Physics and Chemistry of the Earth 36 (2011) 54–60

Contents lists available at ScienceDirect

Physics and Chemistry of the Earth

journal homepage: www.elsevier.com/locate/pce

Comparing the index-flood and multiple-regression methods using L-moments

H. Malekinezhad[a,*], H.P. Nachtnebel[b], A. Klik[c]

[a] Natural Resources Faculty, Yazd University, Yazd, Iran
[b] Institute of Water Management, Hydrology and Hydraulic Engineering, BOKU University, Vienna, Austria
[c] Institute of Hydraulics and Rural Water Management, BOKU University, Vienna, Austria

ARTICLE INFO

Article history:
Received 20 July 2009
Received in revised form 3 June 2010
Accepted 17 July 2010
Available online 24 July 2010

Keywords:
Regional flood frequency analysis
Multiple-regression
Namak-Lake basin
Index-flood
L-moments

ABSTRACT

In arid and semi-arid regions, the length of records is usually too short to ensure reliable quantile estimates. Comparing index-flood and multiple-regression analyses based on L-moments was the main objective of this study. Factor analysis was applied to determine main influencing variables on flood magnitude. Ward's cluster and L-moments approaches were applied to several sites in the Namak-Lake basin in central Iran to delineate homogeneous regions based on site characteristics. Homogeneity test was done using L-moments-based measures. Several distributions were fitted to the regional flood data and index-flood and multiple-regression methods as two regional flood frequency methods were compared.

The results of factor analysis showed that length of main waterway, compactness coefficient, mean annual precipitation, and mean annual temperature were the main variables affecting flood magnitude. The study area was divided into three regions based on the Ward's method of clustering approach. The homogeneity test based on L-moments showed that all three regions were acceptably homogeneous. Five distributions were fitted to the annual peak flood data of three homogeneous regions. Using the L-moment ratios and the *Z*-statistic criteria, GEV distribution was identified as the most robust distribution among five candidate distributions for all the proposed sub-regions of the study area, and in general, it was concluded that the generalised extreme value distribution was the best-fit distribution for every three regions.

The relative root mean square error (RRMSE) measure was applied for evaluating the performance of the index-flood and multiple-regression methods in comparison with the curve fitting (plotting position) method. In general, index-flood method gives more reliable estimations for various flood magnitudes of different recurrence intervals. Therefore, this method should be adopted as regional flood frequency method for the study area and the Namak-Lake basin in central Iran. To estimate floods of various return periods for gauged catchments in the study area, the mean annual peak flood of the catchments may be multiplied by corresponding values of the growth factors, and computed using the GEV distribution.

1. Introduction

An inadequate understanding of the probabilistic behaviour of extreme flows may have significant economical impacts on the design of a hydraulic structure. Underestimation of flood discharges will lead to increased flood risk, while overestimation of flood discharges will lead to increased construction costs. Therefore, a large number of studies have focused on the reduction of uncertainties with flood quantile estimates. Flood frequency analysis is essentially a problem of information scarcity at the arid and semi-arid regions. Particularly, in these regions, the records lengths are usually too short to ensure reliable estimates of low-exceedance probability quantiles needed in many practical problems. In wet conditions and in arid areas rainfall comes as high intensity storms, which causes flash floods (Farquharson et al., 1992). More than 75% of Iran is located in arid and semi-arid region and despite the little amount of annual precipitation; often large floods occur (Malekinezhad, 2005). Small numbers of gauging stations and short annual flood series are the most important difficulties that designers and engineers face in the hydrologic projects and water resources management. 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 is called regional flood frequency analysis. Regionalisation usually involves two steps; the identification of homogeneous regions and the determination of a regional estimation method for different flood frequencies.

Several methods are available to perform a regional analysis. One of the first steps in a regional analysis is to define the region itself. The definition of a region depends on the quantities to be

* Corresponding author.
E-mail address: hmalekinezhad@yazduni.ac.ir (H. Malekinezhad).

1474-7065/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pce.2010.07.013

estimated. Regional flood frequency analysis involves the identification of groups (or regions) of hydrologically homogeneous catchments and the application of a regional estimation method in the identified homogeneous regions. Reliable estimation of extreme flows with given return period is crucial at sites where for example dam constructions, reservoir management, low flow management, or every other hydraulic structure is needed. In practice, however, data are collected only at a limited number of sites and it therefore frequently happens that no streamflow data are available at sites where a hydraulic installation is to be constructed. An inadequate understanding of the probabilistic behaviour of extreme flows at the site may have significant economical impact on the project. Underestimation of flood discharges will lead to increased flood risk, while overestimation of flood discharges will lead to increased construction costs. In cases where no at-site data are available for flood assessment, one may use data from gauged catchments, or, in general, data from catchments with similar hydrologic regimes.

