J. Agr. Sci. Tech. (2011) Vol. 13: 1183-1196

Regionalization Approach for Extreme Flood Analysis Using L-moments

H. Malekinezhad[1*], H. P. Nachtnebel[2], and A. Klik[3]

ABSTRACT

Flood frequency analysis is faced with the problem of data and information limitation in arid and semi-arid regions. Particularly in these regions, the length of records is usually too short to ensure reliable quantile estimates. More than 75% of Iran is located in arid and semi-arid regions and despite the low annual precipitation, often large floods occur. One way to provide more information is to use many records from a region with similar flood behaviour, rather than only at-site data. This research is aimed to delineate homogeneous regions in the study area for further hydrological studies. Estimating regionalized parameters and identification of the best-fit distributions are the other specific objectives of the research. Several watershed attributes in relation to flood were characterized, among which the main characteristics were found by factor analysis. Later, preliminary identification of homogeneous regions was carried out using cluster analysis and region-of-influence approaches. The homogeneity test was done by H-statistic, a testing method based on L-moments. The results of this test showed that a subdivision of selected watersheds into homogenous groups is necessary. Therefore, three homogenous regions were formed. The Z-statistic based on L-moments and L-moment ratio diagrams were applied for identification of the best-fit distribution in each homogenous region. In the regionalization procedure five three-parameter distributions i.e. Generalized Logistic (GLO), Generalized Extreme Value (GEV), Generalized Pareto (GPA), three-parameter Lognormal (LN3), and Pearson type III (PE3) were fitted to the three homogeneous regions and the best-fit distributions were identified using L-moments approach. The results of goodness-of-fit analysis for the three regions indicates that the GEV, LN3 for the regions (1) and (2), and GLO and GEV distributions for the region (3) give acceptably close fits to the regional average L-moments. In general, the GEV distribution could be adopted as the appropriate distribution for the study area.

Keywords: Homogeneity test, L-moments, L-moments ratio diagram, Regionalization, Regional flood frequency.

INTRODUCTION

Flood frequency analysis is faced with the problem of data and information limitation in arid and semi-arid regions. Particularly in these regions, the length of records is usually too short to ensure reliable quantile estimates. Regional flood frequency analysis (RFFA) involves two major steps: (1) Grouping of sites into homogeneous regions, and (2) Regional estimation of flood quantiles at the site of interest. The performance of any regional estimation method strongly depends on the grouping of sites into homogeneous regions. Geographically contiguous regions have been used for a long time in hydrology, but have been criticized for being of arbitrary character. In fact, the geographical proximity does not guarantee hydrological similarity.

[1] Natural Resources Faculty, Yazd University, Yazd, Islamic Republic of Iran.
[*] Corresponding author, e-mail: hmalekinezhad@yazduni.ac.ir
[2] Institute of Water Management, Hydrology and Hydraulic Engineering, BOKU University, Vienna, Austria.
[3] Institute of Hydraulics and Rural Water Management, BOKU University, Vienna, Austria.

Regional flood frequency analysis is often used to enhance the estimation of flooding probabilities at locations that have short data record length relative to the return periods of interest. In such situations, extreme flow information from a number of sites can be used to compensate for an inadequate temporal representation of the extreme flows at a given location. Regional flood frequency analysis can, therefore, be employed at gauged locations, where information from similar sites that are gauged is used to assist with the characterization of the extreme flow regime at the ungauged sites.

An important requirement for regional flood frequency analysis is identification of the region that is used for the transfer of extreme flow information. A region, in this context, deals with the collection of catchments in a region, not necessarily geographically contiguous, that can be considered to be similar in terms of catchment hydrologic behaviour. The goal of the regionalization process can thus be defined as the identification of catchments that are similar enough to warrant the combination transfer of extreme flow information of sites within the region.

One of the most important and challenging steps in regional flood frequency analysis is the delineation of the homogeneous regions. Researchers have developed a number of regionalization techniques for objective determination of homogeneous regions (see e.g. Stedinger and Tasker, 1985: Acreman and Wiltshire, 1989; Burn, 1990; Burn, 1997; Hosking and Wallis, 1997; Shu and Burn, 2004a; Ouarda *et al*., 2006;and Shu and Ouarda, 2008).

L-moment ratio diagrams have become popular tools for regional distribution identification, and testing of outlier stations. Hosking and Wallis (1993) developed several tests for use in regional studies. They gave guidelines for judging the degree of homogeneity of a group of sites and for choosing and estimating a regional distribution. L-moment diagrams as a tool for identifying a regional distribution have been used in numerous other studies, including Pilon and Adamowski (1992), Vogel and Fennessey (1993), and Vogel *et al.* (1993a,b). An alternative test for homogeneity based on estimated dimensionless 10-year floods was developed by Lu and Stedinger (1992a). Chowdhury *et al.* (1991) compared several goodness-of-fit tests for the regional General Extreme Value (GEV) distribution and found that a new chi-square test based on the L-coefficient of variation and the L-coefficient of skewness outperformed other classical tests. Hussain and Pasha (2009) carried out L-moments based regional flood frequency analysis in Punjab, Pakistan.

In Iran, the application of L-moments on flood study was first used by Eslamian and Chavoshi (2003). They applied this technique to study flood frequency in central catchments of Iran. Eslamian and Feizi (2007) used L-moments for the selection of parent distributions to fit maximum monthly rainfall data of some sites in the Zayandehrood basin, Iran. Modarres (2008) carried out a regional low flow frequency analysis in the north of Iran using L-moments. Modarres (2009) also investigated the spatial variation and regional frequency distribution of annual maximum dry spell length for Isfahan Province in Iran, using both L-moment and multivariate analysis.

The main objectives of this paper are: 1) to determine the homogeneous regions: 2) to estimate regionalized parameters, and 3) identification of the best-fit distributions on the peak floods in several gauging stations in Namak-Lake basin in central Iran.

