MODELING FLOOD ESTIMATION USING FUZZY LOGIC & ARTIFICIAL NEURAL NETWORK

Sattar Chavoshi Borujeni*, Wan Nor Azmin Sulaiman, Latifah Binti Abd Manaf, Md Nasir B Sulaiman, Bahram Saghafian

PhD (GS 18002)
6th Semester

Introduction

Estimates of flood discharge with various risks of exceedance are needed for a wide range of engineering problems: examples are culvert and bridge design and construction floods in major projects. At a site with a long record of measured floods, these estimates may be derived by statistical analysis of the flow series. Alternatively the storm magnitude of an appropriate duration, aerial coverage and return period may be estimated and converted into the flood of a given return period using a rainfall/runoff model such as the unit hydrograph. However, in cases where adequate rainfall or river flow records are not available at or near the site of interest, it is difficult for hydrologists and engineers to derive reliable flood estimates directly and regional studies can be useful. This is particularly true in the case of semi-arid areas, where, in general, flow records are scarce. The problem of assigning a flood risk to a particular flow value is one which has received considerable attention in the literature. The estimation of flood risk through the evaluation of a flood frequency distribution is complicated, however, by the lack of a sufficient temporal characterization of the underlying distribution of flood events. The inadequacies in the data availability necessitate the estimation of the flood risk associated with events which have a return period beyond the length of the historical record. Regional flood frequency analysis can be effective in substituting an increased spatial characterization of the data for an insufficient temporal characterization, although problems exist with the implementation of regional flood frequency analysis techniques.

Keywords: *Flood estimation, Fuzzy Logic, ANN, Homogeneity*

Research Objectives

- General Objective:
- Providing the more accurate flood estimation for further hydrologic design projects. In other words the regional flood estimation is expected to have more accuracy compared to the generally used at-site estimation.
- Specific Objective:
- Investigating the potential application and efficiency of Fuzzy Logic and Artificial Neural Network (ANN) solutions to the problem of flood estimation in ungauged catchments
- Regionalization of the study area based on the pooled hydrologic variables

Research Methodology

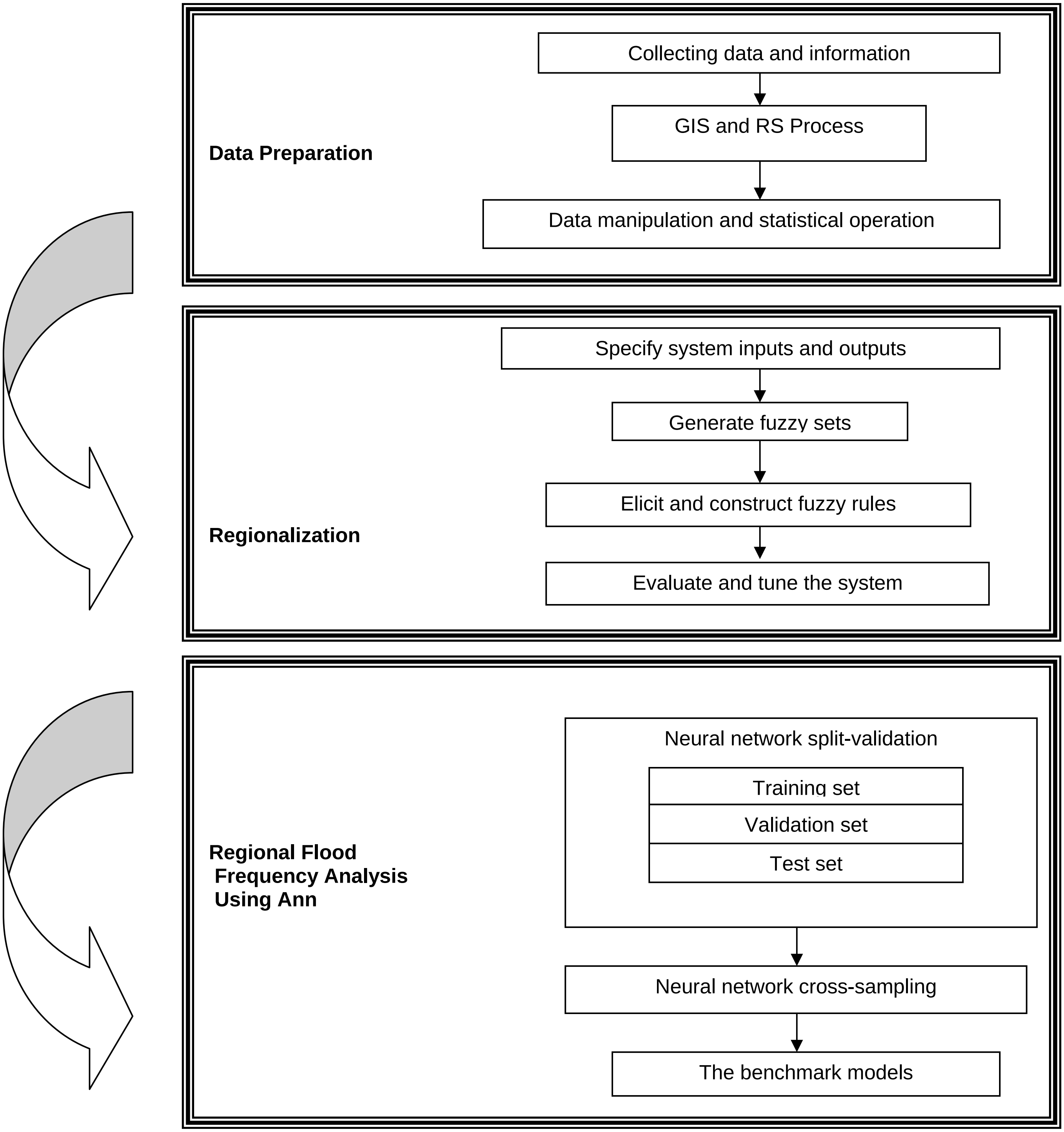
The research methodology as illustrated in figure 1 includes of three main steps of data preparation, regionalization and flood frequency analysis which is explained in detail below.

