CS/AR-16/2000-2001

# DEVELOPMENT OF REGIONAL FLOOD FORMULATE USING L-MOMENTS FOR MIDDLE GANGA PLAINS (SUBZONE 1-F)

आपो हि ष्ठा मयोभुवः

**NATIONAL INSTITUTE OF HYDROLOGY**
**JALVIGYAN BHAWAN**
**ROORKEE - 247 667 (UTTARANCHAL)**

**2000-2001**

# PREFACE

For planning, design and operation of various types of water resources projects, estimation of flood magnitudes and their frequencies has been engaging attention of the engineers the world over since time immemorial. 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 flood estimates directly. In such a situation, the regional flood frequency relationships or the flood formulae developed for the region are one of the alternative methods which may be adopted for estimation of design floods especially for small catchments. Most of the flood formulae developed for different regions of the country are empirical in nature and do not provide flood estimates for the desired return periods. Hence, there is a need for developing the regional flood formulae for estimation of floods of desired return periods for different regions of the country, using the recently developed, improved and efficient techniques of flood frequency analysis.

Regional frequency analysis basically involves substitution of "space for time" where data from different sites in a region are used to compensate for short records at a site and it provides an alternative method for estimation of flood frequency estimates for the gauged and ungauged catchments lying in the region. In this study, discordancy measure ($D_i$) test was carried out for screening the data. Homogeneity of the region has been tested using the L-moment based heterogeneity measure, H. Ten frequency distributions have been considered and based on the recently introduced goodness of fit criteria viz. L-moment ratio diagram and $Z^{Dist}$ statistic; GEV distribution has been identified as the robust distribution for the Middle Ganga Plains (Subzone 1-f). For estimation of floods of desired return periods for gauged catchments, the regional flood frequency relationship has been developed using the L–moment based GEV distribution. Also, for estimation of floods of different return periods for ungauged catchments of the study area, a regional flood formula has been developed by coupling the L-moment based regional flood frequency relationship with the regional relationship between mean annual peak flood and the catchment area.

The study has been carried out by Shri Rakesh Kumar, Dr. C. Chatterjee and Dr. Sanjay Kumar, Scientists of the Institute. Technical assistance has been provided by Shri A.K. Sivadas, Technician. It is expected that the developed regional flood frequency relationship will provide rational flood frequency estimates for the gauged and ungauged catchments of the Subzone 1(f).

**( K.S. Ramasastri)**
**Director**

# CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

Estimation of magnitudes of likely occurrence of floods is of a great importance for solution of a variety of water resources problems such as design of various hydraulic structures, urban drainage systems, flood plain zoning and economic evaluation of flood protection works etc. 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 flood estimates directly. In such a situation, the regional flood frequency relationships or the flood formulae developed for the region are one of the alternative methods which may be adopted for estimation of design flood specially for small catchments. The choice of method primarily depends on the design criteria applicable to the structure and availability of data. As per the Indian 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. However, for design of large and intermediate size dams probable maximum flood and standard project flood are adopted, respectively (NIH, 1992).

In this study, regional flood frequency relationship and flood formula have been developed for the small to moderate size gauged and ungauged catchments of the Middle Ganga Plains (Subzone 1-f). The Subzone 1(f) lies between latitude $24^0$ to $29^0$ North and longitude $80^0$ to $89^0$ East and its total areal extent is 1, 71, 350 $km^2$. It covers parts of Uttar Pradesh, Bihar and West Bengal. The major rivers flowing in this Subzone are Ganga, Yamuna, Gomti, Gandak, Ghagra, Rapti, Kosi including Kamla, Mahananda and others. Annual maximum peak flood data of 11 gauging sites of Subzone 1(f) have been considered.

Screening of the data has been carried out for assessing the suitability of the data for using for regional flood frequency analysis by computing the Discordancy measure ($D_i$) in terms of the L-moments. Also, homogeneity of the region has been tested using the L-moment based heterogeneity measure, H. 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 four parameter Kappa distribution for computing the heterogeneity measure, H. Kappa distribution includes as special cases the GLO, GEV and GPA distributions and it is capable of representing many of the frequency distributions. Based on the homogeneity test, it has been observed that the data of 8 out of 11 sites constitute a homogeneous region. Hence, the data of these eight sites have been used in this study. The record length of the data varies from 23 to 33 years. Catchment areas of these sites vary from 32.9 to 447.8 $km^2$ and the mean annual peak floods range from 24.3 $m^3/s$ to 555.2 $m^3/s$. Comparative regional flood frequency analysis studies have been carried out using the various L-moments based frequency distributions viz. Extreme value (EV1),

General extreme value (GEV), Logistic (LOS), Generalized logistic (GLO), Normal (NOR), Generalized normal (GNO), Exponential (EXP), Generalized Pareto (GPA) and five parameter Wakeby (WAK). L-moments of a random variable were first introduced by Hosking (1990). They are analogous to conventional moments, but are estimated as linear combinations of order statistics. In a wide range of hydrologic applications, L-moments provide simple and efficient estimators of characteristics of hydrologic data and of a distribution's parameters (Stedinger et al., 1992). Hosking (1997) presented state of the art application of L-moments in frequency analysis.

The $Z^{dist}$ statistic values for GEV, GNO, PT-III and GLO distributions are found to be less than 1.64. Its value is found to be minimum i.e. 0.01 for the GEV distribution. Thus, based on the L-moment ratio diagram and $Z^{dist}$ statistic criteria, GEV distribution has been identified as the robust distribution for the study area. For estimation of floods of various return periods for the gauged catchments of the study area, the regional flood frequency relationship has been developed using the L-moment based GEV distribution. Also, for estimation of floods of desired return periods for the ungauged catchments, the regional flood formula has been developed by coupling the regional flood frequency relationship with the regional relationship between mean annual maximum peak flood and catchment area.

When data of all the 11 bridge sites are used instead of data of only 8 bridge sites, without meeting the L-moment based criteria of regional homogeneity; the deviations in growth factors show that the percentage deviations in general increase from 1.5% to 18.5% for the return periods varying from 5 to 1000 years. Thus, excluding the three catchments for meeting the criteria of regional homogeneity leads to under estimation of floods of various return periods for the gauged catchments. In case of the ungauged catchments, the deviations in flood estimates for return periods 25, 50 and 100 years show that there is under estimation for floods of 25, 50 and 100 return periods for lower range of catchment area i.e. 20 to 80 $km^2$, and there is over estimation for larger size catchments varying in areal extent from 80 to 1000 $km^2$.

# Chapter 1

# INTRODUCTION

Estimation of design flood is one of the important components of planning, design and operation of water resources projects. Information on flood magnitudes and their frequencies is needed for design of hydraulic structures such as dams, spillways, road and railway bridges, culverts, urban drainage systems, flood plain zoning, economic evaluation of flood protection projects etc. 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. These estimates are required for the design of culverts, small to medium-sized bridges, causeways and other drainage works, spillways of farm and other small dams and soil conservation works. It is not possible to define precisely what is meant by "small" and "medium" sized, but upper limits of 25 $km^2$ and 500 $km^2$, respectively, can be considered as general guides. 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.

Methods of flood estimation may be broadly divided into five categories viz. (i) flood formulae and envelope curves, (ii) rational formula, (iii) flood frequency analysis, (iv) unit hydrograph techniques and (v) watershed models. The generally adopted methods of flood estimation are based on two types of approaches viz. (i) deterministic approach, and (ii) probabilistic approach. The deterministic approach is based on the hydrometeorological technique, which requires design storm and the unit hydrograph for a catchment. The probabilistic approach is based on the flood frequency analysis using the observed annual maximum peak flood data. Another alternative of estimating the frequency based floods is to carryout frequency analysis of rainfall data and convolute the design excess-rainfall i.e. excess rainfall of the desired frequency with the unit hydrograph or some rainfall-runoff model appropriate to the catchment.

The conventional flood formulae developed for different regions of India are empirical in nature and do not provide flood estimates for desired return periods. A number of studies have been carried out for estimation of design floods for various structures by different Indian 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, 1985) and regional flood frequency studies carried out by RDSO using the USGS and pooled curve methods (e.g. RDSO, 1991) for some of the hydrometeorological

Subzones of India. Besides these, regional flood frequency studies have also been carried out at some of the academic and research Institutions.

Some of the recent studies based on index flood approach include Wallis and Wood (1985), Hosking et al. (1985), Hosking and Wallis (1986), Lettenmaier et al. (1987), Landwehr et al. (1979), Hosking and Wallis (1988), Wallis (1988), Boes et al. (1989), Jin and Stedinger (1989), Potter and Lettenmaier (1990), Farquharson et al. (1992) etc. Based on some of the comparative flood frequency studies involving use of probability weighted moment (PWM) based at-site, at-site and regional and regional methods as well as USGS method, carried out for some of the typical regions of India (NIH, 1995-96) in general, PWM based at-site and regional GEV method is found to be robust. Farquharson et al. (1992) state that GEV distribution was selected for use in the Flood Studies Report (NERC, 1975) and has been found in other studies to be flexible and generally applicable. Use of a Generalised Extreme Value (GEV) distribution as a regional flood frequency model with an index flood approach has received considerable attention (Chowdhury et al., 1991). Karim and Chowdhary(1995) mention that both goodness-of-fit analysis and L-moment ratio diagram analysis indicated that the three-parameter GEV distribution is suitable for flood frequency analysis in Bangladesh while the two-parameter Gumbel distribuion is not.

L-moments of a random variable were first introduced by Hosking (1990). They are analogous to conventional moments, but are estimated as linear combinations of order statistics. Hosking (1990) defined L-moments as linear combinations of the PWMs. 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). Hosking and Wallis (1997) presented the state of art application of L-moments in frequency analysis. The regional flood frequency curves derived by using the L-moment approach have been coupled with the relationship between annual maximum peak floods and catchment area for development of regional flood frequency relationships and flood formulas for the seven Subzones of India (Kumar et al., 1999).

Annual maximum peak flood data of the 11 gauging sites of Middle Ganga Plains (Subzone 1f) are available for this study. After carrying out the L-moment based discordancy measure ($D_i$) test to examine the suitability of the data for carrying out regional flood frequency analysis as well as regional heterogeneity measure (H) test, the data of 8 gauging sites have been used. Among the various frequency distributions, L moment based GEV distribution has been identified as the robust distribution based on the L-moment ratio diagram and $Z^{dist}$ statistic criteria. For estimation of floods of various return periods for the gauged catchments of Subzone 1(f), regional flood frequency relationship has been developed using the robust frequency distribution. Also, for estimation of floods of various return periods for the ungauged catchments of the study area, regional flood formula has been developed.

# Chapter 2

## REVIEW OF LITERATURE

Statistical flood frequency analysis has been one of the most active areas of research since the last thirty to forty years. However, the questions such as (i) which parent distribution the data may follow? (ii) what should be the most suitable parameter estimation technique? (iii) how to account for sampling variability while identifying the distributions? (iv) what should be the suitable measures for selecting the best fit distribution? (v) what criteria one should adopt for testing the regional homogeneity? and many others remain unresolved. The scope of frequency analysis would have been widened if the parameters of the distribution could have been related with the physical process governing floods. Such relationships, if established, would have been much useful for studying the effects of non-stationarity and man made changes in the physical process on frequency analysis. Unfortunately, this has not been yet possible and the solution of identifying the parent distribution still remains empirical based on the principle of the best fit to the data. However, development of geomorphological unit hydrograph seems to be a good effort towards the physically based flood frequency analysis. Inspite of many drawbacks and limitations, the statistical flood frequency analysis remains the most important means of quantifying floods in systematic manner.

As such there are essentially two types of models adopted in flood frequency analysis literature: (i) annual flood series (AFS) models and (ii) partial duration series models (PDS). Maximum amount of efforts have been made for modelling of the annual flood series as compared to the partial duration series. In the majority of research projects attention has been confined to the AFS models. The main modelling problem is the selection of the probability distribution for the flood magnitudes coupled with the choice of estimation procedure. A large number of statistical distributions are available in literature. Among these the Normal, Log Normal, Gumbel, General Extreme Value, Pearson Type III, Log Pearson Type III, Generalized Pearson, Logistic, Generalized Logistic and Wakeby distributions have been commonly used in most of the flood frequency studies. For the estimation of the parameters of the various distributions the graphical method, method of least squares, method of moments, method of maximum likelihood, method based on principle of maximum entropy, method of probability weighted moment and method of L-moment are some of the methods which have been most commonly used by many investigators in frequency analysis literature. Once the parameters are estimated accurately for the assumed distribution, goodness of fit procedures then test whether or not the data do indeed fit the assumed distribution with a specified degree of confidence. Various goodness of fit criteria have been adopted by many investigators while selecting the best fit distribution from the various distributions fitted with the historical data. However, most of the goodness of fit criteria are conventional and found to be inappropriate for selecting a best fit distribution which may provide an accurate design flood estimate corresponding to the desired recurrence interval.

## 2.1 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 (NERC, 1975), or major physical groupings of sites (Matalas et al., 1975). Even though region boundaries may be adjusted after considering model fit; these approaches seem arbitrary and subjective and the resulting regions rarely give the impression of physical integrity.

It is sometimes possible, particularly in small scale studies, to define regions subjectively by inspection of the site characteristics. Schaefer (1990) analyzing annual maximum peak flood data for sites in Washington state formed regions by grouping together sites with similar values of mean annual precipitation.

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 ( Wiltshire, 1985) within-group variation of the sample coefficient of variation (Wiltshire, 1986). The groups are then further divided in an iterative process until a final set of acceptably homogeneous regions is obtained.

Cluster analysis is a standard method of statistical multivariate analysis for dividing a data set into groups and has been successfully used to form regions for regional frequency analysis. A data vector is associated with each site, and sites are partitioned or aggregated into groups according to the similarity of their data vectors. The data vector can include at-site statistics, site characteristics or some combination of the two. 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.

Hosking and Wallis (1997) regard cluster analysis of site characteristics as the most practical method of forming regions from large data sets. The authors state that it has several major variants and involves subjective decisions at several stages. Some suggestions for the use of cluster analysis in regional frequency analysis are also given by the authors.

For regional frequency analysis with an index-flood procedure there is little advantage in using very large regions. 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.

## 2.2 Test of Regional Homogeneity

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. There are various types of homogeneity tests reported in literature e.g. Dalrymple's (1960) homogeneity test (U.S.G.S. test), and the tests proposed by Acreman and Sinclair (1986), Wiltshire (1986), Choudhury, Stedinger and Lu (1991), Hosking and Wallis (1993). 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, L-moment ratio of a combination thereof.

A test statistic H, termed as heterogeneity measure has been proposed by Hosking and Wallis (1993). It compares inter-site variations in sample L-moments for the group of sites with what would be expected of a homogeneous region, the same has been discussed in Section 6.3.

## 2.3 Methods of Regional Flood Frequency Analysis

Cunnane (1988) mentions twelve different regional flood frequency analysis (RFFA) methods. Out of these methods some of the commonly used methods, namely, (i) Dalrymple's Index Flood method, (ii) N.E.R.C. method, (iii) United States Water Resources Council (USWRC) method, (iv) Bayesian method, and (v) Regional Regression based methods as described in literature are briefly described here under.

### 2.3.1 U.S.G.S. method or Darlymple's index flood method

This method is known as the United States Geological Survey (U.S.G.S.) or Darlymple's index flood method. It was proposed by Dalrymple (1960). It is a graphical regional averaging index flood method, which uses unregulated flood records of equal length N, from each of the rivers considered. The homogeneity test of this method is applied at the 10-year return period level and is based on an assumed underlying EV1 population. For each site, a probability plot is prepared and the following steps are followed.

(i) A smooth, eye-judgement curve is used to estimate the Q-T (Quantile-Return Period) relation at each site;

(ii) The quantile value of return period 2.33 years is read off each graph, corresponding to each site;

(iii) The quantile values for the return periods, T = 2, 5, 10, 25, 50, 100 years are read off from each graph, corresponding to each station;

(iv) The quantile values obtained in step (iii) are standardised by dividing by the $Q_{2.33}$ value obtained in step (ii), for the respective sites;

(v) The median of the standardised values from all sites in the region ($X_T$) is computed for each return period considered;

(vi) $X_T$ is plotted against T on EVI (Gumbel) probabilty paper,

(vii) A smooth, eye-guided curve gives the X-T relationship, which is assumed to hold at every site in the region;

(viii) The estimate of $Q_T$ at any site is obtained from : $Q_T = X_T \bar{Q}$, where $\bar{Q}$ is the mean estimated from flood data available at any site or estimated from catchment characteristics, if flood data are not available.

The USGS method for regional flood frequency analysis as given by Dalrymple (1960) and modified to accommodate unequal length of records as described in the following sequential steps.

(i) Select gauged catchment within the region having more or less similar hydrological characteristics.

(ii) Estimate the parameters of EV1 distribution using method of moments.

(iii) Estimate the mean annual flood $\bar{Q}$ at each station.

(iv) Test homogeneity of data using homogeneity test as explained in (NIH, 1995-96).

(v) Establish the relationship between mean annual flood and catchment characteristics.

(vi) Obtain the ratio $Q_T/\bar{Q}$ for different return periods for each site

(vii) Compute mean ratio for each of the selected return period.

(viii) Fit a Gumbel distribution between these mean ratio and return periods or reduced variates either analytically or plotting mean of $Q_T/\bar{Q}$ against return priod (reduced variate) on Gumbel probability paper.

The end result of above sequential steps is a regional flood frequency curve which can be used for quantile estimation of ungauged catchments. For ungauged sites mean annual flood is computed using the relationship established at step (v).

In the above method as compared to original USGS methods, the modification are in terms of (i) estimation of mean annual flood (ii) the replacement of median ratio by the mean ratio $Q_T/\overline{Q}$ (iii) Variable length of data instead of fixed length of data (iv) parameter estimation by method of moments instead of method of least squares.

### 2.3.2 N.E.R.C. method

This method described in the Flood Studies Report, Natural Environmental Research Council (NERC, 1975) involves the following steps of computation and is based on similar general principles of U.S.G.S. method.

(i) Select the gauged catchments in a more or less hydrologically similar region.

(ii) Compute the mean of annual flood for each station of the region, where short records are available, suitably augment the record by regression.

(iii) Establish relationship between mean annual flood and catchment characteristics.

(iv) For each station in the region plot the ranked annual maximum series $Q_i/\overline{Q}$ against reduced variate $y_i$.