MATERIALS AND METHODS

Description of the Study Area

In Iran, a number of extreme floods have occurred over the last 50 years and caused great economic losses. Nowadays, the areas prone to flooding are highly populated and industrialised and develop very fast. Many annual flood series are too short to allow for a reliable estimation of extreme events or there is no flow record available at the site of interest.

The study area refers to the Namak-Lake basin located in Central Basin of Iran, with the geographic coordinates ranging from 48° 20'E to 52° 40'E longitude and 32° 00'N to 36° 30'N latitude (Figure 1). The total area of Namak-Lake basin is 89650 km^2 with minimum and maximum elevation of 800 m and 4375 m above mean see level in Namak Lake and Jajrud Heights, respectively. The Karaj, Jajrud, Shour, Qarechai, and Qomrud are the main rivers of the Namak-Lake basin.

The hydrologic and climatic data, such as annual flood series, were firstly provided for the 71 flow gauging stations of the Namak-Lake basin in the central part of Iran. These data were obtained from the Water Resources Research Organisation of Iran. Most of these gauging stations have short data series with a large number of missing data or have been constructed in recent years with the data length of less than 10 years. Selection of the basins was such that at least 15 years of historical flood data were available. The average number of years of record for the stations was 27 with a range of 15 to 36 years. After preliminary screening of the sites 21 gauging stations and 14 basin characteristics and hydrologic variables (Table 1) were selected for this study.

Factor Analysis

In this study, the Factor Analysis (FA) was used to find the most important independent factors that can summarize and reduce the spatial variations of data into more important factors. Many statistical methods are used to study the relation between independent and dependent variables. Results obtained by FA are necessarily more hypothetical and tentative than when independent variables are observed directly. In particular, it seeks to discover if the observed variables can be explained, largely or entirely, in terms of a much smaller number of variables called factors. Since the fourteen site characteristics have different units, all data sets were standardised. This rescaling effectively gave equal weight to each site characteristic in determining the main variables and clusters. The 21×14 matrix of selected standardised variables was analysed.

Identification of Homogeneous Regions (IHR) Watershed Grouping by Cluster Analysis

Cluster Analysis (CA) is a standard method of statistical multivariate analysis for dividing a data set into groups and has been successfully used to form regions for

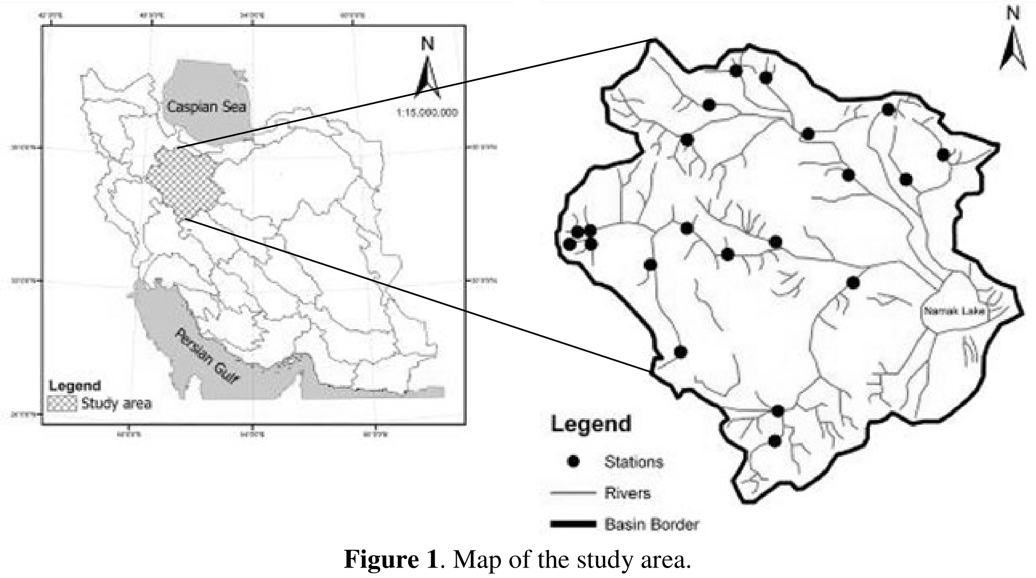

Figure 1. Map of the study area.

[DOR: 20.1001.1.16807073.2011.13.7.1.6]

Table 1. Basin characteristics or at-site variables selected in this study.

Basin characteristics or at-site variables	*symbol*	*Basin characteristics or at-site variables*	*symbol*
Basin area	A	Main river's mean slope	SW
Basin perimeter	P	Time of concentration	Tc
Basin mean elevation	Hm	Mean annual rainfall	Pm
Basin mean slope	Sb	Mean annual maximum daily rainfall	P24
Main waterway length	Lw	Mean annual temperature	Tm
Compactness coefficient	Gr.	Percentage of permeable area	Ppf
Diameter of basin equivalent circle	De	Mean annual peak flood	$\overline{Q}$

regional frequency analysis. A data vector is associated with each site, and sites are partitioned or aggregated into groups according to the similarity of their data vectors. The data vector can include at-site statistics, site characteristics, or some combinations. Clustering techniques have been applied to a wide variety of research problems. CA allows many choices about the nature of the algorithm for combining groups. Each choice may result in a different grouping structure. Hosking and Wallis (1997) recommended using methods that rely on site characteristics only when identifying homogeneous regions, and subsequently using the at-site characteristics to independently test the homogeneity of the proposed regions. They recommended using Ward's method, which is a hierarchical clustering method based on minimizing the Euclidean distance in site characteristics space within each cluster. Soltani and Modarres (2006) used a hierarchical and divisive cluster analysis to categorize patterns of rainfall in Iran.

Many statistical types of software perform cluster analysis method. The Ward's clustering analysis is used for this stage of the study. First, a preliminary determination of homogeneous regions was done by Ward's clustering method and region-of-influence approach as two well known approaches for determination of homogeneous regions in regional flood frequency analysis. Then, the results of these two approaches were compared. At the next steps, the statistical test based on L-moment ratios proposed by Hosking and Wallis (1993) was used for testing the heterogeneity of the proposed regions.