Step 1: Data Preparation

1) Collecting data and information. The required data, listed below will be collected from the Climate Organization, Water Affair Organization and Geographic and Mapping Organization of the country.

- Climatic and hydrologic data including mean daily, monthly and annual rainfall and temperature
  Peak and maximum discharge; Mean daily, monthly and annual discharge
  Satellite images and existing maps i.e. topographic, geologic and vegetation maps

2) GIS and RS Process:


Data Preparation
Collecting data and information
GIS and RS Process
Data manipulation and statistical operation
Regionalization
Specify system inputs and outputs
Generate fuzzy sets
Elicit and construct fuzzy rules
Evaluate and tune the system
Regional Flood Frequency Analysis Using Ann
Neural network split-validation
Training set
Validation set
Test set
Neural network cross-sampling
The benchmark models

- Image processing, digitizing maps and extracting catchment's attributes i.e. area, perimeter, elevation, slope, river network and time of concentration
- Mapping topographic, geologic, vegetation, climatic and hydrologic parameters

3) Data manipulation and statistical operation

Step 2: Regionalization

Regionalization is classification of catchments based on their hydrological similarities. It aims to provide a group of homogenous catchments for regional flood frequency analysis. There are several

methods of regionalization but "The Fuzzy Expert System" is applied in this research. The Fuzzy Expert System is the new method which is recently proposed and is applied for regionalization. The steps for building the FES are as follows. The proposed system intends to fully utilize the linkage between various catchment characteristics such as physiographic characteristics, flood seasonality and geographical distance with the hydrological response of a target site, to ensure that catchments pooled in a group have sufficient similarity in their hydrological characteristics. The proposed system has the capability of utilizing expert knowledge, or common sense that is represented by rules for the catchment grouping process.

2a) Specify system inputs and outputs and define the linguistic variables

Generally, candidates of pooling variables that form the system inputs can be selected from the following three categories of catchment descriptors, flood or rain seasonality, and geographical location. The output of the FES is the similarity (homogeneity) between the targets sites i and the candidate site j.