(v) Select intervals on Y scale (reduced variate scale) like (2.0 to - 1.5), (-1.15 to 1.0), ......................(3.5 to 4.0) and for each interval compute mean on all E $(Y_{(i)})$ and mean of $Q_i/\overline{Q}$ and plot them as a smooth mean curve.

(vi) Use this curve as the regional curve for quantile estimation of ungauged catchments.

### 2.3.3 United States Water Resources Council (USWRC) method

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; because of the variability of at-site sample skew coefficient with a generalized skew coefficient, which is a regional estimate of the log-space skewness. The other notable features of this procedure are treatment of outliers and conditional probability adjustments. Though this procedure attempts to combine regional and at-site flood frequency information, the flood quantiles obtained using this method are quite inferior to those obtained from index flood procedures. This is because, in the USWRC method, regional smoothing is effected only in skewness. In addition to being poor in quantile productive ability, USWRC method is also found to be lacking in robustness as both at-site and regional estimators.

### 2.3.4 Bayesian methods

In the Bayes' Theorem for combining prior and sample flood information it was showed how it could be used to combine regional estimates of $\overline{Q}$ and $C_v$ obtained from catchment characteristics, using bivariate lognormal distribution for $\overline{Q}$ and $C_v$ and site data assumed to be EV1 distributed to give a posterior distribution for $Q_T$. This method involves considerable amount of numerical integration. The Bayesian methods do not have to assume perfect regional homogeneity. In fact, specifying a prior distribution itself, acknowledges heterogeneity. The Bayesian method, in given a posterior distribution of parameters, allows legitimate subjective probability statement to be made about parameters and quantiles and this holds even if a non-informative prior distribution (one which is not based on regional flood information, in this context) is used. This is one of its major advantages (Cunnane, 1987). However, Bayesian flood estimation studies which have used informative prior distributions based on regional regression models (which express the parameters in terms of catchment characteristics), have not been successful, since the regression models are quite imprecise. Nash and Shaw (1965) showed that $\overline{Q}$ estimated from catchment characteristics is only as good as $\overline{Q}$ obtained from one year of at-site flood record or less. This result holds for a catchment located at the centroid of the catchment characteristic space. For other catchments, the result is much worse (Hebson and Cunnane, 1986).

### 2.3.5 Regional regression based methods

Regression can be used to derive equations to predict the values of various hydrologic statistics such as means, standard deviations, quantiles and normalized flood quantiles, as a function of physiographic characteristics and other parameters. Such relationships are useful for estimating flood quantiles at various sites in a region, when little or no flood data are available at or near a site. The prediction errors for regression models of flood flows are normally high. Regional regression models have long been used to predict flood quantiles at ungauged sites, and these predictions compare well with the more complex rainfall-runoff methods.

Consider the traditional log-linear model which is to be estimated by using watershed characteristics such as drainage area and slope.

$$y_i = \alpha + \beta_1 \log(\text{Area}) + \beta_2 \log(\text{slope}) + \ldots + \varepsilon$$

A challenge in analyzing this model and estimating its parameters with available records is that it is possible to obtain sample estimates, denoted by $y_i$ of the hydrologic statistics $y_i$. Thus, the observed error $\varepsilon$ is a combination of: (i) the sampling error in sample estimators of $y_i$ (these errors at different sites can be cross-correlated if the records are concurrent) and (ii) underlying model error (lack of fit) due to failure of the model to exactly predict the true value of the $y_i$'s at every site. Often, these problems have been ignored and standard ordinary least squares (OLS) regression has been employed. (Thomas, and Benson, 1970). Stedinger and Tasker (1985, 1986a, 1986b) have developed a specialized Generalized Least Squares (GLS) regression methodology to address these issues. Advantages of the GLS procedure include more efficient parameter estimates when some sites have short records, an unbiased model-error estimator, and a better description of the relationship between hydrologic data and information for hydrologic network analysis and design (Stedinger and Tasker, 1985; Tasker and Stedinger, 1989). Example are provided by Potter and Faulkner (1987), Vogel and Kroll (1989) and Tasker and Driver (1988). Potter and Faulkner (1987) have used catchment response time as a predictor of flood quantiles. The use of this information reduces the standard errors of regression estimates from regional regression equations. Application of this approach requires estimation of catchment response time at an ungauged site. The cost-effectiveness of this approach remains to be investigated.

### 2.3.6 Improvised index-flood algorithms

The index-flood algorithm originally suggested by Dalrymple (1960) to derive the regional flood frequency curve, was once adopted by the U.S. Geological Survey for flood quantile estimation. Subsequently, it was discontinued, since the coefficient of variation of floods was found to vary with drainage area and other basin characteristics (Stedinger, 1983). However, the index-flood methods came into limelight, once again, in the wake of the new estimation algorithm, Probability Weighted Moments (PWMs), proposed by Greenwood et al. (1979), which helped in reducing the uncertainty in estimating the flood quantiles. The graphical method of Dalrymple (1960) was subsequently improvised by Wallis (1980). The improvised algorithm of Wallis (1980) was an objective numerical method, based on regionally averaged, standardised PWMs. Kuczera (1982a,b) adopted lognormal empirical Bayes estimators, which incorporate the index-flood concept. In Kuczera's work, the log-space mean was estimated using only at-site data, while the log-space variance (denoting the shape parameter that determines the coefficient of variation and coefficient of skew of a longnormal distribution), was assigned a weighted average of at-site and regional estimators. Here, the logarithmic transformation is used to effect normalisation, by means of a simple subtraction of the log space mean, thus avoiding the division by an index-flood estimator in real space (Stedinger, 1983).

Greis and Wood (1981) 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. This is a critical source of uncertainty especially for regions with a large mean CV (Lettenmaier et al., 1987). Hosking et al. (1985b) has given a PWM estimation procedure for the Generalised Extreme Value (GEV) Distribution of Jenkinson (1955). Further, Hosking et al. (1985a) have presented an apprisal of the regional flood frequency procedure followed by the UK Flood Studies Report (FSR)(NERC, 1975), in which they have pointed out that FSR algorithm, at times, can lead to unrealistic upper flood quantile estimates. In fact, the Monte-Carlo simulation studies conducted by Hosking et al. (1985a), indicate that the FSR algorithm may result in high degree of overestimation of flood quantile estimates. The advantages of PWM estimators have been brought out by Landwehr et al. (1979), Hosking et al. (1985a), Wallis (1988) and Hosking (1990). The use of L-moments in selection of regional frequency distribution have been dealt with in Chowdhury et al. (1991), Wallis (1993), Hosking and Wallis (1993), Vogel and Fennessey (1993), and Cong et al. (1993). Further, the unbiasedness of the L-Moment estimators have been well exploited in both regional homogeneity tests and Goodness of Fit test (Lu and Stedinger, 1992; Hosking and Wallis, 1993; Zrinji and

Burn, 1994) which are vital steps in regional frequency analysis. Hosking and Wallis (1988) have studied the impact of cross-correlation among concurrent flows at different sites, on regional index-flood methods. They have concluded that regional analysis is preferable to at-site analysis, even in case of regions with mild heterogeneity and moderate inter-site cross correlation. Furthermore, Hosking et al. (1985a) illustrate the impact of historical information on the precision of computed regional growth curves, in case of regions with large number of gauging stations.

Further, Wallis and Wood (1985) and Potter and Lettenmaier (1990) have found the regional-PWM index-flood estimators to be superior to the variations of the USWRC procedure (USWRC, 1982). Lettenmaier et al. (1987) investigated the performance of eight different GEV-PWM index flood estimators and the effect of regional heterogeneity in a more detailed manner. GEV-PWM index flood quantile estimator was found to be robust and had the least RMSE, when compared with all other at-site as well as regional quantile estimators, for mildly heterogeneous regions. Further, with the increase in the degree of regional heterogeneity or the sample size, a two parameter quantile estimator with a regional shape parameter was found to perform the best method based on standardised L-moments.

## 2.4 Some of the Flood Frequency Studies Carried Out in India

A number of studies have been carried out in the area of regional flood frequency analysis in India. Goswami (1972), Thiru Vengadachari et al. (1975), Seth and Goswami (1979), Jhakade et al. (1984), Venkataraman and Gupta (1986), Venkataraman et al. (1986), Thirumalai and Sinha (1986), Mehta and Sharma (1986), James et al., Gupta (1987) and many others have conducted regional flood frequency analysis for some typical regions in India. 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 Perumal and Seth (1985), Singh and Seth (1985), Huq et al. (1986), Seth and Singh (1987) 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 have been already applied. The Bridges and Structures Directorate of the Research, Designs and Standards Organization, Lucknow has carried out studies for design flood estimation based on regional flood frequency approach for various hydrometeorological sub-zones of India.

A comparative study has been carried out 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, General Extreme Value (GEV) and Wakeby distribution based on PWMs. The mean annual peak flood data of 2 bridge catchments for each sub-zone which were excluded while developing the regional flood frequency curves and these are utilized to compute the at site mean annual peak floods. These at site mean values together with the regional frequency curves of the respective sub-zones were used to compute the floods of various return periods for those 2 test catchments in each sub-zone. The descriptive ability as well as predictive ability of the various methods viz. (i) at site methods, (ii) at site and regional methods, and (iii) regional methods has been tested in order to identify the robust flood frequency method. At site and regional methods viz. SRGEV and SRWAKE have been found to estimate floods of various return periods with relatively less bias and comparable root mean square error as well as coefficient of variation. The regional parameters of the GEV distribution have been adopted for development of the regional flood frequency curves. Floods for these test catchments are also estimated using the combined regional flood frequency curves and respective at site mean annual peak floods. Flood frequency curves developed by fitting the PWM based GEV distribution have been coupled with the relationships between mean annual peak flood and catchment area for developing regional flood formulae for each of the seven sub-zones of India (NIH, 1995-96).

For the above mentioned study area, regional flood frequency relationships developed based on PWM approach have been revised based on the method of L moments (NIH, 1997-98) as briefly summarised below. Regional flood frequency curves are developed by fitting L-moment based GEV distribution to annual maximum peak flood data of small to medium size catchments of the seven hydrometeorological Subzones of zone 3 and combined zone 3 of India. These seven Subzones cover an area of about 10,41,661 $km^2$. Effect of regional heterogeneity is studied by comparing the growth factors of various Subzones and combined zone 3. The flood frequency curves based on probability weighted moment (PWM) approach have been compared with the flood frequency curves based on L Moment approach. Relationships developed between mean annual peak flood and catchment area are coupled with the respective regional flood frequency curves for development of the regional flood formulae.

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. The improvised index-flood procedure suggested by Hosking and Wallis (1993) has been used in the study to find an appropriate regional flood frequency distribution and to obtain regional growth curve for a selected region from U.K. Generalized logistic distribution has been prescribed as the regional flood frequency distribution for the region considered. Index-flood based regional model performed the best when compared to all other models considered in predicting flood quantiles at sites with short record length, which is very vital in any regional study.

Upadhyay and Kumar (1999) applied L-moments approach for regional flood frequency analysis for flood estimation at an ungauged site. The study concludes that at gauged sites, regional flood 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. However, for the sites having sufficiently long records, the difference in accuracy of the at-site and regional estimates is very small. The authors recommended that alongside the discharge data collection at gauging sites, emphasis should be given for collection of data about the physiographic and hydrological characteristics of the catchment. This will improve reliability 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.

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 have been 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.

## 2.5 Current Status

Various issues involved in regional flood frequency analysis are testing regional homogeneity, development of frequency curves and derivation of relationship between MAF and the catchment characteristics. Inspite of a large number of existing regionalisation techniques, very few studies have been carried out with somewhat limited scope to test the comparative performance of various methods. Some of the comparative studies have been conducted by Kuczera (1983), Gries and Wood (1983), Lettenmaier and potter (1985) and Singh (1989). 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 (Cunnane, 1989).

Some of the recent studies based on index flood approach include Wallis and Wood (1985), Hosking et al. (1985), Hosking and Wallis (1986), Lettenmaier et al. (1987), Landwehr et al. (1979), Hosking and Wallis (1988), Wallis (1988), Boes et al. (1989), Jin and Stedinger (1989), Potter and Lettenmaier (1990), Farquharson et al. (1992) etc. Farquharson et al. (1992) state that GEV distribution was selected for use in the Flood Studies Report (NERC, 1975) and has been found in other studies to be flexible and generally applicable. Use of a generalized extreme value (GEV) distribution as a regional flood frequency model with an index flood approach has received considerable attention (Chowdhary et al., 1991). Karim and Chowdhary (1995) mention that both goodness-of-fit analysis and L-moment ratio diagram analysis indicated that the three-parameter GEV distribution is suitable for flood frequency analysis in Bangladesh while the two-parameter Gumbel distribution is not. L-moments of a random variable were first introduced by Hosking (1990). They are analogous to conventional moments, but are estimated as linear combinations of order statistics. Hosking (1990) defined L-moments as linear combinations of the PWMs. 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).

Lu and Stedinger (1992) presented sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. The authors state that for a three-parameter GEV distribution the asymptotic variance of probability weighted moments (PWM) quantile estimators have been derived previously. Their study extended the results to obtain the asymptotic variance of normalized GEV/PWM estimators, which are at-site quantile estimators divided by the sample mean. Monte Carlo simulations provided correction factors for use with small samples. Normalized 10-year flood quattile estimators and their sample variances have been used to construct a regional homogeneity test for GEV/PWM index flood analysis. The new test performed better than the R-statistic test proposed before.

Wang (1996) derived the direct estimators of L moments thus eliminating the need for using probability weighted moments. In another study, 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 an 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 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 unduely reduce the efficiency of high-quantile estimation even if the samples have come from a true GEV distribution.

Rao and Hamed (1997) used regional flood frequency analysis to estimate flood quantiles in Wabash river basin. The parent distribution is identified by analyzing the data from number of stations within the basin. L-moments are used to investigate the feasibility of regional frequency analysis in the basin. Basin is shown to be hydrologically heterogeneous. Basin is divided into smaller sub-regions by using L-moments diagrams. The generalized extreme value distribution is recommended to be the regional parent 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 equipment's).

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 III 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 interval are generally accurate to within a percent or two, as tested by simulations, and are obtained by use of neural network.

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 Generalized Pareto (GP) distribution, in general, performed well.

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 can thus be 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. More recent research shows that method of moments quantile estimators have for $-0.25 < k < 0.30$ smaller root mean square error than L moments and MLEs. Examination of the behaviour of MLEs in small samples demonstrates that absurd values the GEV-shape parameter k can be generated. 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.

L-moments of a random variable were first introduced by Hosking (1990). They are analogous to conventional moments, but are estimated as linear combinations of order statistics. Hosking (1990) defined L-moments as linear combinations of the PWMs. 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). 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) presented first complete account of the L-moment approach to regional frequency analysis. It brings together the results that previously were scattered among academic journals and also includes much new material. The authors comprehensively describe the theoretical background to the subject and provide practical advice to the users.

## 2.6 General Methodology

The main issues involved in regional flood frequency analysis and its generalised approach are mentioned here under:

(i) Regional homogeneity

(ii) Degree of heterogeneity and its effects on flood frequency estimates

(iii) Development of a relationship between mean annual peak flood and catchment characteristics for estimation of floods for the ungauged catchments

(iv) Estimation of parameters of the adopted frequency distributions by efficient parameter estimation approach

(v) Identification of a robust flood frequency analysis method based on descriptive ability or predictive ability criteria

Based on data availability and record length of the available data the following approaches may be adopted for developing the flood frequency relationships:

a. At-site flood frequency analysis

b. At-site and regional flood frequency analysis

c. Regional flood frequency analysis

### 2.6.1 At-site flood frequency analysis

(i) Fit various frequency distributions to the at-site annual maximum peak flood data.

(ii) Select the best fit distribution based on descriptive and predictive ability criteria.

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

### 2.6.2 At-site and regional flood frequency analysis

(i) Screen the data and test the regional homogeneity.

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

(iii) Select the best fit distribution based on descriptive and predictive ability criteria.

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

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

### 2.6.3 Regional flood frequency analysis

(i) Screen the data and test the regional homogeneity.

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

(iii) Select the best fit distribution based on descriptive and predictive ability criteria.

(iv) Develop a regional relationship between mean annual peak flood and catchment and physiographic characteristics for the region.

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

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

Regional Flood Frequency (RFFA) provides a procedure for utilizing the obvious spatial coherence of hydrological variables, as one would do in preparing a rainfall map, and thus all available relevant information is incorporated in the flood estimate. It provides at-site regional flood quantile estimates which are superior to the pure at-site estimates, even if the region is moderately heterogeneous. RFFA can be considered a necessity when one considers the case against complete reliance on at-site estimates alone. Two-parameter distributions are not sufficiently flexible to be able to model all plausible flood-parent distributions. Their parsimony in parameters leads to quantile estimates whose standard errors are not excessively large, but whose bias may be excessively so. Three-parameter distributions, on the other hand, are sufficiently flexible to be relatively unbiased, but this is accompanied by unacceptably large standard error. These facts are true both in the case of homogeneous regions and mildly heterogeneous regions. The gains obtained by RFFA in such cases have been documented by Hosking et al. (1985a). Lettenmaier and Potter (1985), Wallis and Wood (1985), Lettenmaier et al. (1987) and have been reviewed by Lettenmaier et al. (1985). Thus, Regionalisation seems to be the most viable way of improving flood quantile estimation. The performance of Probability Weighted Moments (PWM)-based regional index flood procedure, in particular, is so superior to the currently used institutional methods that no viable argument for the continuation of current practice is evident. Particularly, where the flexibility of using a three-parameter distribution is required, the reduction in the variability of flood quantile estimates achieved by proper regionalisation is so large that at-site estimators should not be seriously considered.

Hosking (1990) has defined L-moments which are analogous to conventional Moments and can be expressed as linear functions of probability weighted moments (PWMs). The basic advantages offered by L-Moments over conventional moments in Hypothesis Testing, and identification of distributions, have opened new vistas in the field of regional flood frequency analysis. In this regard, a very recent and significant contribution has been made by Hosking and Wallis (1993 and 1997), which can be regarded as state-of-the-art approach for regional flood frequency analysis.

