

REGIONALISATION OF FLOOD EXTREMES USING PATTERN ANALYSIS

BY

AHMAD FAKHERI FARD

Civil Engineering Department

Submitted in fulfillment of the requirement of the degree of
Doctor of Philosophy

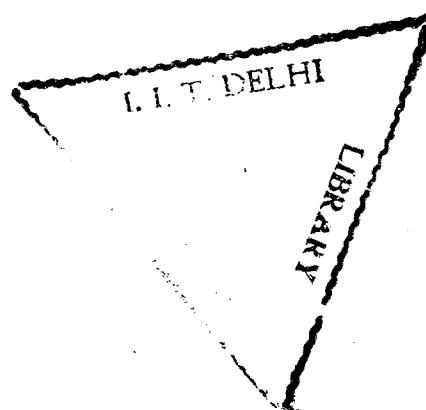
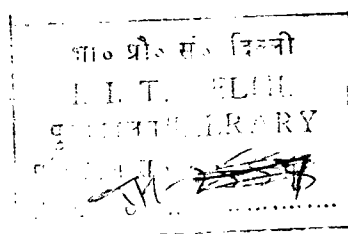
to the



Indian Institute of Technology, Delhi

MAY 1998

TH
627.516:519.28
FAR-R



CERTIFICATE

This is to certify that the thesis titled, " **REGIONALISATION OF FLOOD EXTREMES USING PATTERN ANALYSIS** " being submitted by **Mr. AHMAD FAKHERI FARD**, 94RCE023 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 my supervision and guidance. The thesis work, in my opinion, has reached the standard, fulfilling the requirements for the said 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 or diploma.



Dr. B. P. Parida
Associate Professor
Department of Civil Engg.,
Indian Inst. of Tech.,
New Delhi, 110016, India.

ACKNOWLEDGEMENTS

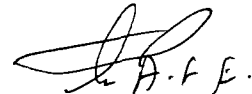
I am grateful and deeply indebted to my guide, Dr. B. P. Parida, Associate professor, Civil Engineering Department, I.I.T Delhi from whom I received systematic guidance, constant help and encouragement throughout the period of research program.

Sincere thanks to my parents Mr Hassan Fakheri Fard and Mrs T. Eskandar Zadeh the best advisers and supporters in my life.

I would like to record my especial heartfelt thanks to my wife Mrs F. Moosavi for her excellent cooperation and support during the entire period of this research and also to my littel son Sina who was a great companion for my wife.

Sincere thanks to my brother Mohmmad and my mother- in- low Mrs T. Ebrahimi for their best cooperation for my travelling to India.

I would like to express my gratefulness to the Government of Iran who awarded scholarship and financially supported me to do this research.

A handwritten signature in black ink, appearing to read 'A.F.F.', with a stylized flourish above it.

Ahmad Fakheri Fard

Abstract

In search for homogeneous regions using pattern analysis, seven currently used hierarchical methods namely Single Linkage, Complete Linkage, Centroid, Median, Weighted Pair Group, Unweighted Pair Group, and Ward as well as non-hierarchical methods such as Forgy's K-mean algorithm as partitioning method and Q-mode Principal Component analysis (PCA), in both rotated and unrotated cases, have been attempted for classification of 88 flood sites situated in central and south-central India with basin characteristics and annual peak flood separately. Homogeneity of the resulted clusters from each method has been tested using S-statistics and appropriate discordant sites have been eliminated from the heterogeneous clusters to achieve perfect/moderate homogeneity. An economical scoring model has been developed using the number of clusters, number of members and the heterogeneity values of each cluster for arriving at the final choice of the regions. Regions identified by the unrotated principal component analysis (UR-PCA) method with the basin characteristics data was found to have the maximum score and as such has been considered as the most economical-fitted method for the study area. Based on the results Pearson Type Three (PTIII) was selected as the best-fitted distribution for study area from among several other choice and growth curves developed. Also a relation between mean annual peak flood($\bar{Q}_p$) and the basin characteristics data has been developed for of the each homogeneous region in order that transfer of information from gauged locations to ungauged ones can be undertaken through corresponding growth curves. These regions were further subjected to tests for internal homogeneity using H-statistics method and the underlying distributions were identified by Index Flood procedure using L-Moments.