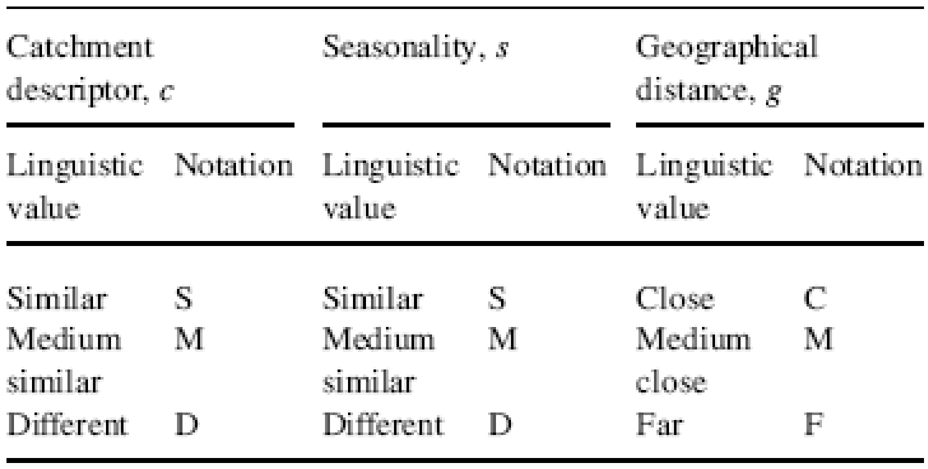
The output of the FES is the similarity (homogeneity) between the target site i and the candidate site j and is denoted using the symbol h: A larger

value in h means a higher similarity and the candidate site that generates this value will have a higher probability of entering the pooling group of the target site. Thus we have defined the three input variables c; s; g, and the output variable h: We classify each of the input variables into three categories and the output variable into seven categories. Generally, better inference results from defining more categories. However, if too many categories are defined, the construction of the rule base may be beyond the capability of the domain expert. The above three input variables and the output variable of the system and their linguistic values are thus described using the notation in Tables 1 and 2, respectively. As an example, we can read from Table 1 that the linguistic variable similarity in catchment descriptor c is associated with the following terms: {similar, medium similar, different}, and the three linguistic values can be represented by the notation {S, M, D}. Similarly, we can read the notations used for the linguistic variables s; g; and h from the two tables.

2b) Generate fuzzy sets

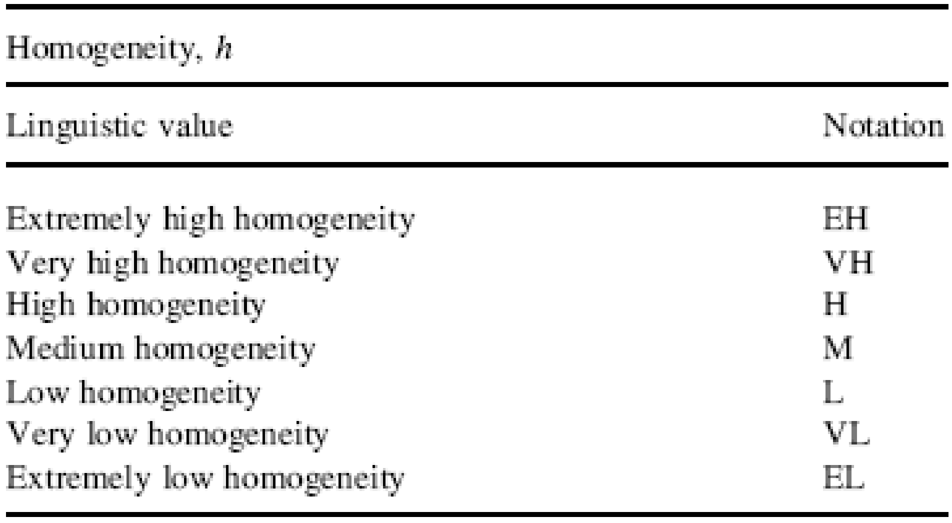
Fuzzy sets are a way of representing nonstatistical uncertainty and approximate reasoning (Cordo´n et al., 2001). Fuzzy set membership occurs by degree over the range over the range [0, 1]. Fuzzy sets can have a variety of shapes. The Gauss membership function is selected in this study; other functions such as triangle, sigmoid and bell function can also be used. The range of the three inputs is normalized to be within the range of [0, 1] by dividing the base numerical values by the corresponding maximum magnitudes.

