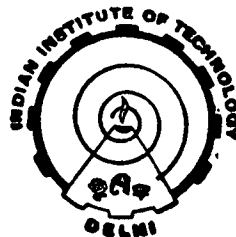


GENERALIZED PARETO DISTRIBUTION FOR FLOOD FREQUENCY ANALYSIS

by

SAAD HAMED SAAD MOHARRAM
Department of Civil Engineering

THESIS SUBMITTED
IN FULFILMENT OF THE REQUIREMENTS OF THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
CIVIL ENGINEERING

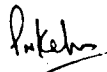


to the
INDIAN INSTITUTE OF TECHNOLOGY, DELHI

JUNE, 1990

CERTIFICATE

This is to certify that the thesis entitled 'GENERALIZED PARETO DISTRIBUTION FOR FLOOD FREQUENCY ANALYSIS' being submitted by Mr. SAAD HAMED SAAD MOHARRAM to the Indian Institute of Technology, Delhi, India, for the award of the degree of DOCTOR OF PHILOSOPHY, is a record of bonafide research work carried out by him under our supervision and guidance. The thesis work, in our opinion, has reached the standard, fulfilling the requirements for DOCTOR OF PHILOSOPHY degree. The research report and the results presented in this thesis have not been submitted, in part or in full, to any other University or Institute, for the award of any degree of diploma.



(Prof. P.N. Kapoor)
Professor
Department of Civil Engg.
Indian Institute of Technology
New Delhi-110 016, INDIA.



(Dr. A.K. Gosain)
Assistant Professor
Department of Civil Engg.
Indian Institute of Technology
New Delhi-110 016, INDIA.

ACKNOWLEDGEMENT

I wish to express my regards and deep sense of gratitude to Prof. P.N. Kapoor, Professor, Department of Civil Engineering, Indian Institute of Technology (I.I.T.), Delhi, for his kind supervision, valuable guidance and continuous encouragement for completing this thesis.

I am deeply indebted to Dr. A.K. Gosain, Assistant Professor, Department of Civil Engineering, I.I.T., Delhi, for his affection, encouragement, supervision and guidance in this thesis.

Sincere thanks are due to Ass. Prof. B.P. Parida and other faculty members of the Department of Civil Engineering, I.I.T., Delhi for all possible help and providing the facilities for conducting this study.

Recognition is due to my wife, Nadia, for her invaluable help in many aspects, and due to my daughter, Raghda, who had endured the neglect during the preparation of this study.

Finally, I express my thanks to Mr. Samsheer S. Dagar for typing the manuscript and Mr. R.V. Aggarwal for preparing the tracings.

Date : 1st June 1990

S.H. Moham

(Saad Hamed Saad Moharram)

SYNOPSIS

GENERALIZED PARETO DISTRIBUTION FOR FLOOD FREQUENCY ANALYSIS

Estimation of hydrologic loading (flood peak) based on preassigned risk, so that a specific service is not interrupted or stopped because of hydrological reasons is central to flood frequency analysis. The selection of design flood for a specific return period is in principle an assessment of the risk involved against the cost of interruption or stoppage of a specific service if a flood greater than the design flood is experienced. Risk is not confined to civil engineering structures alone, but covers a wide field of economic activities. Both socio-political and economic considerations enter into the decision making. The task of identifying a design flood with a specific return period is accomplished by choosing an appropriate probability model. Uncertainty in the flood frequency analysis creeps in because a hydrologist can never be sure about a fitted distribution being the same as nature might have used to generate flood flows and also the data sample may not truly reflect the complete characteristics of the population. To get over this uncertainty, many probability distributions ranging from two-parameter distribution to five-parameter distributions and several parameter estimation techniques need to be examined.

Decision making values in flood frequency analysis usually lie at the tail end of a distribution and selecting a distribution shape from amongst established distributions based on goodness of fit indices is not an easy affair. There has been multiplicity of reasons in justification for various distributions. Chow (1954) thought that causative factors for many hydrologic variables act multiplicatively rather than additively and so the logarithm of the peak floods which are the products of these causative factors should follow the normal distribution. However, if Chow's reasoning about peak flood formation been universally applicable, there was no need of numerous frequency distributions to model the annual peak flood flows. In reality, numerous probability density functions (pdf's) have been tested to see if they fit the annual maximum series of the peak floods. The main problem is that data tend to be asymmetrical and no pdf is universally applicable.

One family of pdf which have been recommended in the Flood Studies Report (Natural Environment Research Council, NERC, 1975) is that of general extreme value distributions (GEV's). A particular example of this is the so called extreme value type-1 (EVI or Gumbel's) distribution. The EVI distribution is the simplest of the family of GEV's and its applicability is limited to the data whose skewness coefficient is in the vicinity of 1.139. There are two more

members of this family: the EV2 and EV3; but these require the estimation of a third parameter, known as the shape factor. Unfortunately, it has not been possible to estimate the third parameter without any uncertainty. Consequently, the Flood Studies Report recommended that the third parameter should be selected according to region, using results derived from regional pooling of data.