L-moment ratio diagrams have become popular tools for regional distribution identification, and for testing outlier stations. Hosking and Wallis (1993) developed several tests to 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). 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.

2. Methods

This study involves four main stages: screening the data and determining the main site and at-site characteristics which affect the flood magnitude and applying them in the regionalisation of the study area, identifying homogeneous regions by cluster analysis and region-of-influence methods, testing the homogeneity of regions, investigating the best-fit distribution for the study area based on L-moments approaches, and comparing index-flood and multiple-regression as the two regional flood frequency method.

2.1. Regionalisation approaches

Regional flood frequency analysis 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 (GREHYS, 1996a,b). Geographically contiguous regions have been used for a long time in hydrology, but have been criticised for being of arbitrary character. In fact, the geographical proximity does not guarantee hydrological similarity.

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 site. An important requirement for regional flood frequency analysis is the identification of the region that is used for the transfer of extreme flow information.

Numerous techniques have been used to identify homogeneous regions for regional flood frequency analysis. 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 minimising the Euclidean distance in site characteristics space within each cluster. In this stage of study, firstly, preliminary determination was carried out by the cluster analysis, and then, the homogeneity of the region is tested by L-moments approach. Cluster analysis is a standard method of statistical multivariate analysis for dividing a data set into groups and has been successfully used to form regions for 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.

2.2. Main basin characteristics

In this study, factor analysis is used for determination of the main variables affecting flood magnitude. Many statistical methods are used to study the relation between independent and dependent variables. Factor analysis is different; it is used to study the patterns of relationship among many dependent variables, with the goal of discovering something about the nature of the independent variables that affect them, even though those independent variables were not measured directly. Thus answers obtained by factor analysis are necessarily more hypothetical and tentative than is true 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.

2.3. Description of the selected distributions

In regional frequency analysis a single frequency distribution is fitted to data from several sites. In general, the region will be slightly heterogeneous, and 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. The chosen distribution needs not be the distribution that gives the closest approximation to the observed data. Even when a distribution can be found that gives a close fit to the observed data, there is no guarantee that future values will match those of the past. In other words, when the data arise from a physical process that can give rise to occasional outlying values far removed from the bulk data. In this study, five three-parameter distributions i.e. generalised logistic (GLO), generalised extreme value (GEV), generalised Pareto (GPA), three-parameter lognormal (LN3), and Pearson type III (PE3) were fitted to the flood data of the three identified homogeneous regions. The three parameters of these distributions are estimated by the L-moments approach.

2.4. Regional flood frequency methods

Two regional analyses were used to develop methods for estimating flood discharges for the basins in Namak-Lake basin in the central part of Iran. The first analysis is index-flood method that has been used widespread in regional flood frequency, and the second analysis, a traditional regression, required the use of generalised least-squares regression to define a set of predictive equations that relate peak discharges for the 2-, 5-, 10-, 25-, 50-, 100- and 200-year recurrence intervals to selected basin characteristics for gauged sub-basins at the main basin.

2.4.1. Index-flood method

The index-flood method, proposed by Dalrymple (1960), is one of the first approaches to regional flood estimation. The basic premise of this method is that a combination of runoff records maintained at a number of gauging stations will produce a more reliable, not a longer, record, and thus will increase the reliability of frequency analysis within a region. The main assumption is that flood at different sites within a region are identically distributed except for a scale factor which is a function of physiographic basin characteristics.

A key assumption of the index-flood method is that flood data at different sites in a homogeneous region has the same distribution, except for a scale parameter or an index factor (Dalrymple, 1960). The scale factor is appointed as an index-flood and is generally taken to be the mean annual flood (Saf, 2010).

There are two major parts of index-flood method. The first is the development of basic dimensionless frequency curve representing the ratio of the flood of any frequency to an index-flood (the mean annual flood). The second is the development of relations between geomorphologic characteristics of drainage areas and the mean annual flood by which to predict the mean annual flood at any point within the region. By combining the mean flood with the basic frequency curve, a regional frequency curve is produced.

