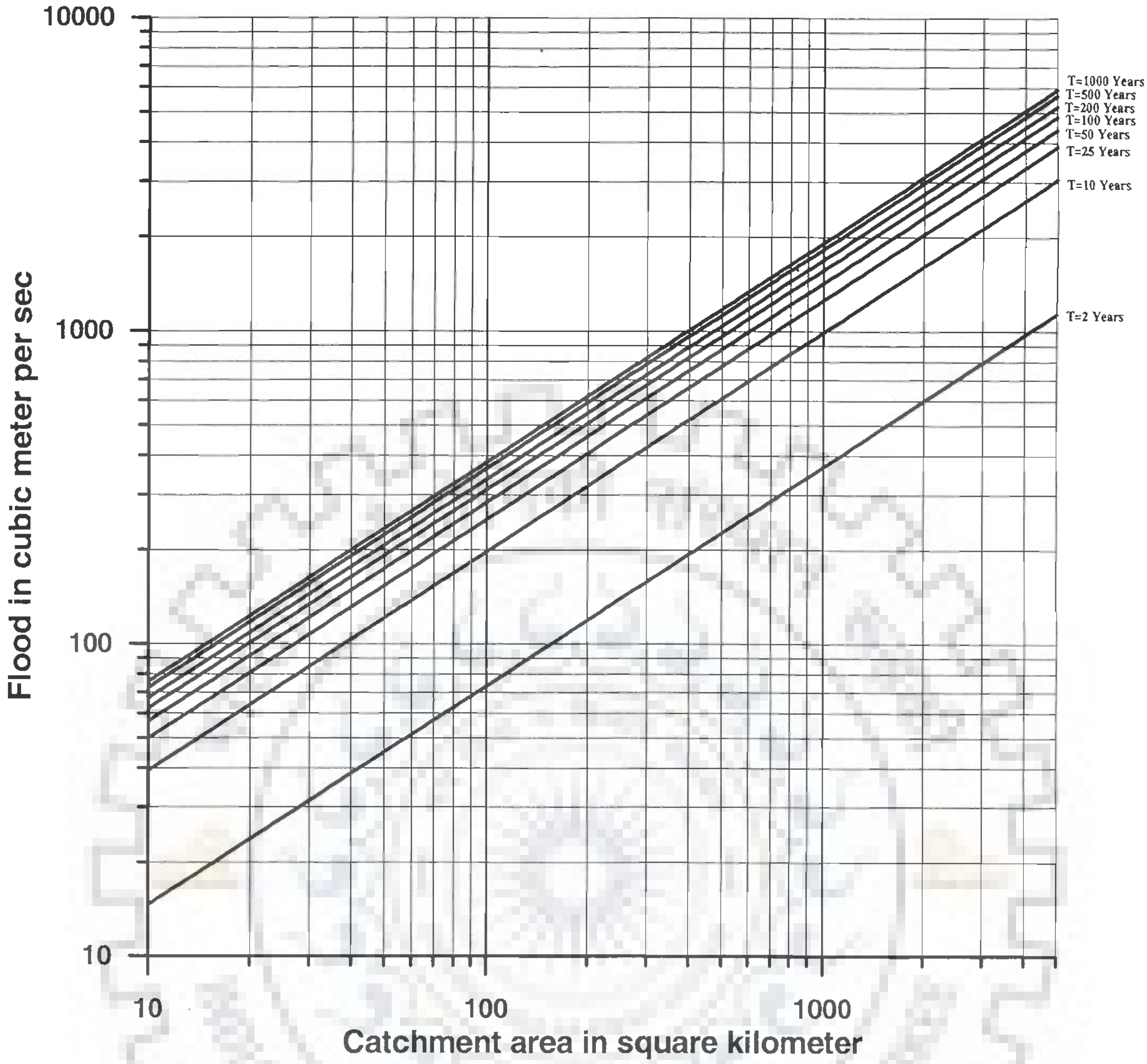


Fig. 5.3.14 Variation of floods of various return periods with catchment area based on L-moments for Krishna and Pennar Subzone 3 (h)

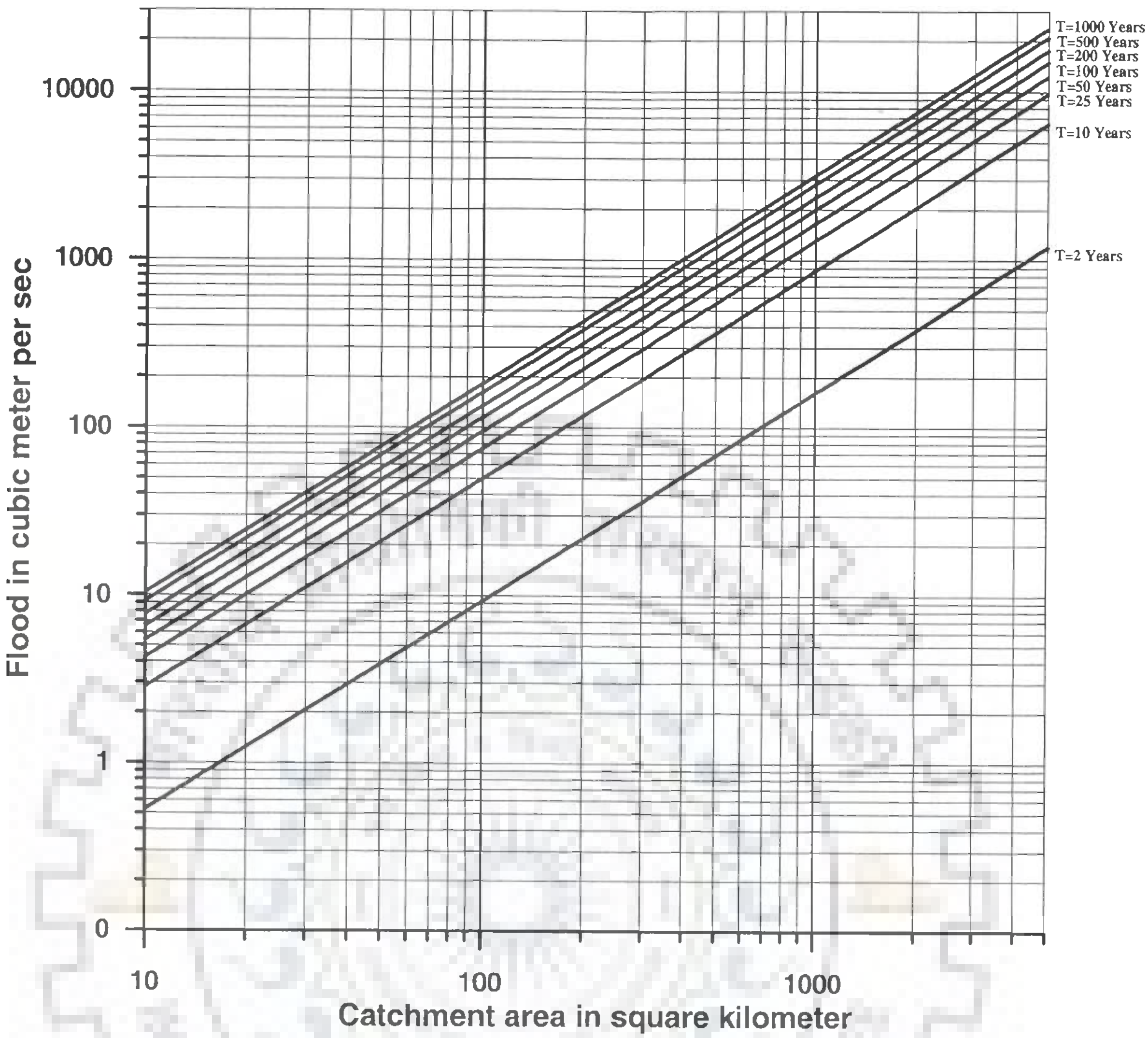


Fig. 5.3.15 Variation of floods of various return periods with catchment area based on L-moments for Kaveri Basin Subzone 3 (i)

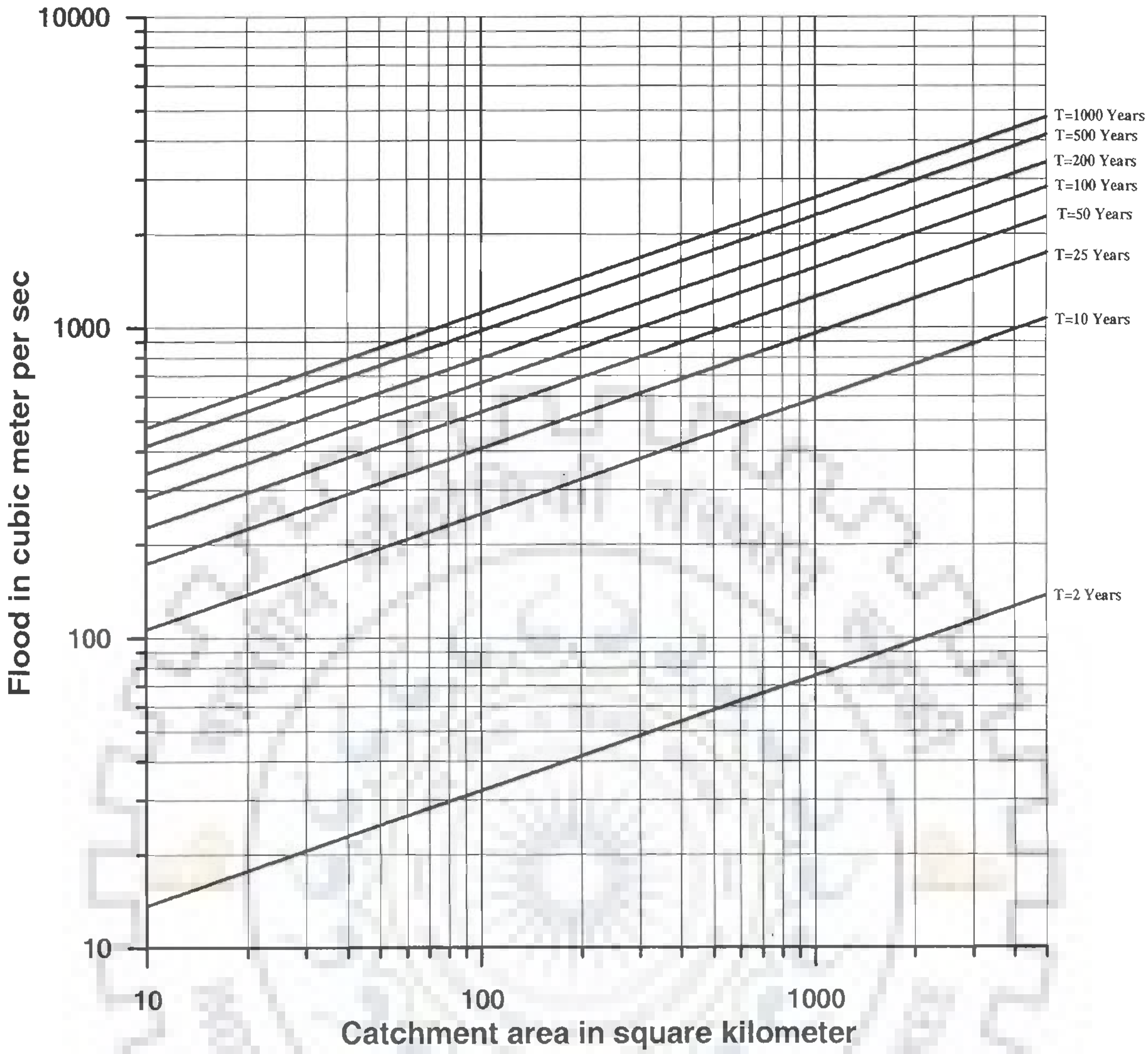


Fig. 5.3.16 Variation of floods of various return periods with catchment area based on L-moments for East Coast Subzone 4 (b)

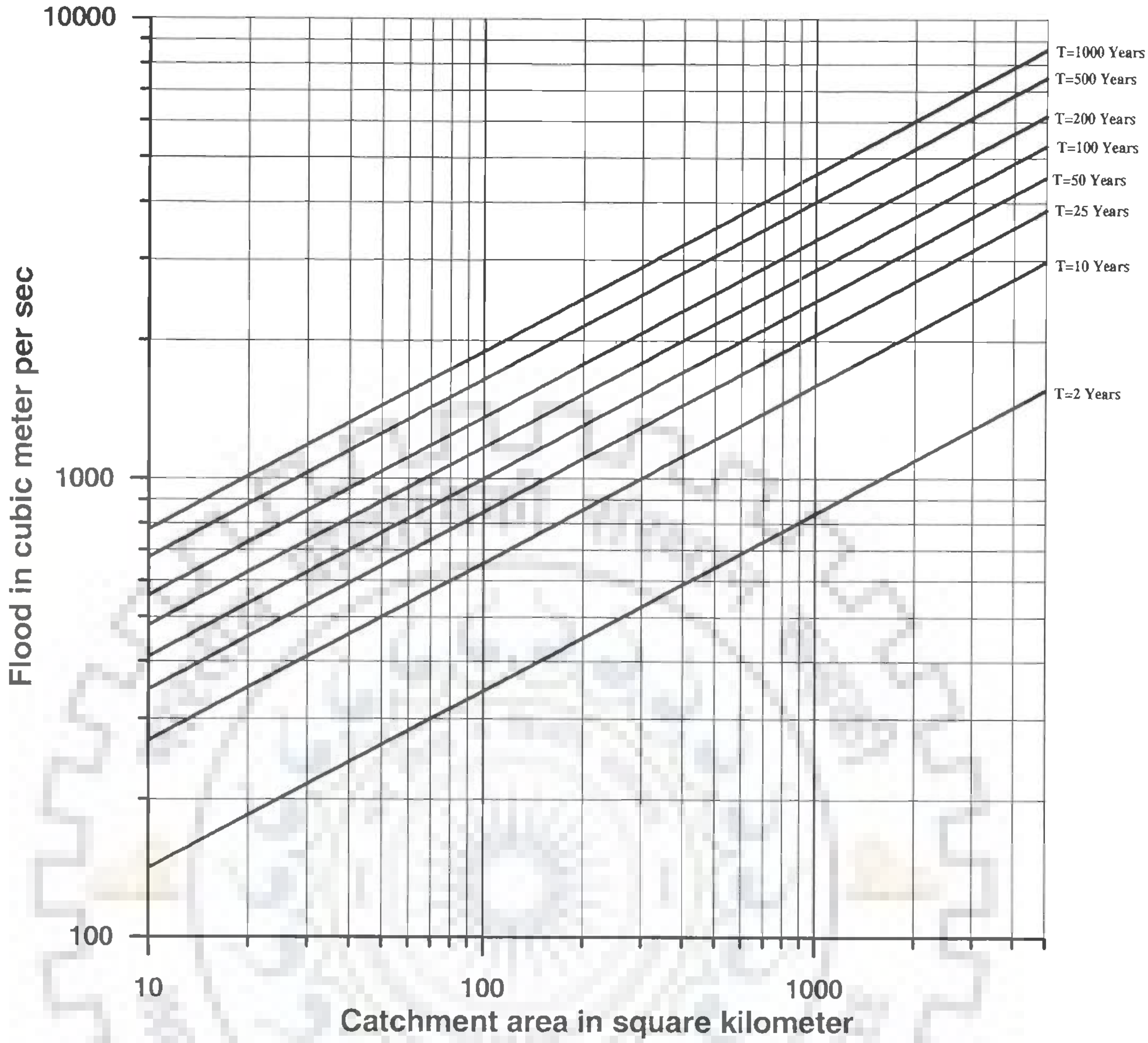


Fig. 5.3.17 Variation of floods of various return periods with catchment area based on L-moments for Sub-Himalayan region Zone-7

5.8 DEVELOPMENT OF REGIONAL FLOOD FREQUENCY RELATIONSHIPS USING ANN AND FIS

Development of regional flood frequency relationships for gauged and ungauged catchments for four Subzones viz. Subzone 3(c), Subzone 3(d), Subzone 3(f) and Zone-7 using ANN and FIS and their comparisons with regional flood frequency relationships developed using L-moments approach are described as follows.