Watershed Grouping by Region-of-influence Approach

In this study, region-of-influence approach was also used for grouping the selected sites, and then the results were compared with the results of cluster analysis and L-moments approach. The variables used for this step include four characteristics obtained by method of factor analysis. The matrix of D_{ij} statistic between every two sites located in the study area was made, and then the first 8 sites that have lower Dij distance with the site of interest were ranked. The unique subset of gauged sites is defined as the N 'nearest' gauge to the gauged site of interest or an ungauged site, where distance between sites i and j is defined by Euclidean distance metric:

$$D_{ij} = \left\{ \sum_{k=1}^{p} \left[\frac{x_{ik} - x_{jk}}{S(x_k)} \right]^2 \right\}^{1/2} \qquad (1)$$

Where,

D_{ij} is the distance between sites i and j in terms of basin characteristics,

p is the number of basin characteristics used to calculate D_{ij} ,

X_k is the kth basin characteristics,

$S(X_k)$ is the sample standard deviation for X_k, and

X_{ik} and X_{jk} are the value of X_k at the sites of i and j.

Homogeneity of the Regions Using H-Statistic

Homogeneity test is aimed to estimate the degree of homogeneity in a group of sites. In this study, H-statistic with the measure of L-CV was used. H-statistic is a statistical test based on L-moment ratios. The H-statistic indicates that a region is acceptably homogeneous when H < 1; possibly heterogeneous when 1 < H < 2 and definitely heterogeneous when H > 2.

Hosking and Wallis (1993) proposed a statistical test based on L-moment ratios for heterogeneity of the proposed regions. The test compares the between-site variation in sample L-Cv with the expected variation for a homogeneous region. The method fits a four parameters kappa distribution to the regional average L-moment ratios. The estimated kappa distribution is used to generate 500 homogeneous regions with population parameters equal to the regional average sample L-moment ratios. The properties of the simulated homogeneous region are compared to the sample L-moment ratios.

$$H = \frac{V_{obs} - \mu_V}{\sigma_V} \tag{2}$$

Where, μ_V and σ_V are the mean and standard deviation of the simulated values of V, respectively, while V_{obs} is calculated from the regional data and is based on a corresponding V-Statistic. For the sample and simulated regions, respectively, V is calculated as a measure of dispersion for the L-CV:

$$V = \frac{\sum_{i=1}^{N} n_i \left(\tau^{(i)} - \bar{\tau}\right)^2}{\sum_{i=1}^{N} n_i}$$

Where, $\bar{\tau}$ and $\tau^{(i)}$ are the group mean of L-CV and, L-CV of observed data at site I; N and n are the number of sites in the pooling group and the data length at each site.

Description of L-moments

L-moments have been widely used in flood frequency analysis. L-moments and L-moment ratios are more convenient than probability-weighted moments because they are more easily interpretable as measures of distributional shape (Hosking, 1994). Similar to ordinary product moments, the purpose of L-moments and probability weighted moments is to summarize theoretical distributions and observed samples. Hence, L-moments can be used for parameter estimation, interval estimation and hypothesis testing (Vogel *et al.*, 1993a). Hosking (1990) introduced L-moments as expectations of linear combinations of order statistics. Certain linear combinations of expectations of order statistics have been shown to be very useful in statistical parameter estimation as compared to other standard methods, e.g., the method of moments, maximum likelihood, and least squares. The main advantage of L-moments is that, being a linear combination of data, they are less influenced by outliers, and the bias of their small sample estimates remains fairly small (Pandey *et al.*, 2001). Unbiased sample estimators (β_i) of the first four probability weighted moments (PWMs) are given as follows (Hosking and Wallis, 1997):

$$\begin{cases} \beta_0 = \frac{1}{n}\sum_{j=1}^{n} X(j) \\ \beta_1 = \sum_{j=1}^{n-1}\left[\frac{n-j}{n(n-1)}\right]X(j) \\ \beta_2 = \sum_{j=1}^{n-2}\left[\frac{(n-j)(n-j-1)}{n(n-1)(n-2)}\right]X(j) \\ \beta_3 = \sum_{j=1}^{n-3}\left[\frac{(n-j)(n-j-1)(n-j-2)}{n(n-1)(n-2)(n-3)}\right]X(j) \end{cases} \tag{3}$$

Where, X(i) represents the ranked annual maximum series (AMS) with X(1) being the highest value and X(n) the lowest value, respectively (Vogel *et al.*, 1993b) . Next, the first four L-moments are given as

[DOR: 20.1001.1.16807073.2011.13.7.1.6]

$$\begin{cases} \lambda_1 = \beta_0 \\ \lambda_2 = 2\beta_1 - \beta_0 \\ \lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{cases} \quad (4)$$

Unbiased sample estimators of the first four L-moments are obtained by substituting the PWM of sample estimators from Eq. (3). The L-mean, λ_1, is a measure of central tendency and the L-standard deviation, λ_2, is a measure of dispersion. Their ratio, λ_2 / λ_1, is termed as the L-coefficient of variation, or L–C.V., τ. The ratio, λ_3 / λ_2, is referred to as τ_3 or L-skewness, while the ratio λ_4 / λ_2 is referred to as τ_4 or L-kurtosis. L-skewness refers to the distribution symmetry with respect to its dispersion from the mean, and L-kurtosis refers to the weight of the tail of a distribution.

$$\tau = \frac{\lambda_2}{\lambda_1}, \ \tau_3 = \frac{\lambda_3}{\lambda_2}, \ \tau_4 = \frac{\lambda_4}{\lambda_2} \quad (5)$$

Description of the Candidate Distributions

In regional frequency analysis, a single frequency distribution is fitted to the observed data from several sites. In general, when a region is slightly heterogeneous, there will be no single "true" distribution that applies to each site. The aim is therefore not to identify a "true" distribution, but to find a distribution that will yield more accurate quantile estimates for each site. There are many families of distributions that might be candidates for being fitted to a regional data set. Their suitability as candidates can be evaluated by considering their ability to produce features of the data that may be of importance in any given application. The delineation of homogeneous regions is closely related to the identification of the common regional distributions that apply within each region. A region can only be considered homogeneous if sufficient evidence can be established that the data at different sites in the region are drawn from the same parent distribution (except for the scale parameter).