Some of the two parameter distributions such as Normal, Exponential and Extreme value type-1 are applicable for a specific skewness and as such refer to fixed shape though these may provide low variability in an estimator. There have been attempts to indirectly account for the third parameter by using a transformation to normality based entirely on the criterion of making the coefficient of skewness near to zero. Based on the above approach, Chander et al. (1978) reported the use of power transformation in the flood frequency analysis. This process ignored the kurtosis of the distribution which governs the tail thickness of the distribution. However, the authors did make attempts to correct for deviation of coefficient of kurtosis away from 3 in the normalized series. Also Cunnane (1985) pointed out that few random samples from normal population have skewness equal to or close to zero.

Boughton (1980) believed that the statistics of the flood data from various catchments strongly demonstrate the need

of three parameter frequency distribution instead of two parameter frequency distribution. After analysing flood data from 78 catchments in Australia, he found that the range of the estimates of coefficient of skewness to extend from +1.43 to -2.26 with a mean value of -0.6. The range is sufficiently large that no two parameter distribution could adequately fit all of the data sets.

Other studies (U.S. Water Resources Council, 1967; Prasad, 1970; NERC, 1975 and Kite, 1977) have tested different probability distributions and their conclusions are in favour of three parameter distribution, such as log Pearson type 3 and GEV, because these fit better to data used. Attempts have been made in the past for correcting the bias in the estimation of coefficient of skewness.

Singh and Sinclair (1972) suggested the use of mixture of two distributions with five parameters to model annual peak flood series. However, Cunnane (1985) discourages the use of mixture of distributions when there is no physical explanation for the need for more than two or three parameters. Houghton (1978a) introduced the five-parameter Wakeby distribution as the one capable of adequately fitting flood records. Although the Wakeby distribution has a versatile shape characteristics to make satisfactory fit for flood records, this advantage alone does not ensure robust estimation of extreme events (Kuczera, 1982b).

To overcome the presence of the outliers and high variability of skewness of historical data, Rossi et al. (1984) suggested the two component extreme value (TCEV) as a model for analysis of annual flood series in Italy. It has four parameters to describe a flood series generated by two distinct independent processes (e.g. Snowmelt and Frontal storms). Ahmad et al. (1988a) examined the Wakeby and TCEV distributions. According to the authors, an ideal distribution for flood frequency analysis must possess the following characteristics: (i) it must reproduce at least as much variability in flood characteristics as is observed in empirical data sets; (ii) it must be insensitive to extreme outliers especially in the upper tail, (iii) it must have a distribution function and an inverse distribution function that can be explicitly expressed in a close form and (iv) it must not be computationally complex nor involve the estimation of a large number of parameters. The Wakeby and TCEV distributions have proved successful in terms of reproductive criteria (i), and includes the separation of skewness in observed and simulated floods. The parameter estimates of Wakeby distribution often have large standard errors which result in wide confidence intervals for the quantile estimates and its distribution function can not be expressed in a closed form giving rise to problems in parameter estimation by maximum likelihood method. Thus the

Wakeby distribution fails to satisfy adequately the criteria (iii) and (iv) as listed above. Similarly, the TCEV fails to perform adequately in terms of criteria (iii) and (iv), since the parameter estimation by maximum likelihood method on selected data can fail to achieve the required convergence. Furthermore, the inverse form of TCEV does not exist and thus the estimates of quantities are difficult to obtain.

Considerable uncertainty exists about the form of the underlying population distribution of flood at any site. Owing to the vast hydrogeological variations possible, it is reasoned that the population distribution may have remarkably wide range of forms for various sites. With the inadequacy of two parameter distributions well established, there is a scope for more three parameter distributions to be tested for performance for flood frequency analysis.

Various parameter estimation methods are in use for estimating the parameters of a frequency model from the past records at specific sites. The method of moments (MOM) which is widely used in hydrology, is subject to some bias and is relatively inefficient. The method of maximum likelihood (ML) provides asymptotically minimum variance estimates. It is used to lesser extent, partly because, the application does not lend itself to easily manipulated algebraic expression (Landwehr et al., 1979b). Another method used is

the least squares (LS), but it may not be preferable as a standard method. Moreover, as a new class of moments, Greenwood et al. (1979) introduced the probability weighted moments (PWM) method as a potential technique for estimating the parameter of distributions which can be written in inverse form. Another alternative method used to estimate the parameters is based on the concept of entropy (Joitte, 1979; and Singh and Singh, 1985), it has not found wide application.

