RN -14

REGIONAL FLOOD FREQUENCY ANALYSIS

SATISH CHANDRA
DIRECTOR

STUDY GROUP

S M SETH

NATIONAL INSTITUTE OF HYDROLOGY
JAL VIGYAN BHAWAN
ROORKEE-247667(UP) INDIA

1984-85

CONTENTS

		Page
	LIST OF FIGURES	i
	LIST OF TABLES	i
	ABSTRACT	ii
1.0	INTRODUCTION	1
1.1	Estimation of Peak Flow Rate	3
1.2	General Probability Concepts	5
1.3	Probability Distributions and Parameter Estimation	7
1.4	Selection of Suitable Distributions	10
1.5	Outlier Problem	13
1.6	Safety, Economy, Risk, Reliability	14
1.7	Regional Flood Studies	18
2.0	REVIEW OF LITERATURE	20
2.1	Method Used by U.S. Geological Survey	20
2.2	U.K. Flood Studies Report	29
2.3	Regional Flood Estimation Procedure in Australia	35
2.4	U.S. Water Resources Council Guidelines	37
2.5	Other Typical Studies Abroad During 1960-80.	39
2.6	Typical Studies in India During 1970-80.	50
2.7	Typical Studies Abroad During 1981-84.	53
2.8	Typical Studies in India During 1981-84.	62
2.9	Typical Studies Abroad During 1985-87.	64
2.9.1	General matters	64
2.9.2	Use of historical information	71
2.9.3	Regional homogeneity	73

2.9.4 Typical regional studies 81

2.10 Typical Studies in India During 1985-1987. 86

2.10.1 General matters 86

2.10.2 Conventional techniques 88

2.10.3 New approaches 93

3.0 REMARKS 98

REFERENCES 101

LIST OF FIGURES

FIGURE	TITLE	PAGE
1	HOMOGENEITY TEST CHART	25

LIST OF TABLES

TABLE	TITLE	PAGE
1	VALUES OF UPPER AND LOWER LIMITS IN TERMS OF GUMBEL'S REDUCED VARIATE Y.	26
2	HOMOGENEITY TEST	26
3	DETERMINATION OF EQUIVALENT MAIN STREAM SLOPE.	32
4	COMPUTED FLOWS FOR SELECTED RETURN PERIOD (STATION PMFS,WATER SURVEY OF CANADA).	44
5	PARTICULARS OF CATCHMENTS AND MEAN ANNUAL FLOODS.	52
6	DETAILS OF CATCHMENTS STUDIED.	63
7	STATISTICAL PARAMETERS FOR FOUR SITES ON RIVER SATLUJ AND BEAS.	92

ABSTRACT

The main objective of hydrological studies of floods is to develop appropriate procedures in order to arrive at desired design variable for the particular structure to be safe under extrmes of floods without leading to overdesign and consequent increase in cost. The flood frequency analysis procedures are based on general statistical and probability concepts and the data used in such analysis has to satisfy the criteria of homogeneity, independence, randomness and time invariance. When only short records are available, regional analysis approach can provide estimates of floods for limited data and ungauged locations, through appropriate regional multiple regression relationships of statistical parameters of the flood series and catchment characteristics. The choice of catchment morphometric characteristics depends on the (i) judgement of the likely predictive success of variables, and (ii) numerical calculation of regression equations and the interpretation of their coefficients. However, these characteristics are often dodged by correlations, resulting in interdependence amongst independent parameters.

The most significant development in this area was bringing out of a manual by U.S. Geological Survey in 1960, which was followed by number of studies including U.K. Flood studies and other typical studies covering general probability considerations, use of historical information, criteria for regional homogeneity etc. Some of the important typical studies in India and abroad during 1960-1987 have been reviewed. Signi-

ficant developments include (i) use of GEV, TCEV and Wakeby distributions, (ii) use of probability weighted moments and Bayesian methodology, (iii) consideration of effect of serial correlation and cross correlation in flood series at different sites, (iv) use of hirarchical approach and paleohydrological information, (v) use of geomorphologic parameters, rainfall areal correlation structure alongwith flood characteristics to delimit hydrologically homogeneous regions, and (vi) development of criterion for selection of physiographic and meteorologic factors.

The application of probability weighted moments technique of parameter estimation appears to work well for situations where records are extremely short and streamflow histories are highly skewed and highly kurtotic. James and Stein estimators give a lower risk with respect to mean squared error than that for the minimum variance, unbiased maximum likelihood estimator. These new developments provide the necessary techniques for combining flood information from gauged sites within a region and using it to infer flood frequencies at ungauged sites.

Most of the regional flood studies in India are based on USGS approach with RDSO data. There is need for taking up systematic regional flood studies using large data base collected by different organisations and agencies. These would enable pooling together of vast information for use in hydrologic analysis and design.

1.0 INTRODUCTION

The planning, design and development of surface water resources require predictions of the amount, the variability and the quality of water in a variety of rivers and streams. The parameters measured, such as streamflow usually describe processes that develop in time, and generally the quantities determined have random fluctuations. The properties of the time series are estimated from records of the historical data and longer the series, the better are estimates of parameters describing the processes. Discharge data is generally obtained by measuring river stages and converting the same using stage discharge relationship. However, many-a-time stage discharge relationship is not defined by a unique functional relation due to typical characteristics of rivers, particularly in case of alluvial rivers. Looped and discontinuous rating curves, insensitive curves and shifting curves are often obtained, which result in errors in estimates of streamflow from stage measurements. Peak discharges, particulrly during flood season are generally estimated using extrapolations of stage discharge relations. Moreover, the measurements of discharges generally tend to become less accurate at higher velocities, where floods etc. may be large and flood flows may spread into flood plains.

The term 'flood' generally means relatively high streamflow that overtops the natural or artificial banks in any reach of a stream or river. This may result in spreading of water over the flood plain and thus will come into conflict with man and his activities. It is, therefore, necessary to study the characteristics of floods, so that they could be controlled. It may be worth mentioning that losses in property and human lives, and the disruptions due to floods are increasing over the years. This is mainly due to encroachment of flood plains and also effect on the river regime due to man's activities (construction of river training works, reservoirs, cutting of forests, urbanisation etc.). Deterministic prediction of floods is feasible only for short time periods, while the probabilistic prediction is the only available prediction for long time periods. This corresponds to the two practial problems of (i) operations requiring short period forecasting, and (ii) planning requiring long period forecasting.

The problem of floods and their computation is one of the main and mot complex problems facing the hydrologists. The optimal development of water resources depends on flood flow control, the design and construction of culverts, bridges, spillways etc., and for taking proper measures for flood control/mitigation. All these problems require accurate and reliable data of floods, and also appropriate procedure for data analysis in order to arrive at desired design variable. However, the requirement of safety of structure under flood conditions should not lead to overdesign and consequent increase in costs. The main objective of study of floods from hydrologic point of view, is to study the run-off process and to estimate the flood flow; and also to develop methods for estimation of floods for design purposes when available data is limited, inadequate or absent. In the present report, the statistical approach of flood frequency analysis is discussed, highlighting the need for standardisation/regionalisation of procedures.

1.1 ESTIMATION OF PEAK FLOW RATE

Many hydrologic design problems require simply the peak flow rate generated by a river system under specified conditions. The rational method of estimating peak flood discharges probably orginated with Mulvaney (1851) in Ireland. Because the method appeared to take into consideration the physical and hydraulic properties of catchments, and their effect on storm rainfall it became known as rational method. The formula for rational method allows the calculation of the maximum or peak rate of discharge from a catchment for a storm of specified rainfall intensity. The rainfall intensity is taken for the selected recurrence interval T (years) and duration equal to the catchment's time of concentration t_c. The rational method is based on the following assumptions:

(a) The rainfall is of uniform intersity over the whole catchment for the duration of the design storm.

(b) The duration of rainfall is equal to the catchment's time of concentration.

(c) The recurrence interval of the peak discharge is equal to that of the rainfall intensity.

The 'time of concentration' is defined as the maximum time taken by water to travel from farthest point of catchment to the outlet under conditions of uniform rainfall. Chow(1961) has discussed many empirical formulae for estimation of time of concentration. One typical formula was recommended for small catchments by Bransby - Williams for overland flow, expressing t_c in minutes = $FL/A^{0.1} \quad S^{0.2}$,where F = 58.5, L = mainstream length in km. and S,slope(m/km).(Institution of Engineers, Australia, 1977).

Empirical formulae proposed for use in varius parts of the world are mathematical representation of envelope curves with additional limitations involving approximations due to curve fitting, and without giving any indication of the scatter of the original data on which they are based. The most common form is of the type Q(in m^3 /sec) = CA^n ,where A = catchment area in km and C, n are coefficients. Some of the commonly used empirical flood formulae in India are Dickens formula (n = 3/4), Ryve's formula (n=2/3) and Inglish formula.

Most of these empirical methods are quite inadequate from of the point of view of providing results consistently within the accuracy required of hydrologic analysis and design.

1.2 GENERAL PROBABILITY CONCEPTS

The flood frequency analysis procedure and results for a particular case, depend to a great extent on available information/data satisfying general probability concepts which form the basis of the approach. Floods being natural phenomena, it is not possible to exactly identify the population set of all possible outcomes. It can only be inferred from examination of a large number of occurences. In addition to identifying the possible events in the parent population, one also requires some idea about the frequency of occurence of each event comprising the population. A complete description of the frequency of occurrence of each event in a parent population is termed as 'frequency distribution'. Knowing the parent population and its distribution, it is possible to determine the probability of occurence of any specified event in the future (HEC, 1975). However, in order for probability theory to be strictly applicable in making probability estimates for future events,the parent population must satisfy the criteria for :

i) Homogeneity: There must be at least one unifying property or characteristic which can be used as a basis for specifying not only elements that are included in the population, but also elements that are excluded from the population.

ii) Independence: The occurrence or non-occurrence of any event must not depend on or be related in any way to the occurence or non-occurrence of any other event in the population.

iii) Randomness: Whether a particular event occurs or does not occur at a given time must be completely a matter of chance.

iv) Stationarily or Time Invariance: The events comprising a population and their associated probabilities of occurrence should not change with time. In other words, the physical processes that underlie the frequency of occurrence of the outcomes cannot change with time.

If any of these four properties are not an attribute of a population, the use of probability theory in making probabilistic estimates of future events may result in erroneous inferences. However, it may be possible to obtain useful probabilistic estimates, even when the population and its distribution, do not conform strictly to the limitations imposed by these properties. This would require sound judgement with proper reasoning in carrying out the analysis ;and inferring the results.

While dealing with natural phenomena like floods, it is almost always impossible to identify specifically each possible outcome in a population and to calculate on a theoretical basis, its relative frequency of occurrence, that is, the true distribution of the parent population is unknown.

For such situations, a general description can be deducted and instead of describing the frequency distribution a priori, it is obtained through analysis of samples. The reliability of the population characteristics inferred from analysis of a sample is directly dependent upon the extent to which the sample is representative of the population which is, in turn, dependent to a large extent on the size of the sample. For example, historical record of annual peak floods of 100 years is a relatively small sample in comparison to size of population (which includes all annual peak floods that have ever occurred or that will occur at that location in future). But from knowledge of meteorology, hydrology and other related sciences, it could be assumed that population inferences from 100 year sample would be relatively reliable. Even where short records are availabke, hydrologic engineering evaluations and decisions cannot be delayed to obtain larger record. It is frequently necessary for such cases to augment the avialable data through consideration of data obtained at other locations in the same region (HEC, 1975).

1.3 PROBABILITY DISTRIBUTIONS AND PARAMETER ESTIMATION

In hydrologic engineering problems such as flood frequency analysis for design flood estimation, the sample of flood series used to draw inferrences about the relative frequency of various events within a population, would have to be very large before one could expect infrequent extreme events to be represented within the sample. There has to be some way of inferring, from the sample information, enough about the population from which the sample was drawn to permit one to estimate the probability of occurrence of events more extreme than those in the sample. This is achieved through identification of and testing with statistical distributions to select a distribution whose characteristics are consistent with known characteristics of natural phenomena under study.

The distributions generally used for modelling the peak flood flows are:

(i) Two parameter log-normal

(ii) Three parameter log-normal

(iii) Extreme value such as Gumbel EV1

(iv) Exponential

(v) Gamma

(vi) Pearson type III

(vii) Log Pearson type III

Recently, the use of the Wakeby distribution with five parameters has also been recommended. This is becoming popular among researchers because of its capabilities to model both the tail ends of the flood series seperately.

Instead of using distributions based on a-priori assumption, many workers have advocated use of transformation to normal distribution as an useful approach. Power transformation or Box-Cox transformation is one such procedure(Box and Cox, 1964). The power transformation is achieved using the transformation formula as follows:

$$z_i = (X_i^{\lambda} - 1)/ \quad , \quad \text{when} \quad \lambda \neq 0$$

$$Z_i = \log X_i \quad , \quad \text{when} \quad \lambda \rightarrow 0 \qquad ...(1)$$

The value of λ generally varies from -1 to +1. The Z_i series is considered to be a near normalized series for that λ which reduces the coefficient of skewness (C_s) to nearly zero. It however, ignores the kurtosis of the distribution which governs the tail thickness of the distribution. Chander et al (1978) have suggested procedure to correct for deviation of coefficient of kurtosis away from 3 in the transformed series. The square root and cube root transformation etc. are all particular cases of Box-Cox transformation.

The distributions like gamma, exponential and Pearson type III have been used generally on the basis of empirical considerations of goodness of fit and not based on a-priori assumption such as for Gumbel EV-1, log-normal etc. Some attempts have also been made to use mixture of two distributions.

For estimation of parameters of statistical distributions, following four techniques are generally used:

(i) Graphical

(ii) Least squares

(iii) Method of moments

(iv) Method of Maximum Likelihood

Rao (1980) used a method of mixed moments to estimate parameters of log Pearson type III distribution. This method attempts to avoid effect of regional variation of skewness co-efficient by intermixing the first two moments of data i.e. mean and variance with the mean of logarithmically transformed data to estimate indirectly co-efficient of skewness of the log transformed series. Such estimate is not subject to distortion due to transformation. For the five parameter Wakeby distribution, Landwehr et al (1979) used probability weighted moments, which avoid using higher order conventional moments to estimate Wakeby parameters.

Though there are many ways of computing statistics from sampled data for use as parameter estimaters, but statistically valid estimators have several specific mathematical and statistical properties. One particular property of importance for hydrologic studies is consistency, which requires that as the size of sample increases, the value of the estimator converges on the true value of the parameter being studied.

1.4 SELECTION OF SUITABLE DISTRIBUTIONS

Many natural phenomena exhibit characteristics which indicate that the normal distribution satisfactorily describes the frequency distribution of outcomes in the parent population. As stated earlier, for flood frequency analysis log normal, Gumbel, log-Pearson type III etc. have been considered appropriate. The selection of suitable distribution for a particular analysis is quite important, since the use of an inappropriate distribution form can produce analysis that grossly distorts the significance of sample data.

In theory, a goodness of fit index should be useful in discriminating between different distributions for the same application. If a single sample say annual flood series is available, the goodness of fit index corresponding to each of the alternative distributions may be calculated and the distribution giving optimum value of the indix is chosen. In practice, a single inference may be sought from the joint study of many samples. However, several alternative indices do not always point to the same distribution. The form of the distribution chosen is dependent not only on the method of estimation but also on the choice of goodness of fit index. The final choice of a particular distribution is thus not unique but is with respect to a stated method of fitting and a stated method of goodness of fit index. The graphical methods are generally used in flood frequency analysis by plotting on specific probability papers to subjectively or qualitatively judge the fit and suitability of that distribution. With the sample sizes, generally available in hydrology, it is always possible to make a mistake by drawing a curve where a line represents the real population and vice versa. Though, the objective methods of goodness of fit tests provide quantitative measures for probabilistic assessment of the best method, they are also not sufficiently sensitive to change in the assumed distribution. A goodness of fit index such as Chi square or Kolmogorov-Smirnov index express the agreement between an observed sample and some theoretically specified population. These tests are however insensitive in the tails of the distribution, whereas the tails or extremes have significant influence on extrapolation of design events for higher recurrence intervals. The use of Chi-square and K-S tests are not encouraged by hydrologists for the purpose of flood frequency analysis.