TABLE OF CONTENTS

CHAPTER 1	INTRODUCTION	Page
1.1	GENERAL PROBLEMS	1
1.2	OBJECTIVES OF THE STUDY	2
1.3	SCOPE OF PRESENTED STUDY	5
1.4	LAYOUT OF THE THESIS	7
CHAPTER 2	SELECTIVE REVIEW OF METHODS OF REGIONALIZATION	
2.1	WHY REGIONALIZATION	9
2.2	CONCEPT OF REGIONAL HOMOGENEITY	10
2.2.1	Cc Index	11
2.2.2	Dalrymple's Homogeneity Test Method	11
2.2.3	Method Of Moment - Cv Test	13
2.2.4	S-statistics Test Method	13
2.2.5	L-moments and H-statistic Test	16
2.3	DISCORDANCY MEASURE	20
2.3.1	Discordancy measure using L-moments	20
2.3.2	Iteratively Elimination of Discordant Site	22
2.4	BRIEF REVIEW OF THE METHODS Of REGIONALISATION	23
2.4.1	Geographical Approaches	23
2.4.2	Analytical Approaches	24
i.	Station Year Method	24

ii.	Index Flood Method	24
iii.	Method of Natural Environmental Research Council (NERC)	25
iv.	United States Water Resources Council (USWRC) Method	26
v.	Two Component Extreme Value (TCEV) Method	27
vi.	Use of L-Moments with Index Flood Procedure	29
2.5	Regional Relations In Ungauged Catchments And Quantile Estimation	30
 CHAPTER 3 THE STUDY AREA AND THE AVAILABLE DATA		
3.1	AREA UNDER STUDY	31
3.2	BASIC DATA	32
3.2.1	Basin Characteristics Data	32
3.2.2	Filling Up of Missing Values of Peak Flood Data	39
 CHAPTER 4 HIERARCHICAL CLUSTERING METHODS		
4.1	BACKGROUND	42
4.2	GENERAL ADVANTAGES	43
4.3	GENERAL STAGES FOR CARRYING OUT THE HIERARCHICAL METHODS	43
4.4	SINGLE LINKAGE METHOD (SLM)	46
4.4.1	Methodology	46
4.4.2	Advantages of SLM	48
4.4.3	Disadvantages of SLM	49
4.4.4	Application of SLM to Basin Characteristics Data	49
4.4.5	Application of SLM to Annual Peak Flood Data	52
4.5	COMPLETE LINKAGE METHOD (CLM)	55
4.5.1	Methodology	55

4.5.2	Advantages of CLM	55
4.5.3	Disadvantages of CLM	56
4.5.4	Application of CLM to Basin Characteristics Data	56
4.5.5	Application of CLM to Annual Peak Flood Data	58
4.6	CENTROID METHOD (CEM)	61
4.6.1	Methodology	61
4.6.2	Advantages of CEM	63
4.6.3	Disadvantages of CEM	63
4.6.4	Application of CEM to Basin Characteristics Data	63
4.6.5	Application of CEM to Annual Peak Flood Data	66
4.7	MEDIAN METHOD (MEM)	66
4.7.1	Methodology	66
4.7.2	Advantages of MEM	69
4.7.3	Disadvantages of MEM	69
4.7.4	Application of MEM to basin characteristics data	69
4.7.5	Application of MEM to Annual Peak Flood Data	70
4.8	UNWEIGHTED PAIR GROUP METHOD (UWPGM)	72
4.8.1	Methodology	72
4.8.2	Advantages of UWPGM	75
4.8.3	Disadvantages of UWPGM	75
4.8.4	Application of UWPGM to Basin Characteristics Data	76
4.8.5	Application of UWPGM to annual peak flood data	76
4.9	WEIGHTED PAIR GROUP METHOD(W.P.G.M.)	78
4.9.1	Methodology	78
4.9.2	Advantages of WPGM	82
4.9.3	Disadvantages of WPGM	82