Table 1) Input linguistic variables



Catchment descriptor, c		Seasonality, s		Geographical distance, g	
Linguistic value	Notation	Linguistic value	Notation	Linguistic value	Notation
Similar	S	Similar	S	Close	C
Medium similar	M	Medium similar	M	Medium close	M
Different	D	Different	D	Far	F

Table 2) Output linguistic variables



Homogeneity, h	
Linguistic value	Notation
Extremely high homogeneity	EH
Very high homogeneity	VH
High homogeneity	H
Medium homogeneity	M
Low homogeneity	L
Very low homogeneity	VL
Extremely low homogeneity	EL

2c) Elicit and construct fuzzy rules

In this study, we make use of a very basic relation between the catchment similarity in catchment descriptors, and the homogeneity between catchments, assuming that other input variables are fixed.

2d) Evaluate and tune the system

Building a FES is an iterative process that involves defining fuzzy sets and fuzzy rules, evaluating the system and then tuning the system to meet the specified requirement. A number of methods have been developed for the tuning purposes that touch every aspect of the FES structural design. In this study, our research is limited to the adjustment of the parameters of the membership functions. The performance of the FES can be improved, without changing the rule base, by properly tuning the fuzzy sets.

Step 3: Regional flood frequency analysis using Ann

3a) Neural network split-validation

It involves dividing available data into three sets:

1) A training set *(It is used to fit ANN model weights for a number of different network configurations and training cycles)*

2) A validation set *(It is used to select the model variant that provides the best level of generalization)*

3) A test set *(It is used to evaluate the chosen model against unseen data)*

3b) Neural network cross-sampling

It is the step to correct for deficiencies in the random division of the sample data sets and to address potential biases arising from urban and rural subsets a cross-sampling technique is also employed.

3c) The benchmark models

A step-wise multiple linear regression (SWMLR) model is used to predict the 10-, 20-, 30-year flood events.

3d) Error measures

In this research he common error measure, “The Mean Squared Relative Error (MSRE)”, is used.

$$MSRE = \frac{1}{n}\sum_{i=1}^{n}(\frac{Q_i - Q}{Q_i})^2$$

Where:

Q_i, The recorded peak discharge

Q, The estimated peak discharge

n, The number of recorded events

Assumption of the research

A common assumption in frequency analysis is that hydrologic extremes (floods or heavy precipitation) are generated by a random process. In summary the following assumptions are considered and tested in this thesis:

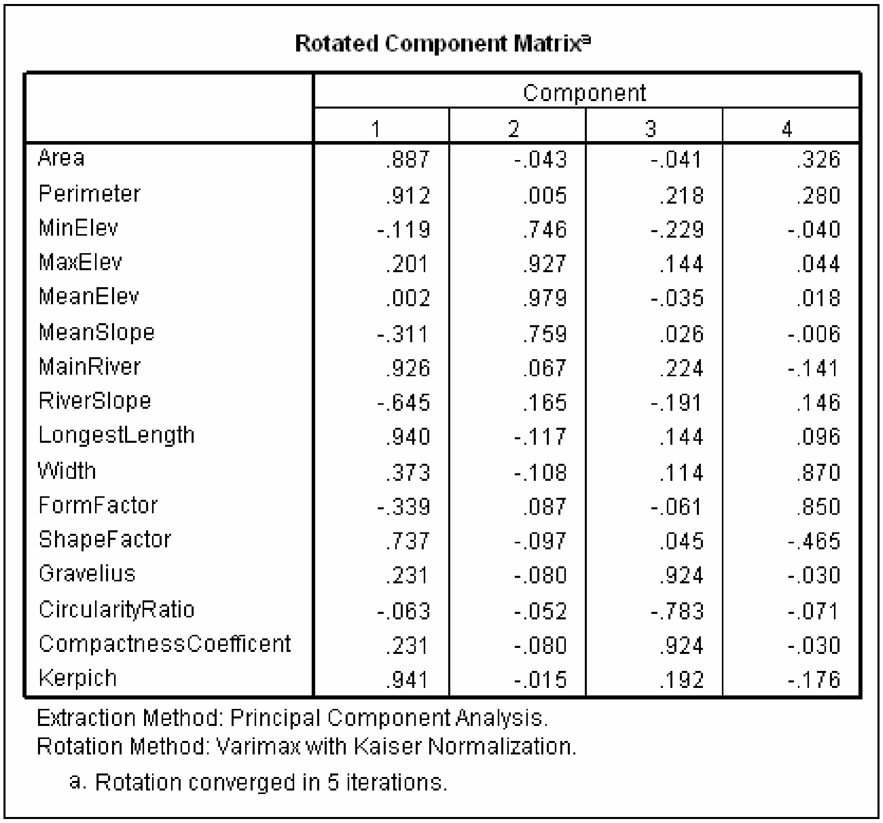
- Data records at each site are identically distributed and independent
- Data observed at different sites are independent
- Sites have the same probability distribution, except for a scale factor