After determination of the main influential variables by FA, discordancy and homogeneity tests were done by L-moment-based measures. Five three-parameter distributions i.e. Generalized Logistic (GLO), Generalized Extreme Value (GEV), Generalized Pareto (GPA), Three-Parameter Lognormal (LN3), and Pearson type III (PE3) were fitted to the three homogeneous regions. The parameters of these distributions were estimated by the L-moments approach.

Goodness-of-fit Measure

Several methods are available for testing the goodness of fit of a distribution to data from a single sample. These include quantile-quantile plots, chi-squared, Kolmogrov-Smirnov, and other general goodness-of-fit tests and tests based on moment or L-moment statistics. Some of these methods can be adapted for using in the regional network. The fit of a postulated regional frequency distribution to each site's data can be assessed by goodness-of-fit statistics calculated at each site, and the resulting statistics then combined into a regional goodness-of-fit statistic. The goodness-of-fit criterion for each of the various distributions is defined in terms of L-moments and is termed the *Z*-statistic (Hosking and Wallis, 1997):

$$Z^{DIST} = \frac{\tau_4^{DIST} - \overline{\tau}_4 + \beta_4}{\sigma_4} \quad (6)$$

Where, DIST refers to a particular distribution, while β_4 and σ_4 are the bias and standard deviation of τ_4 (*L*-kurtosis), respectively, defined as:

$$\beta_4 = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} \left(\tau_4^m - \overline{\tau}_4 \right) \quad (7)$$

$$\sigma_4 = \left[\frac{1}{N_{sim} - 1} \left(\sum_{m=1}^{N_{sim}} \left(\tau_4^m - \overline{\tau}_4 \right)^2 - N_{sim} \beta_4^2 \right) \right]^{1/2} \quad (8)$$

[DOR: 20.1001.1.16807073.2011.13.7.1.6]

Where, N_{sim} is the number of simulated regional data sets generated using a Kappa distribution in a similar way as for the heterogeneity statistic. Note that the subscript *m* denotes the *m*th simulated region. A fit is declared adequate if Z^{DIST} is sufficiently close to zero, a reasonable criterion being $|Z^{DIST}| \leq 1.64$. The performance of Z as a goodness-of-fit measure was assessed by means of Monte Carlo simulation experiments. Data were simulated from homogeneous regions with one of the five three-parameter frequency distributions.

RESULTS AND DISCUSSION

In this study, factor analysis was used for determination of the main variables affecting flood magnitude. The first four factors, accounting for 85.8% of the total variance, were selected and subjected to Varimax Normalised Rotation in FA approach. This method of rotation is widely accepted as the most appropriate type of orthogonal rotation for climatic data. Factor loadings greater than 0.7 (Dinpashoh, 2004) were considered as important loadings. Factor scores for each of the 21 sites were calculated from the standardized variables and the associated factor loadings. Factor analysis technique was utilized to analyze 14 independent variables in 21 basins. It was found that the variables could be summarised in four factors, containing 85.8% of the main data variations. The contribution of each rotated factor has been illustrated in Table 2. The method of principal components and varimax rotation was used to extract the factors loading matrix.

Due to the larger values of correlation coefficients in the factors, Main Waterway Length, Mean Annual Precipitation, Compactness (Gravellius) Coefficient, and Mean Annual Temperature with the factor loadings of 0.954, 0.84, -0.864, and 0.872 are, respectively, the most important variables for the factor (1), factor (2), factor (3), and factor (4), that were selected in the study area. Although basin area is one of the most important physical characteristics in flood frequency analysis, but, as shown in Table 2, the results of the present investigation indicate that the main waterway length has larger factor loading compared to the area and other main characteristics of factor (1). Therefore, main waterway length is selected as the most effective characteristic of factor (1).

After selecting the four most effective physiographical or hydrological variables in relation to the peak flood magnitude, the

Table 2. Rotated Factor Loadings (Varimax Rotation).

Variable	*Factor1*	*Factor2*	*Factor3*	*Factor4*
A	0.918	-0.042	-0.132	0.191
P	0.948	-0.043	-0.208	0.129
Hm	-0.256	0.791	-0.014	0.082
Sb	-0.499	0.626	-0.456	-0.024
Lw	**0.954**	-0.083	-0.093	0.124
Cc.	0.305	0.037	**-0.864**	-0.090
De	0.918	-0.042	-0.132	0.191
Sw	-0.890	0.320	-0.141	-0.016
Tc	0.955	-0.176	-0.006	0.085
Pm	0.176	**0.840**	0.189	-0.241
P24	0.151	0.763	-0.319	0.489
Tm	0.260	-0.109	0.110	**0.872**
Ppf	0.332	-0.619	0.027	0.259
Q2	0.683	0.458	0.425	-0.095

21×4 matrix of sites and related variables, obtained by FA, was subjected to hierarchical clustering based on the Ward's method using Euclidean distance (Fovell, 1997). The Ward's method is one of the most frequently used hierarchical clustering techniques for climatic classification (Kalkstein *et al*., 1987). It seems that there is no general guide for selecting the number of clusters (Fovell, 1997; Dinpashoh *et al*., 2004).

As shown by the dendrogram in Figure 2, 21 sites were divided into 3 homogeneous regions. The dendrogram was cut at the distance of six to define initial homogeneous regions at the study area.There was no significant difference between the results of region-of-influence approach and Ward's method of cluster analysis, confirmed by L-moments homogeneity test. Basins 41-069 and 41-089 have been replaced in regions (2) and (3), representing the only difference between the results of these two approaches. The site 41-089 has an area of 53 km^2 and all of the sites smaller than 100 km^2 are located in region (2). Therefore, the results of cluster analysis were preferred for the next steps in this study.