5.8.1 Development of Regional Flood Frequency Relationships for Gauged Catchments using ANN

The functional form of the model for development of regional flood frequency relationships for gauged catchments using ANN is:

$$Qp = f\ (P)$$

Qp = Annual maximum peak flood

P = Probability of non-exceedance of the annual maximum peak flood estimated by the Weibull's formula i.e. P = m/N+1, where m is the rank of the event when arranged in ascending order and N is the total number of observations of the annual maximum peak floods for a Subzone.

The architecture of ANN model consists of number of layers in a network, the number of neurons in each layer, transfer function of each layer and how the layers connect to each other. In order to improve network performance, many factors like determination of adequate model inputs, data division and pre-processing, the choice of suitable network architecture, selection of network internal parameters, the stopping criteria and model testing are required. Data pre-processing is necessary before they are applied to soft computing models to ensure all variables receive equal attention during the training process (Maier and Dandy, 2000). All the data are transformed, normalized and scaled to remain within a range (0, 1). Data transformations are often used to simplify structure of data so that they follow a convenient statistical model (Sudheer et al., 2007).

The total available data have been divided into training and validation sets prior to the model building, according to the statistical properties like mean and standard deviation of the data sets so that the training and validation data set represents the same population. The set of available data (probability of non-exceedance and the annual maximum peak floods) are divided into two sets: one for system modeling (training) – selected regarding characteristic features - and one for system testing. For each of the Subzones 95% values of the data set were taken for training and remaining 5% dataset for testing. The neural network architecture parameters for the ANN model used for the four Subzones are given in Table 5.12. The values of growth factors estimated by ANN and the robust identified frequency distributions for four Subzones viz. Subzone 3(c), Subzone 3(d), Subzone 3(f) and Zone-7 are given in Tables 5.13.1 to 5.13.4 respectively. The statistical performance indices of ANN and L-moments for four Subzones viz. Subzone 3(c), Subzone 3(d), Subzone 3(f) and Zone-7 are presented in Tables 5.14.1 to 5.14.4 for training datasets and in Tables 5.15.1 to 5.15.4 for validation datasets respectively. The variation of growth factors with return periods estimated by ANN and L-moments are shown in Figs. 5.4.1 to 5.4.4 for the four Subzones. It is observed that the growth factors estimated by ANN provide flat growth curves i.e. relatively lower values of growth factors for higher return periods. Hence, to overcome this limitation of ANN regional flood frequency estimation has been attempted using FIS as discussed in the following Section.

Table 5.12 ANN Architecture parameters for regional flood frequency estimation

No. of Input neurons	1
No. of Hidden neurons	2
No. of hidden layer	1
Learning rate	0.01
Momentum factor	0.1
Transfer function of hidden layer	Sigmoidal
Transfer function of output layer	Linear
Training Algorithm	Back propagation
Training Cycles, epoch	1000
Training Goal	0.0001

Table 5.13.1 Growth factors for ANN and L-moments for Upper Narmada and Tapi Subzone 3 (c)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
ANN	0.852	1.970	2.561	2.852	3.162	3.427	3.763	3.924
L-moments	0.847	1.483	2.454	2.848	3.234	3.614	4.108	4.477

Table 5.13.2 Growth factors for ANN and L-moments for Mahanadi Subzone 3(d)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
ANN	0.856	1.980	2.564	2.848	3.113	3.303	3.660	3.979
L-moments	0.870	1.821	2.331	2.723	3.125	3.538	4.105	4.552

Table 5.13.3 Growth factors for ANN and L-moments for Lower Godavari Subzone 3(f)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
ANN	0.875	1.935	2.230	2.543	2.704	2.858	3.208	3.452
L-moments	0.886	1.849	2.308	2.637	2.955	3.265	3.665	3.962

Table 5.13.4 Growth factors for ANN and L-moments for Sub-Himalayan region Zone-7

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
ANN	0.900	1.885	2.302	2.586	2.962	3.372	3.837	4.248
L-moments	0.911	1.753	2.240	2.646	3.097	3.599	4.355	5.006

Table 5.14.1 Statistical performance indices of ANN and L-moments for training for Upper Narmada and Tapi Subzone 3 (c)

Statistical indices	ANN	L-moments
EFF	98.17	91.72
CORR	0.99	0.94
MAE	3.10	8.64
RMSE	9.07	12.38

Table 5.14.2 Statistical performance indices of ANN and L-moments for training for Mahanadi Subzone 3 (d)

Statistical indices	ANN	L-moments
EFF	98.24	94.14
CORR	0.99	0.92
MAE	2.97	7.48
RMSE	8.96	8.78

Table 5.14.3 Statistical performance indices of ANN and L-moments for training for Lower Godavari Subzone 3 (f)

Statistical indices	ANN	L-moments
EFF	98.32	92.86
CORR	0.99	0.92
MAE	3.32	10.96
RMSE	7.99	8.92

Table 5.14.4 Statistical performance indices of ANN and L-moments for training for Sub-Himalayan region Zone-7

Statistical indices	ANN	L-moments
EFF	98.12	93.47
CORR	0.99	0.92
MAE	3.50	6.13
RMSE	4.16	4.79

Table 5.15.1 Statistical performance indices of ANN and L-moments for validation for Upper Narmada and Tapi Subzone 3 (c)

Statistical indices	ANN	L-moments
EFF	98.10	90.03
CORR	0.99	0.92
MAE	3.69	14.40
RMSE	9.11	11.07

Table 5.15.2 Statistical performance indices of ANN and L-moments for validation for Mahanadi Subzone 3 (d)

Statistical indices	ANN	L-moments
EFF	99.75	93.04
CORR	0.99	0.89
MAE	2.39	14.26
RMSE	3.34	11.32

Table 5.15.3 Statistical performance indices of ANN and L-moments for validation for Lower Godavari Subzone 3 (f)

Statistical indices	ANN	L-moments
EFF	99.06	93.24
CORR	0.99	0.89
MAE	2.89	11.96
RMSE	5.60	9.54

Table 5.15.4 Statistical performance indices of ANN and L-moments for validation for Sub-Himalayan region Zone-7

Statistical indices	ANN	L-moments
EFF	99.05	91.53
CORR	0.99	0.89
MAE	4.72	17.40
RMSE	8.06	12.07

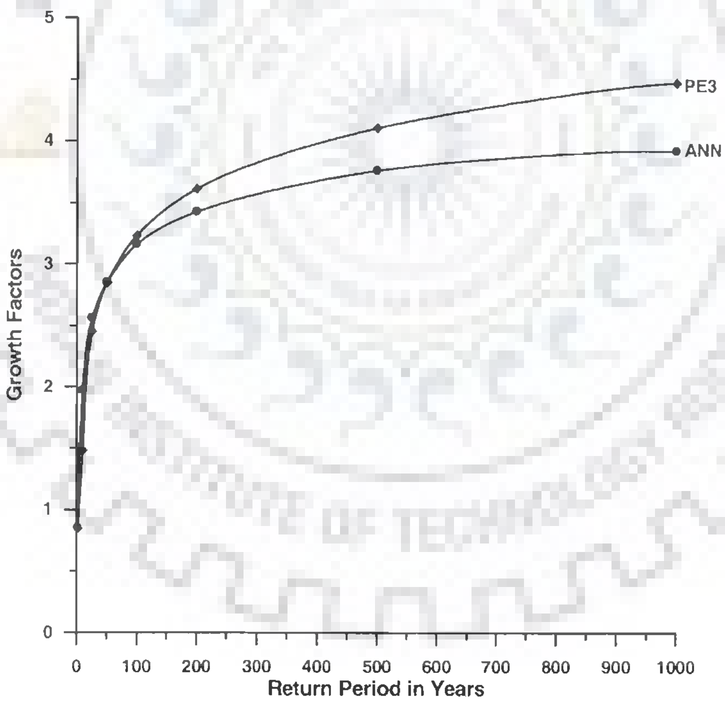


Fig. 5.4.1 Variations of growth factors with return period for ANN and L-moments for Upper Narmada and Tapi Subzone 3(c)

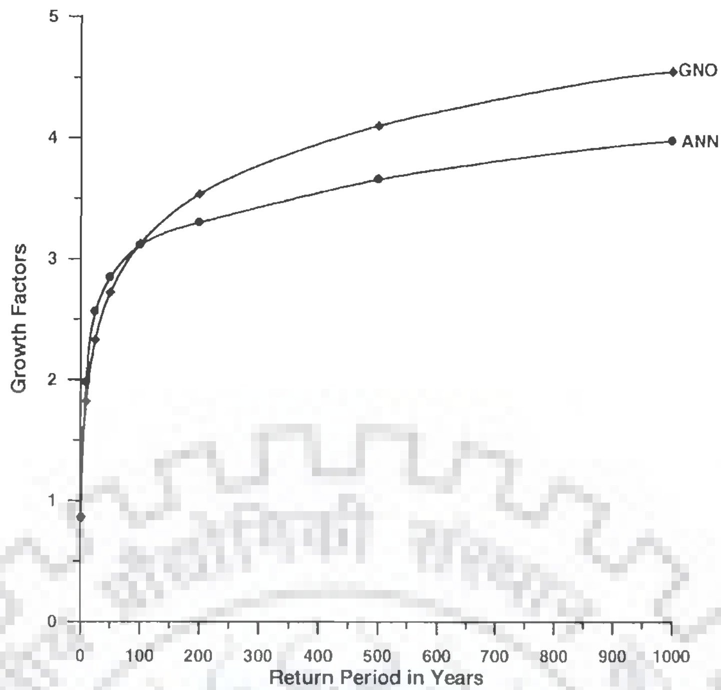


Fig. 5.4.2 Variations of growth factors with return period for ANN and L-moments for Mahanadi Subzone 3(d)

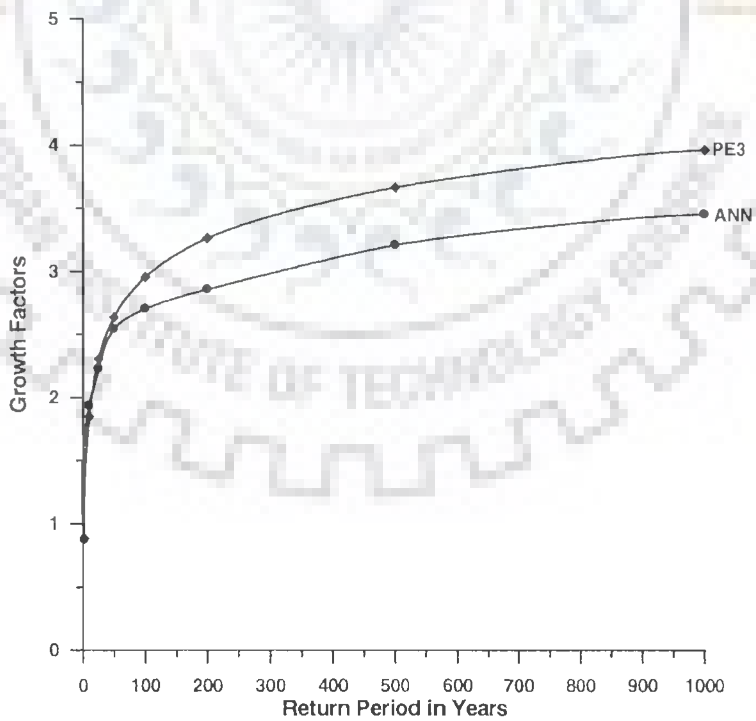


Fig. 5.4.3 Variations of growth factors with return period for ANN and L-moments for Lower Godavari Subzone 3(f)

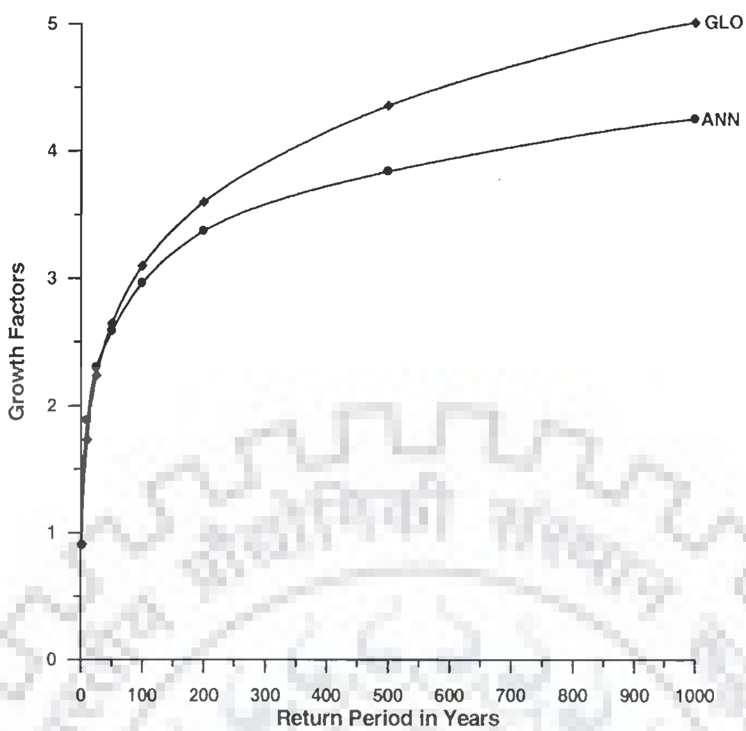


Fig. 5.4.4 Variations of growth factors with return period for ANN and L-moments for Sub-Himalayan region Zone–7

5.8.2 Development of Regional Flood Frequency Relationships for Gauged Catchments using FIS

The functional form of the model for development of regional flood frequency relationships for gauged catchments using FIS is same as that of the ANN approach described in Section 5.8.1. A Takagi-Sugeno Fuzzy Inference System (FIS) has been developed using the subtractive clustering algorithm integrated with a linear least squares estimate algorithm for the regional flood frequency estimation model. The FIS model has been developed based on the assumption that the cluster estimation method when applied to a collection of input and output data produces cluster centers where each cluster center is in essence a prototypical data point that represents a characteristic behavior of the system. Hence, each cluster center can be used as the basis of a rule that illustrates the system behavior. The algorithm for subtractive clustering is as follows. In this case also, the input-output vector has been fixed as the same as that of ANN models. The major parameter that needs to be identified in FIS model is the clustering radius. The cluster radius specifies the range of influence of the cluster centre of the each input and output dimension. Assuming that the cluster radius falls within the hyper box of unit dimension, a smaller cluster radius will yield more clusters in a data and hence a greater number of rules. Simultaneously, it increases the model complexity and decreases parsimony.

The steps involved in the development of the FIS model are as follows: (i) probability values serve as input and annual maximum peak flood values as the output variables for the fuzzy system, (ii) the set of available data (probability of non-exceedance and annual maximum peak flood values) are divided into two sets: one for system modeling (training) in which 95% of the values that are used – selected regarding characteristic features - and the other for system testing, in which 5% values of the input data set, (iii) input variables are subdivided into clusters, (iv) independent models are then generated for all input variables determined according to step-(i), (v) for each model a rule base is defined containing as much rules as the input variable possesses membership functions, (vi) the premises contain one membership function, and the model performance is judged based on the theoretical error criteria, (vii) the input variables showing dependencies of the probability and annual maximum peak flood are collected in premises and then the maximum number of rules is built and the parameters are then optimized, (viii) further improvements of results can solely be achieved if new rules are inserted for regions with maximum error values. Again this step is repeated together with optimization steps as mentioned above until there is no further improvement, (ix) the influence of the input variable to the passing times can be recognized within the rule base. Assumptions can be made, that these dependencies exist also for other (higher) probabilities values. Based on this additional rules are built for probability levels that are not available or yet displayed in the training set. By this way, an interpretation of acquired knowledge is made possible relative simply. From the interpretation of the rule base new rules can be derived, which can describe extreme situations beyond events and can predict such situations with higher accuracy.