Result and discussion

1. Data reduction

In order to select the most important factors in terms of hydrological homogeneity, the extracted characteristic of catchments were analyzed by the method of Factor Analysis using SPSS software. A number of 18 variables were studied including area; perimeter; minimum, maximum and mean elevation; mean slope, length and slope of Main River; time of concentration; form factor; Gravelious coefficient; length and width of longest route; circularity ratio and compactness factor. Four out of 18 variables were selected as the main factors which are perimeter, mean elevation, form factor, Gravelious coefficient. The method of Principle Component Analysis (PCA) with variomax rotation was used.

Table 3. Data reduction results for the studied parameters



Rotated Component Matrix[a]

	Component			
	1	2	3	4
Area	.887	-.043	-.041	.326
Perimeter	.912	.005	.218	.280
MinElev	-.119	.746	-.229	-.040
MaxElev	.201	.927	.144	.044
MeanElev	.002	.979	-.035	.018
MeanSlope	-.311	.759	.026	-.006
MainRiver	.926	.067	.224	-.141
RiverSlope	-.645	.165	-.191	.146
LongestLength	.940	-.117	.144	.096
Width	.373	-.108	.114	.870
FormFactor	-.339	.087	-.061	.850
ShapeFactor	.737	-.097	.045	-.465
Gravelius	.231	-.080	.924	-.030
CircularityRatio	-.063	-.052	-.783	-.071
CompactnessCoefficent	.231	-.080	.924	-.030
Kerpich	.941	-.015	.192	-.176

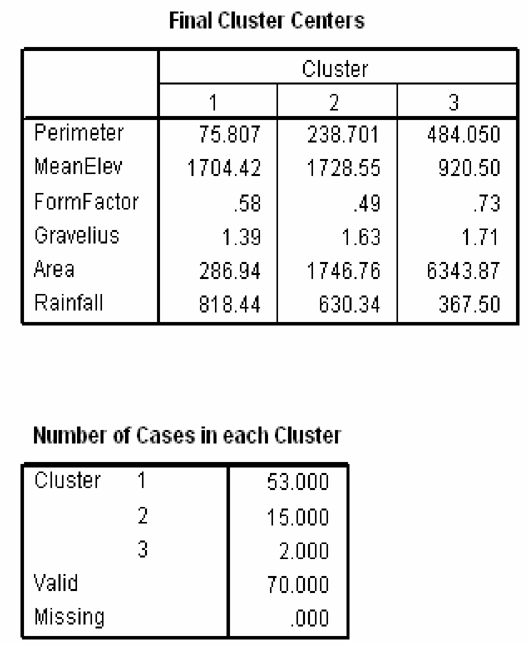
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

2. Classification of Sites

Cluster analysis aims to partition a set of objects into similar sub-sets. For these clusters the within-group dissimilarity should be minimised while maximizing dissimilarity between clusters. Catchments are partitioned using catchments characteristics so that an ungauged catchment can hopefully be allocated using the same catchment characteristics to one of the clusters. This study uses K-Means cluster analysis as the common method. Results indicate one main group of sites as well as two other smaller groups. Moreover hierarchical cluster analysis was used which is based on the assumption that the number of clusters is not known a priori. Ward's minimum variance linkage method together with the Euclidean distance similarity measure is used (Table 4 and Figure 2). Result obtained by this method is similar to the K-Means cluster analysis.