## 2.7 Effect of Regional Heterogeneity on Quantile Estimates

Cunnane (1989) mentions that regional flood estimation methods are based on the premise that standardized flood variate, such as $X = Q/E(Q)$ has the same distribution at every site in the chosen region. Serious departures from such assumptions could lead to biased flood estimates at some sites. Those catchments whose $C_v$ and $C_s$ values happen to coincide with the regional mean values would not suffer such a bias. If the degree of heterogeneity present is not too great its negative effect may be more than compensated for by the larger sample of sites contributing to parameter estimates. Thus $X_T$ estimated from M sites, which are slightly heterogeneous may be more reliable than $X_T$ estimated from a smaller number, say M/3, more homogeneous sites, especially if flow records are short. Hosking et al. (1985a) studied the effect of regional heterogeneity on quantile estimates obtained by a regional index flood method. A heterogeneous region of 20 stations (j = 1, 2......20) is specified, whose flood populations are GEV distributed with parameters varying linearly, thus reflecting a transition from small to large catchments. This simulation study has shown that the regional algorithms give relatively more stable quantile estimates, compared to at-site estimators. Further, Lettenmaier et al. (1985), using heterogeneous GEV data bases (qualitatively similar to those of Hosking et al., 1985a), as compared the two parameter Gumbel at-site estimator with a variety of regional estimators. The clear conclusion from this study is that if record lengths at individual sites are <30 years, at-site quantile estimates are less reliable than regional estimates, even when the regional heterogeneity is found to be moderate. Lettenmaier and Potter (1985) have used a regional flood distribution at each site depend on the logarithm of the catchment area. This offers the advantage of a controlled simulation study, that has been used to impose heterogeneity on the flood generating populations. They have compared the performance of eight estimators, out of which at-site estimators are two and remaining are regional estimators. They found that the index-flood regional estimators had lower root mean square error than the at-site estimators, even under conditions of moderate heterogeneity.

Stedinger and Lu (1995) examined the performance of at-site and regional GEV(PWM) quantile estimators with various hydrologically realistic GEV distributions, degrees of regional heterogeneity, and record lengths. The main importance of this study is that, it evaluates the performance of the above-said estimators, for different possible hydrologic regions, assuming realistic parameters. They have concluded that the index-flood quantile estimators perform better than other estimators, when regional heterogeneity is small to moderate. Further, they conclude that, for sites with sufficient record length, with significant lack of fit, the shape parameter estimator is preferable. For estimating quantiles at sites with long record length (n>T), the use of at-site GEV (PWM) estimator is suggested from their study.

Hosking and Wallis (1997) mention that when the region is heterogeneous, it is possible that a test makes use of the at-site L-moments might enable better discrimination between distributions. The regional average gives a sufficient summary of the data when the region is homogeneous, but this is no longer the case for a heterogeneous region. For the heterogeneous region the authors consider it more important that the chosen distribution be robust to heterogeneity than that it achieves the ultimate quality of fit. The authors tend to prefer the Wakeby distribution for heterogeneous regions, and also state that in a large investigation there may be many regions, and the choice of frequency distribution for one region may affect the others. If one distribution gives an acceptable fit for all or most of the regions, then it is reasonable to use this distribution for all regions even though it may not be the best for each region individually.

Hence, on the basis of recent studies, it may be concluded that dividing the catchment data set into various parts, for obtaining more internal homogeneity of regions is not necessary or quite useful. On the other hand, more reliable flood frequency estimates may be obtained by considering a few larger and slightly heterogeneous regions, comprising of the larger number of catchments, than many homogenous regions, each with only a smaller number of catchments.

## 2.8 Application of L-Moments as a Parameter Estimator

Some of the commonly used parameter estimation methods for most of the frequency distributions include:

(i) Method of least squares
(ii) Method of moments
(iii) Method of maximum likelihood
(iv) Method of probability weighted moments
(v) Method based on principle of maximum entropy
(vi) Method based on L-moments

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, the parameters are estimated by equating the sample moments to these of the theoretical moments of the 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.

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. In addition, the variance of these estimates is found to increase. Further, a notable drawback with conventional moment ratios such as skewness and coefficient of variation is that, for finite samples, they are bounded, and will not be able to attain the full range of values available to population moment ratios (Kirby, 1974). Wallis et al. (1974) have been shown that the sample estimates of conventional moments are highly biased for small samples and the same results have been extended by Vogel and Fennessey (1993) for large samples ($n>1000$) for highly skewed distributions.

Hosking (1990) has defined L-moments, which are analogous to conventional moments, and can be expressed in terms of linear combinations of order statistics, i.e., L-statistics. L-moments are capable of characterising a wider range of distributions, compared to the conventional moements. A distribution may be specified by its L-moments, even if some of its conventional moments do not exist (Hosking, 1990). For example, in case of the generalised pareto distribution, the conventional skewness is underfind beyond a value of 155, (shape parameter = 1/3), while the L-skewness can be defined, even beyond that value. 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. The advantages offered by L-moments over conventional moments in hypothesis testing, boundedness of moment ratios and identification of distributions have been discussed in detail by Hosking and Wallis (1997). Stedinger et al. (1993) have described the theoretical p[roperties 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 therein. Hosking (1990) has 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). Examples for the usage of L-moment ratio diagrams are found in the works of Wallis (1988), Hosking and Wallis (1987a, 1991), Vogel et al. (1993).

Exact analytical forms of sampling properties of L-moments are extremely complex to obtain. Hosking (1990) has derived approximate analytical forms for the sampling properties of same probability distributions, using asymptotic theory. It is to be noted that even these approximate analytical forms are not available for some of the important distributions, of then used in water resources applications, such as generalised normal (Long normal-3 parameter) distribution and Pearson-3 (three parameter Gamma) distribution Further, the sampling properties obtained from the asymptotic theory using first order approximation, give reliable approximation to finite sample distributions, only when sample size is considerable (Hosking et al., 1985b; Hosking, 1986; Chowdhury et al., (1991). But, often, hydrologic records are available for only short periods. Hence, it is necessary to investigate the sampling properties of L-moments for sample size, for which Monte-Carlo simulation provides a viable alternative. In recent literature (Hosking, 1990; Vogel and Fennessey, 1993; Stedinger et al., 1993), it is stated that L-moment estimators in general, are almost unbiased. However, a detailed investigation of the sampling properties of L-moments has been attempted so far. It is to be noted that sample estimators of L-moments are always linear combinations of the ranked observations, while the conventional sample moment estimators require squaring and cubing the observations respectively, which in turn, increases the weightages to the observations away from the mean, thus resulting in considerable bias. However, a detailed comparison of the sampling properties between conventional moment estimators and L-moment estimators has not been attempted so far.

Utilising the desirable properties of the L-moments such as unbiasedness of the basic moments and normality of the asymptotic distributions of the sampling properties. Hosking and Wallis (1993) have defined a set of regional flood frequency measures namely,i) Discordancy measure ii) Heterogeneity measure and iii) Goodness of fit (GOF) measure. They have suitably incorporated these measures in the modified index flood algorithm suggested by Wallis (1980). This has resulted in a very versatile and efficient regional flood frequency procedure, which has been discussed in detail by Hosking and Wallis (1993). The tests suggested by them for regional heterogeneity and goodness of fit are the most powerful, out of the available tests.

The various regional flood frequency distributions coupled with PWM-based index flood procedure, the different at-site estimators (2-parameters and 3-parameter) and the regional shape parameter based models of various distributions together provide a wide range of choice for the selection of the most competitive flood frequency models for the region/site in question. In such situations, regional Monte-Carlo simulation technique will be very much useful in evaluating the performance efficiency of the different alternative models. A further advantage of adopting the Monte-Carlo simulation technique is that regional data can be easily generated according to the pattern of the real-world data of the region and in addition the true flood quantiles are also known, thus enabling the evaluation of the relative performance between the different models (estimators). A few such regional Monte-Carlo simulation exercises have been carried out in order to establish the performance of regional estimators under different conditions of heterogeneity. Lettenmaier et al. (1987) consider GEV regional population, for a hypothetical region of 21 sites, with their CV, Skewness and length of record varying linearly across the sites. However, in a real world situation, these variations may not be linear as assumed. They considered regions with k=0.15 and an average coefficient of variation = 0.5, 1.0, 1.5 and 2.0. Out of the cases considered, only CV=0.5 represents the realistic regional flood frequency distributions, since the other cases of CV give rise to considerable percent of negative flows in the simulation study. Further, their assumption of mean = 1.0 for all sites creates a source of uncertainty in flood quantile estimates, particularly for regions, where the mean CV is large (Stedinger and Lu, 1994).

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, they 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.

Stedinger and Lu (1994) presented the performance of at-site and regional GEV (PWM) quantile estimators through a comprehensive Monte-Carlo simulation study using hydrologically realistic GEV distributions and varying degrees of heterogeneity, and record lengths. The authors evaluated the performance of these estimators for different possible hydrologic regions, using regional average standardised performance measures. Their Monte-Carlo analysis considers a wide range of realistic values of mean CV and coefficient of variation of CV to represent the different hydrologic regions and different degrees of heterogeneity, respectively.

# Chapter 3

## PROBLEM DEFINITION

For design of various types of hydraulic structures, information on flood magnitudes and their frequencies is needed. 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 flood estimates directly. In such a situation, the flood formulae developed for the region are the alternative method for estimation of design flood. Most of the flood formulae developed for different regions of the country are empirical in nature and do not provide flood estimates for the desired return period and there is a need for developing regional flood frequency relationships and flood formulae for estimation of floods of different return periods for the gauged and ungauged catchments of the country. L-moments of a random variable were first introduced by Hosking (1990). Hosking (1990) defined L-moments as linear combinations of the PWMs and presented their state of art applications in the area of frequency analysis. In a wide range of hydrologic applications, L-moments provide simple and efficient estimators of characteristics of hydrologic data and of a distribution's parameters (Stedinger et al., 1992). The objectives of this study are:

(a) To screen the data using discordancy measure ($D_i$) test for examining suitability of the data for flood frequency analysis.

(b) To test regional homogeneity using the available annual maximum peak flood data of Subzone 1(f).

(c) To carryout comparative regional flood frequency analysis studies employing some of the commonly adopted frequency distributions using L-moments approach, and to identify robust regional flood frequency distribution based on L-moment ratio diagram and $Z^{dist}$ statistic criteria.

(d) To develop regional flood frequency relationship for estimation of floods for different return periods for the gauged catchments of the study area using the robust frequency distribution.

(e) To develop regional relationship between mean annual peak floods and physiographic characteristics for estimating the mean annual peak flood for the ungauged catchments of Subzone 1(f).

(f) To couple the regional relationship between mean annual peak flood and physiographic characteristics with the regional flood frequency relationship for developing the regional flood formula for estimation of floods of various return periods for ungauged catchments of Subzone 1(f).

# Chapter 4

## DESCRIPTION OF THE STUDY AREA

As described in CWC (1985) the Middle Ganga Plains Subzone (1-f) lies between latitude $24^0$ to $29^0$ North and longitude $80^0$ to $89^0$ East. The total areal extent of the Subzone (1-f) is about 1, 71, 350 $km^2$. It covers parts of Uttar Pradesh, Bihar and West Bengal. It is bounded in the North by Nepal, in the South by Chambal, Sone, and Hoogly basins (Subzones 1(c), 1(d) and 1(g)), in the East by Bangladesh and in the West by Upper Indo-Gangetic Plains Subzone 1(e). Fig. 1 shows location of the Subzone 1 (f). Some of the important cities in the Subzone are Kanpur, Lucknow, Allahabad, Gorakhpur, Varanasi in Uttar Pradesh and Darbanga, Saharsa, Purnea and Katihar in the state of Bihar and Siliguri and Malda in West Bengal.

The main rivers flowing in this Subzone are the Ganga, Yamuna, Gomti, Gandak, Ghagra, Rapti, Kosi including Kamla and Mahananda. Table 1 gives the drainage areas of the rivers flowing in the Subzone.

**Table 1: Various rivers/tributaries draining Subzone 1(f) and their drainage areas**

| Sl. No. | River/Tributary | Drainage area ($km^2$) |
|---|---|---|
| 1 | Yamuna | 7530 |
| 2 | Ganga | 14400 |
| 3 | Ghagra | 35950 |
| 4 | Gomati | 25270 |
| 5 | Gandak | 26380 |
| 6 | Rapti | 14160 |
| 7 | Kosi including Kamala | 17900 |
| 8 | Mahananda | 16830 |
| 9 | Others | 12930 |
| **Total area** | | **171350** |

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 Tarai area exceeds 150 m. In the plain area the elevation lies between 150 m to 175 m and goes on decreasing eastwards to Bangladesh. The rivers in this Subzone have meandering tendency with wide and shallow channels.


**Fig. 1: Location of Middle Ganga Plains Subzone 1(f)**

The mean annual rainfall in this Subzone varies from 800 mm to 1200 mm in the plains and goes upto 2000 mm in the foothill regions lying in the North of the Subzone. The major portion of the rainfall is received during the five monsoon months viz. June to October from the Southeast monsoon. The Subzone 1 (f) experiences extremes of hot and cold climate. Towards the Northwest Tarai region, the mean annual temperature varies from $22.5^0$ C to $25^0$ C. For the rest of the Subzone the mean annual temperature varies from $25^0$ C to $27.5^0$ C.

Major portion of the Subzone has alluvial soils of recent origin except the Tarai region and the plains on the North-Eastern side between Rapti and Kosi rivers where Tarai and Calcarious alluvium soils are encountered respectively. The plains of the Subzone are fertile and cultivable. Most of the portions of the Subzone are irrigated and most of the land in the Subzone is arable and well irrigated. Forests are present in a part of Tarai portion of the Subzone.

# Chapter 5

## DATA AVAILABILITY FOR THE STUDY

Annual maximum peak flood data of 11 gauging sites lying in the Subzone 1 (f) and varying over 11 to 33 years in record length, have been used. As shown in Table 2, catchment areas of these sites vary from 27.45 to 712 $km^2$ and mean annual peak floods of these sites vary from 24.29 to 555.21 cubic meter per second.

**Table 2: Bridge number, catchment area and record length for the 11 bridge sites of Subzone 1(f)**

| Sl. No. | Bridge Number | Catchment Area ($km^2$) | Mean Annual Peak Flood ($m^3/s$) | Record Length (Years) |
|---|---|---|---|---|
| 1 | 85 | 136.70 | 72.00 | 11 |
| 2 | 59 | 54.39 | 97.48 | 33 |
| 3 | 30 | 447.76 | 490.50 | 30 |
| 4 | 160 | 150.40 | 70.31 | 32 |
| 5 | 3 | 32.89 | 24.29 | 31 |
| 6 | 177 | 712.00 | 239.93 | 14 |
| 7 | 60 | 130.00 | 138.70 | 27 |
| 8 | 24 | 69.75 | 59.31 | 26 |
| 9 | 48 | 27.45 | 83.31 | 26 |
| 10 | 141 | 59.83 | 79.39 | 23 |
| 11 | 104 | 234.11 | 555.21 | 29 |

# Chapter 6

# METHODOLOGY

The following aspects of methodology used for development of L-moment based regional flood frequency relationship for gauged catchments as well as development of regional flood formula for estimation of floods of various return periods for ungauged catchments are discussed as follows.

(i) Probability weighted moments (PWMs) and L-moments,
(ii) Data screening,
(iii) Test of regional homogeneity,
(iv) Frequency distributions used,
(v) Goodness of fit measures, and
(vi) Development of relationship between mean annual peak flood and catchment area.

## 6.1 Probability Weighted Moments (PWMs) and L-moments

L-moments of a random variable were first introduced by Hosking (1990). Hosking and Wallis (1997) state that L-moments are an alternative system of describing the shapes of probability distributions. Historically they arose as modifications of the probability weighted moments' (PWMs) of Greenwood et al. (1979).

### 6.1.1 Probability weighted moments (PWMs)

Probability weighted moments are defined by Greenwood et al. (1979) as:

$$M_{i,j,k} = \int_0^1 x(F)^i \, (F)^j (1-F)^k \, dF \qquad (1)$$

where, $F = F(x) = \int_{-x}^{x} f(x)\,dx$ is the cumulative density function and x (F) is the inverse of it; i, j, k are the real numbers. The particularly useful special cases of the PWMs $\alpha_k$ and $\beta_j$, are:

$$\alpha_k = M_{1,0,k} = \int_0^1 x(F)\,(1-F)^k \, dF \qquad (2)$$

$$\beta_j = M_{1,j,0} = \int_0^1 x(F)\,(F)^j \, dF \qquad (3)$$

These equations are in contrast with the definition of the ordinary conventional moments, which may be written as:

$$E(X^r) = \int \{x(F)\}^r \, dF \tag{4}$$

The conventional moments or "*product moments*" involve higher powers of the quantile function x(F); whereas, PWMs involve successively higher powers of non-exceedance probability (F) or exceedance probability (1-F) and may be regarded as integrals of x(F) weighted by the polynomials $F^r$ or $(1-F)^r$. As the quantile function x(F) is weighted by the probability F or (1-F), hence these are named as probability weighted moments. The PWMs have been used for estimation of parameters of probability distributions as described in Chapter 2.

However, PWMs are difficult to interpret as measures of scale and shape of a probability distribution. This information is carried in certain linear combinations of the PWMs. These linear combinations arise naturally from integrals of x(F) weighted not by polynomials $F^r$ or $(1-f)^r$ but by a set of orthogonal polynomials (Hosking and Wallis, 1997).

### 6.1.2 L-moments

Hosking (1990) defined L-moments as linear combination of probability weighted moments. In general, in terms of $\alpha_k$ and $\beta_j$, L-moments are defined as:

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^{r} p^*_{r,k} \, \alpha_k = \sum_{k=0}^{r} p^*_{r,k} \, \beta_k \tag{5}$$

where, $p^*_{r,k}$ is an orthogonal polynomial (shifted Legender polynomial) expressed as:

$$p^*_{r,k} = (-1)^{r-k} \, {}^rC_k \, {}^{r+k}C_k = \frac{(-1)^{r-k}(r+k)!}{(k!)^2 (r-k)!} \tag{6}$$

L-moments are easily computed in terms of probability weighted moments (PWMs) as given below.

$$\lambda_1 = \alpha_0 = \beta_0 \tag{7}$$

$$\lambda_2 = \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0 \tag{8}$$

$$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{9}$$

$$\lambda_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\,\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 + \beta_0 \tag{10}$$

The procedure based on PWMs and L-moments are equivalent. However, L-moments are more convenient, as these are directly interpretable as measures of the scale and shape of probability distributions. Clearly $\lambda_1$, the mean, is a measure of location, $\lambda_2$ is a measure of scale or dispersion of random variable. It is often convenient to standardise the higher moments so that they are independent of units of measurement.

$$\tau_r = \frac{\lambda_r}{\lambda_2} \quad \text{for} \quad r = 3, 4 \tag{11}$$

Analogous to conventional moment ratios, such as coefficient of skewness $\tau_3$ is the L-skewness and reflects the degree of symmetry of a sample. Similarly $\tau_4$ is a measure of peakedness and is referred to as L-kurtosis. These are defined as:

L-coefficient of variation (L-CV), ($\tau$) $= \lambda_2 / \lambda_1$

L-coefficient of skewness, L-skewness ($\tau_3$) $= \lambda_3 / \lambda_2$

L-coefficient of kurtosis, L-kurtosis ($\tau_4$) $= \lambda_4 / \lambda_2$

Symmetric distributions have $\tau_3 = 0$ and its values lie between -1 and +1. Although the theory and application of L-moments is parallel to that of conventional moments, L-moment have several important advantages. Since sample estimators of L-moments are always linear combination of ranked observations, they are subject to less bias than ordinary product moments. This is because ordinary product moments require squaring, cubing and so on of observations. This causes them to give greater weight to the observations far from the mean, resulting in substantial bias and variance.