The clustering radius is identified through a trial-and-error procedure by varying the clustering radius from 0.04 to 0.5 with an increment of 0.05. For each cluster radius, statistical performance indices were computed for training and testing by the FIS model. The model output that provided the maximum EFF and CORR as well as minimum RMSE and MAE for training and testing was selected for estimation of the growth factors for various return periods for each of the Subzones. The best values of the model output are captured for cluster radius of 0.04, 0.05, 0.05 and 0.04 for Subzone 3(c), Subzone 3(d), Subzone 3(f) and Zone-7 respectively. The values of growth factors for FIS and L-moments for Subzone 3(c), Subzone 3(d), Subzone 3(f) and Zone-7 are given in Tables 5.16.1 to 5.16.4 respectively. The statistical performance indices of FIS and L-moments for training for Subzone 3(c), Subzone 3(d), Subzone 3(f) and Zone-7 are given in Tables 5.17.1 to 5.17.4 and for validation are provided in Tables 5.18.1 to 5.18.4 respectively. The variations of growth factors with return periods estimated by FIS and L-moments are shown in Figs. 5.5.1 to 5.5.4 for the four Subzones.

Table 5.16.1 Growth factors for FIS and L-moments for Upper Narmada and Tapi Subzone 3 (c)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
FIS	0.827	1.857	2.442	2.944	3.510	3.971	4.338	4.480
L-moments	0.847	1.483	2.454	2.848	3.234	3.614	4.108	4.477

Table 5.16.2 Growth factors for FIS and L-moments for Mahanadi Subzone 3(d)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
FIS	0.828	1.862	2.464	2.863	3.366	3.833	4.238	4.402
L-moments	0.870	1.821	2.331	2.723	3.125	3.538	4.105	4.552

Table 5.16.3 Growth factors for FIS and L-moments for Lower Godavari Subzone 3(f)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
FIS	0.929	1.861	2.289	2.589	2.839	3.018	3.385	3.683
L-moments	0.886	1.849	2.308	2.637	2.955	3.265	3.665	3.962

Table 5.16.4 Growth factors for FIS and L-moments for Sub-Himalayan region Zone-7

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
FIS	0.871	1.775	2.181	2.606	3.041	3.426	4.072	4.681
L-moments	0.911	1.753	2.240	2.646	3.097	3.599	4.355	5.006

Table 5.17.1 Statistical performance indices of FIS and L-moments for training for Upper Narmada and Tapi Subzone 3 (c)

Statistical indices	FIS	L-moments
EFF	99.85	91.72
CORR	0.99	0.94
MAE	0.95	8.64
RMSE	2.62	12.38

Table 5.17.2 Statistical performance indices of FIS and L-moments for training for Mahanadi Subzone 3 (d)

Statistical indices	FIS	L-moments
EFF	99.82	94.14
CORR	0.99	0.92
MAE	1.06	7.48
RMSE	2.88	8.78

Table 5.17.3 Statistical performance indices of FIS and L-moments for training for Lower Godavari Subzone 3 (f)

Statistical indices	FIS	L-moments
EFF	99.91	92.86
CORR	0.99	0.92
MAE	1.20	10.96
RMSE	1.87	8.92

Table 5.17.4 Statistical performance indices of FIS and L-moments for training for Sub-Himalayan region Zone-7

Statistical indices	FIS	L-moments
EFF	99.86	93.47
CORR	0.99	0.91
MAE	0.96	6.13
RMSE	2.19	4.79

Table 5.18.1 Statistical performance indices of FIS and L-moments for validation for Upper Narmada and Tapi Subzone 3 (c)

Statistical indices	FIS	L-moments
EFF	98.81	90.03
CORR	0.99	0.91
MAE	1.23	14.40
RMSE	8.76	11.07

Table 5.18.2 Statistical performance indices of FIS and L-moments for validation for Mahanadi Subzone 3 (d)

Statistical indices	FIS	L-moments
EFF	99.69	93.04
CORR	0.99	0.89
MAE	0.81	14.26
RMSE	1.19	11.32

Table 5.18.3 Statistical performance indices of FIS and L-moments for validation for Lower Godavari Subzone 3 (f)

Statistical indices	FIS	L-moments
EFF	99.84	93.24
CORR	0.99	0.89
MAE	1.00	11.96
RMSE	2.32	9.54

Table 5.18.4 Statistical performance indices of FIS and L-moments for validation for Sub-Himalayan region Zone-7

Statistical indices	FIS	L-moments
EFF	98.50	91.53
CORR	0.99	0.89
MAE	1.87	17.40
RMSE	6.61	12.07

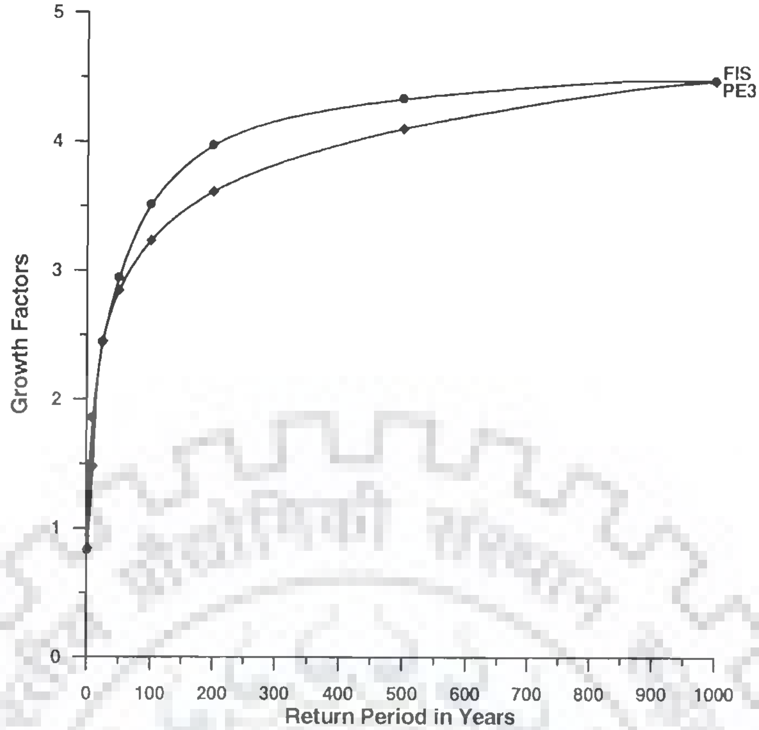


Fig. 5.5.1 Variations of growth factors with return period for FIS and L-moments for Upper Narmada and Tapi Subzone 3(c)

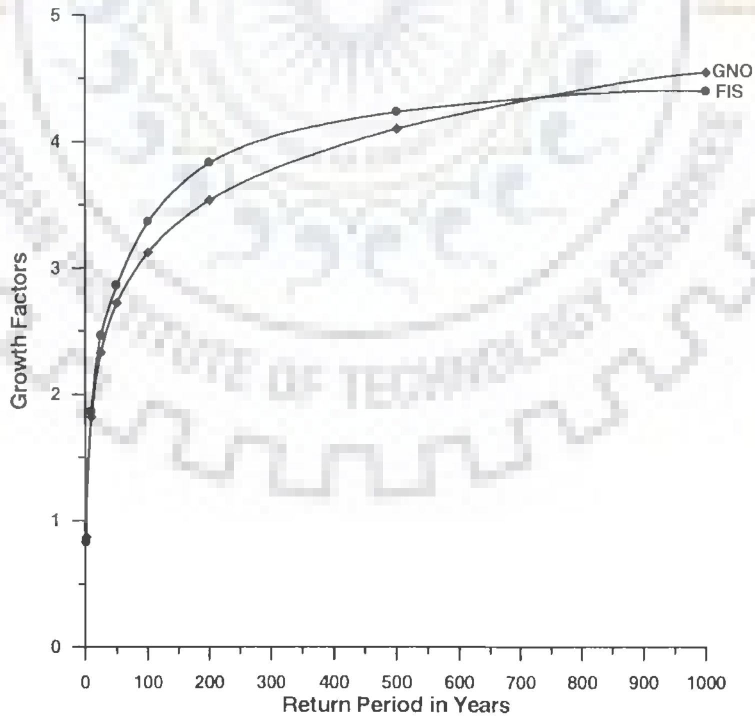


Fig. 5.5.2 Variations of growth factors with return period for FIS and L-moments for Mahanadi Subzone 3(d)

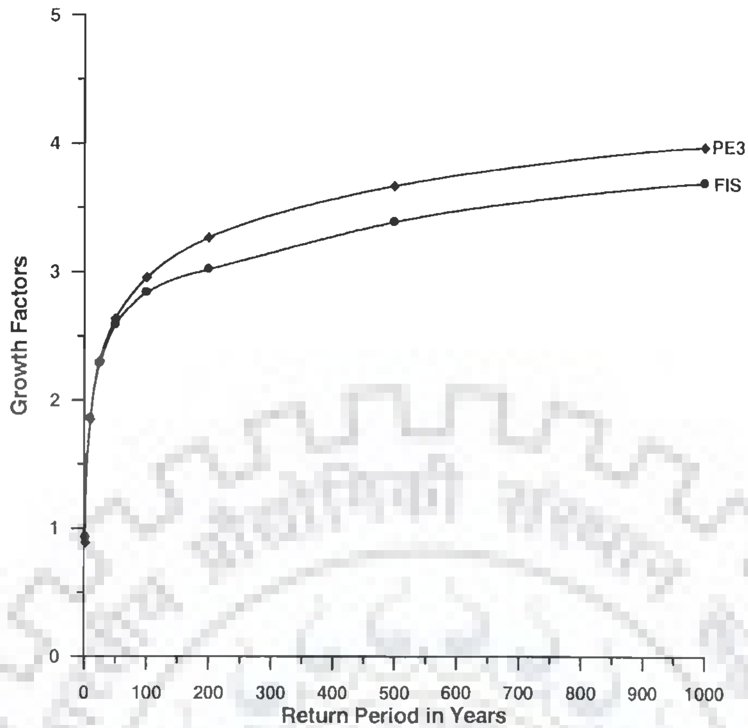


Fig. 5.5.3 Variations of growth factors with return period for FIS and L-moments for Lower Godavari Subzone 3(f)

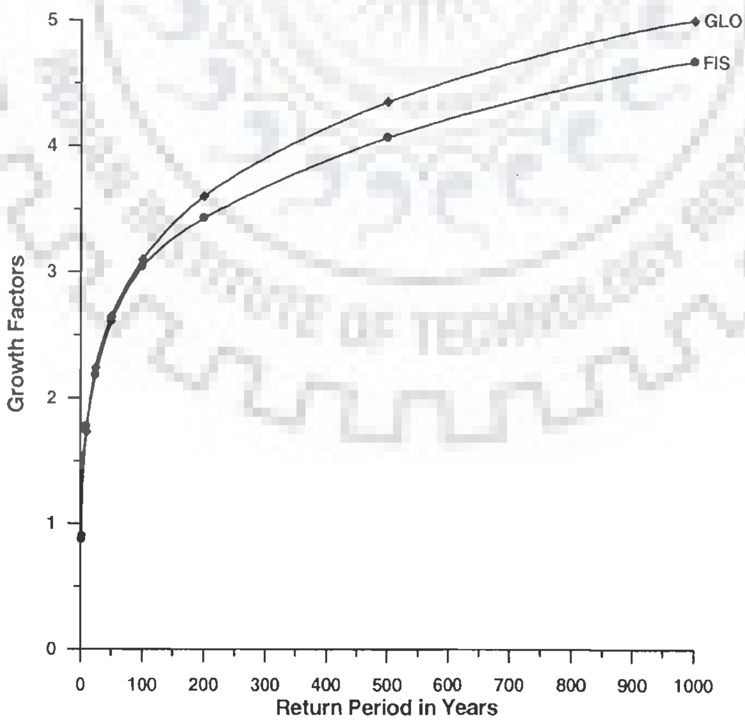


Fig. 5.5.4 Variations of growth factors with return period for FIS and L-moments for Sub-Himalayan region Zone–7

5.9 COMPARISON OF ANN, FIS AND L-MOMENTS

The growth factors developed using ANN, FIS and L-moments have been compared based on the statistical performance indices described earlier. The values of growth factors for ANN, FIS and L-moments for Subzone 3(c), Subzone 3(d), Subzone 3(f) and Zone-7 are given in Tables 5.19.1 to 5.19.4 respectively. Tables 5.20.1 to 5.20.4 provide the values of statistical performance indices for ANN, FIS and L-moments for training datasets for the four Subzones. Tables 5.21.1 to 5.21.4 give the values of statistical performance indices for ANN, FIS and L-moments for the datasets used for validation for the four Subzones. The variations of growth factors with return periods estimated by ANN, FIS and L-moments are shown in Figs. 5.6.1 to 5.6.4 for the four Subzones. Based on the statistical performance indices the performance of FIS is found to be better than that of ANN. For estimation of floods of various return periods for the gauged catchments based on the better identified soft computing technique i.e. FIS, the values of growth factors are summarized in Table 5.22 for the four Subzones.

Table 5.19.1 Growth factors for ANN, FIS and L-moments for Upper Narmada and Tapi Subzone 3 (c)

Distribution	**Return period (Years)**							
	2	10	25	50	100	200	500	1000
	Growth factors							
ANN	0.852	1.970	2.561	2.852	3.162	3.427	3.763	3.924
FIS	0.827	1.857	2.442	2.944	3.510	3.971	4.338	4.480
L-moments	0.847	1.483	2.454	2.848	3.234	3.614	4.108	4.477

Table 5.19.2 Growth factors for ANN, FIS and L-moments for Mahanadi Subzone 3(d)

Distribution	**Return period (Years)**							
	2	10	25	50	100	200	500	1000
	Growth factors							
ANN	0.856	1.980	2.564	2.848	3.113	3.303	3.660	3.979
FIS	0.828	1.862	2.464	2.863	3.366	3.833	4.238	4.402
L-moments	0.870	1.821	2.331	2.723	3.125	3.538	4.105	4.552

Table 5.19.3 Growth factors for ANN, FIS and L-moments for Lower Godavari Subzone 3 (f)

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
ANN	0.875	1.935	2.230	2.543	2.704	2.858	3.208	3.452
FIS	0.929	1.861	2.289	2.589	2.839	3.018	3.385	3.683
L-moments	0.886	1.849	2.308	2.637	2.955	3.265	3.665	3.962

Table 5.19.4 Growth factors for ANN, FIS and L-moments for Sub-Himalayan region Zone-7

Distribution	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
ANN	0.900	1.885	2.302	2.586	2.962	3.372	3.837	4.248
FIS	0.871	1.775	2.181	2.606	3.041	3.426	4.072	4.681
L-moments	0.911	1.733	2.240	2.646	3.097	3.599	4.355	5.006

Table 5.20.1 Statistical performance indices of ANN, FIS and L-moments for training for Upper Narmada and Tapi Subzone 3 (c)

Statistical indices	ANN	FIS	L-moments
EFF	98.17	99.85	91.72
CORR	0.99	0.99	0.94
MAE	3.10	0.95	8.64
RMSE	9.07	2.62	12.38

Table 5.20.2 Statistical performance indices of ANN, FIS and L-moments for training for Mahanadi Subzone 3 (d)

Statistical indices	ANN	FIS	L-moments
EFF	98.24	99.82	94.14
CORR	0.99	0.99	0.92
MAE	2.97	1.06	7.48
RMSE	8.96	2.88	8.78

Table 5.20.3 Statistical performance indices of ANN, FIS and L-moments for training for Lower Godavari Subzone 3 (f)

Statistical indices	ANN	FIS	L-moments
EFF	98.32	99.91	92.86
CORR	0.99	0.99	0.92
MAE	3.32	1.20	10.96
RMSE	7.99	1.87	8.92

Table 5.20.4 Statistical performance indices of ANN, FIS and L-moments for training for Sub-Himalayan region Zone-7

Statistical indices	ANN	FIS	L-moments
EFF	98.12	99.86	93.47
CORR	0.99	0.99	0.92
MAE	3.50	0.96	6.13
RMSE	4.16	2.19	4.79

Table 5.21.1 Statistical performance indices of ANN, FIS and L-moments for validation for Upper Narmada and Tapi Subzone 3 (c)

Statistical indices	ANN	FIS	L-moments
EFF	98.10	98.81	90.03
CORR	0.99	0.99	0.92
MAE	3.69	1.23	14.40
RMSE	9.11	8.76	11.07