4.9.4	Application of WPGM to Basin Characteristics Data	82
4.9.5	Application of WPGM to Annual Peak Flood Data	83
4.10	WARD METHOD (WAM)	85
4.10.1	Methodology	85
4.10.2	Advantages of WAM	91
4.10.3	Disadvantages of WAM	91
4.10.4	Application of WAM to Basin Characteristics Data	92
4.10.5	Application of WAM to Annual Peak Flood Data	92
4.11	SUMMARY	96
 CHAPTER 5 K-MEAN ALGORITHMS		
5.1	INTRODUCTION	98
5.2	FORGY'S K-MEAN ALGORITHM	98
5.2.1	Advantages of Algorithm	99
5.2.2	Disadvantages of Algorithm	99
5.2.3	Application of K-mean Algorithm to Basin Characteristics Data	99
5.2.4	Application of K-mean Algorithm to Annual Peak Flood Data	106
5.3	SUMMARY	112
 CHAPTER 6 PRINCIPAL COMPONENT ANALYSIS (PCA)		
6.1	INTRODUCTION	113
6.2	ADVANTAGES AND DISADVANTAGES OF Q-MODE PCA	114
6.3	GENERAL MODEL OF PRINCIPAL COMPONENT ANALYSIS	115
6.3.1	R-mode PCA	115
6.3.2	Q-mode PCA	120

6.4	ITERATIVE METHOD FOR CALCULATION OF EIGENVECTORS AND EIGENVALUES OF A CORRELATION MATRIX	122
6.4.1	Methodology	122
6.4.2	Application Of Q-mode PCA To Basin Characteristics Data	124
6.4.3	Application Of Q-mode PCA To Annual Peak Flood Data	126
6.5	PCA AS A CLUSTERING METHOD	127
6.6	SELECTION OF PRINCIPAL COMPONENTS	141
6.6.1	Selection Of Principal Components For Basin Characteristics Data	141
6.6.2	Selection Of Principal Components For Annual Peak Flood Data	145
6.7	ROTATION OF PRINCIPAL COMPONENTS	146
6.7.1	Methodology	146
6.7.2	Rotation Of Eigenvectors And Correlation Loadings Of Basin Characteristics Data	152
6.7.3	Rotation Of Eigenvectors And Correlation Loadings Of Peak Flood Data	153
6.8	IDENTIFICATION OF CLUSTERS	153
6.8.1	Identification Of Clusters Based On Basin Characteristics Data	160
	i. Using Unrotated Loadings.	160
	ii. Using Rotated Loadings	160
6.8.2	Identification Of Cluster Based On Annual Peak Flood Data	162
	i. Using Unrotated Loadings	162
	ii. Using Rotated Loadings	166
6.9	SUMMARY	169

CHAPTER 7 ECONOMICAL SCORING OF METHODS

7.1	INTRODUCTION	170
7.2	SCORING MODEL	171
7.3	ANALYSING THE METHOD SCORES AND SELECTION OF The ECONOMICAL-FITTED METHOD	184
7.4	DELINEATION OF HOMOGENEOUS REGIONS	187

CHAPTER 8 REGIONAL DISTRIBUTIONS

8.1	GOODNESS-OF-FIT AND SELECTION OF ADEQUATELY FITTED DISTRIBUTION(S) FOR A REGION	196
8.2	DETERMINATION OF REGIONAL PARAMETERS FOR A SELECTED DISTRIBUTION AND DEVELOPING GROWTH CURVE FOR A HOMOGENEOUS REGION	198
8.3	ADEQUATELY FITTED DISTRIBUTIONS FOR THE RESULTING HOMOGENEOUS REGIONS	200
8.4	DEVELOPMENT OF GROWTH CURVES FOR THE RESULTING HOMOGENEOUS REGIONS	203
8.5	DEVELOPMENT OF REGIONAL RELATION FOR COMPUTATION OF INDEX FLOOD	204

CHAPTER 9 SUMMARY AND CONCLUSION

9.1	CONCLUSIONS	212
9.2	SPECIFIC CONTRIBUTIONS	219
9.3	SCOPE FOR FUTURE WORK	219

REFERENCES	220
------------	-----

BRIEF BIO-DATA OF AUTHORS	227
---------------------------	-----