## 6.2 Data Screening

In flood frequency analysis, the data collected at various sites should be true representative of the annual maximum peak flood measured and must be drawn from the same frequency distribution. The first step in flood frequency analysis is to verify that the data are appropriate for the analysis. The preliminary screening of the data must be carried out to ensure that the above requirements are satisfied. Errors in data may occur due to incorrect recording or transcription of the data values or due to shifting of the gauging site to a different location as well as due to changes in the measuring practices or as a result of water resources development activities. Tests for outliers and trends are well established in the statistical literature (e.g., Barnett and Lewis, 1994; W.R.C., 1981; Kendall, 1975). For comparison of data observed from different sites, some techniques such as double mass plots or quantile-quantile plots are commonly used.

Hosking and Wallis (1997) mention that in the context of regional frequency analysis using L-moments, useful information can be obtained by comparing the sample L-moment ratios for different sites, incorrect data values, outliers, trends and shifts in the mean of a sample can all be related to L-moments of the sample. A convenient amalgamation of the L-moment ratios into a single statistic, a measure of discordancy between L-moment ratios of a site and the average L-moment ratios of a group of similar sites, has been termed as "discordancy measure", $D_i$.

### 6.2.1 Discordancy measure

The aim of the discordancy measure is to identify those sites from a group of given sites that are grossly discordant with the group as a whole. Discordancy is measured in terms of the L-moments of the data of the various sites as defined below (Hosking and Wallis (1997)). Suppose that there are N sites in the group. Let $u_i = [t^{(i)} \; t_3^{(i)} \; t_4^{(i)}]^T$ be a vector containing the t, $t_3$ and $t_4$ values for site i: T denotes transposition of a vector or matrix. Let

$$\bar{u} = N^{-1}\sum_{i=1}^{N} u_i \qquad (12)$$

be the (unweighted) group average. The matrix of sums of squares and cross products is defined as:

$$A = \sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T \qquad (13)$$

The discordancy measure for site i is defined as:

$$D_i = \frac{1}{3}N(u_i - \bar{u})^T A^{-1}(u_i - \bar{u}) \qquad (14)$$

The site i is declared to be discordant if $D_i$ is larger than the critical value of the discordancy statistic $D_i$ given in Table 3.

For a discordancy test with significance level $\alpha$ an approximate critical value of $\max_i D_i$ is $(N-1)Z/(N-4+3Z)$, where Z is the upper $100\alpha/N$ percentage point of an F distribution with 3 and N-4 degrees of freedom. This critical value is a function of $\alpha$ and N, where $\alpha = 0.10$. $D_i$ is likely to be useful only for regions with $N \geq 7$.

**Table 3: Critical values of discordancy statistic, $D_i$ (adapted from Hosking and Wallis, 1997)**

| No. of sites in region | Critical value | No. of sites in region | Critical value |
|---|---|---|---|
| 5 | 1.333 | 10 | 2.491 |
| 6 | 1.648 | 11 | 2.632 |
| 7 | 1.917 | 12 | 2.757 |
| 8 | 2.140 | 13 | 2.869 |
| 9 | 2.329 | 14 | 2.971 |
| | | ≥ 15 | 3 |

## 6.3 Test of Regional Homogeneity

A test statistic H, termed as heterogeneity measure has been proposed by Hosking and Wallis (1993). 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 variation of L-moment ratio is measured as the standard deviation (V) of the at-site LCV's weighted proportionally to the record length at each site. To establish what would be expected of a homogeneous region, simulations are used. A number of, say 500 data regions are generated based on the regional weighted average statistics using a four parameter distribution e.g. Kappa or Wakeby distribution. The inter-site variation of each generated region is obtained and the mean ($\mu_v$) and standard deviation ($\sigma_v$) of the computed inter-site variation is obtained.

Let the proposed region has N sites with site i having record length $n_i$ and sample L-moment ratios $t^{(i)}$, $t_3^{(i)}$, and $t_4^{(i)}$. The regional average L-CV, L-Skewness and L-Kurtosis are weighted proportionally to the sites' record length for example, $t^R$ mentioned below. The various steps involved in computation of heterogeneity measure (H) are mentioned below.

(i) Compute the weighted regional average L moment ratios

$$t^R = \sum_{i=1}^{N_s} n_i \, t^{(i)} \Big/ \sum_{i=1}^{N_s} n_i \qquad (15)$$

The value of $t_3^R$ and $t_4^R$ can also be computed similarly by replacing $t^{(i)}$ by $t_3^{(i)}$, and $t_4^{(i)}$.

(ii) Compute the weighted standard deviation of at site LCV's ($t^{(i)}$)

$$V = \left[ \sum_{i=1}^{N} n_i \, (t^{(i)} - t^R)^2 \Big/ \sum_{i=1}^{N} n_i \right]^{1/2} \qquad (16)$$

(iii) Fit a general 4-parameter distribution (Kappa or 4 parameter Wakeby etc.) to the regional average L-moment ratios, $t^R$, $t_3^R$ and $t_4^R$.

(iv) Simulate a large number of regions say 500 having same record lengths as the observed data of the proposed region.

(v) Repeat steps 1 and 2 for each of the 500 simulated regions and calculate the weighted standard deviations for each simulated region and take it as $v_1, v_2, v_3, \ldots\ldots v_{500}$.

(vi) Compute the mean ($\mu_v$) and standard deviation ($\sigma_v$) of the values obtained in step (v).

(vii) Compute the Heterogeneity measure H as given below.

$$H = \frac{V - \mu_V}{\sigma V} \tag{17}$$

The criteria established by Hosking and Wallis (1993) for assessing heterogeneity of a region is as follows.

| | |
|---|---|
| If $H < 1$ | Region is acceptably homogeneous. |
| If $1 \leq H < 2$ | Region is possibly heterogeneous. |
| If $H \geq 2$ | Region is definitely heterogeneous. |

## 6.4 Frequency Distributions Used

The following commonly adopted 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).

### 6.4.1 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:

$$x(F) = u - \alpha \ln(-\ln F) \tag{18}$$

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

### 6.4.2 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:

$$x(F) = u + \alpha \{1 - (-\ln F)^{k}\} / k ; \qquad k \neq 0 \qquad (19)$$

$$= x(F) = u - \alpha \ln(-\ln F) \qquad k = 0 \qquad (20)$$

Where, u, $\alpha$ 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.

### 6.4.3 Logistic distribution (LOS)

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

$$x(F) = u - \alpha \ln \{(1-F) / F\} \qquad (21)$$

Where, u and $\alpha$ are location and scale parameters respectively.

### 6.4.4 Generalized logistic distribution (GLO)

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

$$x(F) = u + [\alpha [ 1 - \{(1-F) / F\}^{k} ] / k; \qquad k \neq 0 \qquad (22)$$

$$x(F) = u - \alpha \ln \{(1-F) / F\}; \qquad k = 0 \qquad (23)$$

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

### 6.4.5 Generalized Pareto distribution (GPA)

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

$$x(F) = u + \alpha \{ 1 - (1-F)^{k} \} / k; \qquad k \neq 0 \qquad (24)$$

$$x(F) = u - \alpha \ln(1-F) \qquad k = 0 \qquad (25)$$

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

### 6.4.6 Generalized normal distribution (GNO)

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

$$F(x)=\phi\left[-k^{-1}\log\{1-k(x-\xi)/\alpha\}\right] \tag{26}$$

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

### 6.4.7 Pearson Type-III distribution (PT-III)

The inverse form of the Pearson type-III distribution is not explicitly defined. Hosking and Wallis (1997) mention that the Pearson type-III 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 $\mu$, the standard deviation $\sigma$, and the skewness $\gamma$. 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 $\mu$, $\sigma$ and $\gamma$. 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 $\mu$ and standard deviation $\sigma$. 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 $\Gamma(.)$ is Gamma function. If $\gamma > 0$, then the range of x is $\xi \le x < \infty$ and the cumulative distribution function is:

$$F(x) = G\left(\alpha, \frac{x-\xi}{\beta}\right)\Big/\Gamma(\alpha) \tag{27}$$

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

$$F(x) = 1 - G\left(\alpha, \frac{\xi-x}{\beta}\right)\Big/\Gamma(\alpha) \tag{28}$$

### 6.4.8 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).

$$x(F)=\xi+\alpha\left[1-\left\{(1-F)^{h}/h\right\}^{k}\right]/k \qquad (29)$$

where, $\xi$ is the location parameter, $\alpha$ is the scale parameter.

When h = -1, it becomes Generalized logistic (GLO) distribution; h = 0 is the Generalized extreme value (GEV) distribution; and h = 0 is the Generalized Pareto (GPA) distribution. 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).

### 6.4.9 Wakeby distribution (WAK)

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

$$x(F)=\xi+\frac{\alpha}{\beta}\left\{1-(1-F)^{\beta}\right\}-\frac{\gamma}{\delta}\left\{1-(1-F)-\delta\right\} \qquad (30)$$

where, $\xi$, $\alpha$, $\beta$, $\gamma$, and $\delta$ are the parameters of the Wakeby distribution.

## 6.5 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. A number of methods are available for testing goodness of fit of the proposed flood frequency analysis models. These include Chi-square test, Kolmogorov-Smirnov test, descriptive ability tests and the predictive ability tests. Cunnane (1989) has brought out a comprehensive description of the descriptive ability tests and the predictive ability tests. Apart from the aforementioned tests the recently introduced L-moment ratio diagram based on the approximations given by Hosking (1991) and the goodness of fit or behavior analysis measure for a frequency distribution given by statistic $Z_i^{dist}$ described below, are also used to identify the suitable frequency distribution.

### 6.5.1 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).

### 6.5.2 $Z_i^{dist}$ statistic as a goodness-of-fit measure

In this method also the objective is to identify a distribution which 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 goodness-of-fit measure for a distribution is given by statistic $Z_i^{dist}$.

$$Z_i^{dist} = \frac{\left(\bar{\tau}_i^R - \tau_i^{dist}\right)}{\sigma_i^{dist}} \tag{31}$$

where $\bar{\tau}_i^R$ - weighted regional average of L-moment statistic i, $\tau_i^{dist}$ and $\sigma_i^{dist}$ are the simulated regional average and standard deviation of L-moment statistics i for a given distribution.

The distribution giving the minimum $|Z^{dist}|$ value is considered as the best fit distribution. When all the three L-moment ratios are considered in the goodness-of-fit test, the distribution that gives the best overall fit when all the three statistics are considered together is selected as the underlying regional frequency distribution. According to Hosking (1993), distribution is considered to give good fit if $|Z^{dist}|$ is sufficiently close to zero, a reasonable criteria being $|Z^{dist}| \leq 1.64$.

Let the homogeneous region has $N_s$ sites with site i having record length $n_i$ and sample L-moment ratios $t_i$, $t_{3i}$ & $t_{4i}$. Steps involved in computation of statistic $Z_i^{dist}$ are:

i. Compute the weighted regional average L-moment ratios.

$$t^R = \frac{\sum_{i=1}^{N_S} n_i t_i}{\sum_{i=1}^{N_S} n_i} \tag{32}$$

The values of $t_3^R$ and $t_4^R$ are computed similarly by replacing $t_i$ by $t_{3i}$ and $t_{4i}$ respectively.

ii. Fit the candidate distribution to the regional average L-moment ratios $t^R$, $t_3^R$ and $t_4^R$ and mean = 1.

iii. Use the fitted distribution to simulate a number of regions, say 500, having same record length as the observed data.

iv. Repeat step 1 for each simulated region and the weighted regional average for the simulations are taken as $t_1^R$, $t_2^R$ ... $t_{500}^R$ and similarly for $t_3^R$ & $t_4^R$.

v. Compute the mean $\left(\tau_i^{dist}\right)$ and standard deviation $\left(\sigma_i^{dist}\right)$ for the values computed in step 4 above for each L-moment statistic i.

vi. Goodness-of-fit measure $Z_i^{dist}$ is computed as $Z_i^{dist} = \frac{\bar{\tau}_i^R - \tau_i^{dist}}{\sigma_i^{dist}}$ (33)

vii. Repeat the steps 2 to 6 for each of the distributions. Distribution giving the minimum $\left|Z_i^{dist}\right|$ value for the L-moment statistics is identified as the best fit distribution.

## 6.6 Development of Relationship Between Mean Annual Peak Flood and Catchment Characteristics

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,

$$\bar{Q} = a\, A^b S^c D^d R^e \qquad (34)$$

Here, $(\bar{Q})$ 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 for the catchments, a, b, c, d, and e are the regional coefficients to be estimated using the mean annual peak floods of the gauged catchments and A, S, D and R which are the physiographic and climatic characteristics of the gauged catchments of the region.

# Chapter 7

## ANALYSIS AND DISCUSSION OF RESULTS

The annual maximum peak flood data of the 11 bridge sites are available for carrying out the study. Screening of the data has been carried out using the discordancy measure, $D_i$. Homogeneity of the region has been tested using the heterogeneity measure, H. Goodness of fit has been tested using the L-moment ratio diagram as well as $Z^{dist}$ statistic for identifying the robust distribution. Regional flood frequency relationship and regional flood formula have been developed for estimation of floods of various return periods for gauged and ungauged catchments for Middle Ganga Plains (Subzone 1-f), as described below.

## 7.1 Screening of Data using Discordancy Measure Test

The objective of the discordancy measure ($D_i$) test is to identify those sites from a group of given sites that are grossly discordant with the group as a whole. Discordancy measure has been computed in terms of the L-moments of the data for all the 11 bridge sites of the Subzone 1(f), as discussed in Section 6.2.1 and the same are given in Table 4. As per Table 3 given in Section 6.2.1, the critical value for the discordancy statistic $D_i$ for the 11 sites is 2.632. It is observed from Table 4 that the $D_i$ values for all the 11 sites are less than the critical $D_i$ value of 2.632. Hence, as per the discordancy measure test, data of all the 11 sites may be utilised for carrying out the flood frequency analysis.

**Table 4: $D_i$ values for the 11 bridge sites of Subzone 1(f)**

| S. No. | Bridge Number | Sample Size (Years) | $D_i$ Value |
|---|---|---|---|
| 1 | 85 | 11 | 2.10 |
| 2 | 59 | 33 | 1.95 |
| 3 | 30 | 30 | 1.73 |
| 4 | 160 | 32 | 0.41 |
| 5 | 3 | 31 | 0.04 |
| 6 | 177 | 14 | 0.36 |
| 7 | 60 | 27 | 2.59 |
| 8 | 24 | 26 | .22 |
| 9 | 48 | 26 | .25 |
| 10 | 141 | 23 | .77 |
| 11 | 104 | 29 | .69 |

## 7.2 Regional Homogeneity Test

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 for computing the heterogeneity measure H, using the four parameter Kappa distribution. Kappa distribution includes as special cases the GLO, GEV and GPA distributions and it is capable of representing many of the distributions. Its L-moments can be chosen to match the group average L-CV, L-skewness and L-kurtosis of the observed data (Hosking and Wallis, 1997).

The heterogeneity measure for Subzone 1(f) using the data of 11 sites was computed and the same was found to be greater than 1.0. Based on the statistical properties (L-moment ratio) one by one three sites of the region were excluded till H value less than 1.0 was obtained. Thus, the region comprising of 8 sites was identified as the homogenous region. The values of heterogeneity measure computed by carrying out 500 simulations using the Kappa distribution based on the data of 8 sites are given in Table 5.

**Table 5: Heterogeneity measures for Subzone 1(f)**

| S. No. | Heterogeneity measures | Values |
|---|---|---|
| 1. | **Heterogeneity measure (H1)** | |
| | (a) Observed standard deviation of group L-CV | .0462 |
| | (b) Simulated mean of standard deviation of group L-CV | .0385 |
| | (c) Simulated standard deviation of standard deviation of group L-CV | .0108 |
| | (d) **Standardized test value H (1)** | **0.71** |
| 2. | **Heterogeneity measure H (2)** | |
| | (a) Observed average of L-CV / L-Skewness distance | 0.0976 |
| | (b) Simulated mean of average L-CV / L-Skewness distance | 0.0812 |
| | (c) Simulated standard deviation of average L-CV / L-Skewness distance | 0.0185 |
| | (d) **Standardized test value H (2)** | **0.89** |
| 3. | **Heterogeneity measure (H3)** | |
| | (a) Observed average of L-Skewness/L-Kurtosis distance | .1333 |
| | (b) Simulated mean of average L-Skewness/L-Kurtosis distance | .0962 |
| | (c) Simulated standard deviation of average L-Skewness/L-Kurtosis distance | .0211 |
| | (d) **Standardized test value H (3)** | **1.76** |

The discordancy measure of the 8 bridge sites of the Subzone 1(f) whose data have been identified to constitute the homogeneous region has been computed and are given in Table 6. As per Table 3 given in Section 6.2.1, the critical value for the discordancy statistic $D_i$ for the 8 sites is 2.140. It is observed from Table 6 that the $D_i$ values for the 8 sites are less than the critical $D_i$ value of 2.140. Hence, as per the discordancy measure test, data of these 8 sites have been used for development of the regional flood frequency relationship and the regional flood formula for Middle Ganga Plains (Subzone 1-f). The details of catchment area, sample size and sample statistics for the 8 sites which form the homogeneous region are given in Table 7.