The goodness of fit indices such as D-index based on probability or x-y plot provide a measure of difference between the observed sample points and the points corresponding to particular distribution at respective plotting position. A major drawback in these methods is that no allowance is made for the fact that not all members of an ordered sample are subject to the same amount of sampling variation. U.K. Flood Studies Report of NERC (1975) while commenting on these indices, suggests that when only small samples are available and fitted distribution is to be extrapolated for return periods upto 100 years, a two parameter distribution is a reasonable approximation, and provides more sensible results than one involving the estimation of a third parameter. However, the use of a two parameter distribution implies the acceptance of a fixed coefficient of skewness for the data concerned, but adopting a three parameter distribution may be lead to the selection of a distribution having different tail pattern from that of population distribution. The USWRC (1981) has therefore, suggested the use of regional coefficient of skewness to fit the log-Pearson type III distribution rather than inferring the same from analysis of a single sample. The analysis of a site data in conjunction with the regional information would improve the reliability of estimators. Beard (1962) had suggested use of analytical adjustment techniques for extending the record at a site or to adjust frequency statistics using longterm record of nearby site, when there is correlation between recorded values at two locations.

1.5 OUTLIER PROBLEM

Frequency analysis of annual peak flood series is often unrealistic due to the presence of one or more very large observations, or very small observations. The statistical parameters estimated in the presence of such high or low outliers get badly distorted, thereby significantly affecting the results of flood frequency analysis. It has to be decided as to whether to retain, modify or delete these outliers. All procedures for treating outliers ultimately require judgement involving both mathematical and hydrologic considerations.

Natural outliers occur because of the inherent variability in flood data. Errors in the model, measurement or data processing may also originate outliers in a flood series. Studies on the suitability of probability distributions on a regional scale could reduce the model related errors. USWRC (1981) have provided procedures for the detection and treatment of high and low outliers.

1.6 SAFETY, ECONOMY, RISK, RELIABILITY

When the flood frequency analysis approach to hydrologic design is adopted, one of the first steps is to decide the frequency (recurrence interval) with which the design flood of proposed structure is to be exceeded. It requires consideration of socio-economic consequence of such exceedence and the cost of reducing the frequency of such exceedences by increasing the flood handling capacity of the structure.

CWC (1972) recommended criteria for design flood of dams and other hydraulic structures specified as follows(Publication No.258 Sept. 1972),:

(a) In the case of major and medium dams with storages more than 50,000 acre feet (6167 ha.m):-

The probability method when applied to derive design floods for long recurrence intervals several times larger than the length of data has many limitations. In certain cases, however, like that of very large catchments where unit hydrograph method is not applicable and where the sufficient long term discharge data is available, the frequency method may be the only course possible. In such cases the design flood to be adopted for major structures should have a frequency of not less than once in 1,000 years. Where annual flood values of adequate length are available they are to be analysed by the Gumbel's method, and where the data is short, either partial duration method or regional frequency technique is to be adopted as a tentative approach; and the result verified and checked by hydrological approaches.

Sometimes, when the flood data is inadequate, frequency analysis of recorded storms is made and the storm of a particular frequency applied to the unit hydrograph to derive the flood; this flood usually has a return period greater than that of the storm.

While planning there may be some projects where there is hardly any discharge data available. In such circumstances for preliminary studies, the peak flood may be estimated by empirical formulae.

Since the probability method and the empirical approach have their limitations and would give only the design peak discharge and not the complete design flood hydrograph, the Commission are of the view that these methods be used provisionally and every effort should be made to collect the required hydrological data at site and obtain the design flood by rational method (unit hydrograph method) before the project design are finalised.

In the case of barrages and minor dams with less than 50,000 acre feet (6167 ha.m) storage:-

In the case of permanent barrages, and minor dams with less than 50,000 acre ft. (6167 hectare metres) storage, the standard project flood or a 100-yr flood, whichever is higher, is to be adopted.

In the case of weirs, aqueducts etc:-

For pick-up weirs, a flood of 50-100 years frequency should be adopted according to its importance, and level condition.

Waterways for canal aqueducts should be provided to pass a 50-100 years flood, but their foundations and freeboard should be for a flood of not less than 100 years return period. In case of cross drainage works which carry highways or railways, waterways provided should also satisfy the respective standard code of practice of highways or railways.

It is further stated in CWC(1972) that each site is individual in its local conditions, and evaluation of causes and effects. While, therefore, the norms, mentioned herein above may be taken as the general guidelines, and hydrologists,and, the designer would have the discretion to vary the norms, and the criteria in special cases, where the same are justifiable on account of assessable and acceptable local conditions; these should be recorded, and have the acceptance of the competent authority.

One possible aid to an arbitrary selection of a design flood in the absence of economic considerations is to consider the probability that the structure will have its capacity exceeded one or more times during the design life of the structure. There are some situations in which the structure for which the design flood is to be determined has a clearly defined finite life, e.g a temporary coffer dam. However, a permanent dam dam will not be removed at the end of a specific period and would continue to be subjected to floods.

In these situations, the structural design life can be expected to be much more than economic design life.

When the possibility of more than one flood in anyone year is allowed and the occurrence of floods exceeding any specific value is treated as a poisson arrival process, then the probability P of one or more exceedence of design capacity during the design life L of a structure whose design flood recurrence interval is T year is expressed as follow:

$$P = 1 - \exp(-L/T) \qquad \text{...(2)}$$

$$T = L/\log_e(1-P) \qquad \text{...(3)}$$

Nearly similar values are obtained by the following relationship also:

$$P = 1-(1-1/T)^L \qquad \text{...(4)}$$

These relationships can be used to make subjective decesion about an acceptable "risk of failure" or probability that design capacity of a structure will be exceeded one or more times during its life.

1.7 REGIONAL FLOOD STUDIES

Flood frequency analysis for a station with long record can be based almost exclusively on the record of that station alone. But generally annual flood series available at site of interest is short, the choice of appropriate probability distribution cannot be based on the sample alone and prior knowledge about the form of distribution, has to be used. Such prior knowledge includes:

(i) any one of the distributions viz. Gumbel, log-normal, log Pearson type III, Pearson type III etc.

(ii) a regional curve consisting of the mean distributions of all recorded floods appropriately scaled by specific parameters such as mean flood.

(iii) when there is no record available at a site, a regional curve together with an estimate of mean annual flood obtained by appropriate relationship of mean annual flood with catchment characteristics in the hydrologically homogeneous region.

The choice of catchment morphometric characteristics depends on the (i) judgement of the likely predictive success of variables, based on the results of other studies and theory, and (ii) the numerical calculation of regression equations and the interpretation of their coefficients. Some of the physiographic/morphometric factors which may influence the mean annual flood at a given site are catchment area, channel storage, artificial or natural storage in lakes or ponds, stream slope, land slope, stream density, stream pattern, elevation, orographic position, underlying geology, soil cover, cultivation etc. However, these characteristics are often dodged by correlations, resulting in presence of interdependence amongst independent parameters involved in multiple regression analysis. The achievement of an optimal level or regional information transfer is also faced with problems of deciding about regional homogeneity.

This review note examines significant developments in the area of flood frequency analysis with emphasis on regional frequency analysis upto 1984-85 and also in subsequent years upto 1986-87, both in India and abroad (since this report has been finalised in 1987 for printing).

2.0 REVIEW OF LITERATURE

The most significant development in the area of Regional flood frequency analysis was bringing-out of a manual of Flood Frequency Analysis by the U.S. Geological Survey. Other significant developments were studies carried-out by Natural Environmental Research Council of the United Kingdom, Regional Flood Estimaion in Australia and bringing-out of guidelines for flood frequency analysis by the U.S. Water Research Council. Other typical studies have been carried-out in this area both in India and abroad covering general aspects, use of historical information, criteria for regional homogeneity etc. The review in the following sections has accordingly been described under separate heads.

2.1 METHOD USED BY U.S. GEOLOGICAL SURVEY

The U.S. Geological Survey in 1960 brought out a manual of flood frequency analysis describing the methods and practices of Geological Survey under Geological Survey Water Supply paper 1543-A by T. Dalrymple. The method suggested by USGS was recommended for applicability to a region of any size, a river basin or a state, so long as the region could be assumed as hydrologically homogeneous. The proposed analysis provided for development of two curves. The first curve expressed the graphical relationship between peak flood discharge and time, showing variation of peak discharge, expressed as a ratio to the mean annual flood, with recurrence interval. The second curve/relationship relates the mean annual flood to the size of drainage area alone or to the size and other significant basin characteristics. Using these twc curves/relationships, the procedure could be adopted for defining the frequency curve for any place in the region. This would involve, (a) measurement of drainage area and other appropriate basin characteristics from maps, (b) selection of mean annual flood corresponding to appropriate drainage area and/or other factors from the second curve/relationship, (c) determination of ratio of peak discharge to mean annual flood for selected recurrence intervals from first curve/relationship, and (d) multiplication of these ratios by the mean annual flood and plotting peak flood discharges for different return periods/frequency to define the frequency curve.

This procedure recommended by U.S. Geological Survey (Dalrymple 1960) also included detailed procedure for carrying out the homogeneity test, which was developed by Langbein. This test which has been widely used for homogeneity test in flood frequency analysis, is being briefly discussed below:-

Homogeneity Test

The test developed by Langbein is based on the presumption that if the records in a group do not differ from one another by amounts that cannot reasonably be expected by chance, then the records may be considered as 'homogeneous' in a statistical sense. A chart was suggested for the homogeneity test and the basis for construction of the chart was as follows:-

i. If a thousand year record were divided into hundred records of 10 years each, in which the highest flood in each of the 10 year periods would represent an estimate of a 10 year flood, these 10 years floods not all be the same and the difference that exist for these estimates would be mainly due to chance. Such a study would indicate what ranges of variation could be expected and what variations would represent results other than by chance. Similar test could be run by studying a two thousand year record in which hundred records of 20 years each could be taken. In that case, in each record the second highest flood would represent the 10 year flood. The spread of the 10 years flood as estimated for these 20 years records would be less than in the previous case, when thousand year record was considered. From such studies, the kind of spread due to chance could be determined.

ii. There are no record of such lengths, however, it is possible to calculate from theory, the distribution of floods in a thousand year or ten thousand year record in terms of probability. Such calculations are the basis of the test for homogeneity proposed by Langbein.

iii. Assuming Gumbel distribution to be applicable to floods, the procedure involved calculations of the spread to be expected and to set a limit to the spread that will be acceptable. The standard deviation of the spread of the reduced variate y of the Gumbel distribution is given by:-

$$\sigma_y = \frac{e^y}{n}\sqrt{\frac{1}{T-1}} \qquad ...(5)$$

Where:-

T = recurrence interval of return period

n = number of years of record

y = function of the recurrence interval T

This means that in a large number of different but homogeneous records each n-years long, probably two thirds of the estimates of the T-year flood will be within σ_y of their most probable value of recurrence interval T. The proposed method assumes a wider range of variation and involves use of two standard deviations as the permissible range. This means 95 percent of the estimates will lie within 2 σ_y of the most probable value of T.

iv. The method also assumes that the test be made on the 10 year flood, because this is the longest recurrence interval for which most data records will give dependable estimates.

Since : for T = 10 years,

y = reduced variate of Gumbel distribution for T = 10 years = 2.25,

$e^y = 9.49$,

and as such values for upper and lower limits can be computed as follows:-

Lower limit $T_L = y - 2\ \sigma_y$, ...(6)

Upper limit $T_u = y + 2\ \sigma_y$, ...(7)

$$2\ \sigma_y = \frac{2e^y}{n}\left[\sqrt{\frac{1}{T-1}}\right] \triangleq 2\left[\sqrt{\frac{T}{n}}\right] \quad ...(8)$$

Also

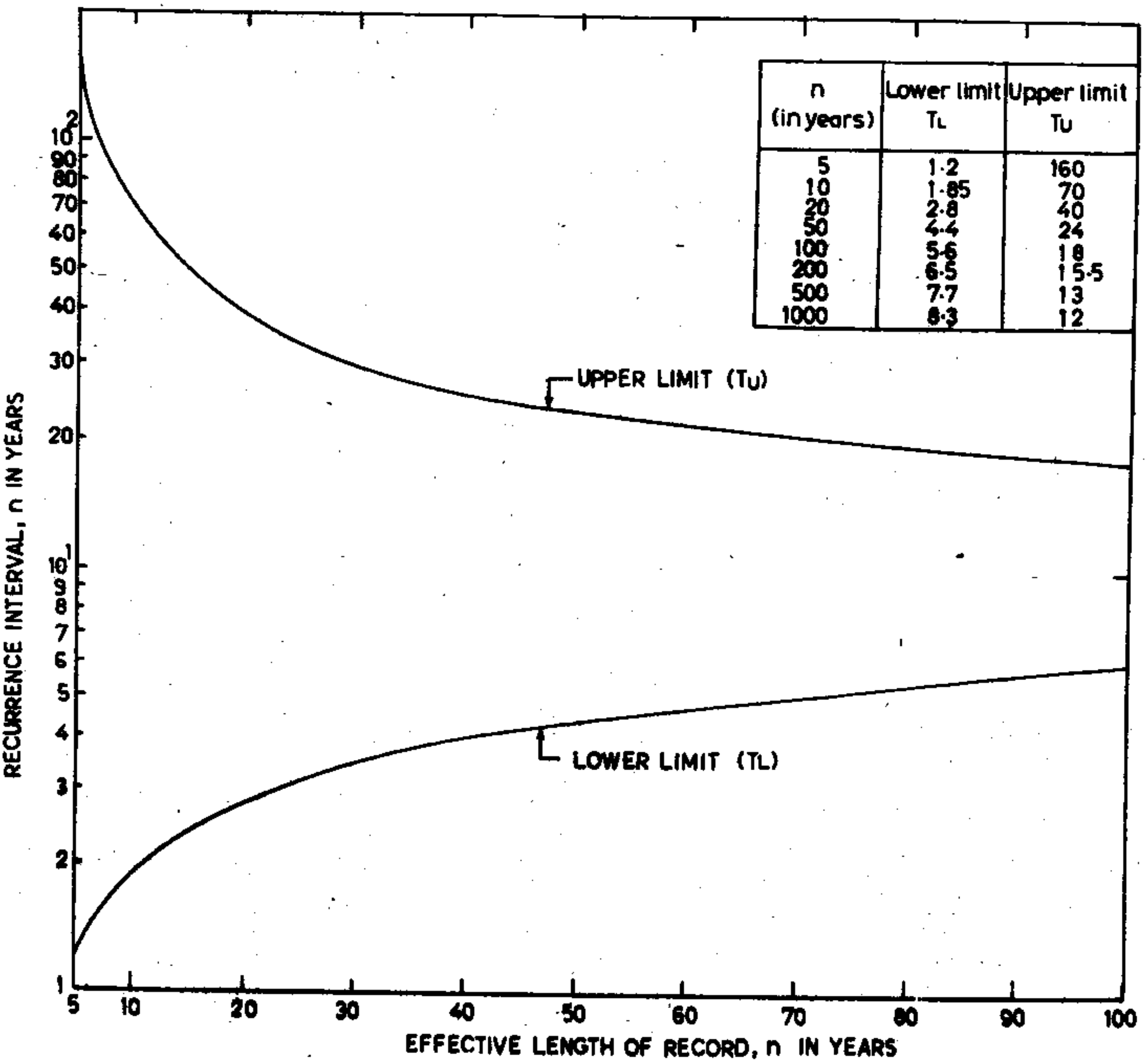
$$2e^y\ \left[\sqrt{\frac{1}{T-1}}\right] = 2 \times 9.49\left[\sqrt{\frac{1}{(10-1)}}\right] = 6.33$$

$$2\ T = 2\sqrt{10} = 6.32$$

$$2\ \sigma_y = 6.33/\sqrt{n} \quad ...(9)$$

The values of y corresponding to record length n varying from 5 years to 1000 years have been computed, and are tabulated giving the upper limit and lower limit for homogeneity test, as reproduced in table....