Table 5.21.2 Statistical performance indices of ANN, FIS and L-moments for validation for Mahanadi Subzone 3 (d)

Statistical indices	ANN	FIS	L-moments
EFF	99.75	99.69	93.04
CORR	0.99	0.99	0.89
MAE	2.39	0.81	14.26
RMSE	3.34	1.19	11.32

Table 5.21.3 Statistical performance indices of ANN, FIS and L-moments for validation for Lower Godavari Subzone 3 (f)

Statistical indices	ANN	FIS	L-moments
EFF	99.06	99.84	93.24
CORR	0.99	0.99	0.89
MAE	2.89	1.00	11.96
RMSE	5.60	2.32	9.54

Table 5.21.4 Statistical performance indices of ANN, FIS and L-moments for validation for Sub-Himalayan region Zone-7

Statistical indices	ANN	FIS	L-moments
EFF	99.05	98.50	91.53
CORR	0.99	0.99	0.89
MAE	4.72	1.87	17.40
RMSE	8.06	6.61	12.07

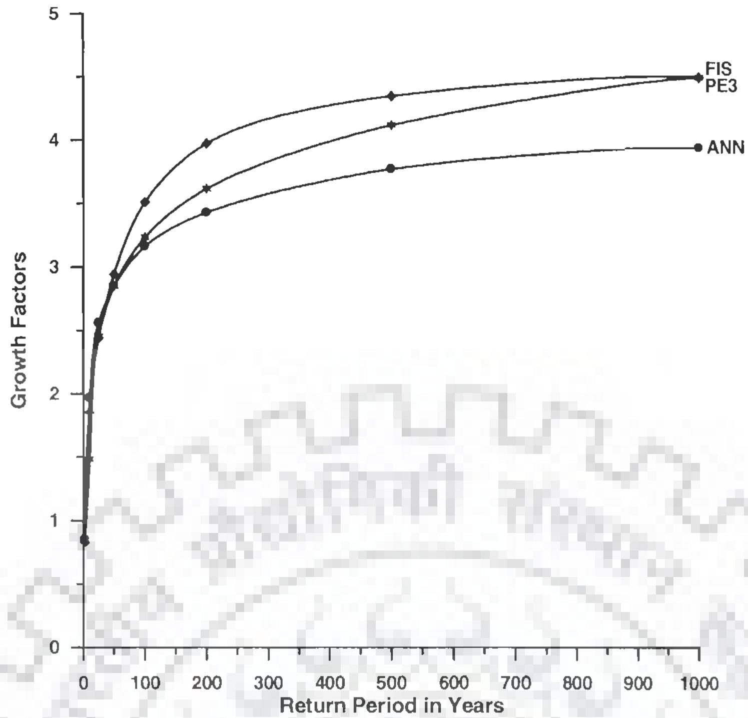


Fig. 5.6.1 Variations of growth factors with return period for ANN, FIS and L-moments for Upper Narmada and Tapi Subzone 3(c)

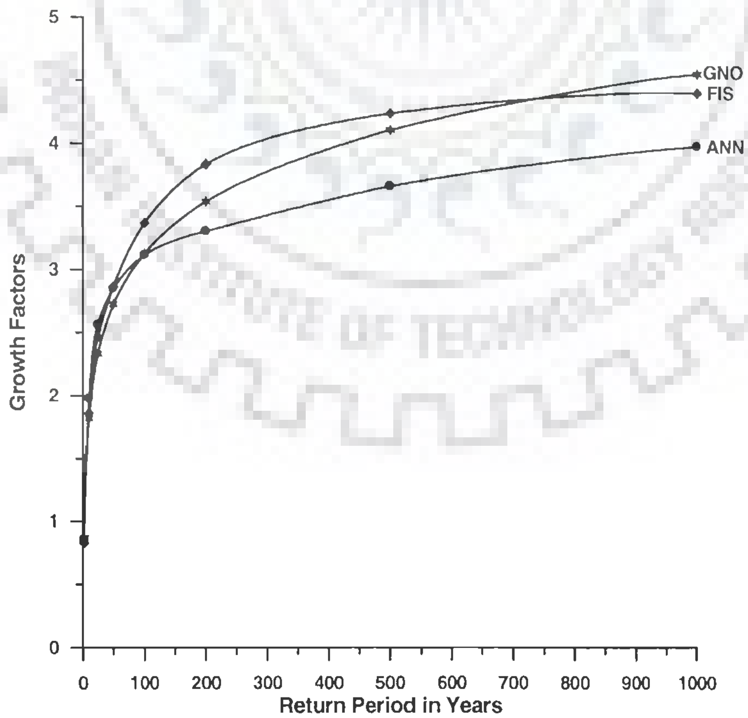


Fig. 5.6.2 Variations of growth factors with return period for ANN, FIS and L-moments for Mahanadi Subzone 3(d)

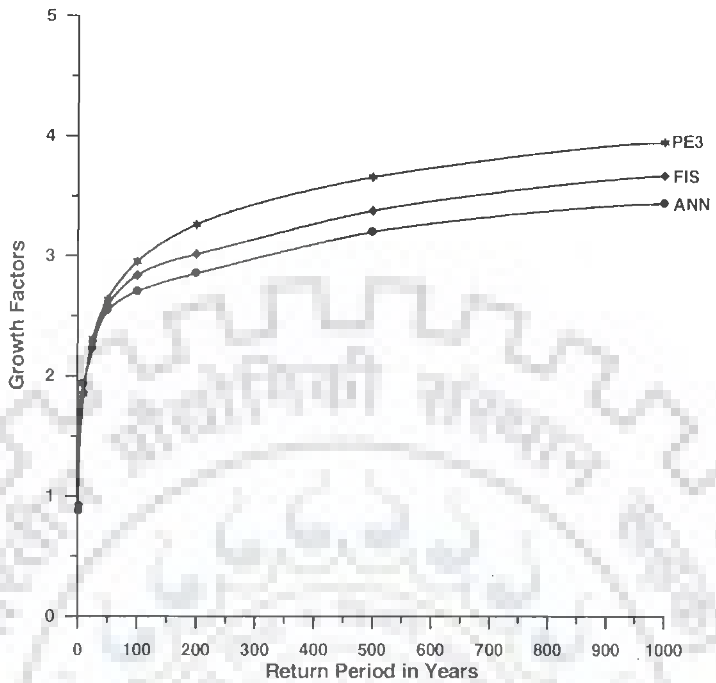


Fig. 5.6.3 Variations of growth factors with return period for ANN, FIS and L-moments for Lower Godavari Subzone 3(f)

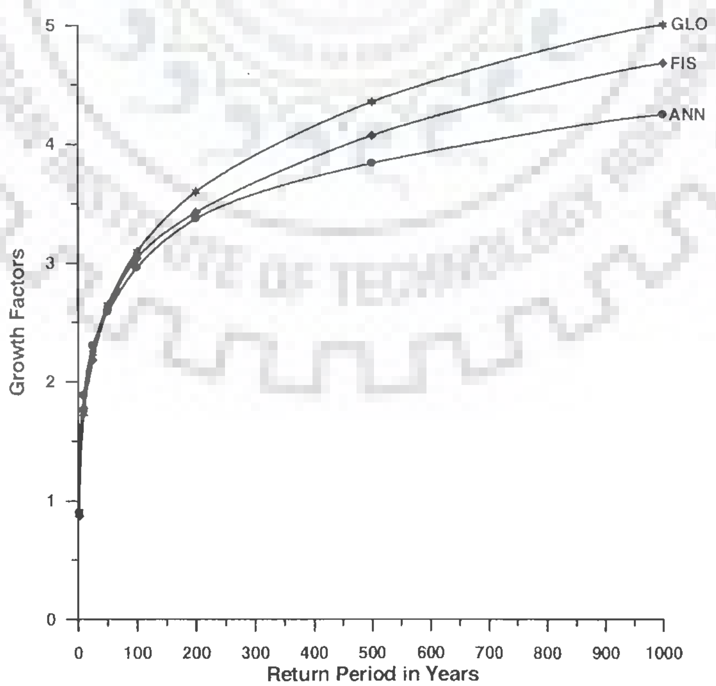


Fig. 5.6.4 Variations of growth factors with return period for ANN, FIS and L-moments for Sub-Himalayan region Zone–7

Table 5.22 Values of growth factors estimated by FIS for four Subzones of India

Subzone	Return period (Years)							
	2	10	25	50	100	200	500	1000
	Growth factors							
3(c)	0.827	1.857	2.442	2.944	3.510	3.971	4.338	4.480
3(d)	0.828	1.862	2.464	2.863	3.366	3.833	4.238	4.402
3(f)	0.929	1.861	2.289	2.589	2.839	3.018	3.385	3.683
Zone-7	0.871	1.775	2.181	2.606	3.041	3.426	4.072	4.681

5.10 DEVELOPMENT OF REGIONAL FLOOD FREQUENCY RELATIONSHIPS FOR UNGAUGED CATCHMENTS USING FIS

As discussed above based on the statistical performance indices the performance of FIS technique has been found to be better than that of the ANN and L-moments. Hence, for development of regional flood frequency relationships for ungauged catchments for the four Subzones the growth factors estimated using the FIS have been coupled with the regional relationships between mean annual peak floods and catchment areas for the respective four Subzones. In this manner the regional flood frequency relationships have been developed for estimation of floods of various return periods for the ungauged catchments of the four Subzones based on FIS. The values of C_T (equation 5.2) and 'b' for FIS for the four Subzones are given in Table 5.23.

Table 5.23 Values of regional coefficients 'C_T' for FIS and 'b' for four Subzones of India

S. No.	Sub-zone	Coeff. 'b'	Return Period (Years)							
			2	10	25	50	100	200	500	1000
			'C_T'for four Subzones							
1.	3 (c)	0.547	19.392	43.545	57.262	69.034	82.306	93.116	101.722	105.052
2.	3 (d)	0.863	2.086	4.690	6.207	7.212	8.479	9.655	10.676	11.089
3.	3 (f)	0.676	9.581	19.192	23.606	26.700	29.279	31.125	34.910	37.983
4.	Zone 7	0.387	55.393	112.885	138.705	165.734	193.398	217.883	258.967	297.698

CHAPTER 6

CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

6.1 CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH WORK

In the present study regional flood frequency relationships have been developed based on the L-moments approach for gauged and ungauged catchments of the 17 Subzones of India. The applicability of soft computing techniques viz. ANN and FIS in regional flood frequency estimation has also been investigated. The results of the study would be useful for the practitioners especially engaged in planning, development and management of water resources projects. The analysis and results reported in the present work leave sufficient scope for further investigations, which could not be taken up owing to time constraint and are briefed along with the conclusions as follows.

i. Regional flood frequency analysis has been carried out based on L-moments approach, using the annual maximum peak flood data of 261 catchments of the 17 Subzones covering about 79% of the geographical area of India. After conducting the L-moments based Discordancy statistic (D_i) test for screening the data for suitability for regional flood frequency analysis and testing the regional homogeneity employing the heterogeneity measure (H) data of 196 streamflow gauging sites have been used in the study.

ii. Twelve frequency distributions viz. Extreme value (EV1), Normal (NOR), General extreme value (GEV), Logistic (LOS), Generalized logistic (GLO), Generalized normal (GNO), Uniform (UNF), Exponential (EXP), Generalized

Pareto (GPA), Pearson Type-III (PE3), Kappa (KAP) and five parameter Wakeby (WAK) have been used in the study. The regional parameters of the frequency distributions have been estimated using the L-moments approach. Based on the L-moment ratio diagram as well as $|Z_i^{dist}|$ –statistic criteria, robust frequency distributions have been identified for the 17 Subzones. It is observed that out of the 17 Subzones PE3 distribution is the robust distribution for 7 Subzones, GNO for 3 Subzones, GEV for 3, GPA for 3 and GLO for 1 Subzone of India.

iii. For estimation of floods of various return periods for gauged catchments of the 17 Subzones regional flood frequency relationships have been developed based on the respective robust identified frequency distributions.

iv. For estimation of floods of various return periods for ungauged catchments the robust identified L-moments based regional flood frequency relationships of the 17 Subzones have been coupled with the respective regional relationships developed between mean annual peak floods and catchment areas and regional flood frequency relationships have been developed. The tabular and graphical forms of these regional flood frequency relationships have also been prepared for estimation of floods of various return periods for ungauged catchments.

v. The regional flood frequency relationships have also been developed for gauged catchments using the soft computing techniques viz. ANN and FIS for four Subzones viz. Subzone 3(c), Subzone 3(d), Subzone 3(f) and Zone-7. The performances of ANN, FIS and L-moments have been compared based on the statistical performance indices viz. CORR, EFF, RMSE and MAE. Based on the comparison of ANN, FIS and L-moments the potential of applicability of FIS in regional flood frequency estimation has been established. Regional flood frequency relationships have also been developed for ungauged catchments for four Subzones by coupling the regional relationships between mean annual peak floods and catchment areas with the respective regional flood frequency relationships developed using FIS.

vi. As the regional flood frequency relationships have been developed using the data of catchments ranging in areal extent from 6 km^2 to 2,297 km^2; therefore, the developed regional flood frequency relationships may be expected to provide reliable flood frequency estimates for the catchments of the respective 17 Subzones, lying nearly in the same range of areal extent, as those of the input data. Further the statistical performance indices viz. CORR, EFF, RMSE and MAE of the relationships developed between mean annual peak floods and catchment areas for the 17 Subzones give the degree of accuracy of the regional relationships and the results of the study are subject to these limitations.

vii. The developed regional flood frequency relationships may be refined for obtaining more accurate flood frequency estimates, when the annual maximum peak flood data for some more streamflow gauging sites become available and physiographic as well as the climatic characteristics other than catchment area are also used for development of the regional flood frequency relationships.

viii. More studies are required to be taken up for evaluation of applicability of the soft computing techniques in regional flood frequency estimation.

REFERENCES

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, 84, 365-380.

Ali, S. and Singh, R. (2001) Regional runoff frequency curve for small watersheds in the hirakud catchment of eastern India. Applied Engineering in Agriculture. Vol. 17(3): 285-292.

Anderson, M.L., Mierzwa, M.D. and Kavvas, M.L. (2000) Probabilistic seasonal forecasts of droughts with a simplified coupled hydrologic-atmospheric model for water resources planning. Stochastic Environmental Research and Risk Assessment, 14(4-5), 263-274.

Anmala, J., Govindaraju, R.S. and Zhang, B. (2000) Comparison of ANNs and other empirical approaches for predicting runoff. Journal of Water Resources, Planning and Management (ASCE), 126(3), pp. 156-166.

Aqil, M., Kita, I., Yano, A. and Nishiyama, S. (2006) A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff. Journal of Hydrology, 337, 22-37, doi: 10. 1016/j.jhydrol. 2007.01.013.

ASCE Task Committee on Application of ANNs in hydrology (2000a) ANNs in hydrology I: Preliminary concepts. Journal of Hydrologic Engineering (ASCE), 5(2), 115-123.

ASCE Task Committee on application of ANNs in hydrology (2000b) ANNs in hydrology II: Hydrologic Applications. Journal of Hydrologic Engineering (ASCE), 5(2), 124-137.

Baldwin, J.F. (1996) (editor) Fuzzy logic. John Wiley and Sons, Chicester.

Barnett, V. and Lewis, T. (1994) Outliers in statistical data. 3rd ed., Wiley, Chichester, U.K.

Bell, F.C. (1968) Estimating design floods from extreme rainfall. Colorado, State Univ., U.S.A. Hydrology Paper No. 29.

Benson, M.A. (1962) Evolution of methods for evaluating the occurrence of floods. U.S. Geol. Surv. Water-Supply Paper: 1550-A.

Bhadra, A., Panigrahy, N., Singh, R., Raghuwanshi, N.S., Mal, B.C. and Tripathi, M.P. (2008) Development of a geomorphological instantaneous unit hydrograph model for scantily gauged watersheds. Env. Modelling & Software, 26, 1013-1025.

Bhatt, V.K. (2003) Estimation of extreme flows for ungauged catchments. P.hd. Thesis, Indian Institute of Technology Roorkee.