The generalized Pareto (GP) distribution, a three parameter distribution, was introduced by Van Montfort and Witter (1985) and (1986) as a model applicable for rainfall series using the maximum likelihood estimates. Moreover, Hosking and Wallis (1987) developed the GP parameter estimates by deriving both methods of moments and probability weighted moments in which the case of lower bound is known to be zero. It was decided to explore the possibility of its application as a candidate distribution in flood frequency analysis. With this in mind the objectives of the study are set as follows:

1. To prepare brief state-of-the-art report on flood frequency analysis.

2. To formulate equations for the parameter estimation of GP distribution using method of moment, method of maximum likelihood and probability weighted moment method where the case of lower bound is not equal to zero. The

formulation based on least squares method has also been done.

3. To study the performance of GP distribution in comparison to the other commonly used distributions.

4. To evaluate the performance of these methods of parameter estimation in terms of commonly used criteria, such as the bias, root mean square error, etc., using Monte Carlo simulation.

Analytical equations for parameter estimation using ML, MOM and PWM methods have been modified with respect to the lower bound of series as a third parameter. Also equations of the LS method have been formulated. The performance of the methods of parameter estimation have been evaluated. The PWM and LS methods are decided to be the best techniques when $c < 0$ and $c > 0$, respectively. Performance of the GP distribution in flood frequency analysis has been compared with the other distributions, such as GEV, log-Pearson type 3, log-logistic, log-Boughton and power transformation. The GP distribution performs reasonably well as compared to the other distributions.

CONTENTS

	Page No.
CERTIFICATE	i
ACKNOWLEDGEMENT	ii
SYNOPSIS	iii
CONTENTS	xi
LIST OF TABLES	xiv
LIST OF FIGURES	xvi
LIST OF SYMBOLS	xviii
LIST OF NOTATIONS	xxi
Chapter 1 INTRODUCTION	1
1.1 General	1
1.2 Statement of the Problem	2
1.3 Objectives of the Study	6
1.4 Prologue	7
Chapter 2 LITERATURE REVIEW	9
2.1 General	9
2.2 Commonly used Frequency Distributions	12
2.2.1 Log Normal Distribution	13
2.2.1.1 2-parameter Log-Normal	13
2.2.1.2 3-parameter Log-Normal	15
2.2.2 Pearson Type 3 Distribution	20
2.2.3 Log Pearson Type 3 Distribution	22
2.2.4 General Extreme Value Distribution	25
2.2.5 Log-Boughton Distribution	32
2.2.6 Log-Logistic Distribution	33
2.2.7 Two-Component Extreme Value Distribution	36

2.2.8	Wakeby Distribution	38
2.2.9	Power Transformation Method	40
2.3	Standard Error of Estimates	41
2.4	Evaluation of Frequency Distributions	45
2.5	Limitations of Frequency Distributions	49
2.6	Methods of Regional Flood Frequency Analysis	52
2.6.1	Regional Regression Method	53
2.6.2	Index-Flood Method	55
2.6.3	Method Based on Standardized PWM's	56
2.6.4	Method Based on Power Transformation	57
2.7	Remarks on Regional Analysis.	58
Chapter 3	GENERALIZED PARETO DISTRIBUTION	60
3.1	Introduction	60
3.2	Generalized Pareto Distribution	62
3.3	Methods of Parameter Estimation	63
3.3.1	Method of Maximum Likelihood	64
3.3.2	Method of Moments	65
3.3.3	Method of Probability Weighted Moments	70
3.3.4	Method of Least Squares	71
3.4	Variances and Covariances of Estimators	73
3.5	Data Analysis	77
3.5.1	Verification of Skewness	81
3.5.2	Flood Frequency Analysis	81
3.5.3	Comparison of Estimation Methods	93
3.6	Conclusions	97
Chapter 4	PERFORMANCE COMPARISON OF GP DISTRIBUTION	99
4.1	Introduction	99

4.2	Analysis of Flood Frequency Distributions	100
4.2.1	Generalized Pareto Distribution	101
4.2.2	Other Flood Frequency Distributions	101
4.3	Comparison of Distributions	104
4.3.1	Procedure	104
4.3.2	Discussion of Results	110
4.4	Conclusions	117
CHAPTER 5	COMPARATIVE STUDY FOR GP DISTRIBUTION USING MONTE CARLO SIMULATION	129
5.1	Introduction	129
5.2	Parameter Estimation Methods	130
5.3	Monte Carol Experiments	134
5.4	Discussion of Results	137
5.5	Conclusions	145
Chapter 6	SUMMARY, CONCLUSIONS AND FUTURE WORK	148
6.1	Summary and Conclusions	148
6.2	Scope of Future Work	152
Appendix (A)	Moments for GP Distribution	153
Appendix (B)	Probability Weighed Moments for GP Distribution	155
Appendix (C)	Derivation of Least Squares Method for GP Distribution	156
REFERENCES		158