These limits plotted in the form of a homogeneity test chart are given in figure 1.



n (in years)	Lower limit T_L	Upper limit T_U
5	1·2	160
10	1·85	70
20	2·8	40
50	4·4	24
100	5·6	18
200	6·5	15·5
500	7·7	13
1000	8·3	12

FIG.1 -HOMOGENEITY TEST CHART

TABLE 1

Value of upper and lower limits in terms of Gumbel's reduced variate y.

Length of record, (in years)	Lower Limit T_L $(y-2\sigma_y) = 2.25-[\frac{6.33}{\sqrt{n}}]$	Upper Limit T_u $(y + 2\,\sigma_y) = 2.25 + (\frac{6.33}{\sqrt{n}})$
5	1.2	160
10	1.85	70
20	2.8	40
50	4.4	24
100	5.6	18
200	6.5	15.5
500	7.7	13
1000	8.3	12

For the use of homogeneity test chart, the annual flood records for all stations in a region are assembled, and 10 years floods are estimated for each of the stations. Each 10 year flood is divided by the mean flood to get the 10 year ratio and an average of these ratios for different stations is obtained. The data for homogeneity test is tabulated as follows:

TABLE 2 HOMOGENEITY TEST

S.No.	Name of Station	Drainage area (km^2)	Mean Annual Flood $Q_{2.33}$ (m^3/sec)	10 year flood Q_{10} (m^3/sec)	Ratio $(\frac{Q_{10}}{Q_{2.33}})$	$Q_{2.33}$ x average ratio
1	2	3	4	5	6	7

Average ratio= ________

T corresponding to Q in col.7	Effective Period of record (adjusted) (years)
8	9

The adjusted period of record in column 9, is the number of years of actual record plus one-half the number of years the record is extended using correlation procedure for use only in assignment of correct order numbers to the peaks.

The data from cols. 8 and 9 of the table is plotted on the chart shown in fig. 1, wherein the upper and lower limits are indicated. All points plotting within these limits indicate that the concerned stations are homogeneous and could be grouped together to define a regional frequency curve. If some points plot outside the limits, then corresponding stations are excluded, average ratio for ($Q_{10}/Q_{2.33}$) is computed again for remaining stations, and the procedure is repeated.

For regional frequency curve, Dalrymple(1960) recommends plotting of preliminary frequency curves of discharge v/s recurrence interval for all stations in a homogeneous region. The ratios of several floods of different recurrence intervals to the mean annual flood are tabulated for each station and median values of these ratios are computed for different recurrence intervals. Each median flood ratio is then plotted to its corresponding recurrence interval on a frequency chart (Gumbel plotting paper) and an average frequency curve is plotted. An adjustment procedure has been suggested to deal with cases when more than one base period is suggested in the available length of flood records.

The application of the regional frequency curve to an ungauged basin in the homogeneous region involves estimation of the mean flood for that basin. The development of a suitable relationship between mean annual flood, and physiographic and meteorologic characteristics is recommended for this purpose by Dalrymple (1960). The physiographic factors which may influence the mean annual flood include:-(a) size of drainage area, (b) channel storage, (c) artificial or natural storage in lakes or ponds, (d) slope of streams, (e) land slope, (f) stream density, (g)stream pattern,(h) elevation (i)aspect, (j) orographic position,(k) underlying geology, (l) soil cover, (m) cultivation and others. The meteorologic factors are concerned with the magnitude and distribution pattern of precipitation, and include:

(a) type of region, whether humid or arid, (b) storm directions (c) storm patterns, (d) storm volumes, (e) precipitatioin intensities, (f) effect of snowmelt, (g) extent of ice jams, and others. The evaluation, treatment and use of the meteorologic elements in regression relationships is generally less certain and there is scope for considerable investigation. However, in the physiographic factors, size of basin is the most important and readily available factor. Rainfall intensity of appropriate duration and frequency would also be more directly related to peak discharges.

For dealing with situations when reasonably short record (2-6 years), bearing an uncommon sequence of drought years is available alongwith a longterm record of a station nearby, Dalrymple (1960) suggested use of partial flood series of the short-term station and its comparison with corresponding period series for long-term station, to estimate mean annual flood for short-term station. The need for special treatment for large rivers was also emphasised, in view of likely change in both the flood frequency relation and the mean annual flood at gauging stations along the stream.

2.2 U.K. FLOOD STUDIES REPORT

This comprehensive report brought out by NERC (1975) in five volumes deals with hydrological studies, meteorological studies, flood routing studies, hydrological data and maps on the basis of work carried out in United Kingdom at the Institute of Hydrology, Meteorological Office and Hydrologic Research Station with the cooperation of Irish Office of Public Works and Meteorological Surveys, Soils Surveys and other organisations. The report presents methods, results and the analysis of flood records from the British Isles. The main objective of the studies has been to develop and present methods of estimating the floods, which will be exceeded at a given site on an average once in T-years and also the possible limit to the flood or the 'probable maximum flood'. Two main approaches have been used in the study: the first was based on statistical analysis of flood series at a gauging station, and the second on investigations of rainfall and resulting runoff at selected stations.

In order to provide a statistical framework within which floods can be analysed, the properties of the various distributions, the methods of fitting distributions to sample data including graphical fitting and fitting by moments; and maximum likelihood have been described and studied. Certain distribution such as log normal, and Pearson type III are used for maximum flood on empirical grounds while the extreme value distribution applied by Gumbel have some theoretical basis. It was seen that since no firm theoretical basis for choice between distributions, the comparison of performance of different distributions depends on the index of fit and on the plotting positions used. Though, the three parameter distributions were found to be better or more flexible than the two parameter distributions, no clear cut choice resulted from comparative study. The studies also included use of peaks over threshold flood series and time series approach for flood frequency analysis; and explain the relationship between design life, risk of failure and design return period.

In many practical cases, estimates of the flood regime have to be made from short periods of records, but since sampling error of a short record is high, it may be desirable to improve the estimates by adding further information from adjacent long term records. A detailed study was made of this technique of extending the short term record by co-relation with adjacent long term record. It was found that the improvement using extended data regression of flood estimates on catchment characteristics was small due to limitations of the regresion model.

For regional applications, extensive studies were carried out for flood estimation from catchment characteristics, so that where no record is available at a site, a preliminary estimate may be made from relations between floods and catchment characteristics. Though, the results of such indirect method of estimation are generally less reliable than the estimate based on direct analysis of record at a site, it provided useful method for estimation of floods at numerous sites; where structures are to be constructed for the development of water resources or for communication purposes. A detailed study of various catchment characteristics including morphometric characteristics as well as rainfall characteristics was carried out, and hydrologically relevant characteristics were chosen, keeping in view that they should be un-correlated as far as possible and should be capable of being measured simply for a large number of catchments.

The catchment indices thus selected and included were as follows:-

1. Catchment area (km²): (AREA)
2. Taylor-Schwarz (TAYSLO) and 10-85% (as S1085) slope in m/km were both tried, though S1085 is easily calculated but as it depends on the two points on the profile, it may be more affected by measurement errors. The Taylor-Schwarz slope is based on the square root of the gradients and it uses the fact that velocity for each reach of a sub-divided main-stream is related in the manning's equation to the square root of slope. The index is equivalent to the slope of a uniform channel having the same length as the longest water course and an equal time of travel. A typical calculation of this index is reproduced here from flow studies:-

TABLE 3

DETERMINATION OF EQUIVALENT MAINSTREAM SLOPE

Elevation (ft)	Elevation (ft)	Length (ft)	S_i	S_i	$1/ S_i$
515	-	-	-	-	-
590	75	52800	0.001420	0.0377	26.15
610	20	52800	0.000379	0.0195	51.28
630	20	52800	0.000379	0.0195	51.28
670	40	52800	0.000758	0.0275	36,36
740	70	52800	0.001328	0.0364	27.47
910	170	52800	0.003220	0.0567	17.64
					210.54

TAYSLO = (Number of equal length reaches/$\Sigma(1/\sqrt{S_i})$)²

= $(6/210.54)^2$ = 0.00081.

3. Drainage network: The stream frequency, (STMFRQ) used as an index, which is defined as the number of junctions per square km of catchment.

4. Climate: On the basis of available information two rainfall induces were considered. The first was average rainfall for a standard period 1916-50, (SAAR), which was easily available from published maps. The other index was (RSMD) based upon daily rainfall and soil moisture deficit. The soil moisture deficit is calculated from a water balance between daily rainfall and the Penman estimate of actual transpiration

5. Soil: The proportion of catchment covered by each class of soil was determined from the soil map and a composite runoff index, (SOIL) was derived from the formula as below:

$$SOIL = \frac{(0.15S_1+0.30S_2+0.40S_3+0.45S_4+0.50S_5)}{(S_1+S_2+S_3+S_4+S_5)} \quad ...(9)$$

where, S_1 , S_2, S_3 , S_4, S_5 denote the proportions of the catchment covered by each of the soil classes 1-5 The factors in this formula were derived from a consideration of the storm runoff data.

. Land Use: The proportion of the catchment in an urban area (URBAN) was used as a parameter.

. Lakes:- The proportion of the catchment draining through a lake was used as the index, (LAKE) for including the effect of lakes.

The mean annual flood and the coefficient of variation as obtained from annual maximum series were chosen for regression with catchment characteristics. Preliminary regression test using multiple regression analysis approach were carried out to answer the following questions:-

(i) Which estimate of the mean annual flood is basically related to catchment characteristics?

(ii) Should all available records be used or should they be selected according to length of record or grade?

(iii) Should all sets of records be grouped together and can they usefully be divided according to size of catchment or alternatively according to region.

These tests led to following decisions:

(i) An estimate of the mean annual flood derived whenever possible by extending the record by co-relation with nearby record(parameters BESMAS) should be used, as a measure of the mean annual flood where available and the arthmatical mean in other cases.

(ii) Short term and low grade stations should be included in the data as their exclusion did not change the results achieved.

(iii) There was no case for treating small and large catchments separately.

(iv) The study area was to be divided into seven regions, six of them having same coefficient but different intercepts.

The multiple regression analysis was carried out for different regions using mean annual flood as the dependent variable and catchment characteristics described above as the independent variables. A typical relationship obtained for central region of British Isles is as below:-

$$\bar{Q} = 0.0213\ (\text{AREA})^{0.94}\ (\text{STMFRQ})^{0.27}\ (\text{S1085})^{0.16} (\text{SOIL})^{1.23}\ (\text{RSMD})^{1.03}\ (1+\text{LAKE})^{-0.85} \quad \ldots(10)$$

2.3 REGIONAL FLOOD ESTIMATION PROCEDURE IN AUSTRALIA

Irish (1977) has described the status of regional flood estimation in Australia. He has mentioned that there is little uniformity in regional flood frequency techniques developed so far in Australia compared to that in U.S.A. He also mentions about large initial effort required for development of a regional flood estimation technique and relevant relationships. Some of the techniques and procedures currently in use in Australia and Papua New Guinea have been listed. The author recommends that these methods should not be used for recurrence intervals of more than 100 years. Ward (1968) presented a precedure using regional flood estimation technique that can be used on limited information such as the discharge records from a single stream gauging station. The procedure is based on a modification of the Myer formula as given below:

$$q_{10} = C_{10} \ A^{0.8} \qquad ...(11)$$

where

q_{10} = Mean daily discharge in m³/sec with a recurrence interval of 10 years

C_{10} = Coefficient based on average recurrence interval of 10 years, and

A = Catchment area (km²)

The average recurrence interval of 10 years has been chosen arbiteraly and the mean discharge value has been used, since at many gauging sites only one gauge reading is taken per day. The procedure involves the following steps:-

1. Measure the catchment area of the stream at the required site from available toposheets.
2. Considering the shape of the catchment, rainfall intensity, vegetation cover, soil type and any other comparable features, select another stream (or streams) whose catchment displays similar characteristics and assess value of C_{10}.
3. Using Myer formula, calculate the mean daily discharge with a recurrence interval of 10 years, q_{10} for the required site.
4. Select a number of observed floods for the stream used in step 2 and determine the average ratio of instantaneous peak discharge to mean daily discharge using

available data for that stream.

5. Multiply the value of q_{10} by the average ratio obtained in step 4 to obtain an estimate of the peak discharge with recurrence interval of 10 years (Q_{10}).

6. The peak discharge for other average recurrence intervals can then be estimated by applying the same ratio to the 10 year peak discharge as applicable for the stream chosen in step.2.

2.4 U.S. WATER RESOURCES COUNCIL GUIDELINES

In a pioneering attempt to promote a consistent approach to flood-flow frequency determination, the U.S. Water Resources Council in Dec.1967, published Bulletin no.15."A uniform Technique for Determining Flood Flow Frequencies". An extension and update of this bulletin was published in March 1976 as bulletin no.17," Guidelines for Determining Flood Flow Frequency", which was subsequently revised in June 1977 as bulletin 17A and in Sept. 1981 as bulletin 17-B of the Hydrology Committee. The guide (US WRC, 1981) is divided into following six broad sections;

(i) Information to be evaluated: Four categories of flood data are recognized viz. systematic records, historic data, comparison with similar watersheds, and flood etimation from precipitation. Their use to define the flood potential is briefly discussed.

(ii) Data Assumptions such as natural trends, randomness of events, watershed changes, mixed populations and reliability of flow estimates are briefly discussed.

(iii) Determination of the frequency curve: Besides basic guide for analysis of annual flood series, procedures are also recommended for conversion of annual to partial duration flood series. The log-Pearson type III distribution is recommended as the basic distribution for defining the annual flood series. The method of moments is used to determine the statistical parameters of the distribution from station data and generalised relations are used to modify the station skew coefficient.

(iv) Reliability applications: Procedure for computing confidence limits to the frequency curve are provided alongwith those for calculating risk and for making expected probability adjustments.

(v) General: It provides discussion on non-conforming special situations, plotting positions and suggested future studies.

(vi) Appendix: It provides references, symbols, tables and the computational details.

The additional needed studies identified by the work group included the following:

(a) Selection of distributions and fitting procedures.

(b) Identification and treatment of mixed distributions.

(c) Treatment of outliers both as to identification and computational procedures.

(d) Alternative procedures for treating historic data.

(e) More adequate computation procedures for confidence limits to the Pearson type III distributions.

(f) Procedures to incorporate flood estimates from precipitation into frequency analysis.

(g) Guides for defining flood potentials for ungauged watersheds and watersheds with limited data.

(h) Guides for defining flood potentials for watersheds altered by urbanization and by reservoirs.

The skew co-efficient of the station record is sensitive to extreme events and as such, it is difficult to obtain accurate skew estimates from small samples. The guide recommends pooling regional information from nearby sites to estimate generalized skew estimates, and weighting the station skew estimates with appropriate generalized estimate. Three methods are involved in the analysis procedure for developing generalized skew coefficients,: (i) skew isolines drawn on a map, (ii) skew prediction equation, and (iii) the mean of the station skew values. Under the assumption that the generalized skew is unbiased and independent of station skew, the mean-square error (MSE) of the weighted estimate is minimised by weighting the station and generalized skew in inverse proportion to their individual mean square errors.