**Table 6: $D_i$ values for the 8 bridge sites of Subzone 1(f)**

| S. No. | Bridge Number | Sample Size (Years) | $D_i$ Value |
|---|---|---|---|
| 1 | 59 | 33 | 1.63 |
| 2 | 30 | 30 | 1.24 |
| 3 | 160 | 32 | 0.28 |
| 4 | 3 | 31 | 0.89 |
| 5 | 60 | 27 | 2.08 |
| 6 | 24 | 26 | 0.08 |
| 7 | 141 | 23 | 0.55 |
| 8 | 104 | 29 | 1.26 |

**Table 7: Catchment area, sample statistics and sample size for the 8 bridge sites of Subzone 1(f)**

| S. No. | Bridge Number | Catchment Area (A) | Mean Annual Peak Flood ($\bar{Q}$) | Standard Deviation (SD) | Coefficient of Variation (CV) | Coefficient of Skewness (CS) | Sample Size (Years) (SS) |
|---|---|---|---|---|---|---|---|
| 1 | 59 | 54.39 | 97.48 | 52.85 | 0.542 | -0.233 | 33 |
| 2 | 30 | 447.76 | 490.5 | 277.93 | 0.567 | 0.322 | 30 |
| 3 | 160 | 150.4 | 70.31 | 37.68 | 0.536 | 0.861 | 32 |
| 4 | 3 | 32.89 | 24.29 | 16.99 | 0.699 | 0.915 | 31 |
| 5 | 60 | 130 | 140.56 | 73.16 | 0.52 | 3.503 | 27 |
| 6 | 24 | 69.75 | 59.31 | 33.99 | 0.573 | 0.441 | 26 |
| 7 | 141 | 59.83 | 79.39 | 47.04 | 0.593 | 0.682 | 23 |
| 8 | 104 | 234.19 | 555.21 | 422.62 | 0.761 | 2.542 | 29 |

## 7.3 Identification of Regional Frequency Distribution

The choice of an appropriate frequency distribution for a homogeneous region is made by comparing the moments of the distributions to the average moments statistics from regional data.

The aim of goodness-of-fit measure or the behaviour analysis 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. In this study, the L-moment ratio diagram and $Z_i^{dist}$ have been used as goodness of fit measures for identifying the regional distribution. The regional averages of L-moment statistics for Subzone 1(f) are given below.

The L-moment ratio diagram based on approximations provided by Hosking (1991) has been used to identify the suitable regional flood frequency distribution. As shown in Fig. 2, the GEV distribution lies closest to the point defined by the regional average values of L-skewness i.e. $\tau_3$ = 0.2077 and L-kurtosis i.e. $\tau_4$ = 0.1494, and the same is identified as the regional distribution, as per this criteria.

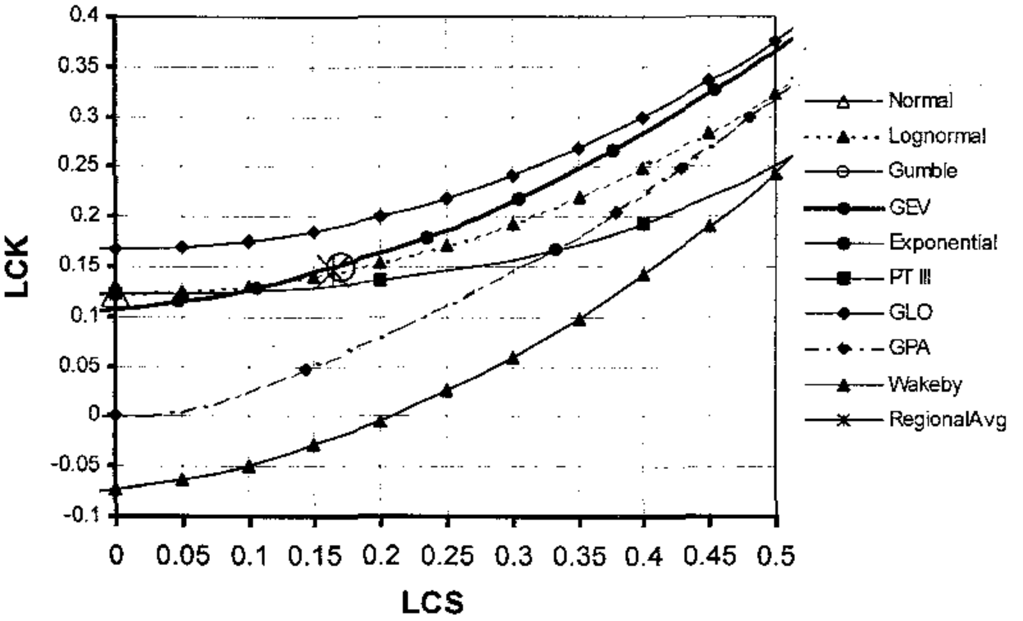


**Fig. 2 L-moment ratio diagram for Subzone 1(f)**

The $Z^{dist}$ –statistic for the various three parameter distributions is given in Table 8. From Table 8 it is observed that the $Z^{dist}$ –statistic value is lower than 1.64 for the four distributions viz. GEV, GNO, PT-III and GLO. Further the $Z^{dist}$ –statistic is found to be the lowest for GEV distribution i.e 0.01; which is very close to 0.0. Thus, the L-moment ratio diagram as well as $Z^{dist}$ –statistic criteria ascertain that the GEV distribution is the robust distribution for Subzone 1(f).

**Table 8: $Z_i^{dist}$ Statistic for various distributions for Subzone 1(f)**

| S. No. | Distribution | Z-Statistic |
|---|---|---|
| 1 | **GEV** | **0.01** |
| 2 | GNO | -0.14 |
| 3 | PT-III | -0.62 |
| 4 | GLO | 1.58 |
| 5 | GPA | 3.40 |

The values of regional parameters for the various distributions which have $Z^{dist}$ –statistic value less than 1.64 as well as the five parameter Wakeby distribution are given in Table 9. As even for heterogeneous regions, it is important to use a distribution that is robust to moderate heterogeneity in the at-site frequency distribution. It is therefore preferred to use Wakeby distribution for heterogeneous regions. Further, the Wakeby distribution which has five parameters, more than most of the common distributions 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.

**Table 9: Regional parameters for the various distributions for Subzone 1(f)**

| Distribution | Parameters of the Distribution | | | | |
|---|---|---|---|---|---|
| **GEV** | **$u = 0.734$** | **$\alpha = 0.468$** | **$k = 0.010$** | | |
| GNO | $\xi = 0.906$ | $\alpha = 0.544$ | $k = -0.337$ | | |
| PT-III | $\mu = 1$ | $\sigma = 0.588$ | $\gamma = 0.994$ | | |
| GLO | $\xi = 0.915$ | $\alpha = 0.308$ | $k = -0.164$ | | |
| WAK | $\xi = 0.109$ | $\alpha = 1.708$ | $\beta = 2.525$ | $\gamma = 0.362$ | $\delta = 0.108$ |

## 7.4 Development of Regional Flood Frequency Relationship for Gauged Catchments

As discussed in Section 7.3, the GEV distribution has been identified as the robust distribution for the study area. The form of the regional frequency relationship for GEV distribution is expressed as:

$$\frac{Q_T}{\overline{Q}} = u + \alpha\, y_T \tag{35}$$

Here, $Q_T$ is T-year return period flood estimate, u and $\alpha$ are the parameters of the GEV distribution and $Y_T$ is GEV reduced variate corresponding to T-year return period i.e.

$$y_T = \left[1 - \left\{-\ln\left(1 - \frac{1}{T}\right)\right\}^k\right] \Big/ k \qquad (36)$$

The values of regional parameters of the GEV distribution foe Subzone 1(f) are mentioned below.

$k = 0.010$, $u = 0.734$ and $\alpha = 0.468$

Substituting values of these regional parameters in equations (35) and (36), the regional flood frequency relationship for estimation of floods of various return periods for the gauged catchments of Subzone 1(f) is expressed as:

$$Q_T = \left[47.534 \; - 46.8\left(-\ln\left(1 - \frac{1}{T}\right)\right)^{0.01}\right] * \bar{Q} \qquad (37)$$

For estimation of flood of desired return period for a small to moderate size gauged catchment of Subzone 1(f), the above regional flood frequency relationship may be used. Alternatively, floods of various return periods may also be obtained by multiplying the mean annual peak flood of the catchment ($\bar{Q}$) by the corresponding value of growth factors given in Table 10.

**Table 10: Values of growth factors ($Q_T/\bar{Q}$) for various distributions for Subzone 1(f)**

| **Distribution** | **Return Period (Years)** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **2** | **5** | **10** | **25** | **50** | **100** | **200** | **500** | **1000** |
| | **Growth Factors/Quantile Estimates** | | | | | | | | |
| **GEV** | **0.906** | **1.431** | **1.776** | **2.209** | **2.527** | **2.84** | **3.151** | **3.557** | **3.862** |
| GNO | 0.906 | 1.435 | 1.777 | 2.203 | 2.516 | 2.826 | 3.136 | 3.549 | 3.864 |
| PT-III | 0.904 | 1.446 | 1.788 | 2.2 | 2.493 | 2.775 | 3.048 | 3.4 | 3.659 |
| GLO | 0.915 | 1.393 | 1.728 | 2.197 | 2.589 | 3.023 | 3.505 | 4.231 | 4.857 |
| WAK | 0.929 | 1.411 | 1.731 | 2.18 | 2.549 | 2.947 | 3.375 | 3.993 | 4.503 |

The variation of growth factors obtained for GEV, GNO, PT-III, GLO and WAK distributions is shown in Fig. 3.

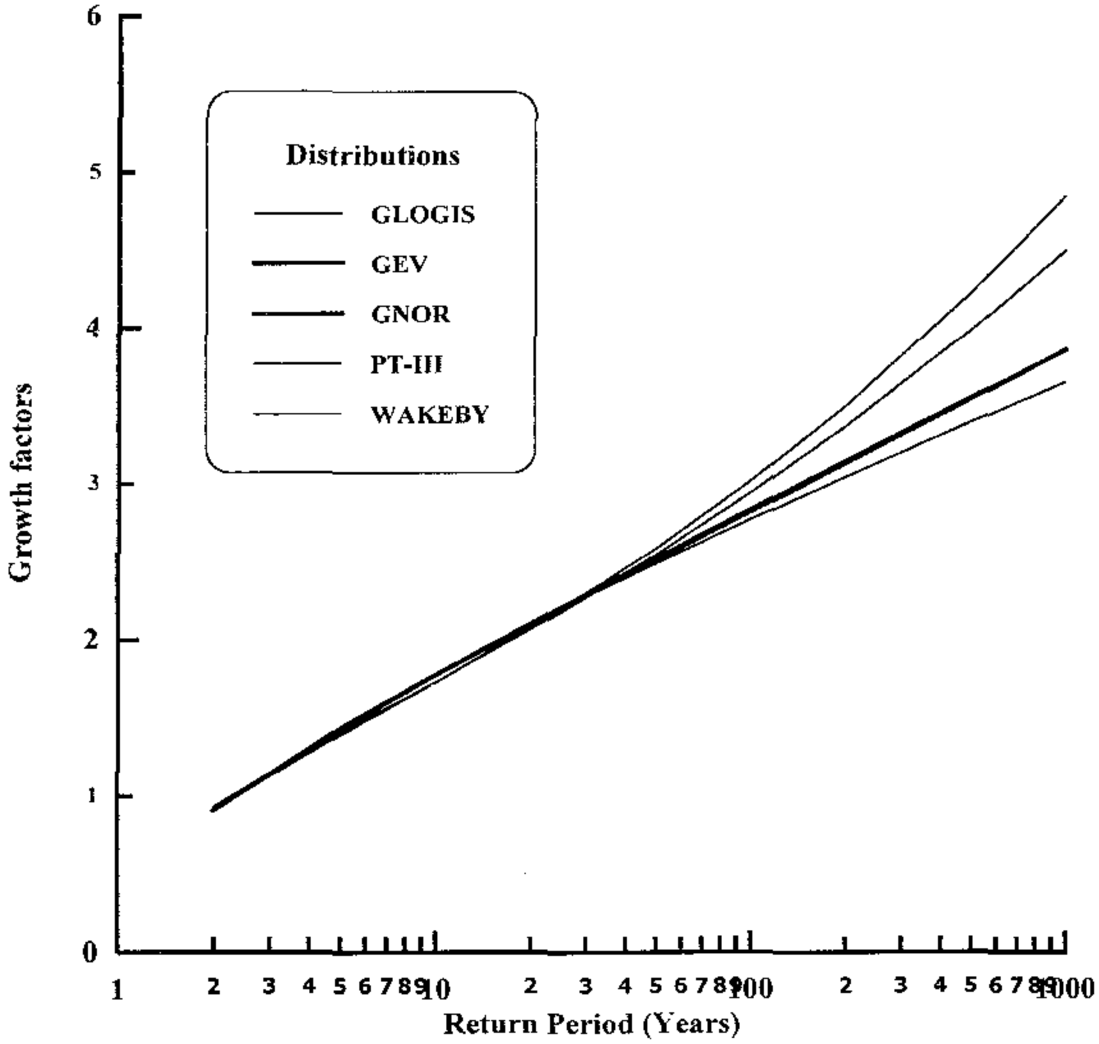


**Fig. 3: Variation of growth factors for various return periods for Subzone 1(f)**

## 7.5 Development of Regional Relationship between Mean Annual Peak Flood and Catchment Area

For estimation of T-year return period flood at a site, the estimate for mean annual peak flood is required. For ungauged catchments at-site mean can not be computed in absence of the observed flow data. In such a situation, a 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. As catchment areas of the various bridge sites were the only physiographic characteristics available; hence, in this study a regional relationship has been developed in terms of catchment area for estimation of mean annual peak flood for ungauged catchments.

Fig. 4 shows the variation of mean annual peak floods with catchment area for the 8 gauging sites of the study area. The regional relationship between $\bar{Q}$ ($m^3$/sec) and A ($km^2$) developed for the region in log domain using least squares approach is given below.

$$\bar{Q} = 0.733\ (A)^{1.084} \qquad (38)$$

for this relationship the correlation coefficient is, r = 0.879, coefficient of determination, $r^2 = 0.774$ and the standard error of the estimates is obtained as 0.545.

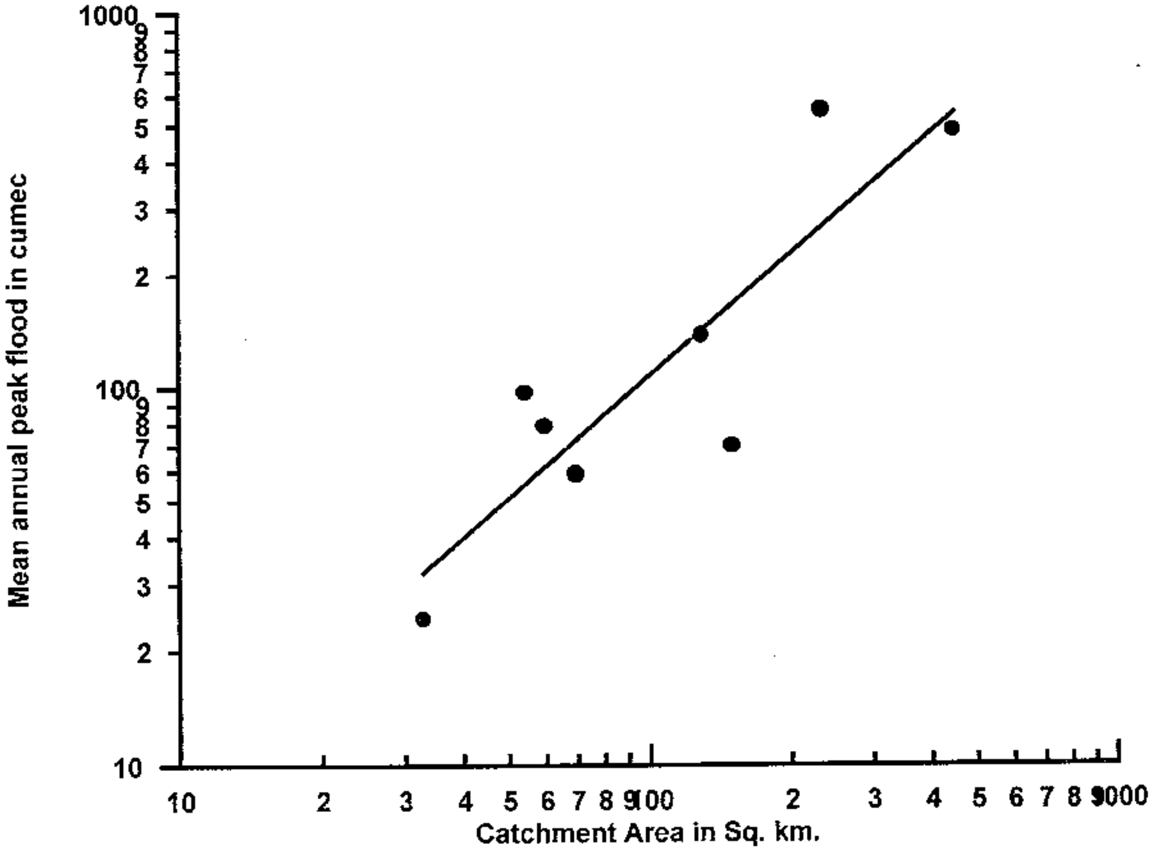


**Fig. 4 Variation of mean annual peak flood with catchment area for Subzone 1(f)**

## 7.6 Development of Regional Flood Formula for Ungauged Catchments

For development of regional flood formula for estimation of floods for various return periods for ungauged catchments, the regional flood frequency relationship given in equation (37) has been coupled with the regional relationship between mean annual peak flood and catchment area, given in equation (38). Derivation of the regional flood formula is given in Appendix-I.

The developed regional flood formula for ungauged catchments of Subzone 1 (f) is expressed as:

$$Q_T = \left[ 34.842 - 34.304 \left\{ -\ln\left(1 - \frac{1}{T}\right)\right\}^{0.01} \right] A^{1.084} \tag{39}$$

Here, $Q_T$ is flood estimate in $m^3/s$ for T year return period, and A is catchment area in $km^2$.

The values of floods of various return periods ($Q_T$) computed using the developed regional flood formula for different catchment areas are given in Table 11. Graphical representation of the developed regional flood formula is illustrated in Fig. 5.