Bhattacharjya, R.K., Datta, B. and Satish, M.G. (2007) Artificial neural networks approximation of density dependent saltwater intrusion process in coastal aquifers. Journal of Hydrologic Engineering, 12(3), 273-282.

Bhunya, P.K, Berndtsson, R., Ojha, C.S.P. and Mishra, S.K. (2007) Suitability of Gamma, Chi-square, Weibull and Beta distributions as synthetic unit hydrographs, Vol-334, 28-38, Journal of hydrology, Elsevier.

Bhunya, P.K., Berndtsson, R., Singh, P.K. and Hubert, P. (2008) Comparison between Weibull and gamma distributions to derive synthetic unit hydrograph using Horton ratios, Vol-44, WR006031, Water Resources Research, AGU.

Bobee, B. and Ashkar, F. (1991) The gamma family and derived distributions applied in hydrology. Water Resources Publications, Littleton, CO.

Boes, D.C., Heo, J.H. and Salas, J.D. (1989) Regional flood quantile estimation for a Weibull model. Water Resources Research, 25(5), 979-990.

Brown, M. and Harris, P. (1994) Application of a neuro-fuzzy technique for rainfall-runoff modelling. Proceeding of Aesthetics in the Constructed Environment (ASCE), New York, 68-73.

Burn, D.H. (1989) Cluster as applied to regional flood frequency. Journal of Water Resources Planning and Management, 115, 567-582.

Burn, D.H. and Goel, N.K. (2000) The formation of groups for regional flood frequency analysis. Hydrol. Sci. J. 45(1), 97-112.

Central Water Commission, (1982) Flood estimation report for Mahanadi Subzone 3(d). Directorate of Hydrology (Small Catchments), New Delhi.

Central Water Commission (1985) Flood estimation for Middle Ganga Plain Subzone 1(f). Report No. GP/10/1984, Directorate of Hydrology (Small Catchments), New Delhi.

Cervantes, J.E., Kavvas, M.L. and Delleur, J.W. (1983) Cluster model of flood analysis. Water Resources Research, 19(1), 209-224.

Chander, S., Sapolia, S.K. and Kumar, A. (1978) Flood frequency analysis by power transformation. Journal of Hydraulics Division (ASCE), 104(11), 1495-1505.

Chang, L. C., J. Chang and Tsai, Y.H. (2005) Fuzzy exemplar-based inference system for flood forecasting. Water Resources Research, 41, W02005.

Chatterjee, C., Forster, S. and Bronstert, A. (2008) Comparison of hydrodynamic models of different complexities to model floods with emergency storage areas. Hydrological Processes, 22, 4695-4709 in Wiley InterScience, doi:10.1002/hyp.7079.

Chebana, F. and Ouarda, T.B.M.J. (2007) Multivariate L-moment homogeneity test. Water Resources Research, 43, W08406.

Cheng, C.T., Oub, C.P. and Chauc, K.W. (2002) Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall–runoff model calibration. Journal of Hydrology, 268, 72–86.

Chokmani, K. and Ouarda, T.B.M.J. (2004) Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. Water Resources Research, 40, W12514.

Chow, V.T. (1954) The log probability law and its engineering applications. Proceeding of Am. Soc. Civil. Engrs., November, vol. 80 paper no. 536, pp. 1-25.

Chow, V.T. (1964) (Editor in Chief), Handbook of applied hydrology. Mcgraw Hill Book Company.

Chowdhury, J.U., Stedinger, J.R. and Lu, L.H. (1991) Goodness of fit tests for regional generalized value extreme value flood distributions. Water Resources Research, 27(7), 1765-1776.

Cong, S. Li., Vogel, J.L. and Schaake, J.C. (1993) Identification of the underlying distribution form of precipitation by using regional data. Water Resources Research, 29(4), 1103-1111.

Coulibaly, P., Anctil, F. and Bobee, B. (2000) Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. Journal of Hydrology, 230, 244-257.

Cunderlik, J.M. and Burn, D.H. (2006) Switching the pooling similarity distances: mahalanobis for euclidean. Water Resources Research, 42, W03409.

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

Cunnane, C. (1989) Statistical distributions for flood frequency analysis. W.M.O. No. 718. Operational Hydrology Report No. 33, Geneva.

Dalrymple, T. (1960) Flood Frequency Analyses. Water Supply Paper, 1543-A, U.S., Geological Survey, Reston, Va.

Dalrymple, T. (ed.) (1960) Flood Frequency Analysis. Manual of Hydrology, pt 3, Flood flow Techniques, U.S. Geological Survey, Water Supply Paper 1543-A.

Durrans, S.R., Eiffe, M.A., Thomas Jr, W.O. and Goranflo, H.M. (2003) Joint seasonal/annual flood frequency analysis. Journal of Hydrologic Engineering, vol. 8(4), 181-189.

Farquharson, J.R. (1992) Regional flood frequency analysis in arid and semi arid areas. Journal of Hydrology, Vol. 138, 487-501.

Filliben, J.J. (1975) The probability plot correlation test for normality. Technometrics, vol. 17(1), 111-117.

Forster, H.A. (1924) Theoretical frequency curves and their application to engineering problems. Trans. Am. Soc. Civil Engrs., vol. 87, 142-173.

Forster, S., Kneis, D., Gocht, M., Bronstert, A. (2005) Flood risk reduction by the use of retention areas at the Elbe river. International Journal of river basin management 3(1): 21-29.

Forster, S., Chatterjee, C. and Bronstert, A. (2008) Hydrodynamic simulation of the operational management of a proposed flood emergency storage area at the

Middle Elbe River. River research and application: River. Res. Applic. in Wiley Inter Science doi: 10.1002/rra.1090.

Fuller, W.E. (1914) Flood flows. Trans. Am. Soc. Civil Engrs., vol.77, 625.

Garbrecht, J.D. (2006) Comparison of three alternative ANN designs for monthly rainfall-runoff simulation. Journal of Hydrologic Engineering, 11(5), 502-505.

Genest, C., Favre, A.C., Beliveau, J. and Jacques C. (2007) Metelliptical copulas and their use in frequency analysis of multivariate hydrological data. Water Resources Research, 43, W09401.

Gill, P.R., Murray, W., Wright, M.H. (1981) The Levenberg-Marquardt method. Practical Optimization. London: Academic Press, pp. 136-137.

Goswami, A. (1972) Regional flood frequency analysis of Brahmputra basin in North-East India. Journal of Institution of Engineers (India), Civil Engineering Division, Vol. 52, 169-174.

Goyal, R. and Arora, A.N. (2007) Multivariate statistical analysis of remote sensing data. First National Conf. on GVDRP-2007, National Institute of Design, Ahmedabad, 28-30 November.

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

Gries, N.P. and Wood, E.F. (1983) Regional flood frequency estimation and network design. Water Resources Research, Vol. 19, No. 4, 1167-1177.

Griffis, V.W. and Stedinger, J.R. (2007a) Evolution of flood frequency analysis with bulletin 17. Journal of Hydrologic Engineering, 12(3), 283-297.

Griffis, V.W. and Stedinger, J.R. (2007b) Log-pearson type 3 distribution and its application in flood frequency analysis, I: distribution characteristics. Journal of Hydrologic Engineering, 12(5), 482-491.

Gumbel, E. J. (1941) Probability interpretation of the observed return periods of floods. Trans. Am. Geophys. Union, vol. 21, 836-850.

Gumbel, E. J. (1941) The return period of flood flows. Ann. Math. Statist., June, vol. XII(2), 163-190.

Gumbel, E. J. (1949) The statistical forecast of floods. Bull. 15, the Ohio Water Resources Board, Columbus, Ohio.

Gumbel, E. J. (1954) Statistical theory of droughts. Proceeding of Am. Soc. of Civil Engrs., May, vol. 80(439), 1-19.

Haykin, S. (1994) Neural Networks: A comprehensive foundation. New York Macmillan.

Hazen, A. (1914) Discussion on flood flows by W. E. Fuller. Trans. Am. Soc. Civil Engrs., vol.77, p.625.

Hazen, A. (1921) Discussion of probable variations in yearly runoff by L.S. Hall. Trans. Am. Soc. Civil Engrs., vol.84, 214-224.

Hebson C. S. and Cunnane C. (1986) Assessment of use of at-site and regional flood data for flood frequency estimation. Hydrologic frequency modelling ed. By V.P. Singh D. Reidel Publishing Company, 433-448.

Hosking, J.R.M. (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistic. J. Royal Stat. Soc., Series B, 52(2), 105-124.

Hosking, J.R.M. and Wallis, J.R. (1997) Regional frequency analysis - an approach based on L-moments. Cambridge University Press, New York.

Huq, S.M., Nair R. and Sil, S.S. (1986) Frequency flood formulae for country wide application. Proceedings of 53rd Annual R & D Session, CBIP, Bhubaneswar, 405-416.

Iacobellis, V. and Fiorentino, M. (2000) Derived distribution of floods based on the concept of partial area coverage with a climatic appeal. Water Resources Research, 36(2), 469 - 482.

Interagency Advisory Committee on Water Data, (1982) Guidelines for determining flood flow frequency. Bull. 17B of the Hydrology Sub-Committee, Office of Water Data Coordination, Geological survey, U.S. Dept. of Interior, Washington, D.C.

Jacoby, S.L.S., Kowalik, J.S. and Pizzo, J.T. (1972) Interative methods for nonlinear optimization problems. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Jain, A. and Indurthy, S.K.V.P. (2003) Comparative analysis of event based rainfall - runoff modeling techniques - deterministic, statistical and ANNs. Journal of Hydrologic Engineering, ASCE, 8 (2), 93-98.

Jain, S.K., Singh, V.P. and Genuchten, M. Th. Van (2004) Analysis of soil water retention data using artificial neural networks. Journal of Hydrologic Engineering, 9(5), 415-420.

Jakhade, G.S., Murti, A.S. and Sethuraman, S. (1984) Frequency floods in Brahmputra valley - regional flood frequency approach. CBIP, Journal, New Delhi, 41(1), 41-48.

Jenkinson, A.F. (1955) The frequency distribution of the annual maximum (or minimum) of meteorological elements. Quar. J. Royal Meteor. Soc., 81, 158-171.

Jin, M. and Stedinger, J.R. (1989) Flood frequency analysis with regional and historical information. Water Resources Research, 25(5), 925-936.

Jingyl, Z. & Hall, M.J. (2004) Regional flood frequency analysis for the Gan-Ming river basin in China. Journal of Hydrology, 296, 98-117.

Karim, M.A. and Chowdhury, J.U. (1995) A comparison of four distributions used in flood frequency analysis in Bangladesh. Hydrol. Sci. J., 40 (1), 55-66.

Kaur, S. (1988) Moderating outliers in rainfall frequency analysis. International Conference on Hydrology of Extremes, organized by National Institute of Hydrology, Roorkee, Dec. 1-3.

Kavvas, M.L. (2003) Nonlinear hydrologic process: conservation equation for determining their means and probability distributions. ASCE Journal of Hydrologic Engineering, Vol. 8, No. 2, 44-53.

Kendall, M.G. (1975) Rank correlation methods. Charles Griffin, London.

Keskin, M.E. and Ozlem, T. (2006) Artificial neural network models of daily pan evaporation. Journal of Hydrologic Engineering, 11(1), 65-70.

Kirby, W. (1974) Algebraic boundness of sample statistics. Water Resources Research, 10(2), 220-222.

Kisi, O. (2007) Streamflow forecasting using different artificial neural network algorithms. Journal of Hydrologic Engineering, 12(5), 532-539.

Kite, G.W. (1977) Frequency and risk analysis in hydrology. Water Resources Publications, Fort Collins CO.

Kite, G.W. (1988) Frequency and risk analysis in hydrology. Water Resources Publication, Littleton, Colo.

Kjeldsen, T.R. and Jones, D.A. (2006) Prediction uncertainty in a median-based index flood method using L-moments. Water Resources Research, 42, W07414.

Kuester, J.L. and Mize J.H. (1973) Optimization techniques with fortran. McGraw-Hill Book Company: New York St. Louis San Francisco Dusseldorf London Mexico Panama Sydney Toronto Johannesburg Kuala Lumpur Montreal, New Delhi, Rio de Janeiro, Singapore.

Kuczera G. (1982) Robust flood frequency models. Water Resources Research, 15(5), 315-324.

Kumar, M., Raghuwanshi, N.S. and Singh, R. (2009) Development and validation of GANN model for evapotranspiration estimation. No. 10.1061/(ASCE) 1084-0699, 14:2(131).

Kumar, M., Raghuwanshi, N.S. and Singh, R., Wallender, W.W. and Pruitt, W.O. (2002) Estimating evapotranspiration using ANN. J. Irrig. Drain. Eng., 128(4), 224-233.

Kumar, R and Singh, R.D. (1996) Development of regional flood frequency relationships and flood formulae for various Subzones of Zone 3 of India. Technical Report, TR (BR)-149, Roorkee.

Kumar, R. and Chatterjee, C. (2005) Regional flood frequency analysis using L-moments for North Brahmaputra Region of India. Journal of Hydrologic Engineering (ASCE), 10(1), 1-7.

Kumar, R., Chatterjee, C. and Kumar, S. (2003) Regional flood formulas using L-moments for small watersheds of Sone Subzone of India. Journal of Applied Engineering in Agriculture, American Society of Agricultural Engineers. 19(1), 47-53.

Kumar, R., Chatterjee, C., Kumar, S., Lohani, A.K. and Singh, R.D. (2003) Development of regional flood frequency relationships using L-moments for Middle Ganga Plains Subzone 1(f) of India. International Journal of Water Resources Management, Kluwer Academic Publishers, Netherlands, August, 17(4), 243-257.

Kumar, R., Singh, R.D. and Seth, S.M. (1999) Regional flood formulas for seven Subzones of Zone 3 of India. Journal of Hydrologic Engineering (ASCE), July, 4(3).

Kurothe, R.S. (1995) Development of physically based flood frequency model. P.hd. Thesis, Department of Hydrology, University of Roorkee, Roorkee.

Kurothe, R.S., Goel, N.K. and Mathur, B.S. (1997) Derieved flood frequency Distribution for negatively correlated rainfall intensity and duration. Water Resour. Res., 33, 2103-2107.

Kurothe, R.S., Goel, N.K. and Mathur, B.S. (2001) Derivation of curve number and kinematic wave based flood frequency distribution. Hydrol. Sci. J., 46(4), 571-584.

Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979a) Probability weighted moments compared with some traditional techniques of estimating gumbel parameters and quantiles. Water Resources Research, 15(6), 1361.

Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979b) Estimation of parameters and quantiles of Wakeby distributions: 2 unknown lower bounds. Water Resources Research, 15(6).

Larson, C.L. and Reich, B.M. (1972) Relationship of observed rainfall and runoff recurrence interval. Proceedings of Second Int. Symp. on Hydrology, Colorado, U.S.A.

Lettenmaier D.P., Wallis J.R. and Wood E.F. (1987) Effect of regional heterogeneity on flood frequency estimation. Water Resources Research, 23(2), 313-323.

Lettenmaier, D.P. and Potter, K.W. (1985) Testing flood frequency estimation methods using a regional flood generation model. Water Resources Research, 21, 1903-1914.

Levenberg, K. (1944) A method for the solution of certain non-linear problems in least squares. The Quarterly of Applied Mathematics 2: 164-168.

Lohani, A.K. (2007) ANN and fuzzy logic in hydrological modelling and flood forecasting. Ph.D. thesis, Dept. of Hydrology, Indian Institute of Technology Roorkee.

Lu, L.H. and Stedinger, J.R. (1992) Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. Journal of Hydrology, 138, 223-245.

Maidment, D.R. (1992) Handbook of hydrology. Mc Graw-Hill, Inc., New York.

Maier, H.R. and Dandy, G.C. (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and application. Environ. Modell. Software, 15, 101-124.

Mamdani, E.H. (1977) Application of fuzzy logic to approximate reasoning. IEEE trans. Comput. 26, 1182-1191.

Marquardt, D.M. (1963) An algorithm for Least-Squares estimation of nonlinear parameters, J. Soc. Indust. Appl. Math., 11, 431-441.