Recent advances in regional frequency analysis include the use of L-moments together with the index-flood method, as reported by Hosking and Wallis (1997). The methodology has been applied successfully in modeling floods in a number of case studies from USA (Vogel et al., 1993a), Australia (Pearson et al., 1991), Southern Africa (Mkhandi and Kachroo, 1997) and South Africa (Kjeldsen et al., 2002). Heinz and Stedinger (1998) generalised the index-flood procedure by employing regression with physiographic information to refine a normalized *T*-year flood estimator.

Suppose that data are available at N sites, with site i having sample size n_i and observed data Q_{ij}, $j = 1, \ldots, n_i$. $Q_i(F)$, $0 < F < 1$, be the quantile function of frequency distribution at site i. The key assumption of an index-flood procedure is that the sites form a homogeneous region, that is, that the frequency distributions of the N sites are identical apart from a site-specific scaling factor, the index-flood.

$$Q_i(F) = \mu_i q(F), \quad i = 1, \ldots, N \quad \text{or } Q_T^i = \mu_i q_T \tag{1}$$

Q_T^i is the flood quantile corresponding to a T-year return period, at a given site i. The index-flood is naturally estimated by $\mu_i = \bar{Q}_i$, the sample mean of the data at site i. Other location estimators such as the median or a trimmed mean could be used instead. μ_i is supposed to be the mean of the at-site frequency distribution, and $q(F)$ is the regional quantile of non-exceedance probability F and q_T is the regional quantile of return period T.

In a more general setting, the two steps in the index-flood method are, in modified version, common to all regional flood estimation procedures. The first part of the analysis is to identify sites which seems sufficiently similar to the target site to provide a basis for information transfer. In practice, one can employ different similarity measures and classification techniques. The second part of analysis is to perform the information transfer, i.e. to actually infer flood quantiles at the target site using data from the sites identified in the first part of analysis.

2.4.2. Multiple-regression method

The most commonly used relation between the flow statistics (represented here by the flood-quantile Q_T of return period T-years) and the watershed characteristics $(A, B, \ldots, M)$ is the power-form function (Pandey and Nguyen, 1999). The multiple-regression model can be expressed in the following form,

$$Q_T = \alpha\, A^a B^b C^c \ldots M^m \tag{2}$$

where α is regression constant defined by regression analysis and $a, b, c, \ldots$, m are regression coefficients defined by regression analysis. This form of the multiple-regression model is achieved by linear regression of the logarithms of the variables.

Multiple-regression analysis was used to estimate the relation between flood discharges for given frequencies and drainage-basin characteristics for three determined homogeneous regions. The multiple-regression technique is a means of determining flood peak magnitude for a given recurrence interval of the sites with low available data and transferring flood-peak characteristics from sites where observed data are available to ungauged locations. The relation is presented by flood-frequency equations.

The regression equations are used to relate the most significant drainage-basin characteristics (independent variables) to flood-peak characteristics (dependent variables; $Q_2, Q_5, \ldots, Q_{100}$).

2.5. Plotting position

This method involves fitting of an assumed probability distribution to observed data. The sample data are arranged in either ascending or descending order of magnitude. Each data point is assigned a rank starting with 1. Many plotting-position formulas are available; all of these formulas can be expressed as special cases of

$$P(m) = \frac{m - a}{N + b} \quad \text{or } T(m) = \frac{N + b}{m - a} \tag{3}$$

where m is the mth ranked data, $P(m)$ is the exceedance probability or non-exceedance probability of mth data for descending or ascending arrangements, respectively, N is the data size, and a and b depend on the type of formula.

The flood series were analysed for all four data sets using the direct curve fitting method mentioned earlier. The observed probabilities were computed from the Gringorten plotting-position formula (Singh et al., 2005)

$$P(m) = \frac{m - 0.44}{N + 0.12} \tag{4}$$

where m is the mth descending ranked observation in the data set. A theoretical distribution was fitted to the values obtained by Eq. (4). The plotting-position probabilities obtained by the direct curve fitting method were compared with those obtained from the index-flood and multiple-regression methods.

2.6. Relative root mean square error (RRMSE)

Relative root mean square error measure (Zrinji and Burn, 1994) was used to evaluate and compare the relative merits of the site estimation of different recurrence interval. This measure is defined as

$$\text{RRMSE} = \left[\frac{1}{NS}\frac{1}{J}\sum_{i=1}^{NS}\sum_{t=1}^{J}\left(\frac{\hat{Q}_t^i - Q_t^i}{Q_t^i}\right)^2\right]^{1/2} \tag{5}$$

where $\hat{Q}_t^i$ and Q_t^i are the estimated and true value for the extreme flow for the tth recurrence interval at site i, respectively, and NS is the number of sites in each homogeneous region.