Table 3 shows that most of the large sites are located in region (1), with the range of 1655 to 2768 km^2. The intermediate sites are located in region (3), with only one site i.e. 41-045, with an area of 1902 km^2, which is between the lowest and the highest range values of region (1). Most of the small sites are located in region (2) and have an areas less than 200 km^2.. In general, the area of the site has interactive relations with many of its other characteristics like perimeter, basin slope, and the main channel length and slope.

The degree of homogeneity for the sites all together and for each identified homogeneous region is shown in Table 4. The value of H-statistic for all 21 sites was 3.29, indicating that the whole region was not homogeneous, therefore, a subdivision of the sites into homogeneous regions was necessary. The results of H-statistics showed that the absolute values of H-statistic for all three identified regions were less than 1 and, therefore, all three regions were acceptably homogeneous.

Estimation of Distribution Parameters Using L-moments

In the L-moments approach, the three parameters (location, scale, shape) of each

Table 3. The final subdivision of the sites into three homogeneous groups.

Group (1)		*Group (2)*		*Group (3)*	
Site code	area (km^2)	Site code	area (km^2)	Site code	area (km^2)
41-047	2768.0	41-035	165.0	41-009	850.0
41-059	1655.0	41-041	41.0	41-045	1902.0
41-067	2520.0	41-043	161.7	41-049	255.0
41-073	1925.0	41-063	96.0	41-051	420.0
Total area = 8868.0		41-089	53.0	41-069	450.0
		41-101	718.0	41-079	560.0
		41-109	196.0	41-095	360.0
		41-117	421.0	41-131	416.0
		41-219	104.0	Total area = 5213.0	
		Total area = 1955.0			

Table 4. The degree of homogeneity for each region.

Region	$\overline{\tau}$	*V*	*H*	*Degree of homogeneity*
All sites	0.46	0.31	3.29	Definitely heterogeneous
1	0.48	0.09	- 0.31	Acceptably homogeneous
2	0.42	0.08	0.46	Acceptably homogeneous
3	0.49	0.14	0.78	Acceptably homogeneous

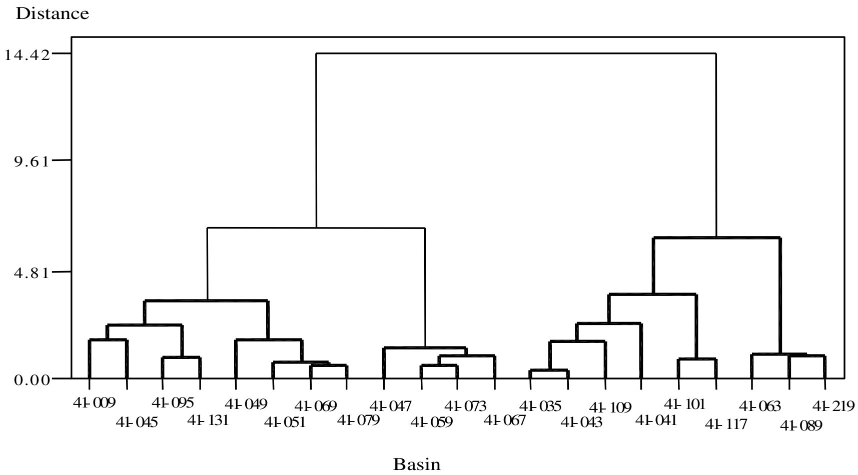

Figure 2. Dendrogram of clustered basins by Ward's method.

probability distribution in regional flood frequency analysis are obtained by the regional averages of the L-moments and L-moment ratios. These parameters are needed for estimating the peak flow of each recurrence interval of interest by the selected best-fit distribution. The results of the estimated regional location, scale and shape parameters for the five distributions fitted to the standardised data of each homogeneous region are given in Table 5.

The values of the parameters depend on the shape of density and cumulative functions of different distributions and their parameter estimation formulas.

The Z-statistic was used as a goodness-of-fit measure for identification of the common regional distribution that applies within each region. This statistic is defined in terms of L-moment ratios. The five three-parameter

Table 5. The estimated regional parameters for five distributions fitted to the standardized data of regions (1), (2) and (3).

distribution		*parameters*		
		Location	*Scale*	*Shape*
Region(1)	GLO	0.69	0.33	-0.44
	GEV	0.49	0.40	-0.38
	GPA	0.16	0.65	-0.22
	LN3	1.11	0.50	-0.94
	PE3	1	0.63	2.64
Region(2)	GLO	0.75	0.30	-0.41
	GEV	0.50	0.34	-0.34
	GPA	0.26	0.61	-0.16
	LN3	1.03	-0.40	-0.87
	PE3	1	0.56	2.44
Region(3)	GLO	0.64	0.34	-0.52
	GEV	0.41	0.34	-0.48
	GPA	0.15	0.53	-0.37
	LN3	1.44	0.58	-1.15
	PE3	1	1.24	3.22

distributions were fitted to the three homogeneous regions. The results of Z-statistic is summarised in Table 6.

As shown in Table 6 GEV with the absolute value of 0.85 for Z-statistic has the best goodness-of-fit with the data at region (1) and LN3 and GP are the other acceptable distributions. The Z value of these distributions is less than 1.64, and the GEV has the lowest value of Z-statistic. The goodness-of-fit measures of GLO and PE3 are greater than 1.64 and these distributions should not be considered as regional distribution.

For the region (2), the Z value for GEV is 0.43 and this distribution has the best goodness-of-fit with the data. Also, the LN3 is the second acceptable distribution. The Z value of these distributions is less than 1.64 and the GEV has the lowest value of Z statistic. The Z values of GPA and PE3 distributions are too high, therefore, they should not be considered as regional distribution.