Table 4. Results of K-Means Cluster analysis in the study area



Final Cluster Centers

	Cluster		
	1	2	3
Perimeter	75.807	238.701	484.050
MeanElev	1704.42	1728.55	920.50
FormFactor	.58	.49	.73
Gravelius	1.39	1.63	1.71
Area	286.94	1746.76	6343.87
Rainfall	818.44	630.34	367.50

Number of Cases in each Cluster

Cluster	1	53.000
	2	15.000
	3	2.000
Valid		70.000
Missing		.000

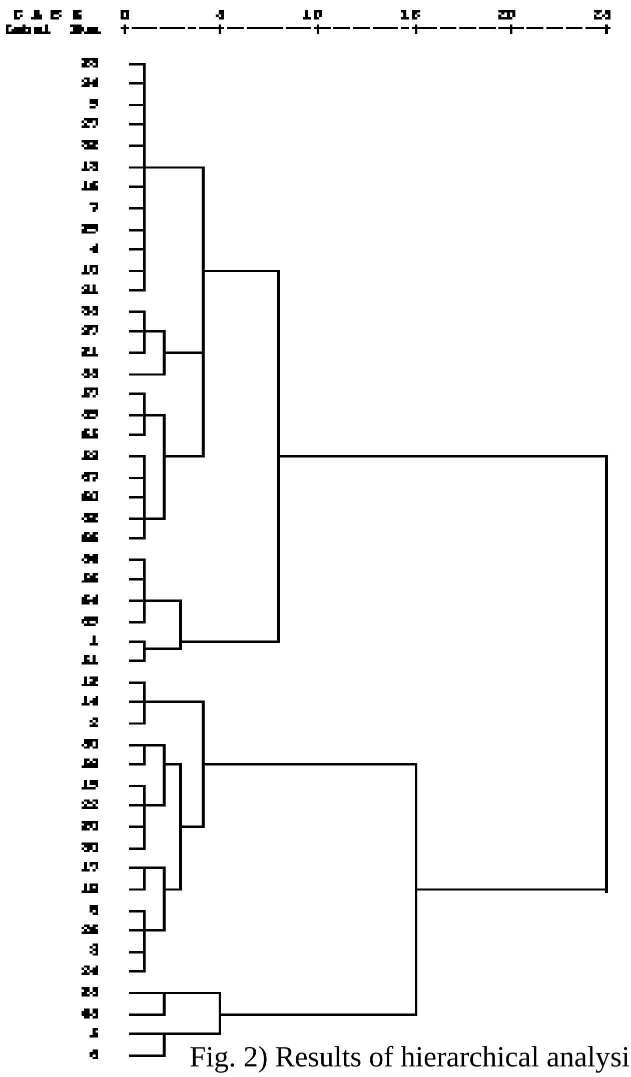


Fig. 2) Results of hierarchical analysis

3. Fuzzy Clustering

Fuzzy clustering generalizes partition clustering methods (such as k-means and medoid) by allowing an individual to be partially classified into more than one cluster. The distribution of membership of a catchment among the fuzzy clusters specifies the strength with which the catchment belongs to each region. Recent studies have shown that a soft membership function is essential for finding high-quality clustering. Hardening the results obtained by fuzzy algorithm produces better hard clustering solutions than those obtained by using the K-means algorithm.

Table 4. Fuzzy clustering results in the study area

Clusters	Distance	Silhouette	F(U)	Fc(U)	D(U)	Dc(U)
2	21.894526	0.241165	0.5000	0.0000	0.4997	0.9993
3	14.596351	-1.000000	0.3333	0.0000	0.6735	1.0103
4	10.947263	-1.000000	0.2500	0.0000	0.7594	1.0125
5	8.757810	-0.985714	0.2000	0.0000	0.8106	1.0133

- **Verifying results of catchments grouping using L-Moment technique**

The final step in classification is to verify the results. Validation of the results of cluster analysis aims at ascertaining whether the results are hydrologically sensible. This step is implemented using L-moment technique which was presented by Hosking and Wallis (1997). They presented the heterogeneity measure which compares the between-site variations in sample l-moments for the group of sites with what would be expected for a homogeneous region. The values of different heterogeneity measures H_1, H_2, H_3 are found 2.32, 0.63 and -1.32, respectively. Therefore the study region demonstrates acceptable homogeneity.