3. Data base and description of the study area

The study area refers to some sub-basins of the Namak-Lake basin located in central basin of Iran, with the geographic coordinates from 48°20′E to 52°40′E longitude and 32°00′N to 36°30′N latitude (Fig. 1). The total area of Namak-Lake basin is 89,650 km^2 with minimum elevation of 800 and maximum elevation of 4375 m above sea level in Namak Lake and Jajrud heights, respectively.

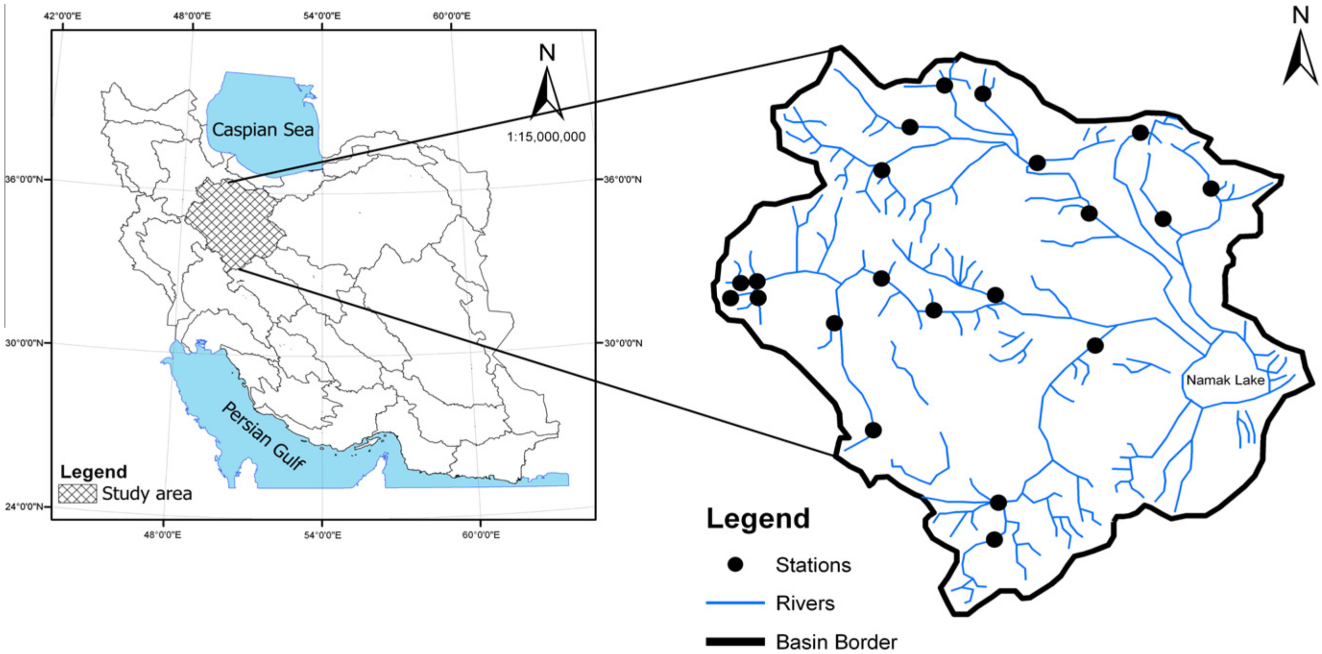

Fig. 1. The map of Namak-Lake basin in Iran.

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 gauging stations of the Namak-Lake basin in the central part of Iran. These data were obtained from the Water Resources Researches Organisation of Iran. Most of gauging stations in this part have short data series with a large number of missing data or have been constructed in recent years with the data length less than 10 years. The selection of basin was made so that at least 15 years of runoff data were available. The average number of years of record for the stations was 27 years with a range from 15 to 36 years. After the preliminary screening of the sites 21 gauging stations were selected for this study.