For the region (3), GLO with the absolute value of 0.76 for Z is the best distribution for regional flood frequency analysis. GEV and PE3 have a Z value less than 1.64 and are also acceptable. The other two distributions have a Z value greater than 1.64 and should not be considered as regional distribution.

The results of goodness-of-fit analysis for the three regions indicate that the GEV, LN3 for the regions (1) and (2), and GEV and GLO distributions for the region (3) give acceptably close fits to the regional average L-moments. In general, the GEV distribution could be adopted as the appropriate distribution for the study area.

L-moments Diagrams

The suitability of the distributions for describing the annual peak flood series in the study area was also investigated through L-moment diagrams according to the procedure described by Vogel and Fennessey (1993). In L-moment diagram, the sample points, L-skewness Versus L-kurtosis, should be distributed above and below the theoretical line of a suitable distribution.

The values of the τ_3 versus τ_4 for the different sites located at the study area are plotted in Figure 3, where the relationship

Table 6. Goodness-of-fit analysis (Z^{DIST}) for five different frequency distributions when applied to the three groupings of sites.

Distribution	*Homogeneous region*		
	(1)	*(2)*	*(3)*
GLO	2.54	1.22	- 0.76
GEV	- 0.85	0.43	- 0.91
GPA	1.52	- 2.03	- 3.18
LN3	1.02	- 0.83	- 2.73
PE3	2.03	1.61	- 1. 36

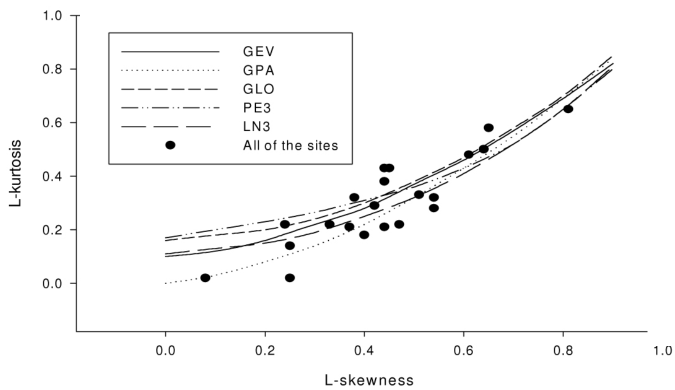

Figure 3. L-moments diagrams for all the 21 sites.

between population L-skew and L-kurtosis and the theoretical lines for the five distributions are shown. As depicted in this figure, there is no evidence of tendency of the sites toward the theoretical line of distributions. AMS in region (3) have higher skewness and kurtosis than the AMS in region (2), and the regional average L-kurtosis in region (1) is lower than the other two regions. The same results can be concluded using L-moments ratio diagram for the three regions (Figure 4a-c). Apparently, the empirical points for different data series of sites in the region (1) are well distributed around the theoretical

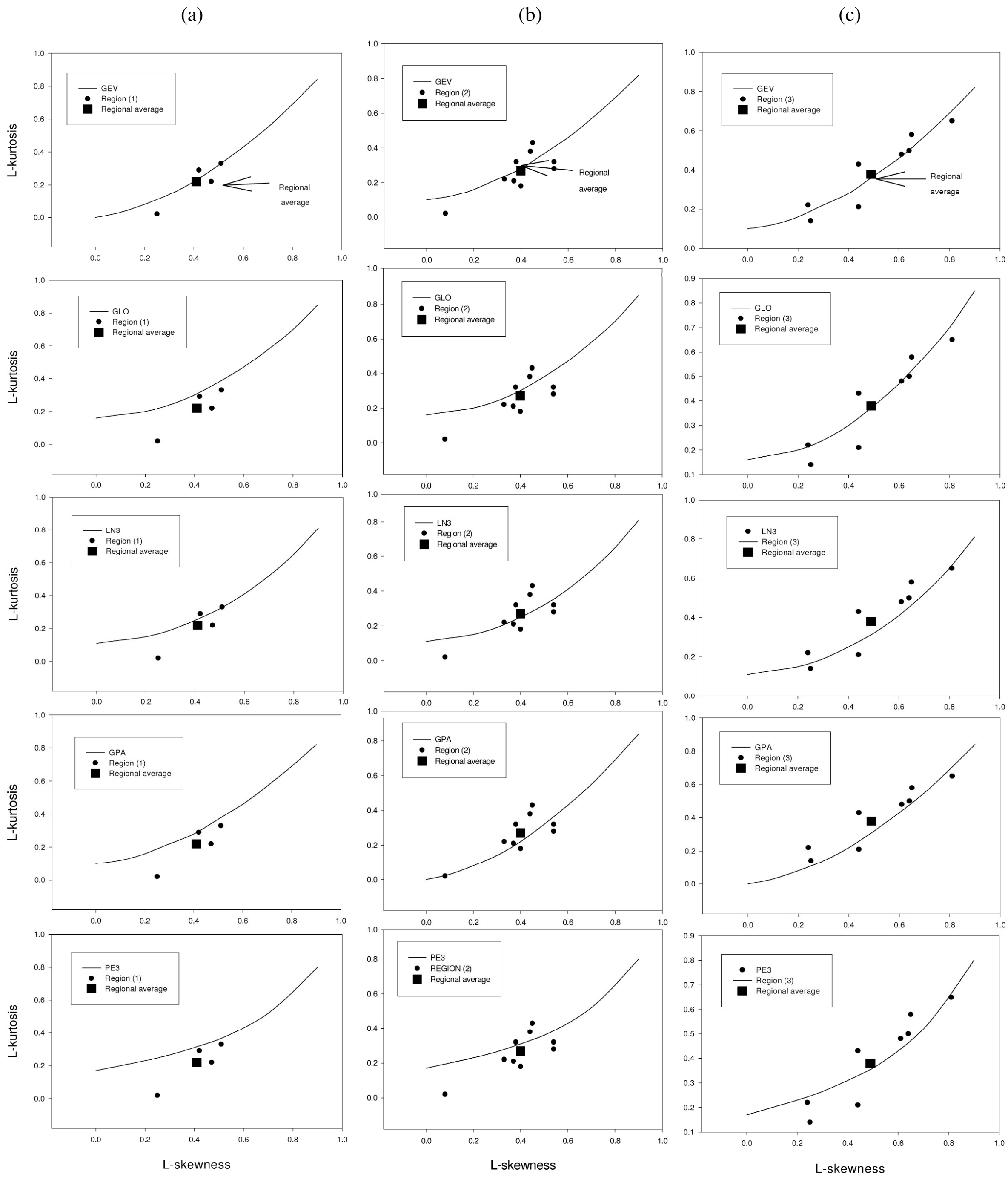

Figure 4. L-moments diagrams for five distributions in regions (a), (b) and (c).