Martins, E.S. and Stedinger J.R. (2000) Generalised maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resources Research, 36(3), 737 - 744.

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

McCuen, R.H. (1993) Microcomputer applications in statistical hydrology. Prentice Hall, Englewood Cliffs, NJ.

Mehta, H.L. and Sharma, D.K. (1986) Flood frequency analysis. Proceedings of Workshop on Flood Estimation in Himalayan Region, Roorkee (U.P.), 13-23.

Merz, R. and Bloschl, G. (2005) Flood frequency regionalization – spatial proximity vs. catchment attributes. Journal of Hydrology, 302, 283-306.

Mohan, S. (2007) Artificial neural network modelling, State of Art Report, INCOH, National Institute of Hydrology, Roorkee. No: INCOH/SAR-28/2007.

Moss, M.E. (1969) Maximization of net benefit from a stream gauge. Proc. of Fiftieth Annual Meeting of the American Geophysical Union, Washington, D.C.

Mukerji, A., Chatterjee, C. and Raghuwanshi, N.S. (2009) Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. Journal of Hydrologic Engineering, Vol. 14, No. 6. doi: 10.1061/ (ASCE)HE. 1943-5584.0000040.

National Institute of Hydrology, (1984-85) Methodology for estimation of design storm. TN-12, Roorkee.

Natural Environmental Research Council, (1975) Flood studies report. 1-Hydrological Studies, London.

National Institute of Hydrology, (1990-91) Regional flood frequency analysis for Godavari basin Subzone-3(f). Technical Report No. TR-59, Roorkee.

National Institute of Hydrology, (1994-95) Development of regional flood formula for Mahanadi Subzone 3(d). Technical Report TR (BR)-134, Roorkee.

National Institute of Hydrology, (1994-95) Regional flood frequency analysis for Upper Narmada and Tapi subzone-3(c). Technical Report TR (BR)-133, Roorkee.

National Institute of Hydrology, (1997-98) Regional flood frequency analysis using L-moments. Technical Report TR (BR)-1/97-98, Roorkee.

Nash J.E. and Shaw B.L., (1965) Flood frequency as a function of catchment characteristic. Proc. of Symposium on river flood hydrology, Inst. Civ. Engnrs., London, 115-136.

National Institute of Hydrology, (1992) Hydrologic design criteria. Course Material of Regional Course on Project Hydrology, Roorkee.

National Research Council, (1988) Estimating probabilities of extreme floods-methods and recommended research. National Academy Press, Washington, D.C.

Natural Environmental Research Council (1975) Flood studies report. Vol. I, Hydrological Studies, London.

Nauck, D, and Kruse, R. (1997) A neuro-fuzzy method to learn fuzzy classification rules from data. Fuzzy sets and Systems, 89, 277-288

Nayak, P.C. and Sudheer, K.P. (2007) Fuzzy model identification based on cluster estimation for reservoir inflow forecasting. Hydrological Processes, 22, 827-841.

Nayak, P.C., Sudheer, K.P. and Jain, S.K. (2007) Rainfall-runoff modeling through hybrid intelligent system. Water Resources Research, Vol. 43, W07415, doi: 10.1029/2006WR004930.

Pal, S.K. and Mitra, S. (1999) Neuro-fuzzy pattern recognition-methods in soft computing. John Wiley & Sons, INC., New York.

Panigrahi, D.P. & Mujumdar, P.P. (2000) Reservoir operation modeling with fuzzy logic. Water Resources Management, 14, 89-109.

Parida, B.P. and Moharram, S.H. (1999) Choice of generalized pareto distribution as a potential candidate for flood frequency analysis. Hydrology Jour., 22(1-4), 1-13.

Parmeswaran, P.V., Singh, J.P, Prasad, J. and Prasad, H.J.S. (1999) Flood frequency studies of Upper Godavari basins in Maharashtra. Proc. of the National Workshop on, "Challenges in the Management of Water Resources and Environment in the Next Millennium: Need for Inter-institute Collaboration", Civil Engg. Dept., Delhi College of Engg., Delhi.

Perumal, M. and Seth, S. M. (1985) Regional flood frequency analysis using power transformation - a case Study. Proceedings of Seminar on Flood Frequency Analysis, New Delhi, 93-108.

Pilgrim, D.H. and Cordery, I. (1992) Flood runoff. Handbook of Hydrology (Mc Graw-Hill, Inc, New York,).

Pilon, P. J. and Adamowski, K. (1992) The value of regional information to flood frequency analysis using the method of L-moments. Canadian Journal of Civil Engineering, 19, 137-147.

Potter, K.W. and Walker, J.F. (1981) A model of discontinuous measurement error and its effect on the probability distribution of flood distribution of flood discharge measurements. Water Resources Research. 17. 1505-9

Potter, K.W. and Lettenmaier, D.P. (1990) A comparison of regional flood frequency estimation methods using resampling method. Water Resources Research, 26, 415-424.

Prasad, T. (1971) Discussion of "Risks in hydrologic design of engineering projects". Proc. of ASCE, 97 (HY1), 201-202.

Raghuwanshi, N.S., Singh R. and Reddy, L.S. (2006) Runoff and sediment yield modeling using ANNs: Upper Siwane River, India. Journal of Hydrologic Engineering, 11(1), 71-79.

Rajasekaran, S. and Vijayalakshmi Pai, G.A. (2004) Image recognition using simplified fuzzy art map augmented with a moment based feature extractor. Pattern Recognition, Artificial Intelligence, 14(8), 1081-1094.

Rao, A.R. and Hamed, K.H. (2000) Flood frequency analysis. CRC Press, Washington, D.C.

Research Design and Standards Organization (1991) Estimation of design discharge based on regional flood frequency approach for Subzones 3(a), 3(b), 3(c) and 3(e). Bridges and Floods Wing Report No. 20, Lucknow.

Robson, A. and Reed, D. (1999) Statistical procedures for flood frequency estimation. Flood Estimation Report Handbook, Institute of Hydrology, UK.

Sankara subramanian, A. (1995) Application of L-moments in regional flood frequency analysis. Hyd. Res. & Water Res. Engg., Civil Engg. Dept. I.I.T., Madras.

Schaefer, M.G. (1990) Regional analysis of precipitation annual maxima in Washington State. Water Resources Research, 26, 119-131.

Seth, S. M. and Singh, R.D. (1987) Flood estimation in India using regional frequency analysis. Proceedings of National Symposium on Hydrology, Roorkee, 1, 1-15.

Shi, Y. and Mizumoto, M. (2001) An improvement of neuro-fuzzy learning algorithm for tuning fuzzy rules. Fuzzy Sets and Systems, 118, 339-350.

Sikka, A.K. and Selvi, V. (2005) Experimental examination of rational runoff coefficient for small agricultural and forest watersheds in the Nilgiris, 215 Agricultural Engineering Journal, The Institution of Engineers (India), Vol. 86.

Singh, R. D. (1989) Flood frequency analysis using at site and regional data. M.Sc. (Hydrology), Dissertation, International P.G. Course in Hydrology, Galway.

Singh, R.D. and Seth, S.M. (1985) Regional flood frequency analysis for Mahanadi basin using Wakeby distribution. Proc. of Seminar on Flood Frequency Analysis, 109-119.

Singh, V.P. (1994) Elementary hydrology, frequency analysis. Dept. of Civil Engg. Louisiana State University, Prentice-Hall of India Pvt. Ltd. 800-859.

Sivapalan, M., Takeuchi K., Franks, S.W., Gupta, V.K., Karambiri Lakshmi, V., Liang, X., McDonnell, Mendioudo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Unlenbrook, S. and Zahe, E. (2003) 'IAHS decade on predictions of ungauged basins (PUB), 2003-2012 : Shaping an exciting future for hydrological sciences', Hydrolol Sc. J, 48, 857-880, 2003.

Stedinger, J.R. (1983) Estimating a regional flood frequency distribution. Water Resources Research, 19(2), 503-510.

Stedinger, J.R., Vogel, R.M. and Foufoula-Georgiou, E. (1993) Frequency analysis of extreme events. Handbook of Hydrology, ed. D.R. Maidment, McGraw-Hill, New York, NY, pp. 18.1-18.66

Stedinger, J.R., Vogel, R.M. and Foufoula - Georgiou, E. (1992) Frequency analysis of extreme events. In: Maidment, D.R. (Editor in Chief), Handbook of Hydrology, Mc Graw- Hill, Inc., New York, 18.1-18.66.

Stephens, M. (1974) E.d.F. statistics for goodness of fit. Journal of Amer. Statistical Assoc., 69, 730-737.

Strupczewski, W.G., Kochanek, K. and Singh, V.P. (2009) Statistics of Extremes in Hydrology. International Conference "Water, Environment, Energy and Society" (WEES-2009), New Delhi, 12-16 January, 2009.

Subramanya, K. (1990) Engineering hydrology. Tata McGraw-Hill, New Delhi.

Sudheer, K.P., Nayak, P.C., Ramasastri, K.S. (2003) Improving peak flow estimates in artificial neural network river flow models. Hydrological Processes 17: 671-686.

Takagi, T. and Sugeno, M. (1985) Fuzzy identification of systems and its application to modelling and control. IEEE Trans Syst. Man. Cybernet;15: 116-132.

Tasker G.D. and Stedinger J.R. (1989) An operational GLS model for hydrologic regression. Journal of Hydrology, 111, 361-375.

Thirumalai, S. and Sinha, P.B. (1986) Flood estimation for small and medium catchments in Krishna basin by Chow's approach modified by Nash. Proceedings of Seminar on Flood Estimation and Control, Roorkee, III- 43-51.

Thiruvengadachari, S., Nayak, S.V. and Ranganna, G. (1975) Flood in Cauvery basin: magnitude and frequency. Proceedings of National Symposium on Hydrology, Roorkee, G-17-21.

Thomas D.M. and Benson M.A. (1970) Generalisation of streamflow characteristic from drainage-basin characteristics. U.S.G.S., Water supply paper no. 1975.

Thomas, R.B. (1971) Errors in streamflow estimates from continuous stage records. Proc. of Symposium on Statistical Hydrology, Tucson, Arizona.

U.S.W.R. (1982) Guidelines for determining flood flow frequency. Bulletin # 17B, U.S. Govt. printing office, Washington, D.C.

Upadhyay, P.C. and Kumar, A. (1999) Use of L-moments in flood estimation at ungauged sites. Proc. of the National Workshop on, "Challenges in the Management of Water Resources and Environment in the Next Millennium: Need for Inter-institute Collaboration", Civil Engg. Dept., Delhi College of Engg., Delhi.

Upadhyay, B.S., Kaur, S., Mishra, M.S. and Mittal, M.K. (1990) Space correlation structure of rainfall over India. Mausam, Vol. 41.

Varshney, R.S. (1979) Engineering hydrology. Nem Chand & Brothers, Roorkee.

Venkatraman, R. and Gupta, P. N. (1986) Flood estimation for small and medium catchments in Sub-Himalayan region. Proceedings of Workshop on Flood Estimation in Himalayan Region, Roorkee, 123-140.

Venkatraman, R., Thirumalai, S. and Gupta, P.N. (1986) Regional flood frequency analysis for small and medium catchments in Lower Godavari basin - A critical review. Proceedings of Seminar on Flood Estimation and Control, Roorkee, II-58-77.

Vogel, R.M., editor (1995) U.S. National Report to International Union of Geodesy and Geophysics 1991-1994-Part 2- Contributions in Hydrology, American Geophysical Union, Washington, D.C., 933-1147.

Vogel, R.M. and Fennessey, N.M. (1993) L-moments should replace product moments diagrams. Water Resources Research, 29(6), 1745-1752.

Wallis, J.R. and Wood E.F. (1985) Relative accuracy of log-pearson III procedures. Journal of Hydrologic Engineering (ASCE), 111(7), 1043-56.

Wallis, J.R., Matalas, N.C. and Slack, J.R. (1974) Just a movement. Water Resources Research, 10(2), 211-219.

Wang, Q.J. (1996) Using partial probability weighted moments to fit the extreme value distributions to censored samples. Water Resources Research, 32(6), 1767-1771.

Water Resources Council (1981) Statistical methods in hydrology. The Iowa State University Press, Ames, Iowa.

Whitley R. and Hromadka T. V. (1999) Approximate confidence intervals for design floods for a single site using a neural network. Water Resources Research, 35(1), 203 - 209.

Wiltshire, S.E. (1986a) Regional flood frequency analysis I: Homogeneity Statistics. Hydrological Sciences Journal, 31, 321-333.

Wiltshire, S.E. (1986b) Regional flood frequency analysis II: Multivariate classification of drainage basins in Britain. Hydrological Sciences Journal, 31, 335-346.

Wu, Jy S., P.E., Han, J., Annambhotla, S. and Bryant, S. (2005) Artificial neural networks for forecasting watershed runoff and stream flows. Journal of Hydrologic Engineering, 10(3), 216-222.

Xiong, L.H. and Shamseldin, A.Y. (2001) A nonlinear combination of the forecasts of rainfall-runoff models by the first order Takagi-Sugeno fuzzy system. Journal of Hydrology, 245, 196-217.

Yen, B.C. (1971) Risks in hydrologic design of engineering projects. Proc. of ASCE, 96(HY4), 959-966.

Yen, B.C. and Ang, A.H.A. (1971) Risk analysis in design of hydraulic projects. Proc. of Symp. of Stochastic Hydraulics, Univ. Pittsburgh, 694-709.

Yevjevich, V. (1972) Probability and statistics in hydrology. Water Resources Publications, Fort Collins, CO.

Yi, S. Y. and Chung, M. J. (1993) Identification of fuzzy relational model and its application to control. Fuzzy Sets and Systems, 59 (1), Elsevier North-Holland, Inc.

Zadeh, A.L. (1973) Outline of a new approach to the analysis of complex systems and decision and decision processes. IEEE Transactions on Systems, Man & Cybernetics, 3, 28-44.

Zafirakou-Koulouris, A., Vogel, R.M., Craig, S.M. and Habermeier, J. (1998) L-moment diagrams for censored observations. Water Resources Research, 34(5), 1241-1249.

Zhang, L. and Singh, V.P. (2006) Bivariate flood frequency analysis using the copula method. Journal of Hydrologic Engineering, 11(2), 150-164.

Zhang, L. and Singh, V.P. (2007) Trivariate flood frequency analysis using the Gumbel–Hougaard copula. Journal of Hydrologic Engineering, 12(4), 431-439.

Zimmermann, H.J. (1991) Fuzzy set theory-and its applications. Allied Pub. Ltd. New Delhi.

Zrinji Z. and Burn D.H. (1994) Flood frequency analysis for ungauged sites using a region of influence approach. Journal of Hydrology, 153, 1-21.

APPENDIX 3.1

Brief descriptions of the 17 subzones are presented as follows.

1. Chambal Subzone 1 (b)

The Chambal Subzone 1 (b) lies approximately between 73^0 20' and 79^0 east longitudes and 22^0 30' and 27^0 15' north latitudes. This covers major parts of Rajasthan and Madhya Pradesh and small portion of Uttar Pradesh. The Chambal is the principal tributary of the Yamuna and other important rivers of the Subzone are Banas from the left bank and Kali Sindh, Parbati, Kunu and Kunwari from the right bank. The river Chambal rises in the Vindhya range near Mhow in the Indore district of Madhya Pradesh at an elevation of 854 m. Thereafter, it flows in a generally northernly direction for a length of about 320 km in Madhya Pradesh upto its border with Rajasthan. The river then flows through Rajasthan and receives its right bank tributaries Kali/Sindh and Parbati. After its confluence with Prabati, the Chambal forms a common boundary between Madhya Pradesh and Rajasthan. Banas, the major left bank tributary joins the Chambal in this reach near the village Rameshwar. The river thereafter forms the common boundary between Madhya Pradesh and Uttar Pradesh before its enters Uttar Pradesh. After flowing for about 46 km in Uttar Pradesh, the Chambal outfalls into the Yamuna. The total length of the river from its source to confluence with Yamuna is about 960 km of which 320 km are in Madhya Pradesh, 226 km in Rajasthan. 251 km from the common boundary between Madhya Pradesh and Rajasthan, 117 km from the common boundary tween Madhya Pradesh and Uttar Pradesh and the balance 46 km area in Uttar Pradesh. From the source down to its junction with Yamuna, the Chambal has a total fall of about 732 m of which 244 m is the first few km and 122 m in a distance of about 100 km from Courashigarh fort to Kota city. For the rest of its course, the river passes through the flat fertile areas of Malwa Plateau and later in Gangetic Plains. There are mainly three types of soil viz. medium black soil, mixed red and black soil, alluvial soil. Other types of soil are red and yellow soil, gray-brown soil, deep-black soil, laterite soil and skeletal soil. The arable land in the Subzone is about 52%, forest cover 23%, grass land scrub 19% and the remaining portions are waste land urban area.