4. Results

4.1. Identification of homogeneous regions (IHR)

Cluster analysis is a standard method of statistical multivariate analysis for dividing a data set into groups and has been successfully used to form regions for regional frequency analysis. Regionalisation approaches such as cluster analysis require selection of variables that are used to define the similarity (or dissimilarity) for the catchments (Burn, 1997). Hosking and Wallis (1997) recommended using methods that rely on site characteristics only when identifying homogeneous regions, and subsequently using the 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 minimising the Euclidean distance in site characteristics space within each cluster. Many of the statistical types of software involve clustering analysis methods. In this study, first, a preliminary determination of homogeneous regions is done by Ward's clustering method for determination of homogeneous regions (Fig. 2). At the next steps, the statistical test based on L-moment ratios proposed by Hosking and Wallis (1993) is used for testing the heterogeneity of the proposed regions.

Total area of 21 selected sites is 16,036 km^2. The identified homogeneous regions (1), (2), and (3) include 8868, 1955, and 5213 km^2, respectively. Most of the large sites are located in the region (1), with the range of 1655–2768 km^2 and Most of the small sites are located in the region (2), and have the areas smaller than 200 km^2. The intermediate area sites are located in the region (3). In general, the area has interactive relation with many of the other site characteristics like perimeter, basin slope, main channel length and main channel slope. The results of identification of homogeneous regions in this study show, the area is the most important characteristic affecting homogeneous regions.

4.2. Homogeneity of the regions

L-moments and L-moment ratios are the bases for all the stages of the L-moments approach, such as; identification of unusual sites (discordancy test), homogeneity test, the goodness-of-fit measure for determination of the best distribution for each homogeneous region, and parameter estimations [location (ξ), scale (σ), and shape (k)] for the five selected distributions. Therefore, the annual peak flood series of basins were ranked in descending order and then, the first four probability weighted moments (β_1, β_2, β_3, and β_4) were calculated for each basin. PWMs are needed for obtaining the L-moments and L-moment ratios of basins. The aim of homogeneity test is to estimate the degree of homogeneity in a group of sites. In this study the 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 the region is acceptably homogeneous when $H < 1$; possibly heterogeneous when $1 < H < 2$ and definitely heterogeneous when $H > 2$.

The results of homogeneity test based on L-moments approach for whole study area show the selected sites should be divided to some homogeneous sub-regions. The results show that the absolute values of H-statistic for all three regions are less than 1, and these regions are acceptably homogeneous.

4.3. Identification of the best-fit distribution

Use of the generalised extreme value (GEV) distribution as a regional flood frequency model with an index-flood approach has received considerable attention (Chowdhury et al., 1991). Several

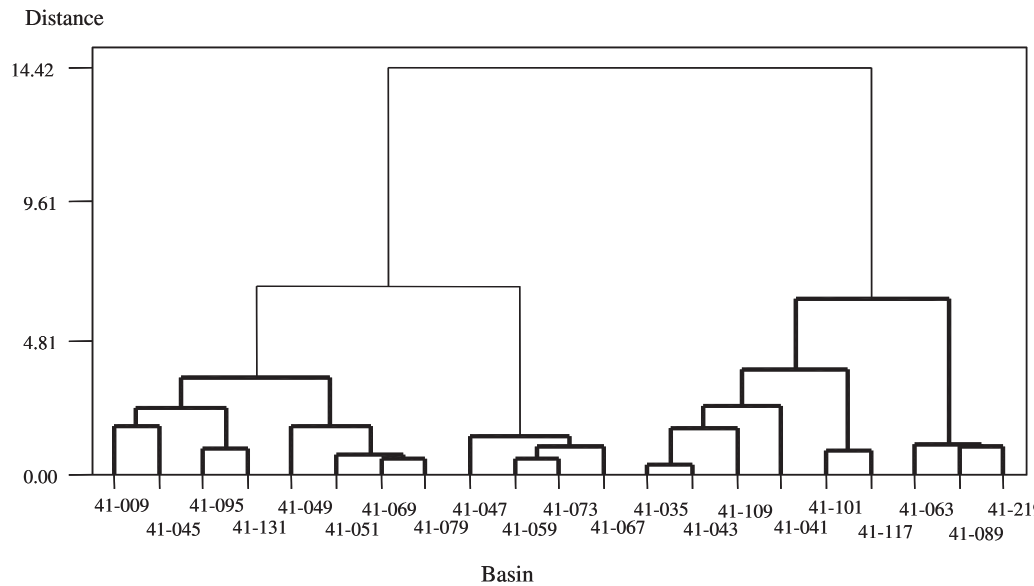

Fig. 2. Dendrogram of clustered basins by Ward's method.

studies on flood frequency analysis have been carried out in different regions and countries using the GEV distribution together with a L-moments approach in USA (Vogel et al., 1993a), Australia (Vogel et al., 1993b), South Africa (Kjeldsen et al., 2002), China (Jingyi and Hall, 2004) and India (Kumar et al., 1999, 2003; Kumar and Chatterjee, 2005).