- **Distribution selection using the goodness-of-fit measure**

After confirming the homogeneity of the study region, an appropriate distribution needs to be selected for the regional frequency analysis. In other words, in a homogeneous region all sites should have the same population L-moments. The selection was carried out by comparing the moments of the candidate distributions to the average moments statistics derived from the regional data. The best fit to the observed data will indicate the most appropriate distribution. A number of five three-parameter distributions, i.e. Generalized Logistic, Generalized Extreme Value, Generalized Pareto, General Normal (LNIII) and Pearson Type III were fitted to the region. The value of ZDIST statistic for the study area for each three-parameter distribution was obtained (Table 5). It can be seen that the first three candidates are acceptable.

Table 5. Z–statistic for various distributions

Distribution	L-KURTOSIS	Z–statistic
GEN. LOGISTIC	L-KURTOSIS= 0.249	Z VALUE = 1.01 *
GEN. EXTREME VALUE	L-KURTOSIS= 0.223	Z VALUE = -0.15 *
GEN NORMAL	L-KURTOSIS= 0.200	Z VALUE = -1.20 *
PEARSON TYPE III	L-KURTOSIS= 0.161	Z VALUE = -3.00
GEN. PARETO	L-KURTOSIS= 0.152	Z VALUE = -3.40

- **Regional flood quantile estimation**

The next step in regional flood frequency is to estimate flood quantiles in the region. In this paper flood quantiles for each distribution is presented at the 90 percent level (Table 6). Moreover estimated parameters of each distribution is presented (Table 7).

Table 6. Quantile estimates with different probability for distributions accepted at the 90% level

Probability	0.010	0.020	0.050	0.100	0.200	0.500	0.900	0.950	0.990	0.999
GEN.LOGISTIC	0.071	0.128	0.228	0.331	0.474	0.821	1.796	2.314	3.995	8.430
GEN. EXTREME VALUE	0.112	0.162	0.247	0.335	0.464	0.810	1.847	2.367	3.884	7.165
GEN NORMAL	0.160	0.195	0.260	0.335	0.453	0.802	1.887	2.402	3.771	6.248
WAKEBY	0.28	0.98	0.290	0.389	0.474	0.786	1.916	2.435	3.719	5.768

Table 7. Parameters of the distribution

Distribution	Parameters		
GEN. LOGISTIC	$\xi=0.821$	$\alpha=0.308$	k=-0.314
GEN. EXTREME VALUE	$\xi=0.652$	$\alpha=0.414$	k =-0.213
GEN NORMAL	$\xi=0.802$	$\alpha=0.539$	k =-0.659
WAKEBY	$\xi=-0.211$	α =21.150	β =39.177
$\gamma=0.646$ $\delta=0.055$			

Significance of finding

- Mapping rainfall using Geostatistical Analysis indicates the Kriging (with Gaussian variogram model type) as the best method

- Data reduction using Principal Component Analysis with variomax rotation results in 4 parameters of perimeter, mean elevation, form factor and Gravelious coefficient as the basin attributes for homogeneity analysis. These factors as well as area and mean annual rainfall were used for basin hydrological classification
-

  The classification analysis using k-mean, hierarchical and fuzzy c-means methods deal with the two main homogeneous regions in the study area. This result is proved by proposed L-moment parameters

- Results of classification analysis will be applied for the next step of the research i.e. flood modeling using artificial neural network

References

1. Chang Shu, Donald H. Burn*. (2004). Homogeneous pooling group delineation for flood frequency analysis using a fuzzy expert system with genetic enhancement. Journal of Hydrology 291,132–149 Chow, V.T., Maidment, D.R., Mays, L.W., (1988). Applied Hydrology. McGraw-Hill, New York.
2. Dalrymple, T., (1960). Flood-frequency analyses. US Geological Survey Water-Supply Paper 1543-A.
3. Hosking, J.R.M., Wallis, J.R., (1997). Regional Frequency Analysis: An Approach Based on L-moments, Cambridge University Press, Cambridge, ISBN 0-521-43045-3.
4. Wiltshire, S.E., (1986). Regional flood frequency analysis II. Multivariate classification of drainage basins in Britain. Hydrological Sciences Journal 31 (3), 335–346.
5. Yi, S.Y., Chung, M.J., (1993). Identification of fuzzy relational model and its application to control. Fuzzy Sets and Systems 59, 25–33.
6. Zhang Jingyia, M.J. Hall. (2004). Regional flood frequency analysis for the Gan-Ming River basin in China. Journal of Hydrology 296, 98–117