2. Sone Subzone 1 (d)

The region defined as Sone Subzone 1 (d) lies in central-eastern part of India. Sone River is one of the major tributaries of the Ganges River flowing in the Subzone 1 (d). Additional major rivers in the region include the Tons, Karmanasa, Punpun and Phalgu. The Sone Subzone 1 (d) region lies between latitudes 22^0 30' to 25^0 45' north and longitudes 80^0 to 86^0 15' east. The Subzone experiences heavy rainfall due to southwest monsoon during June to September. The monsoon rainfall is about 80 to 85% of the annual rainfall. The maximum rainfall is experienced during the months of July and August. The normal annual rainfall of the Sone Subzone generally varies with the decrease in elevation from 1400 mm to 1600 mm in the hills and from 1000 to 1200 mm in the plains. The Subzone is mostly covered with red and yellow soils except the alluvial soils in South Bihar plains and patches of medium black soils, red sandy soils and mixed red and black soils in the South West. Arable land mostly in the plains and also a large number of patches in the remaining part constitute about 45% of the Subzone. Forests cover about 50% of the Subzone and 5 % of the remaining Subzone is mostly grass land, scrub, wasteland marshes and water bodies.

3. Upper Indo-Ganga Plains Subzone 1 (e)

The Upper Indo-Ganga Plains Subzone 1 (e) lies between longitudes 74^0 to 81^0 east and latitude 26^0 to 33^0 north. It is traversed by the Ravi, Beas, Sutlej, Yamuna, Ghaggar, Ganga, Gomti, Sahibi and Banganga and Ramganga rivers. It covers almost entire Haryana, Punjab, Union Territories of Delhi and Chandigarh, Western Uttar Pradesh and eastern boarder areas of Rajasthan. There is a small mountainous area in northern part of Punjab varying in elevation from 450 to 600 m. Areas with elevations less than 150 m are located in the southeast of the Subzone. The general elevation of the remaining area is between 150 to 300 m. The mean annual rainfall in northern parts is 1000 mm. In the middle and southern areas it varies from 600 to 800 mm and in south-western parts and from 300 to 400 mm in the south-western parts. The plains of Yamuna, Ganga, Ramganga, Gomti and upper parts of Ravi, Beas, Sutlej and Ghaggar are covered with recent alluvial soils. The plains in the middle reaches of Beas, Sutlej and Ghaggar are covered with calcareous soils of alluvial origin. The saline and alkaline soils are also found in some parts of the plains covered with alluvial soils in areas lying in the northwest part of the Subzone between Sutlej and Ghaggar. The northwest and southwest portions comprising of 50% of the Subzone are intensely irrigated to an extent of 80%. The intensity of irrigation in 25% of the area is 20% to 60%. The northwestern, southwestern and northeastern areas are covered with forests.

4. Middle Ganga Plains Subzone 1 (f)

The Middle Ganga Plains Subzone 1 (f) lies between latitudes 24^0 to 29^0 north and longitude 80^0 to 89^0 east. It covers parts of Uttar Pradesh, Bihar, Jharkhand and West Bengal. The major rivers flowing in this Subzone are Ganga, Yamuna, Gomti, Gandak, Ghagra, Rapti, Kosi including Kamla, Mahananda and others. The Subzone 1 (f) comprises mostly of plains and a small portion of the foothills of Tarai area in the north. The elevation in the Tarail area exceeds 150 m. In the plains area the elevation lies between 150 m and 75 m and goes on decreasing eastwards to Bangladesh. The rivers Yamuna and Ganga form southern boundary of the alluvial plains for major part of the Subzone 1 (f). The Subzone covers lower portions of Ghaghra, Gandak, Rapti, Kosi, and Mahananda rivers. The mean annual rainfall varies between 800 mm to 120 mm in the plains and goes upto 2000 mm in the portion of foothills in the north of the Subzone. The major portion of rainfall is received between June/July to September/October in the Subzone due to southwest monsoon. Major portion of the Subzone has alluvial soils of recent origin excepting Tarai region and the plains on the northeastern side between Rapti and Kosi rivers where Tarai and Calcarious alluvium soils are encountered respectively. Most of the parts are also irrigated. Forests are seen in a part of Tarai portion of the Subzone. Most of the land in the Subzone is arable and well irrigated.

5. Lower Ganga Plains Subzone 1 (g)

The Lower Ganga Plains Subzone 1 (g) is lies approximately between latitude 21^0 15' to 25^0 45' north and longitudes 84^0 35' to 89^0 east. It was earlier designated as Lower Gangetic Plains including Subarnarekha and other east flowing rivers between Ganga and Baitarani. The river basins included in this Subzone are lower portions of Ganga, Hoogli river system and Subarnarekha. The annual rainfall over the Subzone is of the order of 900 mm over its extreme north west portion and gradually increases to about 1700 mm over extreme south of the Subzone. A large area in central part of the Subzone is covered by red sandy soil. A small portion towards extreme north of the Subzone, is covered by alluvial soil. Red and yellow soil is found in the western parts of the Subzone. The eastern area of the Subzone is almost all covered by alluvial soil along with a small region of red and loamy soil adjoining Berhampore. Mixed red, black and yellow soil, alluvial soil and laterite soil is found, in general, in the southern areas of the Subzone. Deltic alluvial soil is found over the areas in the vicinity of the mouth of Bay of Bengal. The major portion of the area is under cultivation. Rice is the main crop of the region. Other crops grown in the area are Jute and Millets.

6. North Brahmaputra Subzone 2 (a)

The Brahmaputra also known as Tsangpo in Tibet rises at Tamchok Khamdet Chorten in the Chemayung-dung glacier. It has a long course through the comparatively dry and flat region of Southern Tibet, before breaking through the Himalayas below the peak of Nancha Barwa. It is known as the Dihang in the Arunachal Himalayas before it enters the Assam plains. The Dibang and the Lohit join the Dihang from the east near Sadija. After traversing the Assam Valley for 720 km, the Brahmaputra sweeps round the Garo Hills and enters the Rangpur District of Bangladesh near Dhubri. It flows southwards to join the Ganga at Goalundo. The Brahmaputra with a total catchment of 0.94 million km^2 is one of the biggest rivers in the world. The total length of river in India is 885 km. The drainage area of the Brahmaputra basin in India is 1,95,000 km^2 and the Subzone 2 (a) has an areal extent of 1,21,444 km^2. The North Brahmaputra Subzone 2 (a) lies approximately between 88° and $97^{\circ} 20'$ east longitudes and 26° and 29° 25' north latitudes. The Subzone 2 (a) is mostly bounded by international boundaries on all the four sides. It has Bhutan and China on the north, Burma on the east, Nepal on west and Bangla Desh on southwest.

The states covered by this Subzone are Assam (Lower and Upper) part of West Bengal, Sikkim and Arunachal Pradesh. Of the 25 principal north bank tributaries, the Subansiri, the Manas, the Dibang, Dhansiri, Torsa, Testa are a few major ones. The North Bank tributaries have comparatively moderate steep slope, meandering channels almost from the foothills, beds and banks of alluvial soils and comparatively low silt charge. The southwest monsoon and cyclonic storm causes the rainfall in the Subzone from May to October. The normal annual rainfall varies from 2000 mm to 5000 mm. Broadly the soils of the Subzone can be classified as red loamy soil, brown hill soil, terai soil and alluvial soil of recent origin. The red loamy soil is found towards north-east and continues through a belt in the middle of the Subzone up to Itanagar. A small patch of brown hill soil is bound in the northern an western corner of the Subzone. The alluvial soil is also depicted in flood plain covering north eastern part to the west all along the main river Brahmaputra touching important towns of Itanagar, Tezpur, Jalpaiguri. A belt of Terai soil runs through the middle of the Subzone to the west. The Subzone has considerable area under forest.

7. South Brahmaputra Subzone 2 (b)

The drainage area of the Brahmaputra basin in India is 1,95,000 km^2 and the Subzone 2 (b) has an areal extent of 73556 km^2. A number of tributaries drain into the Brahmaputra from north and south, in its course through the State of Assam. There are 15 principal south bank tributaries, the most important among them being the Burhi-Dehing, the Kopili and the Dhansiri. The north bank tributaries are generally large, since their catchments lie in the heavy rainfall zone of the Himalayas. The south bank tributaries of the Brahmaputra in the Assam State are generally smaller than those of the north bank, as their catchments in the Assam Hills are smaller and get less rain. The Brahmaputra, in its course through the Assam plains, divides into many channels and forms numerous braids which enclose islands of which, Majuli, is 1,250 km^2 in area.

The course of the Brahmaputra river in the plains divides the Brahmaputra basin in India into northern and southern Subzones. The Subzone 2 (b) has a variety of soils. Broadly they can be classified as red loamy soil, red and yellow soil, laterite soil and alluvial soils of recent origin. The red loamy soil is found towards the northeast and continues through a belt in the lower half of the Subzone upto Shillong in the southwest. The flood plains covering the Dibrugarh, Mariani, Nowgong, Gauhati and Goalpara districts represent alluvial soils of recent origin. There is a small belt of laterite soils towards southeast. The laterite soils are also found in the southern corner and towards the south-west of the Subzone in the plain portions between foothills and Gauhati and extend throughout the west. The red and yellow soils are normally found towards the south and towards the southwestern corner of the Subzone between Tura and boundary of Bangladesh. The Subzone 2 (b) has considerable area under forest. Although this may have some marked changes in the recent times because of more inhabitations of area towards the northeast, the intensity of the afforestation is maximum in the east and this goes on decreasing as one comes towards west.

8. Mahi and Sabarmati Subzone 3 (a)

The Subzone 3 (a) is traversed by the rivers Mahi, Sabarmati, Saraswati and a large number of coastal streams. The general elevation of this Subzone varies from 0 to 600 m above mean sea level. This Subzone lies in semi-arid region. The Mahi and Sabarmati Subzone 3 (a) lies roughly between 69° to 75° east longitudes and 21° to

25° north latitudes. It covers more than half of Gujarat State and small parts of southern Rajasthan and western Madhya Pradesh States. The rivers flowing in this Subzone are Mahi, Sabarmati, Saraswati and an large number of coastal streams in Kathiawar Peninsula. The Mahi river flows for a total length of 583 km through the states of Madhya Pradesh, Rajasthan and Gujarat before outflowing into the Gulf of Khambhat. The topography of the Subzone 3 (a) is mainly constituted of upper reaches draining the parts of Aravali ranges, Vindhya ranges and Malwa Plateau, Gujarat Plains and Kathiawar Peninsula. The upper reaches of Mahi and Sabarmati rivers vary in elevations from 300 m to 600 m. the general elevation of Gujarat plains vries between 150 m to 300 m and that of Kathiawar peninsula, from 0 m to 150 m along the southern fringes and 150 m to 300 m for the remaining portion with high elevation of 300 m to 600 m in the centre and the southern Gir ranges varying from 150 m to 300 m. The normal annual rainfall varies from 800 mm to 1000 mm over the Mahi basin whereas it varies from 400 mm to 600 mm over the Kathiawar peninsula. The major source of rainfall is southwest monsoon during June to September. About 90 % of the annual rainfall occurs during the monsoon season. The soils in the upper and lower parts of Mahi basin are medium black. The middle part of Mahi basin is covered with red and alluvial soils along with laterite soils. The Sabarmati and Saraswati basins are constituted of grey brown soils. The Kathiawar peninsula is mostly covered with shallow, medium and black soils except the southern coastal belt of alluvial soils and northern coastal areas of deltaic alluvial soils. The Subzone is mostly constituted of arable land interspersed with forests, grassland and scurb.

9. Lower Narmada and Tapi Subzone 3 (b)

The Lower Narmada and Tapi Subzone 3 (b) is located between longitudes of 70^0 30' to 76^0 30' east and latitudes 20^0 30' to 23^0 north. Its total drainage area is about 77,700 km^2. It covers parts of Maharashtra, Gujarat, and Madhya Pradesh. The Subzone comprises of the Kanar, Kayam, Man, Hatori, Hiran, Bhakti, Bhadar, Goi, Korjan, Girna, Bord, Buray, Ganai, and other tributaries. It is a semi-arid region with mean annual rainfall varying from 600 to 1400 mm. The Subzone 3 (b) is traversed by the lower reaches of river Narmada and Tapi and their tributaries. It constitutes about 50% area of the Narmada and Tapi basins. The study area has a complex relief. Plains of medium heights up to 300 m exist on the western and eastern sides and in the centre of the Subzone. Low plateaus in the range of 300-600 m exist in eastern, southern and central parts. High plateaus in the range 600-900 m lie in the northern and also in the southern parts. The Subzone has a continental type of climate, i.e. cold in winter and hot in summer. Most of the rainfall results from southwest monsoon during June to October. Thunder storms also occasionally occur in the region. The main soil group in the Subzone are black soil, and coastal alluvial soils at the mouth of river Tapi. There is a small patch of laterite soil on the western portion of the Subzone. Approximately, 70% of the area of the Subzone is arable land and 25% is forest and rest is grass and waste land.

10. Upper Narmada and Tapi Subzone 3 (c)

The Upper Narmada and Tapi Subzone 3 (c) is located between east longitudes 76° 12′ to 81^0 45' and north latitudes of 20^0 10^0 to 23° 45′. Lying in the northern extremity of the Deccan plateau, the Subzone covers the States of Madhya Pradesh and Maharashtra. The Subzone 3 (c) comprises of upper portion of Narmada and Tapi basins and constitutes about 50% of the entire area of the combined Narmada and Tapi basins. The Narmada, westward flowing river rises near Amarkantak in the Mahaikala range in the Shahdol district of Madhya Pradesh at an elevation of about 1000 meters above sea level. It flows for a length of about 1300 km before it outfalls into the Gulf of Cambay in the Arabian sea. Upper Narmada and its tributaries drains a total area of 62,264 km^2 which form 72% of the area of Subzone. The river Tapi rises near Multai in the Betwa district of Madhya Pradesh and like Narmada it flows westward for a length of about 725 km before outfalling into Gulf of Cambay. The lengths of main Narmada and Tapi rivers in the upper Subzone are 813 km and 219 km, respectively. The upper Subzone covers parts of Madhya Pradesh and Maharashtra States. The important tributaries of Upper Narmada are Burhnar, Banjar, Sher, Shakkar, Dudha, Tawa, and Ganjal along left bank and Hiran, Tendori, Barna, Kolar, Jamner and Datuni along right bank. Purna is the main tributary of Tapi. Upper parts of Purna fall in the upper Subzone 3(c). About 20% area of the Subzone is under scrub and forest and the remaining is cultivable area. The main crops in the Subzone are wheat, millets, pulses, cotton and rice. The Subzone receives most of the rainfall from southwest monsoon. About 90% rainfall is received in months of June to October, July and August being the wettest months. The amount of rainfall varies from 800 mm in southwestern part of the Subzone to more than 2000 mm in the south-central parts of this Subzone. Station Pachmarhi receives the heaviest annual rainfall of more than 2000 mm. The rainfall from the south-central part of the Subzone decreases sharply and then increases to 1600 mm towards both western and eastern parts. Further towards southwest, it decreases to less than 800 mm. The far Eastern part of the Subzone receives rainfall of the order of 1400 mm.