In this study, five three-parameter distributions were selected; generalised extreme value distribution (GEV), generalised Pareto distribution (GPA), generalised logistic distribution (GLO), three-parameter lognormal distribution (LN3), and Pearson type III distribution (PE3). These distributions have been commonly used in hydrological studies and projects. The goodness-of-fit test and L-moments diagram were used for choosing the best-fit distribution in different homogeneous regions. In this section, first, the results of parameters estimation, and then, the results of the best-fit distribution test are given. In the L-moments approach, the three parameters (location, scale, shape) of each probability distribution in regional flood frequency analysis are obtained by the regional averages of the L-moments and L-moment ratios. The estimated regional parameters were obtained for GEV as the best-fit distribution. The Z-statistic was used as a goodness-of-fit measure for the identification of common regional distribution that applies within each region. This statistic is defined in terms of L-moment ratios. The five three-parameter distributions were fitted to the three homogeneous regions.

As it is shown in Table 1 the GEV has the best goodness-of-fit with the data at region (1) and LN3 and GPA 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 GEV is the best-fit distribution for the flood analysis at the region (2); also the LN3 and PE3 are the other acceptable distributions. 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.

Table 1
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

4.4. Index-flood results

Two regional analyses were used to develop methods for estimating peak flood discharges for the basins in Namak-Lake basin in the central part of Iran. The first analysis is index-flood method that has been used widespread in regional flood frequency, and second analysis, a traditional regression, required the use of generalised least-squares regression to define a set of predictive equations that relate peak discharges for the 2-, 5-, 10-, 25-, 50-, 100- and 200-year recurrence intervals to the four obtained hydro-climatic and physical characteristics by factor analysis. The results of two methods are compared by the relative root mean square error.

In this method, the flood quantiles are estimated by calculating the regional growth curve (q_T) and the mean value of annual peak series of each site. The GEV distribution, as the best-fit distribution to regional data, is used for estimating the regional growth curve for different return periods. The results are given in Table 2 for three homogeneous regions.

4.5. Multiple-regression results

The aim is to develop the relationships between Q_T as a dependent variable and main site and at-site characteristics as the independent variables at each homogeneous region. The parameters of generalised extreme value distribution as the best-fit identified distribution i.e. location (ξ), scale (α), and shape (k), by direct use of the data, were calculated for each site in the homogeneous regions. These parameters are needed for estimation of Q_T. The predicted flood magnitude of the 2-, 5-, 10-, 25-, 50-, 100-, and 200-year were separately obtained for each site. The results are shown in Tables 3–5.

Table 2
The values of regional growth curve (q_T) of different frequencies calculated by GEV for three regions.

Non-exceedance probability $F = 1 - \frac{1}{T}$	Return period (T), year	q_T		
		Region (1)	Region (2)	Region (3)
0.50	2	1.10	1.01	0.97
0.80	5	1.99	1.71	1.87
0.90	10	2.83	2.35	2.81
0.96	25	4.28	3.42	4.57
0.98	50	5.74	4.46	6.49
0.99	100	7.71	5.83	9.28
0.995	200	10.20	7.51	13.07

Table 3
The predicted flood magnitudes (m^3/s) by GEV and direct use of the data of each site located at region (1).

Site code	Return period (T), year						
	2	5	10	25	50	100	200
41-047	53.2	81.4	107.4	151.8	195.7	254.0	327.0
41-059	58.07	113.7	168.2	265.5	367.1	513.0	690.8
41-067	119.0	181.0	227.0	290.6	342.6	401.0	462.5
41-073	59.9	119.5	180.2	294.0	417.0	593.8	832.2

Table 4
The predicted flood magnitude (m^3/s) s by GEV and direct use of the data of each site located at region (2).