2.5 Other Typical Studies Abroad During 1960-80.

(i) Yevjevich (1963) advocates treatment of floods from a probabilistic point of view in hydraulic design. He has also mentioned difficulties in using small samples of data. The probability distribution curves of floods taking into account both the errors in computed flood variable and the sampling error, have very large confidence intervals for the standard levels of significance (10% or 5% level) in the region of higher observed flood flows. The extreme parts, as well as the extrapolated parts, of flood probability distribution curves etc. are, therefore, unreliable. Yevjevich suggests use of the upper curve of the confidence interval on 5%or 10% level of significance computed around the flood probability distribution curve of small sample records.

ii) Nash and Shaw (1966) analysed discharge records of 57 catchments of Great Britain and the parameters of the frequency distributions of annual maximum flood peaks were correlated with the area, slope and mean annual rainfall of the catchments. The mean annual maximum discharge and the coefficient of variation of the annual peak flood were used,since these are sufficient to obtain any two parameter frequency distribution viz. Gumbel, log normal etc. The best prediction equations providing highly significant correlations in the study were as follows:-

$$\left.\begin{aligned} \bar{Q} &= 0.009\ A^{0.85}\ R^{2.2} \\ \bar{Q} &= 0.074\ A^{0.75}\ S \\ CV_Q &= 219\ R^{-0.5} \\ CV_Q &= 178\ S^{-0.25} \end{aligned}\right| \qquad \ldots(12)$$

where

$\bar{Q}$ = mean of annual peaks for a catchment

CV_Q = Coefficient of variation of annual peak flood series

A = Catchment area (Sq. miles)

S = mean catchment slope in parts per 10000

R = mean annual rainfall in inches.

The authors also provided method for working out the variance of the estimate of the flood of any given return period using Gumbel and log-normal distributions, and the regression relationships.

(iii) Cole (1966) applied the USGS procedure of regional analysis of flood flows described by Dalrymple (1960) to data of stations in Eastern England for 1939-60 and Wester England and Wales for 1938-62. Regional frequency curves on Gumbel probability paper were plotted as st. lines representing relationships between $(Q_T/\bar{Q})$ and return period T. Typical values from two curves are as follows:-

	Eastern England			Western England and Wales		
1. Return Period	10	50	100	10	50	100
2. $(Q_T/\bar{Q})$	1.45	1.90	2.08	1.48	1.96	2.15

The relationships between mean annual flood (cusecs) and catchment area (sq.miles) were obtained as parellel lines for different groups according to certain catchment characteristics; and defined by $\bar{Q} = CA^{0.85}$. The value of exponent 0.85 compared with value of 0.8 for Ohio and nearly all of Georgia, and 0.65 for Louisiana. This preliminary investigation was considered quite promising. However, it was suggested that to derive full benefit from the method in England and Wales, the length of data record should improve further.

iv) Jennings and Benson (1969) pointed out the difficulty in fitting the log-Pearson type III distribution to a series, if it contains a number of zero values, such case may occur for small arid-region streams. They suggested a method wherein the probability of occurrence of a non-zero peak is combined with the conditioned probability of exceeding a given flood magnitude, given that a non-zero peak has occured. The method has also been found useful for fitting flood series in which information of peak annual floods below a specific stage is lacking.

v) Yevjevich (1972) discussed a wide variety of problems related to floods. These included: (a) properties of basic data for analysis of floods, (b) general properties of floods,(c) extraction of information on floods from rainfall data, (d)extraction of information on floods from runoff data,(e) prediction of floods, (f) probability distribution of floods, and (g) statistical inference for flood events.

(vi) Burges et al (1975) have indicated important advantages of the three parameter log normal distribution for application in hydrology and water resources. They compared two methods for estimation of third parameter a viz. (a) using sample mean,median and standard deviation, and (b) using sample mean, standard deviation and skew. It was found that for skew coefficient greater than 0.5, the skew method of parameter estimation results in lesser variability and smaller bias, in comparison to that for median method.

(vii) Bobee (1975) carried out a mathematical and statistical study which shows the flexibility and limitations of the log-Pearson type III distribution. The various forms of density function and relationships that exist between distribution parametes and moments,coefficient of variation and coefficient of skewness were indicated. It was also shown that in case of $C_s > 0$, the coefficient $(C_s)_1$ is always positive, but when $C_s < 0$, $(C_s)_1$ may be either positive or negative, where C_s and $(C_s)_1$ are coefficients of skewness for original and log transofrmed series, respectively.

He also proposed a technique of parameter estimation that preserves the moments of the observed sample rather than those of the logarithms of the data, thereby giving equal importance to all elements of the sample. The method which gives same weight to the logarithms of the observed values reduced the importance of the larger elements of the sample.

Bobee (1975) also compared the results of flood frequency analysis using proposed technique with those using different distributions, as given by Sangal and Biswas, as given below:

TABLE 4

COMPUTED FLOWS FOR SELECTED RETURN PERIOD
(STATION PMFS,WATER SURVEY OF CANADA)

(in 1000 ft^3/sec)

Distribution	T=50 X_{10}	T=100 X_{100}	T=1000 X_{1000}	T=10000 X_{10000}
1. Gumbel	506	546	675	804
2. Lognormal-2	467	492	572	648
3. Lognormal-3	466	491	570	645
4. Pearson	468	494	574	649
5. Log Pearson-3(WRC)	457	478	542	598
6. Log Pearson-3(proposed method)	466	493	577	657

He recommended the proposed method for LP3 in comparison to WRC method for determination of events with a high return period and suggested comprehensive study of two methods by using generated samples from a log pearson type3 distribution with Monte Carlo simulation technique.

viii) Vicens et al (1975) discussed the three type of uncertainties that may exist in any water resources design problem: natural, parameter and model uncertainities. The use of regional information has been suggested as a technique for reducing parameter uncertainities and the use of Bayesian methodology for combining regional information with at-site historical records. The authors have described the Bayesian analysis of hydrologic models in general and the independent normal process as an example. It was concluded that the use of Bayesian approach is more explicit in considering the parameter uncertainty in inferences and decisions. The use of regional information through informative prior pdf's reduces the parameter uncertainty, especially for short historical records.

(ix) Sokolov et al (1976) have compiled flood flow computation methods from world experience. The authors state the three principal problems involved in determining the probability distribution that most reliably describes the statistical characteristics of a short sample, viz. (a) selection of the appropriate theoretical distribution, (b) determination of the numerical values of the distribution parameters, and (c) evaluation of the accuracy of the parameter values. While commenting on relative eficiencies of methods of parameter estimation, it is stated that when the three parameters of the distribution are computed, the estimate made by the method of moments is less than half as efficient as that made by the method of maximum liklihood.

(x) McCuen et al (1977) have carried out a literature evaluation of flood flow frequency for ungauged watersheds, to identify alternative procedures and to gather information regarding the accuracy, reproducibility, and practicality of the alternatives. The eight categories in which procedures were subdivided, included statistical estimation of peak flood, statistical estimation of moments and index flood estimation. Nearly 240 publications were reviewed including some typical publications involving comparison of procedures for frequency analysis as follows:

(a) Chow (1962) had presented a method for estimating peak runoff for rural watersheds in Illinois, for recurrence intervals of 5,10,25,50, and 100 years. Only topographic and land use data are needed as input. The report also presents a listing of 66 other prediction methods developed before 1960.

(b) Cruff and Rantz(1965) compared three methods: the index flood, multiple regression and moment estimation with four different probability distributions tested using the last method. A humid and sub humid region in California were used as study areas. It was seen that regression approach provides the most reliable results for estimating the 50 years and 100 years peak.

(c) Lara (1974) compares the following four regional flood frequency methods using data collected at 131 gauging stations in Iowa: (1) multiple regression relating the peak discharge for selected return periods to watershed characteristics,(2) the index flood method; and regionalization of the parameters of the (3) log Pearson type III distribution, and (4) log normal distribution. On the basis of split sample tests, it was concluded that both the multiple regression approach and the regionalization of the log-Pearson type III parameters are equally reliable.

The most important conclusion of this literature evaluation was that the literature, while voluminous, does not contain sufficient information to provide a general statement about either specific procedures or general techniques as to their accuracy, reproducibility or practicality.

(xi) Kite (1978) has devoted a full chapter on regional analysis in his book on 'Frequency and risk analysis in hydrology'. He mentions about three deficiencies found by Benson (1962) in the USGS index flood method:
(a) The index flood (mean annual flood) for stations with short periods of record may not be typical.
(b) The homogeneity test is used to determine whether the differences in slopes of frequency curves are greater than may be attributed by chance alone. The 10 year level apparently established homogeneity in some studies but the individual curves indicated wide and sometimes systematic differences at higher levels.

(c) The ratios of any specified flood to the mean annual flood will vary inversely with the drainage area. In general, the larger the drainage area, the flatter the frequency curve. The effect of drainage area is relatively greater for floods of higher recurrence intervals.

The author also mentions about study by Thomas and Benson(1969) wherein both techniques of estimating specific return periods and regional distribution parameters were examined using over 70 parameters of stream flow. The drainage area and mean annual precipitation were indicated as the most important parameters.

In another study, a procedure for using the square-grid technique to estimate events at required return periods on ungauged streams was suggested by Kite (1974). The mean and standard deviation of annual peak flood series were related to the square grid physiographic data for those streams which were gauged. The relationship was extended to ungauged sites assuming a log-normal probability distribution of annual extremes.

Kite (1978) mentions about use of Bayesian analysis for combining a single station analysis with regional analysis information. Bayes theorem states that the posterior probability distribution function (pdf) is proportional to the product of the pdf containing the prior information (e.g. from regional analysis) and the likelihood function of the sample pdf'. The author also mentions about the problem with the use of Bayesian analysis regarding no account being taken of the uncertainty in prior information. He has also discussed the problems associated with serial correlation in data series at a station as well as cross-correlation among simultaneous observations, annual flood peaks etc. at a number of sites in a region, leading to reduction in information content.

(xii) Wallis (1980) has dealt with in detail, various issues related with risk and uncertainties in the evaluation of flood events for the design of hydrologic structures, in his keynote address. Discussing about the log Pearson type III methodology recommended by the U.S. Water Resources Council, he points out problems related with regionalization of log skew and its use for ungauged sites involving geographic shift. He recommends that countries such as Italy where standard procedure does not exist, should consider distributions other than the log Pearson type III.

While commenting on possible causes for so called 'seperation effect', Wallis (1980) sugests use of Wakeby distribution which has stationary distribution with sample functions that exhibit separation. For a ranked sequence of observations (y_i , i = 1,2..n) a biased estimator, F_i, that has empirically been found to be satisfactory over a fairly wide range of distributions is

$$F_i = (i - 0.35)/n \qquad ...(13)$$

This equation is numerically fairly close to Gringorten plotting position formula. For cases of very short sample lengths or the data are considerably less kurtotic than log-normal, the author suggests regional studies of flood frequency analysis, using Wakeby distribution.

Wallis (1980) feels that the 3 parameter general extreme value distribution recommended in the British Flood Studies Report has not had extensive use outside of the U.K. and its regional application is quite specific for U.K. conditions. He suggests further studies to make the approach similarly useful to another region. An important conclusion of this report, that regionally derived flood risk estimates of the extreme quantiles are preferable to at-site estimates, is true even for long records.

2.6 TYPICAL STUDIES IN INDIA DURING 1970-80

(i) Goswami (1972) carried out regional flood frequency analysis for Brahmputra basin in North East India, using USGS procedure. He analysed annual peak flood series data for 25 sites for 1955-70 for catchment areas ranging from 63 to 69230 km². The mean annual flood $\bar{Q}$ (m^3/sec) for 2.33 year return period was graphically related with catchment area A (km^2). This could be represented by following expression:

$$\bar{Q} = 0.523\ A^{0.884} \qquad \ldots(14)$$

The study considers very large catchment of 6[illegible] km² alongwith other small catchments of the same basin which does not seem to be appropriate.

(ii) Thiruvengadachari et al (1975) carried out regional frequency analysis using USGS procedure and annual flood series data for 16 small and medium catchments ranging in size from 133 to 8500 km² in magnitude having exceedance probability of 0.43 as index flood $Q_{0.43}$ (m³/sec) which was related to catchment area A(km²) and mean annual rainfall R(cm) as follows:

$$Q_{0.43} = 0.0055\ A^{0.79}\ R^{1.11} \qquad \ldots(15)$$

The frequency curves were derived only for two ranges of index flood (i) $Q_{0.43}$ greater than 93.8m³/sec and (ii) $Q_{0.43}$ less than 357 m³/sec.

(iii) Seth and Goswami (1979) carried out regional flood frequency analysis for ten tributories of river Brahmputra in North Eastern India with available annual flood series varying in length from 11 to 25 years. Three techniques utilised in the study include:(a)using annual flood series of all stations in the region having more than 10 years of record, (b) extension of short records of some streams by developing suitable relationships with concurrent peak flood records of streams with long record (Beard, 1962), and (c) adjustment of statistical parameters obtained from short records by means of statistical parameters obtained from longer records of neighbouring stations (Beard,1962).

The mean annual floods estimated for 10 sites are tabulated alongwith information about catchment area and length of record etc.

TABLE 5

PARTICULARS OF CATCHMENT AND MEAN ANNUAL FLOODS

S.No.	Name of tributory	Length of data(years)	Stream length(km)	Mean annual flood (m^3/sec)
1.	Manas	17	290	6226
2.	Pagladiya	25	89	705
3.	Nonai	11	37	103
4.	Borolia	18	52	193
5.	Puthimari	23	110	484
6.	Dhansiri	13	70	525
7.	Pachnoi	22	48	209
8.	Belsiri	17	61	285
9.	Gabhru	15	56	272
10.	Jiabharali	13	248	5445

The relationship between mean annual flood Q_M(m^3/sec) and catchment area A (km^2) and length of stream L (km) was expressed as:

$$Q_M = 0.2752\, A^{0.251}\, L^{1.329} \qquad \ldots(16)$$

Multiple correlation coefficient R for this relationship was 0.983.

.7 TYPICAL STUDIES ABROAD DURING 1981-84

(i) Greis and Wood (1981) have investigated the use of probability weighted Moments (PWM) for improving estimates of flood recurrence quantile events in both gauged and ungauged basins. The question of network design in terms of number of sites and number of years of record is also re-examined in light of PWM techniques. Regionalisation of flood frequency quantile estimates is useful for two commonly encountered situations: (a) The length of record at any gauged site may not be of sufficient length to provide accurate parameter estimates. The PWM method could be used to obtain regionalized estimate without relying on regioinal regression. (b) Estimates of flood frequency are frequently required at ungauged sites for which there is no available record.

Quoting from Wallis (1980), the authors have given following six basic steps which are common to the regionalization process:

(a) Calculate the probability weighted moments $M_{(o)}$ and $M_{(1)}$ for each site.

(b) Standardize all moments at each site by dividing them by the mean $M_{(o)}$ for each site. After standardization, the regional $M_{(o)}$ will be identically equal to 1.

(c) Average the moments at all sites to obtain a regional $M_{(o)}$ and $M_{(1)}$.

(d) Estimate regional parameters a and m.

(e) Estimate regional quantiles Q_T (T = 10,100,200...)

(f) 1. From step (e), obtain regionalized at site quantiles for a previously gauged site, and scale the regional quantile by the mean $M_{(o)}$ for each site obtained by PWM.

(g) 2. From step(e), obtain regionalized at-site quantiles for an ungauged site, and scale the regionalized quantile by the mean obtained from other procedure(i.e regional regression).