11. Mahanadi Subzone 3(d)

The Mahanadi Subzone 3(d) is located between longitudes of 80^0 25′ to 87^0 east and latitudes 19^0 15′ to 23^0 35′ north. The Mahanadi Subzone 3(d) comprises of Mahanadi, Brahmani and Baitarani basins. The Mahanadi, Brahmani and Baitarani rivers are peninsular rivers, outfalling into the Bay of Bengal. The major tributaries of Mahanadi river are Seonath, Hasdeo, Mand and Ib joining from north, and Jonk, Ong and Tel joining from south. The total length of Mahanadi river is about 850 km and the river lengths of Brahmani and Baitarani are about 705 km and 333 km, respectively. Its total drainage area is about 1,95,256 km^2 out of which catchment area of Mahanadi is 1,40,628 km^2., which forms about 72% of the total area of the Subzone 3 (d). About 50% of the area of this Subzone is hilly varying from 300 m to 1350 m. Rest of the area lies in the elevation range of 0 to 300 m. The normal rainfall over the region varies from 1200 to 1600 mm. The Subzone receives about 75% to 80% of the annual rainfall from southwest monsoon during the monsoon season from June to September. The red and yellow soils cover major part of the Subzone. The red sandy, submontane and coastal alluvial soils cover the remaining part of the Subzone. The Subzone has an extensive area under forest. Paddy is the main crop grown on the cultivable land. Most of the irrigated area is in Sambalpur district under the canals of the Hirakud project. In the deltaic area around Cuttak, the irrigation is mostly done by inundation canals.

12. Upper Godavari Subzone 3 (e)

The Upper Godavari Subzone 3(e) lies between longitudes 73° 30′ to 78° 45′ east and latitudes 17° 25′ to 20° 35′ north. The Godavari river system in its upper reaches up to Manjra confluence constitutes the Upper Godavari Subzone. The

Godavari river rises in the eastern side of the western ghats at an elevation of 1067 m. It flows for a total length of 584 km in the Subzone before entering the Lower Godavari Subzone 3 (f). The major portion of the Subzone covers a part of Maharashtra State and the minor portions in the southeast of the Subzone cover small parts of Andhra and Karnataka States. The important towns and cities in the Subzone are Nasik, Aurangabad, Parbhani, Bidar, Bir and Nander. The Godavari river originates at an elevation of 1350 m. in the western ghats. The areas in the Subzone along the north western, western and southern boundary vary in elevations from 600 to 900 m. The rest of the area in the Subzone is a plateau ranging in elevation from 300 to 600 m. Along the ghats, the mean annual rainfall decreases from 1600 to 800 mm with the decrease in elevation. Further down upto Aurangabad the mean annual rainfall ranges from 600 to 700 mm. Thereafter, in the rest of the Subzone, mean annual rainfall is of the order of 800 mm with a patch of heavy rainfall of 1000 mm along the eastern periphery. The Subzone experiences the southwest monsoon during June to October with the maximum mean monthly rainfall in July and September. The Subzone is mostly covered with medium black soils with a strip of deep black soils from east to west in the middle and red sandy soils in southeast extremity. Patches of shallow black soils are found in north. The Subzone is covered mostly with arable land with patches of forests along the northern and eastern periphery and grass land scrub mostly along the southwestern portion.

13. Lower Godavari Subzone 3 (f)

The Lower Godavari Subzone 3(f) lies between latitudes of 17^0 to 23^0 north and longitudes of 76^0 to 83^0 east. The Godavari river rises in the eastern side of the Western Ghats at an elevation of 1067 m. Lower Godavari Subzone 3 (f) is a sub-humid region with elevation varying from 150 meters to 1350 meters in its various portions. The Subzone receives about 75% to 80% rainfall of its annual rainfall from southwest monsoon during the period of June to October. The Subzone having a continental type of climate cold in winter and very hot in summer receives most of the rainfall from the southwest monsoon (June to September). A small part of the Subzone on the south-east end gets rain from northeast monsoon (November to December) besides short duration thunder storms. The mean annual rain progressively increases from west to east. Mean monthly rainfall histograms typical of the two cities, Nagpur and Chandrapur of the Subzone and the adjoining Hyderabad indicate sudden rise in rainfall from June to September, covering 80% of the annual total. The broad soil groups in the Subzone are red soils and black soils. The red soils are either classified into red sandy, red loamy and red yellow soils. Black soils are classified as deep black, medium black and shallow black soils. The black soils are clayey in texture and are derived from trap rocks. The texture of the red soils vary considerably from place to place and are derived from all groups. More than 50% of the area is covered by forests. Arable land is of the order of 25%.

14. Krishna and Penner Subzone 3 (h)

The Krishna and Penner Subzone 3 (h) lies between longitudes $73^0 21'$ to $80^0 25'$ east and latitudes of $13^0 7'$ to $19^0 25'$ north. This Subzone is catered by the Krishna and Penner rivers excluding their deltaic strip along the eastern coast. The Krishna river is the largest east flowing river of peninsular India. It rises on the eastern side of western ghats about 60 km south of Pune at an altitude of 1337 m. The river Penner originates in the Chenna Kesabir hill of Nandidoug range in Karnataka state. The elevation range of its various parts varies from 150 m to 600 m. The total drainage area of the Subzone 3(h) is 2,80,881 km^2. The Subzone 3(h) has a continental climate. It is very hot in summer and moderately cold in winter. The Subzone 3(h) receives about 75% to 80% of annual rainfall from southwest monsoon during the monsoon season i.e. from mid June to mid October. The variation of normal annual rainfall over the Subzone 3(h) is from a minimum of 600 mm to a maximum of 2000 mm. The eastern side of the western boundary receives the heaviest rainfall. Two broad soil groups of the Subzone are red soils and black soils. Most of the areas covered by the upper portion of the Subzone are having black cotton soil. The lower portion including northeast side of the Subzone consists of red type of soil. In addition, there are pockets of red and black type of soils. The Subzone is having extensive area under arable land.

15. The Kaveri Subzone 3 (i)

The Kaveri Subzone 3 (i) lies between longitudes 75° 25' to 79°10' east and latitudes 10° to 14° north. The Kaveri river has its origin in the Brahmagiri range of the western ghats in Coorg District of Karnataka State. It flows eastwards for a total length of about 804 km through Karnataka and Tamil Nadu States, before out falling into the Bay of Bengal. The Kaveri river originates almost at the very edge of the western ghats within sight of Arabian sea at a height of about 1355 m and flows eastwards crossing mountain barrier of western ghats. The river falls about 450 m within a course of 8 km from its source. The upper reaches of the Kaveri and its tributaries drain the western ghats before flowing over a wide plateau. The eastern and western ghats fringe the plateau. The Kaveri Subzone has a complex relief. The general elevations of the plateau vary from 900 to 600 m in the northwestern part and 600 to 150 m in the southeastern part interspersed with higher elevations of 3000 to

900 m along the western periphery and inside the Subzone. The Subzone experiences rainfall by both southwest and northeast monsoons during June to September and October to December, respectively. The normal annual rainfall generally varies with the decrease in elevation along the eastern side of the western ghats from about 4000 mm to 1000 mm on the eastern side of the ghats in the Subzone. The remaining portion of the Subzone experiences a normal annual rainfall ranging from 600 mm to 800 mm. The Subzone is generally covered with red sandy soils barring a couple of areas of red loamy soil on the eastern and northern edges. The soil type varies considerably from the above mentioned groups. Arable land constitutes about 6% of the Subzone. About 25% of the Subzone is grass land and scrubs, the rest of it is covered with forests.

16. East Coast Subzone 4 (b)

The eastern coastal belt, comprising of Upper, Lower and South Subzones 4(a), 4(b) and 4(c) lies roughly between 77° to 80° east longitudes and 8° to 20° north latitudes. The eastern coastal belt extends roughly from Mahanadi delta to Kanniya Kumari. The eastern coastal belt covers parts of Orissa, Andhra Pradesh and Tamil Nadu States and Union Territory of Pondicherry. There are large number of small and medium coastal streams besides the outfall reaches of Godavari, Krishna, Kaveri, Vellar, Ponniyar, Pallar and Penner in the eastern coastal belt out falling into the Bay of Bengal and the Indian Ocean. The coastal streams rise in the eastern ghats and overflow their banks during the periods of heavy rainfall in their catchment areas. Similarly the other rivers flowing in the plains also overflow their banks during floods. The rivers flowing into the Bay of Bengal and Indian Ocean are also affected by the sea tides near their outfall reaches. The major deltas of Kaveri and Krishna form parts of Subzone 4(b). The mean annual rainfall along the coastal plains from the coast to eastern ghat varies from 1000 to 1200 mm, whereas the mean annual rainfall in the eastern ghat ranges varies from 1400 to 1600 mm. About two thirds of the annual rainfall occurs in the northern and middle parts of Subzone 4(b) during the period of southwest monsoon from June to September. The southern portions of Subzone 4(b) receive the rainfall from southwest monsoon during June to August and also from northeast monsoon during September to November. The rainfall during the northeast monsoon is higher as compared to the rainfall from the southwest monsoon in the southern portions of the Subzone-4(b). Soils in the eastern coastal belt are mostly coastal alluvial soils, coastal sandy soils and coastal deltaic soils in the deltas of Godavari, Krishna and Kaveri. Besides patches of red loamy soils and red sandy soils are interspersed in the coastal belt. About 70% of east coast belt is arable land. About 10% of the area is covered with grass land and scrub. The remaining 20% of the area in this east belt is covered with forest.

17. Sub-Himalayan Region Zone-7

The study area comprises of small and medium size catchments of the Sub-Himalayan region which has been categorized as one of the 26 Subzones of India. The Himalayan region up to its foot-hills, lying within the great arc passing through Madhopur near Dara Baba Nanak in the north east between 76^0 to 96^0 east longitudes and 26^0 to 32^0 north latitudes has been grouped under Zone-7. This Zone holds a great potential for generation of hydropower but flood estimation for this Zone is proving to be an intractable problem as runoff from this region consists of snow melt as well as rainfall. The areas located in the extreme north and northeast of the zone have elevation ranging between 7500 to 6000 m. The elevation decreases towards south and in the central portion of the zone it varies between 6000 to 4500 m. In the areas adjoining the river banks the elevation varies between 4500 to 600 m. In the plain areas of Uttar Pradesh, Punjab and Himachal Pradesh, the elevation varies between 600 to 300 m. In the northern areas of the zone, skelral soil along with saline and alkali soils are found. The areas around Indus river are covered with mountain-meadow soils. Sub-mountain soils are located in the central northwest to north east areas of the Zone-7. The southern areas are covered with brown hill soils. It has widely varying topographical features; elevation being as low as 300 m over its southern parts and as high as 7500 m in the mountainous parts of the Zone. The areas located in the vicinity of Subzone 1 (e) are covered with tarai soils. Nearly 75% of the area located in north, northeast and southeast of the Zone is a waste land. Small pockets towards south and southwest of the Zone are covered with scrubs. Forests are located in the areas northeast and southeast of the Zone. Rice wheat and millets along with fruits of various kinds are grown over the remaining areas.

APPENDIX 4.1

LEVENBERG-MARQUARDT ALGORITHM

In mathematics and computing, the Levenberg-Marquardt algorithm (LMA) provides a numerical solution to the problem of minimizing a function, generally nonlinear, over a space of parameters of the function (Levenberg, 1944; Marquardt, 1963; Jacoby et al., 1972; Kuester and Mize, 1973; Gill et al., 1981). The algorithm was first published by Kenneth Levenberg, while working at the Frankford Army Arsenal. It was rediscovered by Donald Marquardt who worked as a statistician at DuPont. These minimization problems arise especially in least squares curve fitting and nonlinear programming.

The LMA interpolates between the Gauss-Newton algorithm (GNA) and the method of gradient descent. The LMA is more robust than the GNA, which means that in many cases it finds a solution even if it starts very far off the final minimum. On the other hand, for well-behaved functions and reasonable starting parameters, the LMA tends to be a bit slower than the GNA.

The LMA is a very popular curve-fitting algorithm used in many software applications for solving generic curve-fitting problems.

The primary application of the Levenberg-Marquardt algorithm is in the least squares curve fitting problem: given a set of empirical data pairs of independent and dependent variables, (x_i, y_i), optimize the parameters β of the model curve $f(x, \beta)$ so that the sum of the squares of the deviations

$$S(\beta) = \sum_{i=1}^{m} \left[y_i - f(x_i, \beta) \right]^2$$

becomes minimal.

Like other numeric minimization algorithms, the Levenberg-Marquardt algorithm is an iterative procedure. To start a minimization, the user has to provide an initial guess for the parameter vector, β. In many cases, an uninformed standard guess like $\beta^T = (1,1,\ldots 1)$ will work fine; in other cases, the algorithm converges only if the initial guess is already somewhat close to the final solution.

In each iteration step, the parameter vector, β, is replaced by a new estimate, β + δ. To determine δ, the functions $f(x_i, \beta+\delta)$ are approximated by their linearizations

$$f(x_i, \beta+\delta) \approx f(x_i, \beta) + J_i\delta$$

where $J_i = \dfrac{\partial f(x_i, \beta)}{\partial \beta}$ is the gradient (row-vector in this case) of f with respect to β.

At a minimum of the sum of squares, called S, the gradient of S with respect to β is 0.

Differentiating the squares in the definition of S, using the above first-order approximation of $f(x_i, \beta+\delta)$, and setting the result to zero leads to:

$$(J^T J)\delta = J^T[y - f(\beta)]$$

Where J is the Jocabian matrix whose i-th row equals J_i, and where f and y are vectors with i^{th} component $f(x_i, \beta)$ and Y_i, respectively. This is a set of linear equations which can be solved for δ.

Levenberg's contribution is to replace this equation by a "damped version",

$$(J^T J + \lambda I)\delta = J^T[y - f(\beta)]$$

Where I is the identity matrix, giving as the increment, δ, to the estimated parameter vector, β.

The (non-negative) damping factor, λ, is adjusted at each iteration. If reduction of S is rapid, a smaller value can be used, bringing the algorithm closer to the Gauss-Newton algorithm, whereas if an iteration gives insufficient reduction in the residual, λ can be increased, giving a step closer to the gradient descent direction. Note that the gradient of S with respect to β equals $-2(J^T[y-f(\beta)])^T$. Therefore, for large values of λ, the step will be taken approximately in the direction of the gradient. If either the length of the calculated step, δ, or the reduction of sum of squares from the latest parameter vector, $\beta + \delta$, fall below predefined limits, iteration stops and the last parameter vector, β, is considered to be the solution.

Levenberg's algorithm has the disadvantage that if the value of damping factor, λ, is large, inverting $J^TJ + \lambda I$ is not used at all. Marquardt provided the insight that we can scale each component of the gradient according to the curvature so that there is larger movement along the directions where the gradient is smaller. This avoids slow convergence in the direction of small gradient. Therefore, Marquardt replaced the identity matrix, I, with the diagonal of the Hessian matrix, J^TJ, resulting in the Levenberg-Marquardt algorithm:

$$(J^TJ + \lambda \mathrm{diag}(J^TJ))\delta = J^T[y - f(\beta)].$$

A similar damping factor appears in Tikhonov regularization, which is used to solve linear ill-posed problems, as well as in ridge regression, an estimation technique in statistics.

Various more-or-less heuristic arguments have been put forward for the best choice for the damping parameter λ. Theoretical arguments exist showing why some of these choices guaranteed local convergence of the algorithm; however these choices can make the global convergence of the algorithm suffer from the undesirable properties of steepest-descent, in particular very slow convergence close to the optimum.

The absolute values of any choice depends on how well-scaled the initial problem is. Marquardt recommended starting with a value λ_0 and a factor $v>1$. Initially setting $\lambda=\lambda_0$ and computing the residual sum of squares $S(\beta)$ after one step from the starting point with the damping factor of $\lambda=\lambda_0$ and secondly with λ/v. If both of these are worse than the initial point then the damping is increased by successive multiplication by v until a better point is found with a new damping factor of λv^k for some k.

If use of the damping factor λ/v results in a reduction in squared residual then this is taken as the new value of λ (and the new optimum location is taken as that obtained with this damping factor) and the process continues; if using λ/v resulted in a worse residual, but using λ resulted in a better residual then λ is left unchanged and the new optimum is taken as the value obtained with λ as damping factor.