Site code	Return period (T), year						
	2	5	10	25	50	100	200
41-035	54.7	95.5	132.0	192.8	251.4	327.8	421.2
41-041	0.1	1.5	3.4	5.4	8.6	11.6	14.2
41-043	20.7	31.4	40.3	53.9	66.2	81.2	98.4
41-063	19.6	38.9	59.2	98.7	142.7	207.8	298.0
41-089	22.0	43.6	66.4	110.5	159.8	232.7	333.9
41-101	99.5	154	201.2	277.1	348.0	437.7	544.4
41-109	63.3	111.5	157.3	237.4	318.5	428.7	569.2
41-117	76.8	118.0	154.0	212.3	267.3	337.4	421.3
41-219	23.6	40.5	56.4	84.1	111.8	149.2	196.6

Table 5
The predicted flood magnitudes (m^3/s) by GEV and direct use of the data of each site located at region (3).

Site code	Return period (T), year						
	2	5	10	25	50	100	200
41-009	77.7	150.8	219.4	338.5	458.3	619.8	824.3
41-045	33.5	56.2	77.6	114.6	151.8	202.1	265.7
41-049	30.3	46.3	58.0	73.9	86.7	100.9	115.6
41-051	27.8	58.3	94.4	171.8	266.8	419.8	650.6
41-069	35.5	77.4	125.0	224.0	341.7	525.8	796.0
41-079	47.0	72.3	91.0	117.0	138.2	162.0	187.0
41-095	81.5	116.0	141.3	176.2	204.4	235.9	268.8
41-131	2.6	5.7	9.4	17.1	26.5	41.5	64.1

The relation of flood peaks of selected recurrence intervals to basin and climatic parameters is determined by multiple-regression methods. The resulting relation is the form of

$$Q_T = \alpha L_w^a Gr^b P_m^c T_m^d \tag{6}$$

where Q_T is the peak flood magnitude for T-year return period in cubic meter per second, L_w, Gr, P_m, and T_m are the main water way length (km), compactness (Gravellius) coefficient, mean annual rainfall (mm), and mean annual temperature (°C), respectively. α a, b, c, and d are the regression coefficients and R is the correlation coefficient defined by regression analysis. These coefficients were calculated using multiple-regression for different flood probabilities at each homogeneous region. The coefficients of regression equation for each region are shown in Tables 6–8.

4.6. Comparisons of regional flood analyses

In order to evaluate the performance of the index-flood and multiple-regression methods in comparison with the curve fitting (plotting position) method, the relative root mean square error (RRMSE) measure was applied. The lower value of RRMSE for each method indicates that the method has the better fitness to the data set. The results of RRMSE values indicate that index-flood method gives better results for prediction of flood magnitude of different return periods at regions (2) and (3). The RRMSE values of index-flood and multiple-regression methods are 0.23, and 0.36 for the region (2), and 0.29, 0.53 for region (3), respectively.

As it has been shown in Table 9, the RRMSE values indicate that the index-flood method gives better results for prediction of flood magnitude of different return periods for the regions (2) and (3). For region (1), multiple-regression shows better performance than the index-flood method. In this region, correlation coefficient between Q_T and basin characteristics is very high ($R > 0.995$). In general, the large difference between RRMSE obtained by index-flood and multiple-regression methods show that the index-flood method gives more reliable estimations for various flood magnitudes of

Table 6
The coefficients of regression equation for region (1).

Recurrence interval (T), year	α	a	b	R^2
2	8.13×10^{-3}	1.98	1.23	0.995
5	0.25	1.47	−0.597	0.999
10	2.58	1.09	−1.61	0.999
25	54.95	0.564	−2.79	0.998
50	575.4	0.156	−3.62	0.996
100	7244.4	−0.290	−4.48	0.995
200	81283.1	−0.724	−5.26	0.996

Table 7
The coefficients of regression equation for region (2).

Recurrence interval (T), year	α	a	b	c	d	R^2
2	2×10^{-3}	1.12	−3.38	1.12	0.222	0.91
5	6×10^{-3}	0.926	−3.20	1.10	0.291	0.89
10	0.010	0.776	−3.02	1.13	0.328	0.87
25	0.014	0.571	−2.74	1.21	0.369	0.82
50	0.019	0.41	−2.51	1.28	0.399	0.79
100	0.022	0.24	−2.27	1.36	0.428	0.75
200	0.026	0.07	−2.01	1.44	0.456	0.72

Table 8
The coefficients of regression equation for region (3).

Recurrence interval (T), year	α	a	b	c	d	R^2
2	3.2×10^{-16}	1.44	0.30	6.47	−2.14	0.90
5	3.2×10^{-14}	1.65	0.63	6.06	−3.25	0.88
10	7.9×10^{-13}	1.76	0.84	5.73	−3.91	0.86
25	5×10^{-11}	1.90	1.12	5.31	−4.74	0.83
50	1.3×10^{-9}	2.01	1.33	4.98	−5.36	0.80
100	3.9×10^{-8}	2.11	1.56	4.63	−6.02	0.76
200	1.1×10^{-6}	2.21	1.79	4.28	−6.66	0.73

Table 9
The RRMSE values of index-flood and multiple-regression methods for three homogeneous regions.