**Table 11: Variation of floods of various return periods with catchment area for Subzone 1(f)**

| Catchment Area ($km^2$) | Return periods (Years) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 2 | 5 | 10 | 25 | 50 | 100 | 200 | 500 | 1000 |
| | Floods of various return periods ($m^3/s$) | | | | | | | | |
| 20 | 17 | 27 | 33 | 42 | 48 | 54 | 59 | 67 | 73 |
| 30 | 27 | 42 | 52 | 65 | 74 | 83 | 92 | 104 | 113 |
| 40 | 36 | 57 | 71 | 88 | 101 | 114 | 126 | 142 | 154 |
| 50 | 46 | 73 | 90 | 112 | 129 | 145 | 160 | 181 | 197 |
| 60 | 56 | 89 | 110 | 137 | 157 | 176 | 195 | 221 | 240 |
| 70 | 66 | 105 | 130 | 162 | 185 | 208 | 231 | 261 | 283 |
| 80 | 77 | 121 | 150 | 187 | 214 | 241 | 267 | 301 | 327 |
| 90 | 87 | 138 | 171 | 213 | 243 | 273 | 303 | 342 | 372 |
| 100 | 98 | 154 | 192 | 238 | 273 | 306 | 340 | 384 | 417 |
| 150 | 152 | 240 | 297 | 370 | 423 | 476 | 528 | 596 | 647 |
| 200 | 207 | 327 | 406 | 505 | 578 | 650 | 721 | 814 | 884 |
| 250 | 264 | 417 | 518 | 644 | 736 | 828 | 918 | 1036 | 1125 |
| 300 | 322 | 508 | 631 | 784 | 897 | 1008 | 1119 | 1263 | 1371 |
| 350 | 380 | 601 | 745 | 927 | 1060 | 1192 | 1322 | 1493 | 1621 |
| 400 | 439 | 694 | 861 | 1071 | 1226 | 1377 | 1528 | 1725 | 1873 |
| 450 | 499 | 789 | 979 | 1217 | 1392 | 1565 | 1736 | 1960 | 2128 |
| 500 | 560 | 884 | 1097 | 1365 | 1561 | 1754 | 1946 | 2197 | 2386 |
| 550 | 621 | 980 | 1216 | 1513 | 1731 | 1945 | 2158 | 2436 | 2645 |
| 600 | 682 | 1077 | 1337 | 1663 | 1902 | 2138 | 2372 | 2677 | 2907 |
| 650 | 744 | 1175 | 1458 | 1813 | 2074 | 2331 | 2587 | 2920 | 3170 |
| 700 | 806 | 1273 | 1580 | 1965 | 2248 | 2526 | 2803 | 3164 | 3436 |
| 750 | 869 | 1372 | 1703 | 2118 | 2423 | 2723 | 3021 | 3410 | 3702 |
| 800 | 931 | 1471 | 1826 | 2271 | 2598 | 2920 | 3240 | 3657 | 3971 |
| 850 | 995 | 1571 | 1950 | 2425 | 2775 | 3118 | 3460 | 3906 | 4240 |
| 900 | 1058 | 1672 | 2075 | 2580 | 2952 | 3318 | 3681 | 4155 | 4511 |
| 950 | 1122 | 1773 | 2200 | 2736 | 3130 | 3518 | 3903 | 4406 | 4784 |
| 1000 | 1186 | 1874 | 2326 | 2893 | 3309 | 3719 | 4126 | 4658 | 5057 |

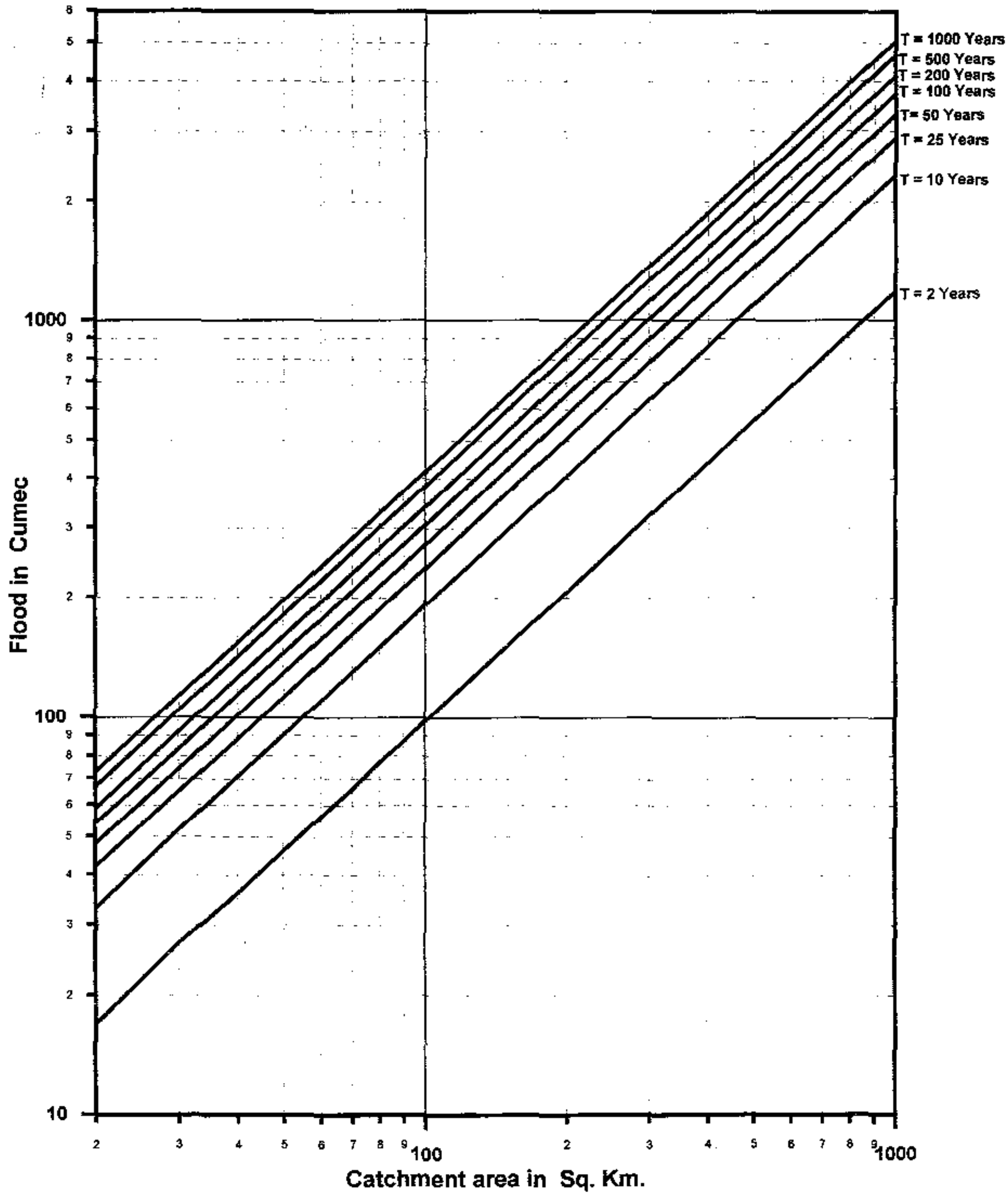


**Fig. 5: Variation of floods of various return periods with catchment area for Subzone 1(f)**

## 7.7 Comparison of Flood Estimates using Data of 8 and 11 Sites

In the regional flood frequency analysis discussed above in Sections 7.1 through 7.6, the annual maximum peak flood data of 8 bridge sites of Subzone 1(f); which constitute a homogeneous region as per the L-moment based regional homogeneity test discussed in Section 6.3 have been used. Regional flood frequency relationship and flood formula have been developed using the data of the 8 sites. As discussed earlier, in all data of 11 bridge sites of Subzone 1(f) are available. Catchment area, sample statistics and sample size for the bridge sites of Subzone 1(f) are given in Table 12.

**Table 12: Catchment area, sample statistics and sample size for the 11 bridge sites of Subzone 1(f)**

| S. No. | Bridge Number | Catchment Area (A) | Mean Annual Peak Flood ($\bar{Q}$) | Standard Deviation (SD) | Coefficient of Variation (CV) | Coefficient of Skewness (CS) | Sample Size (Years) (SS) |
|---|---|---|---|---|---|---|---|
| 1 | 85 | 136.70 | 72.00 | 92.33 | 1.282 | 1.551 | 11 |
| 2 | 59 | 54.39 | 97.48 | 52.85 | 0.542 | -0.233 | 33 |
| 3 | 30 | 447.76 | 490.50 | 277.93 | 0.567 | 0.322 | 30 |
| 4 | 160 | 150.40 | 70.31 | 37.68 | 0.536 | 0.861 | 32 |
| 5 | 3 | 32.89 | 24.29 | 16.99 | 0.699 | 0.915 | 31 |
| 6 | 177 | 712.00 | 239.93 | 208.45 | 0.869 | 1.123 | 14 |
| 7 | 60 | 130.00 | 138.70 | 72.36 | 0.522 | 3.686 | 27 |
| 8 | 24 | 69.75 | 59.31 | 33.99 | 0.573 | 0.441 | 26 |
| 9 | 48 | 27.45 | 83.31 | 71.65 | 0.860 | 1.780 | 26 |
| 10 | 141 | 59.83 | 79.39 | 47.04 | 0.593 | 0.682 | 23 |
| 11 | 104 | 234.11 | 555.21 | 422.62 | 0.761 | 2.542 | 29 |

However, data of 3 sites have been excluded for meeting the criteria of regional homogeneity and this leads to a significant loss of data. Hence, a comparative study has been carried out for examining the deviations in flood frequency estimates computed using the regional flood frequency relationship and flood formula for the gauged and ungauged catchments using the data of 8 sites and 11 sites.

### 7.7.1 Heterogeneity measures using data of 11 sites

The heterogeneity measures (described in Section 6.3) for Subzone 1(f) using the data of 11 sites have been computed and the same are given in Table 13. As the values of heterogeneity measures H(1), H(2) and H(3) are observed to be greater than 1.0, the region is found to be heterogeneous.

**Table 13: Heterogeneity measures for Subzone 1(f) (using data of 11 sites)**

| S. No. | Heterogeneity measures | Values |
|---|---|---|
| 1. | **Heterogeneity measure (H1)** | |
| | (a) Observed standard deviation of group L-CV | .0824 |
| | (b) Simulated mean of standard deviation of group L-CV | .0462 |
| | (c) Simulated standard deviation of standard deviation of group L-CV | .0102 |
| | (d) **Standardized test value H (1)** | **3.55** |
| 2. | **Heterogeneity measure H (2)** | |
| | (a) Observed average of L-CV / L-Skewness distance | .1206 |
| | (b) Simulated mean of average L-CV / L-Skewness distance | .0903 |
| | (c) Simulated standard deviation of average L-CV / L-Skewness distance | .0183 |
| | (d) **Standardized test value H (2)** | **1.66** |
| 3. | **Heterogeneity measure (H3)** | |
| | (a) Observed average of L-Skewness/L-Kurtosis distance | .1400 |
| | (b) Simulated mean of average L-Skewness/L-Kurtosis distance | .1070 |
| | (c) Simulated standard deviation of average L-Skewness/L-Kurtosis distance | .0206 |
| | (d) **Standardized test value H (3)** | **1.60** |

### 7.7.2 Development of regional flood frequency relationship for gauged catchments using data of 11 sites

The $Z^{dist}$–statistic for the various three parameter distributions is given in Table 14. From Table 14 it is observed that the $Z^{dist}$–statistic value is lower than 1.64 for the four distributions viz. GEV, GNO, PT-III and GLO. The $Z^{dist}$–statistic values are found to be comparable and lowest for GEV and GNO distributions.

As GEV has been identified as the robust distribution in the study carried out using the data of 8 sites; therefore, in this comparative study, flood frequency estimates utilizing the data of 8 and 11 bridge sites have been computed and compared using the GEV distribution as described below.

**Table 14: $Z_i^{dist}$ Statistic for various distributions (using data of 11 sites)**

| S. No. | Distribution | Z-Statistic |
|---|---|---|
| 1 | GEV | 0.19 |
| 2 | GNO | -0.11 |
| 3 | PT-III | -0.76 |
| 4 | GLO | 1.61 |
| 5 | GPA | -3.01 |

The values of regional parameters for the various distributions which have $Z^{dist}$–statistic value less than 1.64 as well as the five parameter Wakeby distribution are given in Table 15.

**Table 15: Regional parameters for the various distributions (using data of 11 sites)**

| Distribution | Parameters of the Distribution | | | | |
|---|---|---|---|---|---|
| GEV | $u = .698$ | $\alpha = .488$ | $k = -0.040$ | | |
| GNO | $\xi = .877$ | $\alpha = 0.582$ | $k = -0.404$ | | |
| PT-III | $\mu = 1.000$ | $\sigma = 0.651$ | $\gamma = 1.185$ | | |
| GLO | $\xi = 0.889$ | $\alpha = 0.330$ | $k = -0.196$ | | |
| WAK | $\xi = 0.071$ | $\alpha = 1.591$ | $\beta = 2.701$ | $\gamma = 0.459$ | $\delta = 0.080$ |

The values of regional parameters of the GEV distribution for Subzone 1(f) are mentioned below.

$k = -0.040$ , $u = 0.698$ and $\alpha = 0.488$

Substituting values of these regional parameters in equations (35) and (36), the regional flood frequency relationship for estimation of floods of various return periods for the gauged catchments of Subzone (1f) is expressed as:

$$Q_T = \left[ -11.502 + 12.2 \left( -\ln\left(1 - \frac{1}{T}\right)\right)^{-0.04} \right] * \overline{Q} \qquad (40)$$

The values of growth factors based on GEV distribution computed using the data of 11 sites are given in Table 16. Fig. 6 shows the variation of the growth factors with return period.

**Table 16: Values of growth factors ($Q_T/\overline{Q}$) for various return periods (using data of 11 sites)**

| Distribution | Return Period (Years) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 2 | 5 | 10 | 25 | 50 | 100 | 200 | 500 | 1000 |
| | Growth Factors/Quantile Estimates | | | | | | | | |
| GEV | 0.879 | 1.453 | 1.847 | 2.363 | 2.758 | 3.162 | 3.575 | 4.139 | 4.578 |
| GNO | 0.877 | 1.461 | 1.854 | 2.359 | 2.74 | 3.124 | 3.515 | 4.046 | 4.458 |
| PT-III | 0.875 | 1.478 | 1.872 | 2.356 | 2.705 | 3.043 | 3.373 | 3.801 | 4.119 |
| GLO | 0.889 | 1.414 | 1.794 | 2.342 | 2.813 | 3.345 | 3.951 | 4.887 | 5.715 |
| WAK | 0.897 | 1.441 | 1.82 | 2.345 | 2.769 | 3.216 | 3.688 | 4.354 | 4.89 |

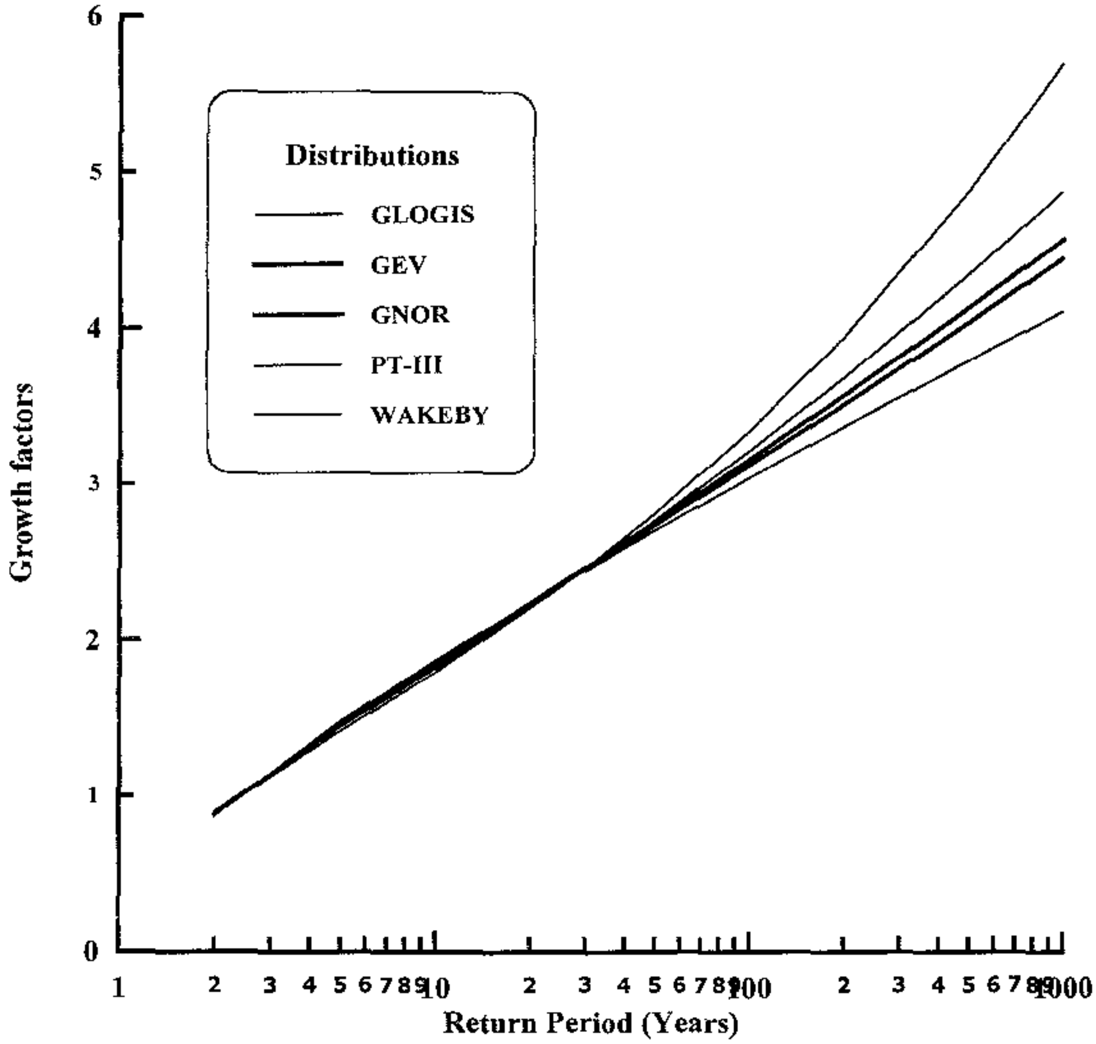


**Fig. 6: Variation of growth factors for various return periods for Subzone 1(f) using data of 11 sites**

### 7.7.3 Development of regional flood formula for ungauged catchments using the data of 11 sites

The regional relationship between $\overline{Q}$ ($m^3$/sec) and A ($km^2$) developed using the data of 11 bridge sites using least squares approach is given below.

$$\overline{Q} = 4.358\,(A)^{0.69} \tag{41}$$

for this relationship the correlation coefficient is, r = 0.761, coefficient of determination, $r^2 = 0.579$.