(ii) McMohan (1982) has made an interesting comparison of hydrological stream flow characteristics between 220 Australian streams and 184 world rivers. He concludes that Australian streams show more variability in terms of annual flow volumes and larger extreme flood events than world rivers. Some typical values of parameters for peak discharges for Australian rivers (172 rivers) and world rivers excluding Australian ones (107 rivers) given by author are as follow:-

(a) mean values and ranges of the index of variability of the peak annual floods (the standard deviation of the logarithm of flows).

<u>Australian rivers</u>	0.35 (0.12 to 1.3)
<u>World rivers</u>	0.15 (0.02 to 0.36)

(b) mean values and range of coefficient of skewness in log domain.

<u>Austrialian rivers</u>	-0.58 (-3.9 to 1.4)

World rivers -0.17 (-4.4 to 6.0)

(c) Serial correlation coefficient of annual peak flood series.

Australian Rivers 0.08 (-0.27 to 0.53)

World rivers 0.13 (-0.39 to 0.58)

The study provides a good idea of methodology for comparing hydrologic streamflow characteristics on regional or national scale.

(iii) Hebson and Wood (1982) derived flood frequency distribution from assumed climatic distribution and the geomorphologic unit hydrograph (GUH) model of catchment response. The authors suggested use of this derived frequency distribution for studying the effect of catchment shape on flood frequency characteristics, and the question of catchment homogeneity and similarity necessary for region flood estimation.

(iv) Zhang (1982) derived a generalized plotting position formula for estimation of the recurrence interval of hydrologic events in an incomplete data sample containing both extraordinarily, large and small historical values. The derivation is based on order statistics and is independent of the probability distribution of the hydrologic variable. It was also shown that Weibull formula is a special case of the generalized formula. The methodology for a discontinuous sample has been explained by the author using an illustrative example.

(v) Kuczera (1983) used power normal (PN) distribution based on Box-Cox transformation for studying the problem of combining site and regional flood data to infer the shape parameter of a flood distribution. It is similar to log Pearson type III distribution in the sense that the parameters μ and σ are moments of the transformed floods and the power transform λ is akin to the LP log skew parameter. A method based on Bayesian concepts is developed in which weights assigned to site and regionalized estimates of λ reflect the relative information content of in site and regional flood data. Two methods of inference were considered to infer from the regional flood data, the mean and variance of the super-population making allowance for unequal site sampling variance and spatial correlation It was seen that by Bayesian approach reliable estimates of the super-population parameters can be obtained with the sensitivity to outliers largely removed and negative variances no longer a possibility. Moreover, the Bayesion approach naturally compensates for the distortive effects of unequal site sampling variance of λ and spatial correlation. It also readily enables full allowance to be made for uncertaity in the PN parameters when inferring the T-year or design floods.

(vi) Greis (1983) made a detailed review of activities in the area of flood frequency analysis during 1979-1982 covering the period following endorsement of the log Pearson type III methodology by the U.S. Water Resources Council in 1977. Three main areas, representing the most exciting and fertile areas for the next quadrennium, were reviewed. These are: (i) distributional analysis, (2) robust parameter estimation, and (3) developments of potential in the area of regionalized and geomorphologic approaches to the flood frequency problem.

Discussing the underlying assumptions of WRC procedure, Greis mentioned about use of parametric statistical inference techniques with assumption that peak annual streamflow data comes from a parent population whose distribution function is known, is analytically expressable, and contains a finite number of parameters. Though a large number of peak, flow distributions have been studied, however, the lack of power of conventional goodness -of-fit tests with respect to typically skewed flood peak distributions contributed to considerable variability in the estimation of design events. In order to deal with the problem of lack of long sample records to give a good estimate of skew, the WRC outlined procedures for obtaining a weighted combination of regional and sample skew estimates. As an alternative to the combined use of original flood data and watershed characteristics, the WRC offered a national regional skew map. It was seen that a large area of the WRC log skew map has negative values which assuming an LP III distribution, implies a finite upper bound, or maximum certain flood, and an infinite negative flood, neither of which is physically meaningful.

The large variability associated with higher moments such as sample skew coefficients estimated from small samples, has important consequences when using these parameters (skew, kurtosis etc.) to choose a model from among a class of distributions. The WRC recommendation of assigning arbitrary weights to sample and regional skew estimates, irrespective of their information content, led to several comparative studies and suggestions for improved skew estimation based on LP III methodology.

The overall usefulness of regional skew maps in log space was examined by Landwehr et al (1978). It was shown that distinctly different skew contours in real space could give rise to identical contours in log space and vice versa; and the problems associated with transforming flow data in real space into log space for flood analysis were highlighted. Greis (1983) suggests use of non parametric methods like the Jacknife and bootstrap methods to estimate the variance of sample skew, on the basis of their superior performance in comparative studies.

Greis (1983) has also highlighed the problems of parameter estimation using method of moments method of maximum likelihood etc.leading to bias and large sampling variations in the estimates of the parameters. The method of mixed moments was introduced to circumvent skew estimation problems. This involves use of the mean and variance of the real data and the mean of the logarithmic data to fit the LP III distribution. Since primary inferences about flood events are dependent upon tail behaviour, the effect of distributional assumptions on tail behaviour were examined with respect to (a) estimated return periods, (b) upper bound, and (c) existence of moments. The distribution which has aroused the maximum interest is the five parameter wakeby distribution with its right hand (flood) and left hand (drought) tails have the most pull towards one distribution or other. The robust statistical estimation tries to "cleanse" the effect of outliers.

The use of regional parameters to estimate design events is an important step for robust estimation. Certain kinds of regionalization involving the use of regional type of information of the combination of regional and site specific information, and also use of regression techniques are being adopted in hydrological studies since a long time; Recently, a new statistical estimation procedure based on the calculation of probability weighted moments (PWM) has been introduced for the estimation of regionalized parameters of flood frequency. A method has been proposed for combining regional and site specific information by simple averaging and scaling. Moments of the regional distribution are obtained by simple averaging of the respective probability weighted moments at each site. It was found to work well for situations where records are extremely short and streamflow histories are highly skewed and kurtoic.

A number of other studies have appeared which apply some sort of regionalisation which can be modeled separately.

Greis(1983) has discussed the problem of 'robustness' in flood estimation. Robust estimates are those estimates which are relatively insensitive or robust to deviations from ideal assumptions. Good parameter estimates are not always sufficient to provide stable estimates of flood recurrence intervals. These will be of less use, if small perturbations in the parameter, no matter how well estimated, produce large deviations in the probability density function. It has been observed that density functions which contain exponential terms are particularly vulnerable to wide fluctuations in extreme events estimated from the tails, and thus, to distributional assumptions. Statistically robust estimators seek to provide the assurance that T-year recurrence estimate under a particular distributional assumption does not vary too much from corresponding estimate for a true distribution.

Typical streamflow records of annual events usually contain only one or two flood events which occur at highest recurrence intervals, and these events usually contain considerable measurement error. These events or 'outliers' usually have large influence on the results of parameter estimation, and also provide to flood frequency analysis, on the lines similar to estimation procedures which were first introduced by James and Stein in 1961. Regional techniques such as James Stein, probability weighted moments etc. have been demonstrated to improve estimates at gauged sites. They can also be useful for combining information from gauged sites within a region and using it to infer flood frequencies at ungauged site. The regression approach provides a useful, but somewhat limited methodology for the transfer of information from gauged to ungauged sites.

In regionalizing streamflow data, the main problem is to specify the region. Greis (1983) mentions about the Geomorphologic Unit Hydrograph (GUH), which is a synthesis of the hydrologic impulse response of a catchment to surface runoff obtained by linking the time varying instantaneous unit hydrograph (IUH) to the geomorphologic parameters of a basin. It is suggested that two basins may be defined to be hydrologically similar, if they have same time invariant IUH when kinematic conditions are kept the same. Methods have also been developed for determining the probability distribution of floods using historical flood data from such disparate sources as photographs, public records, written historical records, tree ring analysis, radio carbon, dating of organics in alluvium deposits etc.

(vii) Landwehr et al (1984) have examined the applicability of the James Stein estimator in regional hydrologic analysis. Regional studies are frequently characterised by relatively short, generally correlated, samples drawn from nonsymmetric and bounded, i.e non-normal populations. The authors considered Webull distribution with coefficient of skewness ranging from 0 to 10 for computer simulation studies and the James Stein estimator with some modifications was shown to be robust. It was also seen that the relative advantage of this method in comparison to other traditional methods is tempered by the cross correlation among the variables and the degree of heterogeneity within the collection that is deemed to be the pertinent hydrologic region.

2.8 TYPICAL STUDIES IN INDIA DURING 1981-84

(i) Jakhade et al (1984) have applied regional flood frequency approach of USGS to analyse the data of fourteen sites (having data for 10 or more years) in the Brahmputra valley divided into two different hydrometeorological zones, as given below:-

TABLE 6

DETAILS OF CATCHMENTS STUDIED

Group(A) Rivers: $Y = 16.6605\, X^{0.5783}$, $r = 0.9765$

$RQ_{10} = 1.5228$, $RQ_{100} = 2.2357$, $RQ_{1000} = 2.9457$

S.No.	Site and River	Catchment area(sq.km)	Length of record(yrs)
1.	Mathanguri-Manos	29424	31
2.	NT Road-Puthimari	940	10
3.	NT Road-Jiabhorali	14197	22
4.	NT Road-Borgong	409	16
5.	Chow Pdhawa Ghat-Subansiri	27000	19
6.	Passighat-Dihang	249586	17
7.	Khowang-BurhiDihang	4923	18

Group(B) Rivers: $Y = 1.1080\, X^{0.7968}$, $r = 0.9628$

$RQ_{10} = 1.389$, $RQ_{100} = 1.9225$, $RQ_{1000} = 2.4409$

S.No.	Site and River	Catchment area(sq.km)	Length of record(yrs)
1.	NH Crossing-Disang	3549	16
2.	NH Crossing-Dikhow	3610	17
3.	NH Crossing-Bhogdi	600	20
4.	Numliagarh-Dhansiri	10242	16
5.	Bharamtal-Kopili	14537	21
6.	Pandu-Brahmputra	424130	27
7.	Srirampur-Sankosh	9920	19

Note:- X = Catchment area in sq.km., Y = mean annual flood in m^3/sec and r = correlation coefficient. RQ_{10}, RQ_{100}, and RQ_{1000} are respectively the mean ratios for 10, 100, 1000 years return period flood to mean annual flood, for the region.

The authors have mentioned about the drawbacks of the study due to inadequate data and data extrapolation; and have recommended that the classification into two groups and corresponding regional flood frequency curves should be used for preliminary planning purposes only for the sites within the Assam valley.

2.9 TYPICAL STUDIES ABROAD DURING 1985-87.

2.9.1 General Matters

(i) Thomas (1985) gives a brief historical review of the development of bulletins of U.S. Water Resources Council and highlights the motivation and justification for the adoption of uniform technique of fitting log Pearson type III distribution to annual peak discharges. Specific techniques for development of regional skew, weighting of regional and station skew, the basis for low and high outlier tests and the basis for the adjustment of frequency curves using historical information are examined by the author.

The author mentions about suggestion of Kuczera regarding use of Box-Cox Power transformation to normalize the annual peak discharges, and the Wakeby distribution described by Houghton. He recommends further evaluation of the parameters of the Wakeby distribution for providing uniform and accurate estimates of flood frequency. Other useful developments cited by author include (a) suggestion of Tung and Mays for weighting of sample mean, standard deviation and skew regionalized estimates of the same parameters to reduce the parameters uncertainty using non-parametric methods like the Jackknife and Boot strap methods, (b) suggestion of Rao for using the method of mixed moments to estimate the parameters of the log Pearson type III distribution. (c) use of maximum like'ihood estimates for fitting Pearson type III distribution, and (d) use of the concept of the probability weighted moments for parameter estimation etc.

He also lists some important areas/topics that need further study as follows:-

1. Evaluation of techniques for testing the homogeneity of the annual peak series: These techniques are needed to determine homogeneous periods of record for watersheds undergoing land use or channel changes. An example of using parametric and non-parametric tests is given by Hirsh and others.
2. Development of procedures to analyze a nonstationary annual peak series if in fact the watershed is undergoing some channel or land use change.
3. Identification and treatment of mixed distributions resulting from two or more types of hydrologic events such as flooding from snowmelt versus thunderstorm activity. Russel and Singh have seperately suggested that two separate distributions be

combined to better fit the distribution of annual peak discharges.

4. Treatment of outliers both as to identification and computational procedures.

5. Alternative procedures for treating historic data. Leese, Condie and Lee, and Cohn and Stedinger suggested censoring theory and maximum likelihood estimates as ways of more efficiently utilizing historic informatioin. The computational disadvantages of the censoring theory/maximum likelihood estimates must be indicated against any improvement in accuracy.

6. Procedures to incorporate flood estimates based on precipitation or data available at nearby watersheds, or both, (for example a regional estimation procedure) into the frequency analysis at a given station.

(ii) Cunnane (1985) discusses various issues which lead to markedly different magnitude-return period (Q-T) relationships when various distributions are used for modelling the flood series. He has discussed in detail a priori theoretical arguments, empirical criteria for distribution choice and methodology for discrimination between candidate distributions. The effects which methods of parameter estimation, treatment of outliers, inclusion of large historical flood values, data transformations and causal compositions of the flood population may have on choice of distribution are also considered. In the conclusion, he emphasizes that distribution choice cannot be based on theoretical arguments alone. Annual maximum floods are the maxima of some unknown number of within-year flood events of unknown, perhaps nonhomogeneous distribution. Observed annual maximum flood series data display positive skewness and large kurtosis, as well as high variability of skewness in comparison to that between similar sized random samples from all previously used flood distributions except Wakeby. Cunnane(1985) recommends that inference about distribution type must be based on many sets of flood data simultaneously. He points out relative insensitivity of traditional goodness of fit tests to distribution tail behaviour. He recommends further research on high outliers, and cautions against use of data transformations and mixtures of distributions without physical justification.

(iii) Gottschalk (1985) describes application of pairwise grouping as well as principal components to Swedish river runoff data in order to analyse the patterns of variation and to delimit regions where hydrological behaviour may be regarded as uniform. The variability in mean values, coefficients of variation of annual/ monthly runoff volumes as well as of annual floods/low flows were analysed. The author suggests pairwise grouping as the appropriate method to use on a national scale with heterogeneous hydrological regimes. Principal components demand homogeneity but within such regions describe very well the trends and variations. Further systematic scientific studies are recommended to draw conclusions about the way physiographical and hydrological model regions are related to each other.

(iv) Adamowski (1985) carried out a simulation study to compare parametric procedures using lognormal with three parameters (LN3), log-Pearson III (LP III), generalized extreme value (GEV) and Wakeby with the non-parametric kernel estimation procedure of flood frequencies. For simulation purposes, the LP III distribution was selected as a parent distribution because it fits a wide range of observations and is mandatory in the USA. The errors between estimates obtained by different methods and the population values were computed using the bias and the root mean square error for each event. The simulation results demonstrated closeness of rms error for the non-parametric method and the LP III distribution (the parent used in simulation). The bias of the non-parametric method was small, sometimes even smaller than that of LP III. The experiment,though limited,suggests viability of non-parametric method as a possible alternative for flood frequency analysis.

(v) Tasker and Stedinger (1986) describe a statistical procedure for estimating generalized skew co-efficients used in WRC guidelines for flood frequency estimation. The method is based on weighted least squares regression in which the weights are determined by seperating residual variance into that due to model error and that due to sampling error. When compared to ordinary least squares regression, this method provides more accurate estimates of the skew model's parameters; and more realistic and accurate estimates of the variance of prediction at a site.