curve of the GEV and LN3 distributions (Figure 4-a). The point representing the regional average of τ_3 versus τ_4 of GEV distribution falls on the theoretical curve of this distribution; therefore, it appears that the GEV distribution would be the best-fit distribution for the region (1), while GPA, GLO, and PE3 distributions should not be considered.

For the region (2), the empirical points for extreme flood data of gauging stations are well distributed around the theoretical curve of the GEV and LN3 distributions (Figure-b). The point representing the regional average of τ_3 versus τ_4 falls on the theoretical curve of these distributions; therefore, it appears that the GEV distribution would be the best-fit distribution for this region. For the region (3), GLO and GEV are acceptable distributions for regional flood analysis (Figure 4-c).

CONCLUSION

The results of this study show that application of L-moments in regionalization approach for estimating the peak flood with a specific recurrence interval is convenient in hydrology and water resources engineering. Particularly in arid and semi-arid regions, the length of records is usually too short to ensure reliable quantile estimates. L-moments are used for homogeneity test, distribution parameters estimation, and selection of the best-fit distribution for a hydrologically homogeneous region.

The Z-statistic, as a goodness-of-fit measure for the identification of common regional distribution that applies within each region, is defined in terms of L-moment ratios. The results indicate that this measure properly identifies the best-fit distribution to extreme floods in a homogeneous region. Also, the results show that the GEV has the best goodness-of-fit with the data in the region (1), and LN3 and GP distributions are the other acceptable distributions (Figure 3).

As shown in Figure 4-b, for region (2), the sample points have better distribution above and below the theoretical lines of the GEV and LN3 distributions, and the regional average points of τ_3 versus τ_4 for these two distributions fall very close to their theoretical curves. For this region, GEV and LN3 distributions could be adopted as the regional distributions. The AMS in this region have lower skewness and kurtosis than the AMS in region (3).

As depicted in Figure 4-c, GEV and GLO could be considered for the region (3). The regional average L-skewness-L-kurtosis points are located on the theoretical lines of these two distributions. Therefore, both distributions could be adopted as the regional distribution. The choice of a suitable standard frequency distribution is often controversial. Comparison of the results between the goodness-of-fit test and L-moment diagrams and the widespread acceptance of GEV distribution indicates that the latter could be selected as the best distribution for the study area.

REFERENCES

1. Acreman, M.C. and Wiltshire, S.E. 1989. The Regions Are Dead; Long Live the Regions. Methods of Identifying and Dispensing With Regions for Flood Frequency Analysis, IAHS Publication No. 187 (1989), pp. 175–188.
2. Burn, D.H. 1997. Catchment Similarity for Regional Flood Frequency Analysis Using Seasonality Measures. *J. Hydrol.*, **202**: 212 – 230.
3. Burn, D.H.1990. Evaluation of Regional Flood Frequency Analysis with a Region of Influence Approach, *Water Resour. Res.*, **26 (10)**: 2257 – 2265.
4. Chowdhury, J. U., Stedinger, J. R. and Lu, L. H. 1991. Goodness-of-fit Tests for Regional Generalized Extreme Value Flood Distribution. *Water Resour. Res.* **27(7)**:1765-1777.
5. Dinpashoh, Y. A., Fakheri-Fard, M. Moghaddam, S. Jahanbakhsh and Mirnia, M. 2004. Selection of Variables For The Purpose of Regionalization of Iran's Precipitation

Climate Using Multivariate Methods. *J. Hydrol.*, **297**:109 - 123.