The regional flood formula developed using the data of 11 sites for estimation of floods of various return periods for the ungauged catchments of Subzone 1 (f) is expressed as:

$$Q_T = \left[ -50.126 + 53.168 \left\{ -\ln\left(1 - \frac{1}{T}\right) \right\}^{-0.04} \right] A^{0.69} \tag{42}$$

Here, $Q_T$ is flood estimate in $m^3$/s for T year return period, and A is catchment area in $km^2$.

The values of floods of various return periods ($Q_T$) computed using the above regional flood formula for different catchment areas are given in Table 17.

**Table 17: Variation of floods of various return periods with catchment area for Subzone 1(f) (using data of 11 sites)**

| Catchment Area ($km^2$) | Return periods (Years) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 2 | 5 | 10 | 25 | 50 | 100 | 200 | 500 | 1000 |
| | Floods of various return periods ($m^3/s$) | | | | | | | | |
| 20 | 30 | 50 | 64 | 81 | 95 | 109 | 123 | 142 | 158 |
| 30 | 40 | 66 | 84 | 108 | 126 | 144 | 163 | 188 | 208 |
| 40 | 49 | 81 | 103 | 131 | 153 | 176 | 198 | 230 | 254 |
| 50 | 57 | 94 | 120 | 153 | 179 | 205 | 231 | 268 | 296 |
| 60 | 65 | 107 | 136 | 174 | 203 | 232 | 263 | 304 | 336 |
| 70 | 72 | 119 | 151 | 193 | 225 | 258 | 292 | 338 | 374 |
| 80 | 79 | 130 | 165 | 212 | 247 | 283 | 320 | 371 | 410 |
| 90 | 85 | 141 | 179 | 230 | 268 | 307 | 347 | 402 | 445 |
| 100 | 92 | 152 | 193 | 247 | 288 | 330 | 373 | 432 | 478 |
| 150 | 121 | 201 | 255 | 326 | 381 | 437 | 494 | 572 | 632 |
| 200 | 148 | 245 | 311 | 398 | 465 | 533 | 602 | 69[illegible] | 771 |
| 250 | 173 | 286 | 363 | 464 | 542 | 621 | 703 | 813 | 900 |
| 300 | 196 | 324 | 412 | 527 | 615 | 705 | 797 | 922 | 1020 |
| 350 | 218 | 360 | 458 | 586 | 684 | 784 | 886 | 1026 | 1135 |
| 400 | 239 | 395 | 502 | 642 | 750 | 859 | 972 | 1125 | 1244 |
| 450 | 259 | 428 | 544 | 697 | 813 | 932 | 1054 | 1220 | 1349 |
| 500 | 279 | 461 | 585 | 749 | 874 | 1002 | 1133 | 1312 | 1451 |
| 550 | 298 | 492 | 625 | 800 | 934 | 1070 | 1210 | 1401 | 1550 |
| 600 | 316 | 522 | 664 | 849 | 991 | 1137 | 1285 | 1488 | 1646 |
| 650 | 334 | 552 | 702 | 898 | 1048 | 1201 | 1358 | 1572 | 1739 |
| 700 | 351 | 581 | 738 | 945 | 1103 | 1264 | 1429 | 1655 | 1830 |
| 750 | 369 | 609 | 774 | 991 | 1156 | 1326 | 1499 | 1735 | 1919 |
| 800 | 385 | 637 | 810 | 1036 | 1209 | 1386 | 1567 | 1814 | 2007 |
| 850 | 402 | 664 | 844 | 1080 | 1261 | 1445 | 1634 | 1892 | 2093 |
| 900 | 418 | 691 | 878 | 1124 | 1311 | 1503 | 1700 | 1968 | 2177 |
| 950 | 434 | 717 | 912 | 1166 | 1361 | 1561 | 1764 | 2043 | 2259 |
| 1000 | 449 | 743 | 944 | 1208 | 1410 | 1617 | 1828 | 2116 | 2341 |

### 7.7.4 Comparison of growth factors based on regional frequency relationships developed for gauged catchments

As discussed in Section 7.4 (Table 10), the growth factors for the commonly used distributions have been computed and based on the L-moment diagram as well as the $Z^{dist}$ statistic the GEV distribution has been identified as the robust distribution when data of 8 bridge sites are considered. Following the same procedure, the growth factors have been computed using the data of 11 bridge sites as discussed in Section 7.7.2 (Table 16). Fig. 7 shows the comparison of the growth factors based on GEV distribution for various return periods computed using the data of 8 and 11 bridge sites. The percentage deviations in growth factors computed using the data of 8 and 11 sites are given in Table 18. It is observed from Fig. 7 and Table 18 that the percentage deviations in growth factors in general increase from 1.5 to 18.5 for return periods of 5 to 1000. The percentage deviations for return periods of 25, 50 and 100 years are 7.0, 9.1 and 11.3.

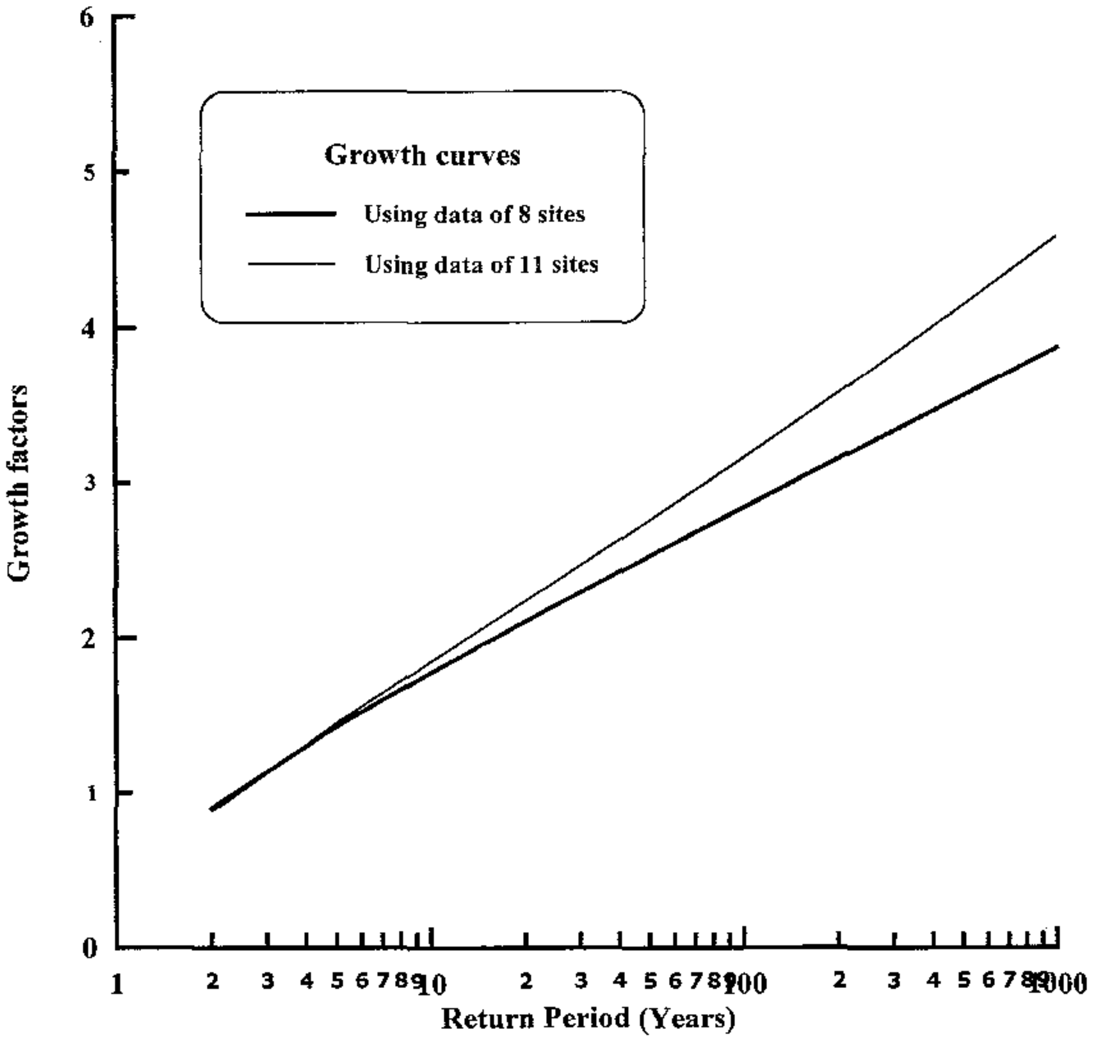


**Fig. 7: Variation of growth factors for various return periods computed using data of 8 and 11 sites**

**Table 18: Percentage deviations in growth factors based on GEV distribution computed using data of 8 and 11 sites**

| Return period (Years) | Growth factor (using 8 sites) | Growth factor (using 11 sites) | Percentage deviation |
|---|---|---|---|
| 2 | 0.906 | 0.879 | -3.0 |
| 5 | 1.431 | 1.453 | 1.5 |
| 10 | 1.776 | 1.847 | 4.0 |
| 25 | 2.209 | 2.363 | 7.0 |
| 50 | 2.527 | 2.758 | 9.1 |
| 100 | 2.840 | 3.162 | 11.3 |
| 200 | 3.151 | 3.575 | 13.5 |
| 500 | 3.557 | 4.139 | 16.4 |
| 1000 | 3.862 | 4.578 | 18.5 |

### 7.7.5 Comparison of floods of various return periods based on regional flood formulae developed for ungauged catchments

As discussed in Section 7.6, the regional flood formula has been developed using the data of 8 bridge sites (Eq. 39). The flood estimates for various return periods computed using Eq. 39 are given in Table 11. Following the same procedure, the regional flood formula has been developed using the data of 11 bridge sites (Eq. 42), as discussed in Section 7.7.3. The flood estimates for various return periods computed using Eq. 42 are given in Table 17. Fig. 8 shows the comparison of the flood estimates for return periods 25, 50 and 100 years computed using the data of 8 and 11 bridge sites. The percentage deviations in flood estimates for return periods 25, 50 and 100 years computed using the data of 8 and 11 sites are given in Table 19. It is observed from Fig. 8 and Table 19 that the percentage deviations in flood estimates in general increase when the catchment area decreases from about 80 $km^2$ to 20 $km^2$. It is also seen that the percentage deviations increase as the catchment area increases from 80 $km^2$ to 1000 $km^2$. For example, the percentage deviations for catchment areas of 20, 80 and 1000 $km^2$ for a return period of 50 years are 68.8, -0.9 and –63.5 respectively.

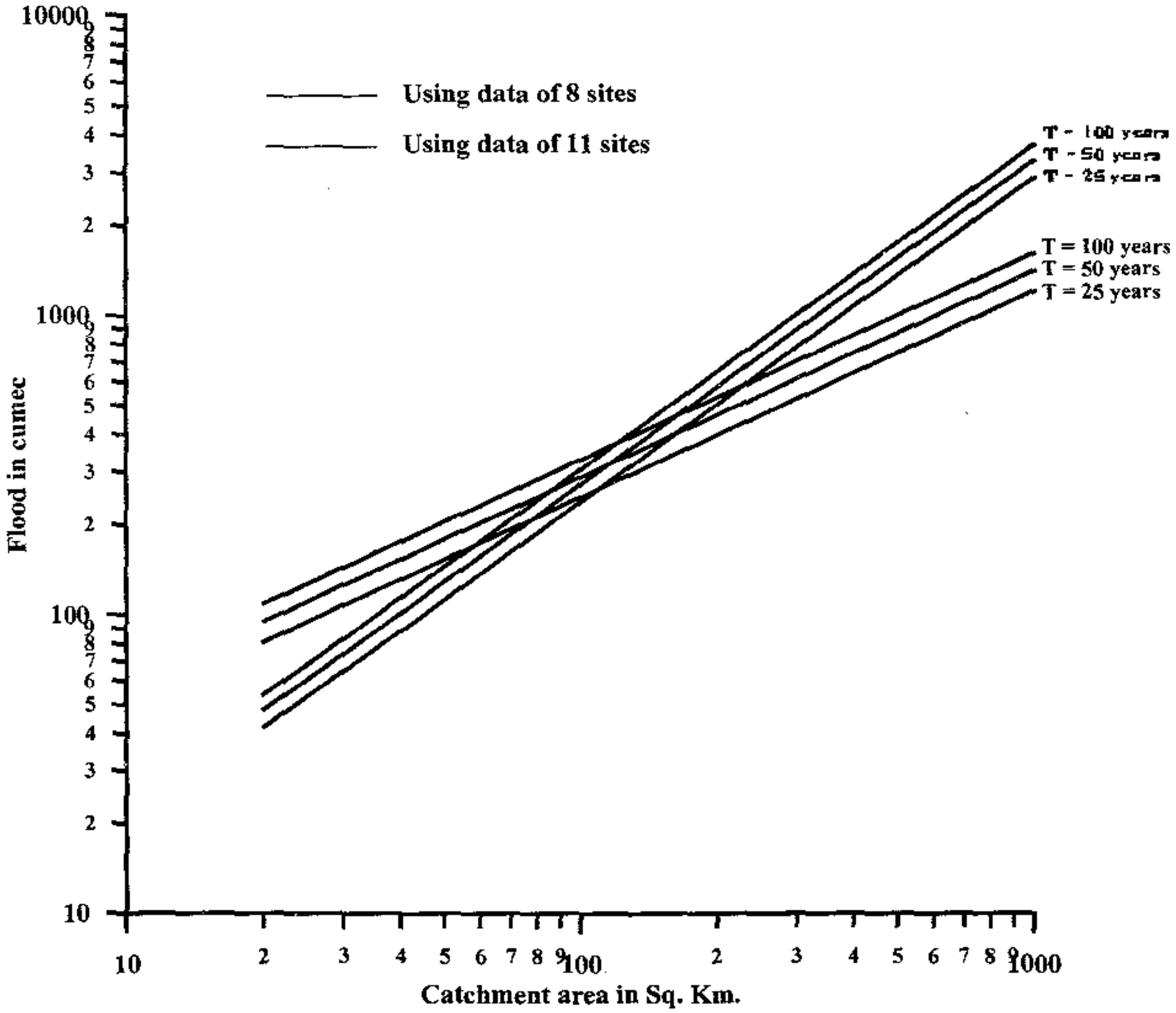


**Fig. 8: Comparison of floods of various return periods computed using data of 8 and 11 sites**

**Table 19: Percentage deviation in flood estimates for ungauged catchments using data of 8 and 11 sites**

| Area (km$^2$) | Return period = 25 years | | | Return period = 50 years | | | Return period = 100 years | | |
|---|---|---|---|---|---|---|---|---|---|
| | 8 sites | 11 sites | Perc. devn. | 8 sites | 11 sites | Perc. devn. | 8 sites | 11 sites | Perc. devn. |
| 20 | 42 | 64 | 52.4 | 48 | 81 | 68.8 | 54 | 95 | 75.9 |
| 30 | 65 | 84 | 29.2 | 74 | 108 | 45.9 | 83 | 126 | 51.8 |
| 40 | 88 | 103 | 17.0 | 101 | 131 | 29.7 | 114 | 153 | 34.2 |
| 50 | 112 | 120 | 7.1 | 129 | 153 | 18.6 | 145 | 179 | 23.4 |
| 60 | 137 | 136 | -0.7 | 157 | 174 | 10.8 | 176 | 203 | 15.3 |
| 70 | 162 | 151 | -6.8 | 185 | 193 | 4.3 | 208 | 225 | 8.2 |
| 80 | 187 | 165 | -11.8 | 214 | 212 | -0.9 | 241 | 247 | 2.5 |
| 90 | 213 | 179 | -16.0 | 243 | 230 | -5.3 | 273 | 268 | -1.8 |
| 100 | 238 | 193 | -18.9 | 273 | 247 | -9.5 | 306 | 288 | -5.9 |
| 150 | 370 | 255 | -31.1 | 423 | 326 | -22.9 | 476 | 381 | -20.0 |
| 200 | 505 | 311 | -38.4 | 578 | 398 | -31.1 | 650 | 465 | -28.5 |
| 250 | 644 | 363 | -43.6 | 736 | 464 | -37.0 | 828 | 542 | -34.5 |
| 300 | 784 | 412 | -47.4 | 897 | 527 | -41.2 | 1008 | 615 | -39.0 |
| 350 | 927 | 458 | -50.6 | 1060 | 586 | -44.7 | 1192 | 684 | -42.6 |
| 400 | 1071 | 502 | -53.1 | 1226 | 642 | -47.6 | 1377 | 750 | -45.5 |
| 450 | 1217 | 544 | -55.3 | 1392 | 697 | -49.9 | 1565 | 813 | -48.1 |
| 500 | 1365 | 585 | -57.1 | 1561 | 749 | -52.0 | 1754 | 874 | -50.2 |
| 550 | 1513 | 625 | -58.7 | 1731 | 800 | -53.8 | 1945 | 934 | -52.0 |
| 600 | 1663 | 664 | -60.1 | 1902 | 849 | -55.4 | 2138 | 991 | -53.6 |
| 650 | 1813 | 702 | -61.3 | 2074 | 898 | -56.7 | 2331 | 1048 | -55.0 |
| 700 | 1965 | 738 | -62.4 | 2248 | 945 | -58.0 | 2526 | 1103 | -56.3 |
| 750 | 2118 | 774 | -63.5 | 2423 | 991 | -59.1 | 2723 | 1156 | -57.5 |
| 800 | 2271 | 810 | -64.3 | 2598 | 1036 | -60.1 | 2920 | 1209 | -58.6 |
| 850 | 2425 | 844 | -65.2 | 2775 | 1080 | -61.1 | 3118 | 1261 | -59.6 |
| 900 | 2580 | 878 | -66.0 | 2952 | 1124 | -61.9 | 3318 | 1311 | -60.5 |
| 950 | 2736 | 912 | -66.7 | 3130 | 1166 | -62.7 | 3518 | 1361 | -61.3 |
| 1000 | 2893 | 944 | -67.4 | 3309 | 1208 | -63.5 | 3719 | 1410 | -62.1 |