(vi) Linsley (1986) mentions about large errors and biases towards overestimates in the procedures and assumptions in common use for estimating flood peak frequency for ungauged streams. Though, various journals in the field of hydrology are full of papers presenting complex methods, while in real world of engineering rational equation suggested by Mulvaney in 1851, unit hydrograph approach of Sherman given in 1932 and regional flood frequency approach of Dalrymple (1950) are widely used. The regional frequency method leaves the volume and shape of the hydrograph undetermined. The regression equation used in regional frequncy procedures have standard errors generally 20% or larger. When logarithmic regressions are used the standard errors for longer return periods may be in the order of 100% or more. Considering that procedures for estimating flood flows which have average errors between 10 and 30%, will produce many estimates with much larger errors and most likely will have a bias as well, making any economic analysis of floods quite unreliable, the author recommends various steps including effective testing of hydrologic methods.

(vii) Arnell et al (1986) emphasised the need for using unbiased plotting positions depending upon the particular distribution and suggested procedures for determining unbiased plotting positions for the general extreme value distributioin. For simpler level of calculation the following expression is suggested:

$$F_r = \frac{r - \alpha}{n - \beta} \qquad \ldots(17)$$

This compares with general formula given by Cunnane (1978) as follows:

$$F_r = \frac{r - \alpha}{n - 2\alpha + 1} \qquad \ldots(18)$$

The parameters α and β are allowed to vary with sample size and the shape parameter of the general extreme value distribution. Here F_r = non-exceedance probability of a flood event, n = sample size and r is the ranking in an ascending order series, r = 1 for largest value in the sample.

(viii) Hirsch and Stedinger (1987) examined the situations when plotting positions are needed where in addition to a systematically recorded annual flood series, one would have a record of any large floods which occured during an extended historical period, if they occured. For such situations, authors state that it is impossible to construct highly accurate estimates of the exceedence probabilities of the largest floods using only their rank, the number of observed historical floods, and the lengths of the historical period and the systematic record.

2.9.2 Use of Historical Information

(i) Hosking and Wallis (1986) used computer simulation to assess the possible increase in accuracy of estimates of extreme floods, when a single paleoflood estimate is included in a single site or regional flood frequency analysis procedure. It was concluded that the use of a paleological maximum event is likely to be worthwhile only when estimating a three parameter flood frequency distribution from a short gauged record from a single site or a small number of homogeneous sites in a regional study. The authors further state that the accumulation of as large a number as possible of more or less homogeneous and independent gauged flood records, followed by a regional analysis using an efficient and robust algorithm such as three parameter GEV/PWM (general extreme value/probability weighted moments) will increase the accuracy of flood estimates to a level which cannot be significantly improved by the inclusion of a small amount of paleological information. The conclusions are subject to assumptions of the study, particularly regarding use of only one estimate of paleological maximum event.

(ii) Stedinger and Cohn(1986) have examined methods for utilizing cultural records of large floods, and other historical and paleoflood information in flood frequency analysis. Paleoflood hydrology is defined as the study of movements of water and sediment in channels before the time of continuous hydrological records or direct measurements. Geological and stratigraphic evidence of catastrophic flood which changed a basin's characteristics requires a somewhat different and more sophisticated statistical framework, because such extraordinary events distort or obliterate physical evidence left by earlier floods and change stage discharge relationship of the flood channel. The authors found maximum likelihood methods flexible, efficient and relatively robust on the basis of a series of Monte Carlo experiments.

2.9.3 Regional Homogeneity

(i) Wiltshire (1985) presented a new method of grouping basins prior to regional flood frequency analysis. This is based on measured basin characteristics such as basin area, average annual rainfall, soil index etc. rather than on geographical regions. Statistical procedures for identifying efficient groupings are described.

(ii) Lettenmaier and Potter (1985) have reviewed different approaches for regional flood estimation also termed as regionalization. These included:

a. the most common method of regionalization which utilizes normalized annual flood series, i .e flood series that have been standardized by one or more at site statistics. The class of index flood methods results when the annual flood series are normalized by the at-site means. For these methods, it is usually assumed that the distribution of the standardized annual flood is identical at all sites in the region. Stedinger had suggested standardization in log space by subtracting the sample mean of the logarithms of the annual flood discharges. Rossi et.al regionalized annual flood series that were assumed to follow a two component extreme value distribution using the at-site estimates of the two parameters of the extreme value distribution.

b. the empirical Bayes method which assumes that the parameters governing the distribution of annual floods at a site in a region come from a specified super population, which has unknown parameters. The parameters are inferred from the flood data themselves,or from relationships which are estimated from flood peak characteristics, and physiographic and climatic factors. An empirical Bayes estimator of an at-site flood characteristics pools information regarding the super population with at-site information. Kuczera evaluated several empirical Bayes estimators using Monte Carlo method. In one study,the five parent distributions were lognormal, while in other study the parents were four standardized Wakeby distributions.

Lettenmaier and Potter (1985) proposed a regional flood generation model, wherein the logarithms of the first two moments of the annual flood distribution at each site depend on the logarithms of the drainage area. This model is parameterized in terms of the regional mean coefficient of variation and the site-to-site coefficient of variation of the coefficient of variation. Eight flood frequency estimation methods, including two at-site viz.(i) maximum likelihood estimation of the parameters of an Ev-1 distribution and (ii) two parameter lognormal distribution, and six regional estimators were tested. The regional estimators included: (i) probability weighted moments with extreme value type 1, (ii) linear empirical Bayes with lognormal, (iii) first modified linear empirical Bayes with lognormal (iv) seasonal modified linear

Bayes with lognormal,(v) linear empirical Bayes with extreme value type 1, and (vi) modified linear empirical Bayes with extreme value type-1.

The Monte Carlo experiments consisted of generation of independent flood sequences of length n at each of the m sites. The performance was judged by identifying the estimator with the lowest root mean square error (rmse) for the various generators as a function of record length n, number of site m and return period T.

The results of study indicated that the regional estimators nearly always had lower (rmse) than that for at-site estimators, suggesting that the better regional estimators were able to extract additional information from the multiple sites, even when the sites were moderately hetrogeneous with respect to the coefficient of variation, except for EV-1 parent for which the at-site EV-1 estimator sometimes had lower rmse. The regional estimators which were evaluated are not always robust and the worst robustness was encountered for LN 2 based estimators. The PWM was the best estimation technique when both regional mean coefficient of variation and site to site coefficient of variation of the coefficient of variation were low. But for high values of these coefficients, the performance of PWM was often worse than that of an at-site estimator.

The authors suggest that for such situations more attention should paid to reducing the variability of estimates of the at-site means and less to estimation of the regional (normalized) flood frequency distribution.

(iii) Stedinger and Tasker (1985) compared the performance of ordinary, weighted and generalised least squares estimators of the parameters of regional hydrological relationships in situations where the avaialable streamflow records(such as peak floods etc.) can be of different and widely varying lengths and concurrent flows at different sites are cross correlated using a Monte Carlo study. The generalised least squares (GLS) procedure provided (a) more accurate parameter estimates, (b) better estimates of the accuracy with which the regression model's parameters were being estimated, and (c) almost unbiased estimates of the model error. A weighted least squares procedure which neglects the cross correlations among concurrent flows does as well as the GLS procedure when the cross correlations among concurrent flows is relatively modest (less than 0.6). The ordinary least squares (OLS) approach however, provides very distorted estimates of the regression model's parameters. Furthermore, it was also seen that the accuracy of regional regression models can be improved by incorporating into the analysis sites with as little as 3 years of record, using GLS approach.

(iv) Wiltshire(1986) mentions about deficiencies in the test procedure for regional homogeneity suggested by Dalrymple (1960). Regions of great sizes are expected to be heterogeneous in terms of station skewness.Some attempts to identify flood regions by methods other than geographical delimitation have been based on statistics of the basin flood frequency distribution or basin characteristics. The author recommends use of both flood information as well as basin characteristics. Two approaches are suggested, viz. (a) based on regional variability of the site coefficient of variation (b) based on a-priori selection of parent distribution. The preferred test is based on the distribution function of the regional parent and its power is shown to depend on region size, record length and choice of parent distribution.

Commenting on various limitations of multivariate techniques like cluster and factor analysis, Wiltshire (1986) presented a procedure based on iterative search through the basin characteristic data-base for classifying basins into distinct homogeneous groups for regional frequency analysis. The procedure optimises statistics that describe the efficiency of grouping and its applicatioin to basins in Britain yielded five groups formed on the basis of basin area,average rainfall and urban fraction. The author also states two weaknesses of the suggested procedure, viz characterization of annual maximum series of each site by the coefficient of variation of the series and non-uniqueness of the resulting solution in terms of basin groupings. The conventional procedure of adopting geographically defined regions enclosed by political, administrative or physiographic boundries, needs to be replaced by a more logical approach for defining the homogeneous region in terms of the basin characteristics which are expected to control flood frequency. The author discussed significant advantages of the proposed method and suggested possible inclusion of statistics other than coefficient of variation for further improvement.

Lettenmaier et.al. (1987) studied effect of regional hetrogeneity on flood frequency estimation using Monte Carlo experiments with the following objectives:

a. to explore the robustness of selected regional flood methods with respect to the assumed form of the regional flood distribution.
b. to explore the robustness of selected regional flood methods with respect to regional hetrogeneity in the at-site flood distributions,
c. to explore the sensitivity of selected regional flood methods to record length, and
d. to explore the performance of regional flood methods that provide for site to site variations in moments higher than the first order.

The results of the experiments showed that (a) two parameter regional flood frequency estimation methods can perform quite well if the assumed distributional form is similar to the population, but that large biases may result if the distributional form is misspecified, (b) the three parameter GEV distribution tends to give extremely variable flood quantile estimates for at site applications. However, when incorporated in a regional estimation procedure,it performs nearly as well as the two parameter estimates,(c)for high values of the regional mean coefficient of variation, the advantage of methods that assume regional homogeneity in moments above the first order is reduced. At very high values of regional mean coefficient of variation, such as might be encountered in arid regions, alternate estimators that accommodate the regional hetrogeneity in higher moments are preferred, and (d) within the range of record lengths that are commonly available, record length dependence of the index flood PWM estimators is likely to be less important than geographic differences in regional mean coefficient of variation.

(vi) Wood and Hebson (1986) have investigated the interrelationships between rainfall and basin characteristics, and flood frequency characteristics, by deriving a scale independent, dimensionless flood frequency curve. Adjacent catchments, subject to apparantly similar climates, can have quite different flood frequency distribution characteristics. This could be due to short records, sampling variations or differences in basin response. Many previous studies had used point rainfall statistics to represent lumped input statistics and involved a large number of parameters. The authors used dimensionless rainfall inputs and a dimensionless basin response function in the proposed model. For the rainfall intensity the probability distribution for dimensionless areal rainfall is derived and depnds upon a geoclimatic scaling factor which is a function of the basin size and the rainfall areal correlation structure. Rainfall duration is scaled by a characteristic basin time, and this same parameter is used to develop a dimensionless geomorphologic unit hydrograph. The flood frequency distribution derived based upon this approach, is independent of spatial and temporal scales. For hydrologic similarity, two basins will have the same dimensionless frequency curve. Further theoretical studies are aimed at to investigate a series of relevant questions, including possibility of use of this approach to group stations for regional flood analysis. Preliminary results have indicated that one rainfall distribution may govern the flood frequency curve at low return periods while another rainfall distribution govern the flood frequncy at long return periods.

2.9.4 Typical Regional Studies

(i) Acreman (1985) described the procedure for development of multiple regression relation for predicting mean annual flood from physiographic and climatological characteristics. Flood data for 168 Scottish basins containing 3071 station years of record were used and following relationship was developed:

$$\bar{Q} = 0.00173\,(AREA)^{0.843}\,(SAAR)^{1.085}\,(STMFRQ)^{0.187}\,(SOIL)^{1.750}\,(100.\,LOCH + 1)^{-0.437} \qquad \ldots(19)$$

Where AREA = basin area(km^2)

SAAR = standard (1916-1950) average annual rainfall (mm).

STMFRQ= Stream frequency (junction/km^2)

SOIL = soil type indix

LOCH = fraction of the basin covered by lakes or reservoirs.

Q = mean of annual maximum flood series(m^3/sec)

The equation has a coefficient of determination of 0.914 and standard error of the estimate of the predicted mean annual flood is 0.147.

(ii) Fiorentiono et al (1985) describe the use of two component extreme value (TCEV) distribution for modelling Italian annual flood series to account both for the presence of flow outlier and for the high variability of the skewness of historical series. This distribution treats the annual flood as the maximum of a Poissonian number of flood peaks, generated by a mixture of two exponential distributions. Because of the data limitations for individual sites, the four parameters could not be evaluated with desired level of accuracy and reliability. The authors recommend use of regional studies for this purpose. The application of this distribution for regional studies using annual flood series for 39 sites in Italy has been discussed by the authors.

(iii) Hosking et al (1985) made a critical appraisal of the regional flood frequency procedure in the UK Flood Studies Report (FSR). The algorithm for estimating the regional flood frequency hazard contained in the 1975 Natural Environment Research Council flood Studies Report (FSR) can occasionally lead to upper quantile estimates that appear unrealistic when compared with engineering judgement. Tests with the FSR algorithm were made for several sets of observed flood sequences and a great variety of synthetic data in a Monte Carlo simulation study. Similar tests were conducted with many other regional and at-site flood frequency estimation procedures including a regional generalized extreme value distribution (GEV) procedure and a regional Wakeby distribution (WAK) procedure, both of which used biased probability weighted moments (PWM) in their formulation. For the Monte Carlo simulations, for which the true quantiles to be estimated were known, it was found that the FSR algorithm yielded quantile estimates that were always more variable, often by a factor of as much as 4 or 5, than those obtained by either the GEV/PWM or WAK/PWM algorithms. Further, the FSR algorithm was found to yield quantile estimates that were nearly always more biased than those obtainable by either the GEV/PWM or the WAK/PWM procedures.

The authors summarised their important findings as follows:

(a) The use of the FSR algorithm with small data bases, short records, or in regions of high CV, or in the presence of "outlier" in the flood sequences, leads to results that are relatively unreliable when compared with those obtainable by other methods.

(b) If data bases are heterogeneous, the FSR growth curves comprise a composite that gives more influence to those sites that have shorter records and higher CVs- the very site where the available information is the least reliable.

(c) It is suggested that as the available data base for the UK can now be expected to be about twice that which was available at the time of the Flood Studies Report , a complete re-analysis with the updated data base should be done. (At present a complete and consistent data base for UK flood data is being prepared).

(d) It is recommended that GEV/PWM algorithm could well be used as a substitute for the FSR algorithm. However, there is a minor caveat that if the regional

Wakeby distribution resulting from WAM/PWM regionalization with the same data base has a skew/kurtosis ratio that departs markedly from that of a GEV distribution, then the assumption that the flood data follow a GEV distribution might well be called into question, and the adequacy of using the GEV/PWM algorithm investigated.

(e) Quantile estimates made by index flood type algorithms, such as the GEV/PWM and WAK/PWM, would benefit greatly from research into better ways of estimating the at-site component of the regional estimates. This is, of course, necessary for ungauged sites, but it would also be most helpful even for gauged sites with long records.

(f) Using sets of catchment that are hydrologically homogeneous with regard to the CVs of their flood sequence, can be expected to result in nearly optimal estimates of flood quantiles. Research into the methodology for defining sets of catchments with equal CV's is needed, and can be expected to pay relatively big dividends in the form of improved flow quantile estimates.

(iv) Fiorentino et al (1987) propose a hierachical approach for regional frequency estimation keeping in view the observation that the hydrologic homogeneity with respect to the shape parameters of the annual flood distribution can be hypothesized in large regions, while as regards the coefficient of variation, the homogeneity is evident only in smaller areas. It is characterized by three investigation levels: (a) identification of homogeneous regions wherein the skewness coefficient is assumed not to vary from site to site, (b) identification of homogeneous subregions wherein the coefficient of variation is assumed to be constant too, and (c) use of regression models to relate a location parameter of the flood distribution to climatic and physiographic factors. The approach is presented with the use of the two component extreme value (TCEV) distribution and validity demonstrated using 28 Italian annual flood series. The authors also show that local disturbance factors can heavily affect the extreme values of flood resulting in anomalous shapes of annual flood series, which are independent of the general characteristics of the basin.