6. Eslamian, S. S. and Chavoshi S. 2003. The Application of L-moments Theory in Flood Frequency Analysis of Central Catchments of Iran. *JWSS* **7(1)**: 1-17.
7. Eslamian, S. S. and Feizi, H. 2007. Maximum monthly rainfall analysis using L-moments for an Arid Region in Isfahan Province, Iran, *J. Appl. Meteor. Climatol.*, **46(4)**:494-503.
8. Fovell, R. G. 1997. Consensus Clustering Of U.S. Temperature and Precipitation Data. *J. Climate*, **10**:1405–1427
9. Hosking, J. R. M. 1994. The Four Parameter Kappa Distribution. IBM, *J. Res. Develop.*, **38**: 251 – 258.
10. Hosking, J. R. M. 1990. L-moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *J. Royal Statist. Soc.,* **52**: 105 - 124.
11. Hosking, J. R. M. and Wallis, J. R., 1997. Regional Frequency Analysis: An Approach Based on L-Moments, Cambridge University Press, London, UK.
12. Hosking, J. R. M. and Wallis J. R. 1993. Some Statistics Useful in Regional Frequency Analysis. *Water Resour. Res.*, **29**: 271 – 278.
13. Hosking, J. R. M. and Wallis J. R., and Wood E. F. 1985b. Estimation of the Generalised Extreme-Value Distribution by the Method of Probability-Weighted Moments. *Technometrics*, **27**: 251 – 261.
14. Hussain, Z. Pasha, G. R. 2009. Regional Flood Frequency Analysis of the Seven Sites of Punjab, Pakistan, using L-moments. *Water Resour. Manage.* **23**: 1917-1933.
15. Kalkstein, L. S. Tan, G. and Skindlov, J. A. 1987. An Evaluation of Three Clustering Procedures for Use in Synoptic Climatological Classification. *J. Climate Appl. Meteor.*, **26**:717–730.
16. Lu, L. H. and Stedinger, J. R. 1992a. Sampling Variance of Normalized GEV/PWM Quantile Estimation and a Regional Homogneity Test. *J. Hydrol.*, **138**: 223 - 245.
17. Lu, L. H. and Stedinger, J. R. 1992b. Variance of Two- and Three-Parameter GEV/PWM Quantile Estimators: Formulae, Confidence Intervals, and a Comparison. *J. Hydrol.*, **138**: 247–267.
18. Modarres, R. 2008. Regional Frequency Distribution Type of Low Flow in North of Iran by L-moments. *Water Resour. Manage.* **22**: 823-841.
19. Modarres, R. 2009. Regional Dry Spells Frequency Analysis by L-Moment and Multivariate Analysis. *Water Resour. Manage,* **24:**2365-2380
20. Ouarda, T. B. M. J. Cunderlik, J. M. St-Hilaire, A. Barbet, M. Bruneau P. and Bobée, B. 2006. Data-based Comparison of Seasonality-Based Regional Flood Frequency Methods, *J. Hydrol.* **330**: 329–339
21. Pilon, P. J., and Adamowski, K. 1992. The Value of Regional Information to Flood Frequency Analysis Using the Method of L-Moments. *Can. J. Civ. Eng.,* **19**:137–147.
22. Shu, C. and Burn, D. H. 2004a. Homogeneous Pooling Group Delineation for Flood Frequency Analysis Using a Fuzzy Expert System with Genetic Enhancement. *J. Hydrol.*, **291**:132–149.
23. Shu, C. and Ouarda, T. B. M. J. 2008. Regional Flood Frequency Analysis at Ungauged Sites Using the Adaptive Neuro-Fuzzy Inference System. *J. Hydrol.*, **349:** 31-43.
25. Soltani, S., Modarres, R. 2006. Classification of Spatio-Temporal Pattern of Rainfall in Iran Using a Hierarchical and Divisive Cluster Analysis. *Journal of Spatial Hydrology*, **6(2)**: 1-12
26. Stedinger, J. R. and Tasker, G. D. 1985. Regional Hydrologic Analysis 1, *Water Resour. Res.*, **21 (9)**: 1421–1432.
27. Vogel, Richard M. and Fennessey Neil M. 1993. L-moments Diagrams Should Replace Product Moment Diagrams. *Water Resour. Res.*, **29**: 1745–1752.
28. Vogel, R. M., Thomas, W. O. and McMahon, T. A. 1993a. Flood-flow Frequency Model Selection in Southwestern United States. *J. Water Resour. Ping. Mgmt.,* **119(3)**: 353–366.
29. Vogel, R. M., McMahon, T. A. and Chiew, F. H. S. 1993b. Flood-flow Frequency Model Selection in Australia. *J. Hydrol.*, **146**: 421–449.

رویکرد منطقه ای به تحلیل فراوانی سیل با استفاده از گشتاورهای خطی

ح. ملکی نژاد، ه.پ. ناخت نبل، و آ. کلیک

چکیده

کمبود داده اساساً یکی از مشکلات تحلیل فراوانی سیل در مناطق خشک ونیمه خشک است. در این مناطق طول دوره آماری کوتاه و یا فقدان آن باعث عدم اطمینان در برآوردهای هیدرولوژیکی می شود. هر چند که بیش از ۷۵ درصد وسعت ایران را مناطق خشک و نیمه خشک با بارندکی اندک تشکیل می دهد ولی سیلابهای بسیار شدیدی در این مناطق رخ می دهد. استفاده توامان از آمار مناطق دارای رفتار هیدرولوژیکی مشابه در مقایسه با بکارگیری داده های موجود فقط در یک ایستگاه یکی از راهکارهای افزایش داده ها در یک تحلیل می باشد. تعیین مناطق همگن و تخمین پارامترهای مورد نیاز در قالب یک تحلیل منطقه ای و تعیین بهترین توزیع آماری جهت برآورد شدت حداکثر سیلاب در چند ایستگاه هیدرومتری در منطقه مرکزی ایران از اهداف اصلی این تحقیق بشمار می روند. مهمترین عوامل موثر بر دبی حداکثر سیلاب با استفاده از تکنیک تحلیل عاملی شناسایی شدند و سپس تعیین مقدماتی مناطق همگن بوسیله روشهای تحلیل خوشه ای و منطقه اثر انجام شد. در ادامه آزمون همگنی بر روی این مناطق همگن اولیه با استفاده از آماره H که یک آماره معتبر بر اساس روش گشتاورهای خطی است، صورت گرفت. نتایج حاصل از آماره H نشان داد که حوضه های انتخابی از نظر هیدرولوژیکی همگن نیستند. بر این اساس در قالب سه منطقه همگن گروه بندی شدند. برای تعیین بهترین توزیع آماری سازگار با هر یک از مناطق همگن مشخص شده از آماره Z استفاده شد که مبتنی بر تکنیک گشتاورهای خطی و نمودار گشتاورهای خطی است. در روند منطقه ای سازی تحلیل سیل، سه منطقه همگن مشخص شد و به کمک تکنیک گشتاورهای خطی پنج توزیع آماری سه پارامتره شامل لاجستیک تعمیم یافته، مقادیر حد تعمیم یافته، پارتو تعمیم یافته، لوگ نرمال سه پارامتره و پیرسون نوع ۳ مورد ارزیابی قرار گرفتند. نتایج نشان داد که بترتیب توزیع های GEV و LN3 در دو منطقه همگن یک و دو و همچنین توزیع های GLO و GEV در منطقه همگن سه بهترین برازش را با داده های حداکثر سیل دارند. بنابراین از بین توزیعهای آماری بررسی شده، GEV بهترین تطابق را با داده ها جهت برآورد سیل دارد.

[DOR: 20.1001.1.16807073.2011.13.7.1.6]