# Chapter 8

## CONCLUSIONS

On the basis of this study following conclusions are drawn.

i. Regional flood frequency analysis has been carried out based on L-moments approach, considering the annual maximum peak flood data of 11 catchments of the Middle Ganga Plains (Subzone 1-f). Discordancy measure ($D_i$) test was carried out and it was found that the data of all the sites are suitable for carrying out the flood frequency analysis. Homogeneity of the region has been tested using the L-moment based heterogeneity measure, H. Based on this test, it has been observed that the data of 8 out of 11 sites constitute a homogeneous region. Hence, the data of these 8 sites have been used in this study.

ii. Various distributions viz. Extreme value (EV1), General extreme value (GEV), Logistic (LOS), Generalized logistic (GLO), Generalized normal (GNO), Exponential (EXP), Generalized Pareto (GP), Kappa (KAP) and five parameter Wakeby (WAK) have been used in the study. The regional parameters of these distributions have been estimated using the L-moments approach. Based on the L-moment ratio diagram as well as $Z^{dist}$ –statistic criteria the GEV distribution has been identified as the robust distribution for the Middle Ganga Plains (Subzone 1-f).

iii. Regional flood frequency relationship has been developed based on the GEV distribution for gauged catchments of the Subzone 1(f). For estimation of floods of various return periods for the gauged catchments of the study area, either the developed regional flood frequency relationship may be used or the mean annual peak flood of the catchment may be multiplied by the corresponding values of the growth factors.

iv. The L-moment based regional flood frequency relationship derived for the GEV distribution has been coupled with the regional relationship between mean annual peak flood and the catchment area and the regional flood formula has been developed for estimation of floods of desired return periods for ungauged catchments of Subzone 1(f). The developed regional flood formula, or its graphical representation may be used for estimation of floods of desired return periods for the ungauged catchments of the study area. Floods of various return periods for different catchment areas may also be obtained from the tabular form of the developed regional flood formula.

v. The conventional empirical flood formulae do not provide floods of various return periods. However, the regional flood formula developed in this study is capable of providing flood estimates for desired return periods.

vi. As the regional flood formula has been developed using the data of catchments ranging from 32.9 $km^2$ to 447.8 $km^2$ in area; therefore, the developed regional flood frequency relationship or formula may be expected to provide estimates of floods of various return periods for the catchments of Subzone 1(f), lying nearly in the same range of areal extent, as those of the input data.

vii. The data of only 8 gauging sites, varying from 23 to 33 years have been used in this study. The relationship between mean annual peak flood and catchment area developed on the basis of available data is able to explain 77.4% of initial variance ($r^2 = 0.774$) and the standard error of the estimates is obtained as 0.545. Hence, the results of the study are subject to these limitations. However, the developed regional flood frequency relationship and the regional formula may be refined for obtaining more accurate flood frequency estimates, when the annual maximum peak flood data for some more gauging sites become available and catchment and physiographic characteristics other than catchment area are also used for development of the regional flood formula.

viii. In case of the gauged catchments, deviations in growth factors computed using the data of 8 and 11 sites show that the percentage deviations in general increase from 1.5% to 18.5% for the return periods varying from 5 to 1000 years. It illustrates that there is over estimation in floods of various return periods when data of all the 11 sites are used without meeting the L-moment based criteria of regional homogeneity and floods of 25, 50 and 100 return periods are over estimated by 7%, 9.1% and 11.3%, respectively. Thus, excluding the three catchments for meeting the criteria of regional homogeneity leads to under estimation of floods of various return periods.

ix. In case of the ungauged catchments, deviations in flood estimates for return periods of 25, 50 and 100 years computed using the data of 8 and 11 sites show that the percentage deviations in flood estimates in general increase when the catchment area decreases from about 80 $km^2$ to 20 $km^2$. It is also seen that the percentage deviations increase as the catchment area increases from 80 $km^2$ to 1000 $km^2$. For example, the percentage deviations for catchment areas of 20, 80 and 1000 $km^2$ for a return period of 50 years are 68.8, -0.9 and –63.5 respectively. Thus, there is an under estimation for floods of 25, 50 and 100 return periods for lower range of catchment area i.e. 20 to 80 $km^2$, and there is over estimation for larger size catchments varying in areal extent from 80 to 1000 $km^2$.

## 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.

Barnett, V. and Lewis, (1994). T. 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.

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

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

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

Chowdhury, J.U., Stedinger, J.R. and Lu, L.H., (1991). Goodness of fit tests for regional generalized value extreme value flood distributions. Water Resour. Res., 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 Resour. Res., 29(4), 1103-1111.

Cunnane C., (1987). Review of statistical models for flood frequency estimation, Regional flood frequency analysis, ed. V. P. Singh, D. Reidel Publishing Company.

Cunnane C., (1988). Methods of merits of regional flood frequency analysis. J. of Hydrol 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.

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

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

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

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

Hosking J.R.M. and Wallis J.R., (1993). Some statistics useful in regional frequency analysis. Water Resour. Res., 29(2), 271-281.

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

Hosking, J.R.M. (1986). The theorey of probablity weighted moments. Res. Rep. RC12210, IBM Res., Yorktown Heights, N.Y.

Hosking, J.R.M. and Wallis, J.R., (1986). The value of historical data in flood frequency analysis. Water Resour. Res., 22(11), 1606-1612.

Hosking, J.R.M. and Wallis, J.R., (1988). The effect of intersite dependence on regional flood frequency analysis. Water Resour. Res., 24(4), 588-600.

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

Hosking, J.R.M., (1991). Approximations for use in constructing L-moment ratio diagrams. Res. Rep., RC-16635, 3, IBM Res. Div., T.J., Watson Res. Cent., Yorktown Heights, N.Y.

Hosking, J.R.M., J.R. Wallis and E.F. Wood, (1985a). Estimation of the genaralized extreme value distribution by the method of probability weighted moments. Technometrics, Vol.27, No.3, 251-262.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F., (1985b). An apraisal of the regional flood frequency procedure in the UK flood studies report. Hydrol. Sci. J., 30(1), 85-109.

Huq, S.M., R. Nair and S.S. Sil, (1986). Frequency flood formulae for countrywide application', Proceedings, 53 Annual R & D Session, CBIP, Bhubaneshwar, 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 Res. Res., Vol 36(2), 469 - 482.

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

Jakhade, G.S., A.S. Murti and S. Sethuraman, (1984). Frequency floods in Brahmputra valley Regional Flood Frequency Approach. CBIP, Journal, New Delhi, Vol. 41,No. 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 Resour. Res., 25(5), 925-936.

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.

Kendall. M.G., (1975). Rank correlation methods. Charles Griffin, London.

Kirby, W., (1974). Algebraic boundness of sample statistics, Water Resources Research, 10 (2), 220-222.

Kuczera G., (1982a). Combining site specific and regional information: An empirical Bayes approach. Water Resour. Res., 18(2), 306-314.

Kuczera G., (1982b). Robust flood frequency models. Water Resour. Res., 15(5), 315-324.

Kuczera, G., (1983). Effect of sampling uncertainty and spatial correlation on an empirical Bayes procedure for combining site and regional information. Journal of Hydrology, Vol. 65, 373-398.

Kumar, R., Singh, R.D. and Seth, S.M, (1999). Regional flood formulas for seven subzones of zone 3 of India, J. of Hydrologic Engg., ASCE, July, Vol. 4, No. 3.

Landwehr, J.M., N.C. Matalas and J.R. Wallis, (1979a). Probablity weighted moments compared with some traditional techniques of estimating Gumbel parameters and quantiles. Water Resources Research, Vol. 15, No. 6, 1361.

Landwehr, J.M., N.C. Matalas and J.R. Wallis, (1979b). Estimation of parameters and quantiles of Wakeby distributions: 1. Known Lower Bounds. Water Res. Res., Vol. 15, No. 6.

Landwehr, J.M., N.C. Matalas and J.R. Wallis, (1979c). Estimation of parameters and quantiles of Wakeby distributions: 2. Unknown Lower Bounds. Water Resources Research, Vol. 15, No. 6, 1373.

Larson, C.L. and B.M. Reich, (1972). Relationship of observed rainfall and runoff recurrence interval, Proc. 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 Resour. Res., 23(2), 313-323.

Lettenmaier, D.P. and K.W. Potter, (1985). Testing flood frequency estimation methods Using a Regional Flood Generation Model. Water Resources Research, Vol.21, 1903-1914.

Lu L.H., and Stedinger, J.R., (1992). Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test, J. of Hydrol., 138, 223-245.

Maidment, D.R., (1993). Handbook of Hydrology, Mc Graw-Hill, Inc., Newyork.

Martins E.S. and Stedinger J.R., (2000). Generalised maximum-likelihood generalised extreme-value quantile estimators for hydrologic data. Water Res. Res., Vol. 36, No. 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.

Mehta, H.L. and D. K. Sharma., (1986). Flood frequency analysis, Proceedings of Workshop on Flood Estimation in Himalayan Region, Roorkee (U.P.), 13-23.

N. I.H., (1984-85). Methodology for estimation of design storm, TN-12, Roorkee.

N.E.R.C., (1975). Flood studies report, 1-Hydrological Studies, London.

N.I.H., (1990-91). Regional flood frequency analysis for Godavari basin Subzone-3(f). Technical Report No. TR-59, Roorkee.

N.I.H., (1994-95). Development of regional flood formula for Mahanad subzone 3(d), Technical Report TR(BR)-134, Roorkee.

N.I.H., (1994-95). Regional flood frequency analysis for upper Narmada and Tapi subzone-3(c), Technical Report TR(BR)-133, Roorkee.

N.I.H., (1995-96). Development of regional flood frequency relationships and flood formulae for various subzones of Zone 3 of India, Technical Report, TR(BR)-149.

N.I.H., (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. 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.

Parida, B.P. and Moharram, S.H., (1999). Choice of generalized Pareto distribution as a potential candidate for flood frequency analysis. Hydrology Journal, vol 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, Oct. 8-9.

Perumal, M. and S. M. Seth., (1985). Regional flood frequency analysis using power transformation - A case Study. Proceedings, Seminar on Flood Frequency Analysis, New Delhi, 93-108.

Pilgrim, D.H. and I. Cordery, (1992). Flood Runoff, Chapter in Handbook of Hydrology, Mc Graw-Hill, Inc., Newyork.

Pilgrim, D.H., I.A. Rowbottom and D.G. Doran, (1987). Development of design procedures for extreme floods in Australia, in V.P. Singh, ed., Application of Frequency and Risk in Water Resources, Reidel, 63-77.

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 Faulkner E.B., (1987). Catchment response times as a predictor of flood quantiles., Water Resour. Res., 23, 857-861.

Potter K.W. and Lettenmaier D.P., (1990). A comparision of regional flood frequency estimation methods using resampling method., Water Resour. Res., 26, 415-424.

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.

Sankarasubramanian, 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 analyses of precipitation annual maxima in Washington State, Water Resources Research, 26, 119-131.

Seth, S. M., (1984-85). Regional Flood Frequency Analysis. National Institute of Hydrology, Roorkee, Report No. RN-14.

Seth, S. M. and B. N. Goswami, (1979). Regional flood frequency analysis with limited data, Proceedings, Third World Congress on Water Resources, Mexico City.

Seth, S. M. and R. D. Singh, (1987). Flood estimation in India using regional frequency analysis, Proceedings, National Symposium on Hydrology, Roorkee, Vol. 1, I-1-15.

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 R. Kumar, (1991), Estimation of discharge hydrograph for an ungauged catchment using unit hydrograph approach. Proc. of 4th National Symp. on Hydrology of Minor Water Resources Schemes, Madras.

Stedinger J.R. and Tasker G.D., (1985). Regional hydrologic regression; I. Ordinary weighted and generalised least squares compared, Water Resour. Res., 21, 1421-1432.

Stedinger J.R. and Tasker G.D., (1986a). Correction to regional hydrologic analysis 1. Ordinary weighted and generalised least squares compared., Water Resour. Res., 22, 844.

Stedinger J.R. and Tasker G.D., (1986b). Regional hydrologic analysis 2. Model error estimation, estimation of sigma and log-pearson type –3 distribution., Water Resour. Res., 22, 1487-1499.

Stedinger J.R., (1983). Estimating a regional flood frequency distribution., Water Resour. Res., 19(2), 503-510.

Stedinger, J.R. and Lu, L.H., (1995). Appraisal of regional and index flood qutantile estimators, Stochastic Hydrol. Hydrau., 9 (1), 49-75.

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, 18.5-18.9.

Tasker G.D. and Stedinger J.R., (1989). An operational GLS model for hydrologic regression., J. of Hydrol., 111, 361-375.

Tasker G.D. and Driver N.E., (1988). Nationwide regression models for predicting urban runoff water quality at unmonitored sites., Water resour. Bull., 24(5), 1091-1101.

Thirumalai, S. and P. B. Sinha., (1986). Flood estimation for small and medium catchments in Krishna basin by Chow's Approach Modified by Nash', Proceedings, Seminar on Flood Estimation and Control , Roorkee, III- 43-51.

Thiruvengadachari, S., S. V. Nayak and G. Ranganna., (1975). Flood in Cauvery basin: Magnitude and Frequency, Proceedings, 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.

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, Oct. 8-9.

Varshney, R.S., (1979). Engineering Hydrology, Nem Chand & Brothers, Roorkee.

Venkatraman, R. and P. N. Gupta, (1986). Flood estimation for small and medium catchments in sub-Himalayan region', Proceedings, Workshop on Flood Estimation in Himalayan Region, Roorkee, 123-140.

Venkatraman, R., S. Thirumalai and P. N. Gupta, (1986). Regional flood frequency analysis for small and medium catchments in lower Godavari basin - A critical review. Proceedings, Seminar on Flood Estimation and Control, Roorkee, II-58-77.

Vogel R.M. and Fennessey N.M., (1993). L-Moments should replace Product Moments Diagrams., Water Resour. Res., 29(6), 1745-1752.

Vogel R.M. and Kroll C.N., (1989). Low-flow frequency analysis using probability –plot correlation coefficients. J. water Resour. Plan. Manage. ASCE, 115(3), 338-357.

Wallis J.R. and Wood E.F., (1985). Relative accuracy of Log-Pearson III Procedures., J. of Hydraul. Engrg., ASCE, 111(7), 1043-56.

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

Wallis J.R., (1980). Risk and uncertainties in the evaluation of flood events for the design of hydraulic structures., Piene e Siccita, ed. E. Guggino, G. Rossi and E. Todini, Catania, Italy, 3-36.

Wallis J.R., (1988). Catastrophes, computing and containments: Living in our restless habitat., Speculation in Sci. and Tech. 11(4), 295-315.

Wallis J.R., (1993). Regional frequency studies using L-Moments., Concise Encyclopedia of Environmental Systems, ed. P.C. Young, Perbamon Press, N.Y.

Wang, Q. J., (1996). Direct sample estimators of L moments. Water Res. Res., Vol. 32, No. 12, 3617 – 3619.

Wang, Q.J., (1996). Using partial probability weighted moments to fit the extreme value distributions to censored samples, Vol. 32, No. 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 Res. Res., Vol. 35, No. 1, 203 - 209.

Wiltshire, S.E., (1985). Grouping basins for Regional flood frequency analysis. Hydrological Sciences Journal, 30, 151-159.

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.

Wiltshire, S.E., (1986c). Identification of homogeneous regions for flood frequency analysis. Journal of Hydrology, 84, 287-302.

Zrinji Z. and Burn D.H., (1994). Flood frequency analysis for ungauged sites using a region of influence approach., J. of Hydrol., 153, 1-21.

**Derivation of the Regional Flood Formula**

The form of regional flood frequency relationship for the GEV distribution is:

$$\frac{Q_T}{\overline{Q}} = u + \alpha\ y_T \qquad (1)$$

where,

$$y_T = \left[1 - (-\ln(1-1/T))^k\right]/k \qquad (2)$$

The conventional Dicken's formula is:

$$Q = c\,A^{0.75} \qquad (3)$$

The form of this formula may be generalized as:

$$Q_T = C_T\ A^b \qquad (4)$$

The form of regional relationship between mean annual peak flood and catchment area is:

$$\overline{Q} = a\ A^b \qquad (5)$$

Dividing Eq. (4) by Eq. (5) the following expression is obtained.

$$\frac{Q_T}{\overline{Q}} = \frac{C_T}{a} \qquad (6)$$

It may be expressed as:

$$C_T = \frac{Q_T}{\overline{Q}}\ a \qquad (7)$$

or, Substituting the value of $\frac{Q_T}{\overline{Q}}$ from Eq. (1)

$$C_T = (u + \alpha\, y_T)a \qquad (8)$$

Substituting the value of $C_T$ in Eq. (4)

$$Q_T = (u + \alpha\, y_T) a\, A^b \qquad (9)$$

Substituting the value of $y_T$ from Eq. (2)

$$Q_T = [ua + a\alpha\, y_T]\, A^b \qquad (10)$$

or,

$$Q_T = \left[ ua + a\alpha[1 - \{-\ln(1 - \frac{1}{T})\}^k]/k \right] A^b \qquad (11)$$

$$Q_T = \left[ ua + \frac{a\alpha}{k} - \frac{a\alpha}{k}\left\{-\ln(1 - \frac{1}{T})\right\}^k \right] A^b \qquad (12)$$

or,

$$Q_T = \left[ a(\frac{\alpha}{k} + u) - \frac{a\alpha}{k}\left\{-\ln(1 - \frac{1}{T})\right\}^k \right] A^b \qquad (13)$$

$$Q_T = \left[ \beta + \gamma\left\{-\ln(1 - \frac{1}{T})\right\}^k \right] A^b \qquad (14)$$

where,

$\beta = a(\alpha/k + u)$ and $\gamma = -\frac{\alpha}{k} a$

# Development of Regional Flood Formulae using L-Moments for Middle Ganga Plains (Subzone 1-f)

| | |
|---|---|
| **Director** | **K. S. Ramasastri** |
| **Coordinator CFMS, Patna** | **R. D. Singh** |
| **Head CFMS, Patna** | **Rakesh Kumar** |

## STUDY GROUP

**Rakesh Kumar**
**Chandranath Chatterjee**
**Sanjay Kumar**

## SUPPORTING STAFF

**A. K. Sivadas**