(v) Farquhar-son et al (1987) collected annual maximum flood data from 1121 gauging stations in 70 different countries, comprising 31000 stations years from various sources and analysed in a uniform manner. For each station the general extreme value distribution was fitted to annual flood series using PWM method. The derived station flood frequency curves were made dimensionless by dividing the annual series by the mean annual flood. Where data were sufficient, the station growth curves for each country or region were combined to produce a flood frequency curves intended to be representative of floods in that region. The authors anticipate that the presented results could be used for planning and pre-feasibility and possibly for design of small schemes and small structures.

(iv) Baker et al (1987) mention about the role of ancient flood records in China after careful interpretation in terms of magnitude and frequency, in formulating design decisions for high risk projects. They introduce the multi disciplinary complexity of paleoflood hydrologic studies involving expertise in geology and geomorphology as well as in hydrology and hydraulics. Physical evidence of large paleofloods provides objective evidence of the likelihood and frequency of larger floods than can be possible through historic flood records. The authors emphasize the need to suitably incorporate paleoflood information in the overall decision process leading to design.

2.10 TYPICAL STUDIES IN INDIA DURING 1985-1987.

These studies broadly fall under three categories:

(A) dealing with general matters

(B) dealing with application of conventional techniques

(C) dealing with application of new techniques

Salient features of these studies are being reviewed in following sections.

2.10.1 General Matters:

(i) Seth and Goel (1985) discussed the general issues of regional frequency analysis and mentioned about the problems due to non-availability of satisfactory methodology for regionalizing flood parameters and data limitations. For a country like India with great variety of hydrometeorological and hydrological conditions, the authors recommended comprehensive studies to develope suitable procedures and guidelines for regional flood estimation, so as to ensure consistent flood estimation within each region.

(ii) Perumal and Seth (1985) have described various distributions used for fitting the peak flood series. The inference regarding the best fit distribution for peak flood series at a particular site is influenced by a-priori assumption regarding the peak flood data, method of parameter estimation, goodness of fit tests, sampling variability, presence of outliers etc. Different hydrologists would arrive at different flood estimates depending on adopted methodology. The authors recommended adoption of uniform procedures for the country as a whole or for different regions.

(iii) Goel and Seth (1985) have discussed various data related problems in flood frequency analysis viz. broken record, incomplete record, zero flood years, presence of outliers etc. The causes of these problems and possible rectification are also discussed.

(iv) Rao and Goel (1986) have discussed various assumptions and limitations of flood frequency analysis. They consider EV-1 distribution as appropriate for annual flood peak series for small Indian catchments, and log-normal and normal respectively for large and very large catchments. This is based on assumption of flood from tributaries to be EV 1 distributed and combining to form peak flood in large basin exceeding over 10000 sq.km. area. In case of very large drainage basins of over 200000 sq. km. with many major tributories, like the Brahmputra at Pandu or the Ganga at Farakka, authors suggest normal distribution as most appropriate. They mentioned that the central limit theorem establishes the asymptotic normality of the sum of random variables, independently to some extent of the nature of distributions of the variabiles.

While discussing regional flood frequency analysis, Rao and Goel (1986) mention about decrease in the values of coefficient of variation of annual peak flood series with increase in the area of the catchment. It is small for large catchments as they represent highly damped systems. The authors recommend that the area range of the catchments chosen for regional analysis has to be necessarily restricted in order to ensure the absence of significant difference in coefficients of variations.

2.10.2 Conventional Techniques

(i) Huq (1985) carried out regional flood frequency studies of Cauvery basin subzone 3(i) using multiple regression technique for developing relationship between frequency flood Q as dependent variable and physiographic parameters like catchment area A(km^2), slope of main stream S (m/km), percent of surface storage area of tanks plus

0.5 percent, S_t , length of main stream L (km), shape factor of the catchment S_h = (L^2/A), 24 hour rainfall in cm of recurrence interval T years I_T as independent variables. The data of 23 catchments ranging in size from 66 to 10619 km² was used for analysis using Gumbel distribution. It was seen that relationship involving A, S and S_t , provided better and reliable estimates. Some typical relationships are as follows:

(a) <u>Return period T = 2 years</u>

$$Q_2 = 0.628\ (A)^{0.902} \quad , r = 0.8844 \qquad \ldots(20)$$

$$Q_2 = 0.38(A)^{0.98}\ (S)^{0.007}(S_t)^{-0.52}, R=0.9238 \qquad \ldots(21)$$

(b) <u>Return period T = 100 years</u>

$$Q_{100} = 5.88(A)^{0.757} \quad , r = 0.8624 \qquad \ldots(22)$$

$$Q_{100} = 1.17(A)^{0.96}(S)^{0.32}(S_t)^{-0.42}, R=0.8925 \qquad \ldots(23)$$

Where r is simple and R is multiple correlation coefficent.

(ii) Venkataraman and Gupta(1986) used USGS method for deriving relationship between 50 year return period flood Q_{50} (m³/sec) and catchment area A(km²) for small and medium catchments in sub-Himalayan region. They used data from 11 railway bridge sites with catchment areas ranging from 6 to 2072 km². The following relationship was obtained:

$$Q_{50} = 15.84\ (A)^{0.75} \qquad \ldots(24)$$

They also developed a regioinal dimensionless frequency curve by plotting median values of ratios of floods of 50,25,10 and 5 years return period to mean annual flood ($Q_{2.33}$). The relationship for mean annual flood $Q_{2.33}$ (m³/sec) with catchment area was obtained as follows:

$$Q_{2.33} = 5.89 \quad A^{0.75} \qquad \text{...(25)}$$

The authors indicated various limitations of the study due to inadequate samples and length of data,filling up of missing data, extension and extrapolation, and the stability of stage discharge curves.

(iii) Venkataraman et al (1986) presented a regional flood study for subzone 3(f) in lower Godavari basin using 26 years data of 14 small and medium railway bridge catchments with areas ranging from 15 to 824 km². At these sites, the daily peak flood stages are observed during monsoon and are converted to discharge values using rating curves. The rating curves are based on regular gauging observations of both stage and discharge Before proceeding for analysis, the length of peak flood data for all 14 sites was brought to common length using some procedure not stated by authors in their paper. They examined choice between (i) general extreme value (GEV) and Gumbel EV1 distribution, (ii) regional pooling of data vis-a-vis averaging of ratios of floods for particular return periods to mean annual flood obtained at individual sites, and (iii) using parameters like catchment area $A(km^2)$, length of main stream (L) statistical slope S, 50 years 24 hour point rainfall R_{24} (cm) and 2 year one hour point rainfall R_1 (cm) for developing regression relations with mean annual flood Q taken as $Q_{2.33}$(m^3/sec), for important hydraulic structures. The authors concluded that for subzone 3(f), the mean annual flood can be expressed satisfactorily as a function of A and R_1 or only A, as follows:

$$Q_{2.33} = 1.23\ A^{0.74}\ R_1^{-0.98} \qquad ...(26)$$

$$Q_{2.33} = 7.10\ A^{0.755} \qquad ...(27)$$

It was further concluded that 50 years return period flood could be obtained from graphical approach as per USGS method. For less important structures,where economic considerations may decide the choice, the authors suggested use of GEV growth curve.

iv) Thirumalai and Sinha (1986) carried out regional frequency analysis for 14 small and medium railway bridge catchments in Krishna Basin, having catchment areas ranging from 30 to 730 km^2., and flood series from 6 to 26 years. Mean annual flood $\bar{Q}$ (m^3/sec) was related with catchment area $A(km^2)$ and 2 year 24 hour point rainfall R_{24}(cm) as follows:

$$\bar{Q} = 3.35\ A^{0.682} \qquad ...(28)$$

$$\bar{Q} = 1.009\ A^{0.664}\ R_{24}^{0.64} \qquad ...(29)$$

They used Chow's approach modified by Nash in the analysis.

(v) Mehta and Sharma (1986) have presented results of flood frequency analysis of annual peak series for four sites on river Satluj and Beas in Himachal Pradesh using normal, gamma, log normal type-3, Pearson type-3 and log Pearson type-3 distributions. Method of moments has been used to evaluate parameters of these distributions and to estimate floods of 10,25,50,75,100,200 and 500 year return periods.Sample statistics on all the four sites considered on these rivers showed high value of coefficient of skewness. The annual maximum peak flood Series of the river Satluj at Rampur for 22 years from 1963 to 1984, Kasol for 17 years from 1964 to 1984 with missing data for 1966, 1969, 1982 and 1983 and Bhakra for 44 years from 1912 to 1955, and the river Beas at Thalot for 44 years from 1941 to 1984 have been used in this study. the statistical parameters in real and log space are as follows:

TABLE 7

STATISTICAL PARAMETERS

River	Station	Sample size (years)	Real space			Log space		
			Mean (m^3/sec)	St.Dev. (m^3/sec)	Skew	Mean (m^3/sec)	St Dev. (m^3/sec)	Skew
Satlaj	Rampur	22	1595	460	0.489	7.335	0.290	0.055
	Kasol	17	1810	724	2.094	7.446	0.334	0.818
	Bhakra	44	3925	1675	0.996	9.191	0.411	0.192
Beas	Thalot	44	1640	891	1.913	7.366	0.444	0.478

The fitting performance of LP3 distribution was found to be better compared to normal, LN3, gamma and P3 distributions. The authors recommended use of LP3 for return periods higher than 100 year and suggested further studies by carrying out flood frequency analysis of other Himalayan rivers.

(vi) James et al (1987) have carried out regional flood frequency analysis using data from nine stations in the Chaliyar river basin of western ghats region. The relationship between mean annual flood $\bar{Q}$ (m³/sec) and area A (km²) was obtained as:

$$\bar{Q} = 3.53\ A^{0.37} \qquad \text{...(30)}$$

(vii) Gupta (1987) has used USGS procedure for studying annual peak flood series of twelve bridge sites with catchment areas varying from 19 km² to 1094 km² for Sub Zone 3(a) Mahi and Sabarmati Basin. A dimensionless curve giving the ratios of return period flood peaks to mean annual flood peak ($Q_{2.33}$) has been developed. The relationship between $Q_{2.33}$ (m³/sec) and catchment area A (km²) is expressed by:-

$$Q_{2.33} = 8.02\ A^{0.63} \qquad \text{...(31)}$$

2.10.3 New Approaches:

(i) Perumal and Seth (1985) suggested a regional analysis approach in which the annual peak flood data of each gauging site is standardized with reference to the site specific mean annual flood, and flood frequency analysis is carried out using combined series of all sites using power transformation method. This study, using data of 18 railway bridge sites in Mahanadi basin subzone 3(d) of size 23 years each and catchment areas varying between 17 to 1150 km^2, indicated that Box-Cox power transformation could be considered as a useful technique for regional frequency analysis. The following regional relationship was obtained for mean annual flood $\bar{Q}$ (m^3/sec) and catchment area A(km^2)

$$\bar{Q} = 4.142\ A^{0.7484} \qquad ...(31)$$

(ii) Singh and Seth(1985) carried out regional frequency analysis for Mahandi basin subzone 3(d) using the five parameter Wakeby distribution. the annual peak flows for 23 years from 1958 to 1980 for 18 bridge catchments used earlier by Gupta and also by Perumal and Seth (1985), were used in this study. The data for 15 sites was used to estimate the regional parameters of the Wakeby distribution with the average probability weighted moments and remaining data of 3 sites was used as test data. The James-Stein corrected means JSM(m^3/sec) for flood series of 15 sites were computed and following relationship was obtained with catchment area A(km^2)

$$JSM = 5.00\ A^{0.71} \qquad ...(33)$$

The floods of 2,5,10,20 and 50 years recurrence intervals were estimated using regional parameters of Wakeby distributioin and JSM. The fitting performance of the regional parameters were tested on the basis of efficiency computed for each of the 15 sites and also for 3 independent sites. Errors of less than 13% were found in estimation of 50 year flood using regional approach with the Wakeby distribution indicating encouraging performance of the suggested methodology.

(iii) Huq et al (1986) have attempted to evolve the frequency flood formulae for countrywide application using the frequency storm rainfall and runoff (flood) models. They used concurrent rainfall runoff data of 219 small and medium catchments located all over the country. For developing relationship for 50 year flood peak in m^3/sec taking into account catchment area $A(km^2)$, statistical or equivalent slope (S) (m/km) of the stream, and 50 year return period 24 hour rainfall R (cm) for use for ungauged catchments upto 5000 km^2 in size. The country was divided into four distinct categories of areas for evolving their respective flood formulae:

(a) alluvial plains of Indus, Ganga and Brahmputra river system with equivalent slope upto 1.5 m/km, and (b) for equivalent slopes above 1.5 m/km. (c) for remaining areas (excluding alluvial plains) with statistical slope upto 3.5 m/km and (d) for statistical slope above 3.5 m/km.

The formulae for Q_{50} (m^3/sec) for these four categories were obtained using multiple regression approach as follows:

<u>Category (a)</u>

$$Q_{50} = 0.765\, A^{0.738}\, S^{0.338}\, R_{24}^{0.713} \qquad \ldots(34)$$

<u>Category(b)</u>

$$Q_{50} = 1.468\, A^{0.594}\, S^{0.425} R_{24}^{0.751} \qquad \ldots(35)$$

loss rate = 0.3 cm/hr ...(35)

Category (c)

$$Q_{50} = 0.921\ A^{0.731}\ S^{0.106}\ R_{24}^{0.927}$$

loss rate = 0.4 cm/hr. ...(36)

Category(d)

$$Q_{50} = 1.932\ A^{0.714}\ S^{0.134}\ R_{24}^{0.717}$$

loss rate = 0.4 cm/hr. ...(37)

The authors also indicated design loss rates adopted in these studies for converting rainfall into excess rainfall for convoluting with synthetic unit hydrograph for the concerned basin to derive flood peak value. It was also mentioned that Q_{25} is about 0.85 times Q_{50}, and Q_{100} is 1.15 times Q_{50} on an average for similar loss rates. The study is subject to various limitations of unit hydrograph assumptions, conversion of point rainfall to areal rainfall, time distribution curves for rainfall, constant loss rates and base flows.

(iv) Recently, Seth and Singh (1987) have further used this approach with the Wakeby distribution using data of catchments for three typical regions viz.

(a) Lower Godavari basin subzone 3(f), (b) Brahmaputra basin and (c) Sub Himalayan region, and obtained encouraging results. The following relationships were obtained:

(a) Lower Godavari subzone 3(f)

$$JSM = 9.02\ A^{0.71} \quad ...(38)$$

(b) Brahmputra basin

$$JSM = 3.32\ A^{0.75} \quad \ldots(39)$$

(c) Sub-Himalayan region

$$JSM = 10.17\ A^{0.63} \quad \ldots(40)$$

Where JSM is James Stein corrected mean annual flood (m^3/sec) and A is catchment area in km^2.

3.0 REMARKS

There is a continuing need to determine the probability of rare floods for their inclusion in risk assessment studies. Most of the empirical methods are quite inadequate from the points of view of prividing results consistently within the accuracy required of hydrologic analysis and design. Hydrologic engineering evaluations and decisions have to be taken mostly with short records at a site, and it becomes necessary to augment the information content through consideration of data obtained at other sites in the same hydrologically homogeneous region. Statistical probability distributions are selected to represent the flood series on a-priori assumption about their suitability and consistency with known characteristics of flood series.

As discussed in previous sections, in recent years significant developments have taken place in application of regional approach to flood frequency analysis. These are:

(i) Use of generalised extreme value (GEV), two component extreme value (TCEV), Wakeby and mixed distributions.

(ii) Use of transformation techniques,method of mixed moment, probability weighted moments and James Stein estimators.

(iii) Use of Bayesian methodology, non-parametric estimators, principal component analysis, generalised least squares procedures and statistically robust estimators.

(iv) Consideration of effect of serial correlation and cross correlation.

(v) Development of methodology for treatment of outliers.

(vi) Use of geomorphologic parameters, rainfall areal correlation structure, alongwith flood characteristics to delimit hydrologically homogeneous regions.

(vii) Use of hierarchical approach for regional frequency estimation for considering effect of size of regions.

(viii) Use of paleohydrological information.

(ix) Development of criterion for selection of physiographic and meteorologic factors affecting mean annual flood for a basin in a hydrologically homogeneous region.

In India, most of the regional flood studies are still based on USGS procedure and are mostly based on RDSO (Railway Design and Standards Organisation, Lucknow) data. Few studies have been attempted using Wakeby distribution and transformation techniques. As seen from equations (14), (20), (25), (27) (28), (31), (32), (38), (39), (40) and table 6, the index n for relation between mean annual flood and basin area of the form $\bar{Q} = CA^n$, varies between 0.58 to 0.9 for different regions. However, for Chaliyar basin, as given in eqn. (30), n=0.37. When length of river is also included as an additional parameter, as shown by eqn. (16), index n = 0.251. There is need for taking up systematic regional flood studies using large data base collected by different state and central agencies, so that flood frequency curves for different regions are readily available for use in planning and prefeasibility for design of hydraulic structures; as well as for design of small schemes and structures.

There is also a continuing need to determine the probability of rare floods for their inclusion in risk assessment studies. Dawdy and Lettenmaier (1987) suggest several approaches to estimate the probability of PMF like floods using parametric and non-parametric approaches, simulation methods and paleoflood analysis. Chow(1982) had sounded a word of caution while commenting an introduction of sophisticated methodologies involving complicated mathematical modelling and massive computer programmes; and providing numerical quantitative answers. Practicing engineers tend to follow cookbook like procedures like Bible without any attempt to understand inherent assumptions and limitations. He advocated more stress on basic concepts, data quality and availability in the area of probability analysis for hydrologic design before proceeding further for introduction of more elaboration in procedures. For a country like India with a wide variety of climate and physiographic situations, it is therefore, desirable that information content of vast data base of whatever quality available with different sources is compiled together for use in hydrologic analysis and design. Regional flood frequency analysis technques have an important role to play in this direction.

REFERENCES

1. Acreman, M.C.(1985), 'Predicting the Mean Annual Flood from Basin Characteristics in Scotland', Hydrological Sciences Journal, Vol.30, pp.37-49.

2. Adamowski, K.(1985), 'Non-parametric Kernel Estimation of Flood Frequencies', Water Resources Research, Vol.21, pp.1585-1590.

3. Arnell, N.W., M. Beran , and J.R.M. Hosking,(1986), 'Unbiased Plotting Positions for the General Extreme Value Distribution', Journal of Hydrology, Vol.86, pp.59-69.

4. Baker, V.R., L.L. Ely, J.E. O'Connor and J.B. Partridge, (1987), 'Paleoflood Hydrology and Design Applications', in Regional Flood Frequency Analysis, Proceedings of the International Symposium on Flood Frequency and Risk Analysis, Baton Rouge, U.S.A., pp.339-354.

5. Beard, L.R.(1962),'Statistical Methods in Hydrology', U.S. Army Corps of Engineers, Sacramento.

6. Benson, M.A.(1962),'Evolution of Methods for Evaluating the Occurrence of Floods', U.S. Geological Survey, Water Supply Paper 1580-A.

7. Bobee, B.(1975),'The Log Person Type III Distribution and Its Application in Hydrology', Water Resources Research.

8. Buges, S.J., D.P. Lettenmaier and C.L. Bates(1975), 'Properties of the Three Parameter Lognormal Probability Distribution', Water Resources Research.

9. Chander, S., S.K. Spolia, and A. Kumar (1978),'Flood Frequency Analysis by Power Transformation', Proc. ASCE, Journal of Hydraulics Div., Vol.104, pp.1495-1504.

10. Chow, V.T.(1982),'Crystal Ball for Urban Storm Drainage Design - The Probability Consideration', Key note Lecture Proceedings, Second International Conference on Urban Storm Drainage, Urbana, Illinois, USA, pp.461-468.

11. Cole, G.(1966),'An Application of the Regional Analysis of Flood Flows', River Flood Hydrology, Institution of Civil Engineers, London, pp.39-58.

12. Cunnane, C.(1985),'Factors Affecting Choice of Distributions for Flood Series', Hydrological Sciences Journal, Vol.30, pp.25-36.

13. Dalrymple, T.(1960),'Flood Frequency Analysis, Manual of Hydrology: Part 3, Flood Flow Techniques', U.S. Geological Survey, Water Supply Paper 1543-A

14. Dawdy, D.R. and D.P. Lettenmaier (1987),'Initiative for Risk Based Flood Design', Journal of Hydraulic Engineering, Proceedings, ASCE, Vol.113, No.8, pp.1041-1054.

15. Farqueharson, F.A.K., C.S. Green, J.R. Meigh and J.V. Sutcliffe (1987),'Comparison of Flood Frequency Curves for Many Regions of the World', in Regional Flood Frequency Analyis, Proceedings of the International Symposium on Flood Frequency and Risk Analysis, Baton Rouge, USA, pp.223-256.

16. Fiorentiono, M., P. Versace and F. Rossi (1985), Regional Flood Frequency Estimation Using the Two Component Extreme Value Distribution', Hydrological Sciences Journal, Vol.30, pp.51-64.

17. Fiorentino, M., S. Gabriele, F. Rosi and P. Versace, (1987),'Hierarchical Approach for Regional Flood Frequency Analysis', in Regional Flood Frequency Analysis, Proceedings of the International Symposium on Flood Frequency and Risk Analysis, Baton Rouge, USA, pp.35-50.

18. Goel, N.K. and S.M. Seth (1985),'Data Related Problems in Frequency Analysis', Proceedings, Seminar on Flood Frequency Analysis, New Delhi, pp.14-30.

19. Goswami, A.(1972),'Regional Flood Frequency Analysis of Brahmputra Basin in North-East India', Journal, Institution of Engineers(India), Civil Engg. Div., Vol.52, pp.169-174.

20. Gottschalk, L.(1985),'Hydrological Regionalisation of Sweden', Hydrological Sciences Journal, Vol.30, pp.65-83.

21. Greis N.P.(1983),'Flood Frequency Analysis: A Review of 1979-1982', Review of Geophysics and Space Physics, Vol.21, No.3, pp.699-706.

22. Greis, N.P. and E.F. Wood(1981),'Regional Flood Frequency Estimation and Network Design', Water Resources Research, Vol.17, pp.1167-1177.

23. Gupta, P.N.(1987),'Regional Flood Frequency Approach Estimation of Flood Peaks for Mahi and Sabarmati Basin(Subzone 3 a)',Proceedings, National Symposium on Hydrology, Roorkee, Vol.I, pp.I-16-26.

24. Hebson, C. and E.F. Wood(1982),'A Derived Flood Frequency Distribution Using Horton Order Ratios', Water Resources Research, Vol.18, pp.1509-1518.

25. HEC(1975),'Hydrologic Engineering Methods for Water Resources Development, Vol.3, Hydrologic Frequency Analysis', U.S. Army Corps of Engineers, Hydrologic Engineering Centre.

26. Hirsch, R.M. and J.R. Stedinger(1987), Plotting Positions for Historical Floods and Their Precision', Water Resources Research, Vol.23, pp.715-727.

27. Hosking, J.R.M., J.R. Wallis and E.F. Wood(1985), 'An Appraisal of the Regional Flood Frequency Procedure in the U.K. Flood Studies Report', Hydrological Sciences Journal, Vol.30, pp.85-109.

28. Hosking, J.R.M. and J.R. Wallis(1986),'Paleoflood Hydrology and Flood Frequency Analysis', Water Resources Research, Vol.22, No.4, pp.543-550.

29. Huq, S.M.(1985),'Regional Flood Frequency Analysis of Cauvery Basin Sub-zone 3(i), Proceedings, 52nd Annual R&D Session, CBIP, Aurangabad, Vol.II, pp.97-108.

30. Huq, S.M., R. Nair and S.S. Sil(1986),'Frequency Flood Formulae for Countryvide Application', Proceedings, 53rd Annual R&D Session, CBIP, Bhubaneshwar, pp.405-416.

31. Irish, J.(1977),'Regional Flood Estimation', in Australian Rainfall and Runoff - Flood Analysis and Design, The Institution of Engineers Australia, pp.116-119.

32. 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, pp.41-48.

33. James, E.J., K.D. Nambudripad, K.E. Sreedharan and R. Mohanan(1987),'Rainfall Runoff Criteria for the Chaliyar Basin of Western Ghats Region - Final Report CBIP Project, CWRDM, Kozhikode, pp.99-103.

34. Jennings, M.E. and M.A. Benson(1969),'Frequency Curve for Annual Flood Series with Some Zero Events or Incomplete Flows', Water Resources Research, Vol.5.

35. Kite, G.W.(1978),'Frequency and Risk Analysis in Hydrology', Water Resources Publication, Fort Collins, USA.

36. Kuczera, G.(1983),'A Bayesian Surrogate for Regional Skew in Flood Frequency Analysis', Water Resources Research, Vol.19, pp.821-832.

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

38. Ladwehr, S.M., N.C. Matalas and J.R. Wallis(1984), 'Note on the Applicability of the James Stein Estimation in Regional Hydrologic Analysis', Water Resources Research, Vol.20, pp.1630-1638.

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

40. Lettenmaier, D.P., J.R. Wallis and E.F. Wood(1987), 'Effect of Regional Heterogeneity on Flood Frequency Estimation', Water Resources Research, Vol.23, pp.313-323.

41. Linsley, R.K.(1986),'Flood Estimates: How Good Are They', Water Resources Research, Vol.22, pp.1595-1645.

42. McCuen, R.H., W.J. Rawls, G.T. Fischer and R.L. Powell, (1971),'Flood Flow Frequency for Ungauged Watersheds-A Literature Evaluation', Agricultural Research Service, U.S. Dept. of Agriculture, pp.1-136.

43. McMahon, T.A.(1982),'World Hydrology! Does Australia Fit? Hydrology and Water Resources Symposium, Melbourne.

44 Mehta, H.L. and D.K. Sharma(1986),' Flood Frequency Analysis', Proceedings, Workshop on Flood Estimation in Himalayan Region, Roorkee(U.P.), pp.13-23.

45. Nash, J.E. and B.L. Shaw(1966),'Flood Frequency as a Function of Catchment Characteristics', River Flood Hydrology, Institution of Civil Engineers, London, pp.115-136.

46. NERC(1975),'Flood Studies Report Volume 1 - Hydrological Studies', Natural Environment Research Council, London.

47. Perumal, M and S.M. Seth(1985),'Need for Uniform Procedure of Flood Frequency Analysis', Proceedings, Seminar on Flood Frequency Analysis, New Delhi, pp.1-13.

48. Perumal, M. and S.M. Seth(1985),'Regional Flood Frequency Analysis Using Power Transfermation - A Case Study', Proceedings, Seminar on Flood Frequency Analysis, New Delhi, pp.93-108.

49. Rao, D..(1980),'Log Pearson Type III Distribution, Method of Mixed Moments', Journal of the Hydraulics Div., ASCE, pp.999-1019.

50. Rao, P.R. and R.S. Goel(1986),'Flood Frequency Analysis - Assumptions and Limitations', Proceedings, Workshop on Flood Estimation in Himalayan Region, Roorkee(U.P), pp.1-12.

51. Seth, S.M. and B.N. Goswami(1972),'Regional Flood Frequency Analysis with Limited Data', Proceedings, Third World Congress on Water Resources, Mexico City.

52. Seth, S.M. and N.K. Goel(1985),'Regional Flood Frequency Analysis', Proceedings, Seminar on Flood Frequency Analysis, New Delhi, pp.31-39.

53. Seth S.M. and R.D. Singh(1987),'Flood Estimation in India Using Regional Frequency Analysis', Proceeding, National Symposium on Hydrology, Roorkee, Vol.1, p.I- 1-15.

54. Singh, R.D. and S.M. Seth(1985),'Regional Flood Frequency Analysis for Mahandi Basin Using Wakeby Distribution', Proceedings, Seminar on Flood Frequency Analysis, New Delhi, pp.109-124.

55. Sokolov, A.A. S.E. Rantz and M.Roche,(1976),'Flood Flow Computation - Methods Compiled from World Experience', Unesco Press, pp.37.

56. Stedinger, J.R. and G.D. Tasker(1985),'Regional Hydrologic Analysis — I. Ordinary, Weighted and Generalised Least Squares Compared', Water Resources Research, Vol.21, pp.1421-1432.

57. Stedinger, J.R. and T.A. Cohn(1986),'Flood Frequency Analysis with Historical and Paleoflood Information', Water Resources Research, Vol.22, No.5, pp.785-793.

58. Tasker, G.D. and J.R. Stedinger(1986),'Regional Skew with Weighted Least Squares Regression', Journal W.R. Planning and Management, Vol.112, pp.225-237.

59. 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, pp.III - 43-51.

60. Thiruvengadachari, S., S.V. Nayak and G. Ranganna(1975), 'Flood in Cauvery Basin: Magnitude and Frequency', Proceedings, National Symposium on Hydrology, Roorkee, pp.G-17-21.

61. Thomas, W.O.(1985),'A Uniform Technique for Flood Frequency Analysis', Jnl. of W.R. Planning and Management, ASCE, Vol.111, pp.321-337.

62. USWRC(1981),'Guidelines for Determining Flood Flow Frequency', Bulletin 17B of the Hydrology Committee, U.S. Water Resources Council, Washington.

63. Vicens, G.J., I.R. Iturbe and J.C. Schaake(1975), 'A Bayesian Framework for the Use of Regional Information in Hydrology', Water Resources Research, Vol.11, pp.405-414.

64. 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 andf Control, Roorkee, pp.II - 58-77.

65. Venkatraman,R.and P.N.Gupta(1986),'Flood Estimation for Small and Medium Catchments in Sub-Himalayan Region, Proceeding, Workshop on Flood Estimation in Himalayan Region, Roorkee, pp 123-140.

66. Wallis, J.R.(1980),'Risk and Uncertainities in the Evaluation of Flood Events for the Design of Hydrologoic Structures', Keynote Address, Seminar on Extreme Hydrological Events: Floods and Droughts, Erice, Italy.

67. Wiltshire, S.E.(1985),'Grouping Basins for Regional Flood Frequency Analysis', Hydrological Sciences Journal, Vol.30, pp.151-159.

68. Wiltshire, S.E.(1986),'RegionalFloodFrequency Analysis I: Homogeneity Statistics', Hydrological Sciences Journal, Vol.31,pp.321-333.

69. Wiltshire, S.E.(1986),'Identification of Homogeneous Regions for Flood Frequency Analysis', Journal of Hydrology, Vol.84, pp.287-302.

70. Wood, E.F. and C.S. Hebson(1986),'On Hydrologic Similarity - 1.Derivation of the Dimensionless Flood Frequency Curve', Water Resources Research', Vol.22 pp.1549-1554..

71. Yevjevich, V.M.(1963),'Flood Characteristics and Spillway Design', IAHR Congress, London.

72. Yevjevich, V.M.(1972),'New Vistas for Flood Investigations', Academia Nazional Del, Lincei, Rome.

73. Zhang, Y.(1982),'Plotting Position of Annual Flood Extremes Considering Extraordinary Values', Water Resouorces Research, Vol.18, pp.859-864.