Region	Index-flood	Multiple-regression
1	0.22	0.13
2	0.23	0.36
3	0.29	0.53

different recurrence intervals. This method should be adopted as the regional flood frequency method for the study area and the Namak-Lake basin in central Iran.

5. Summary and conclusion

The study presented herein reports a regional analysis carried out in central Iran, aimed to evaluate the suitability of L-moments approach and index-flood method for predicting flood discharge at different return periods. A careful screening of data coming from about 71 stream gauging sites was carried out with the use of available extreme floods data for regional flood frequency analysis. The study area was analyzed firstly as a whole and then as three smaller sub-regions using cluster analysis technique.

When data record length is short, the direct application of probability distributions for anticipating the flood occurrence in different return periods does not give reliable results. Application of L-moments technique is a suitable approach for increasing the data length at regional flood frequency analysis. L-moments approach applies simultaneous use of all the data of several homogeneous basins in a hydrologic analysis. The result of this study shows that this technique is an effective approach in discharge estimation of flood peak in basins with missing data or basins with short time data record.

The results of factor analysis technique for determination of the main variables show that the 14 independent variables could be summarised to four factors. The water way length, mean annual precipitation, compactness (Gravellius) coefficient, and mean annual temperature were identified as the most important variables of the four factors.

The hierarchical clustering based on the Ward's method using Euclidean distance is a suitable approach for regionalisation objectives in hydrology. The dendrogram of clustered basins was cut from the distance of six to define the initial homogeneous regions in the study area. The results of homogeneity test based on L-moments approach at whole study area show the selected sites should be divided to some homogeneous sub-regions. Using the L-moment ratios and the *Z*-statistic criteria, GEV distribution was identified as the most robust distribution among five candidate distributions for all the proposed sub-regions of the study area, and in general, it can be concluded that the generalised extreme value distribution is the best-fit distribution for all three regions. The estimated regional growth curves were significantly different for the three different sub-regions. To estimate floods of various return periods for gauged catchments in the study area, the mean annual peak flood of the catchments may be multiplied by corresponding values of the growth factors, computed using the GEV distribution.

The RRMSE values indicate a considerable difference between index-flood and multiple-regression approaches in most of the identified HRs. The RRMSE values of these two methods showed just for very high values of correlation coefficient the multiple-regression gives acceptable estimations. It can be concluded that the high difference between RRMSE obtained by index-flood and multiple-regression methods show that index-flood method gives more reliable estimations for various flood magnitudes of different return periods. This method should be adopted as regional flood frequency method for the Namak-Lake basin in Iran.

References

Burn, D.H., 1997. Catchment similarity for regional flood frequency analysis using seasonality measures. Journal of Hydrology 202 (1–4), 212–230.
Chowdhury, J.U., Stedinger, J.R., Lu, L.H., 1991. Goodness-of-fit tests for regional GEV flood distributions. Water Resource Research 27 (7), 1765–1776.
Dalrymple, T., 1960. Flood Frequency Analyses. Water Supply Paper 1543-A. US Geological Survey, pp. 11–51.
Farquharson, F.A.K., Meigh, J.R., Sutclife, J.V., 1992. Regional flood frequency analysis in arid and semi-arid areas. Journal of Hydrology 138, 487–501.
Groupe de recherche en hydrologie statistique (GREHYS), 1996b. Presentation and review of some methods for regional flood frequency analysis. Journal of Hydrology 186 (1–4), 63–84.
Groupe de recherche en hydrologie statistique (GREHYS),, 1996b. Inter-comparison of regional flood frequency procedures for Canadian rivers. Journal of Hydrology 186 (1–4), 85–103.
Heinz, D.F., Stedinger, J.R., 1998. Using regional regression within index flood procedures and an empirical Bayesian estimator. Journal of Hydrology 210, 128–145.
Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. Water Resource Research 29, 271–278.
Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press, London, UK.
Jingyi, Z., Hall, M.J., 2004. Regional flood frequency analysis for the Gan-Ming River basin in China. Journal of Hydrology 296, 98–117.
Kjeldsen, T.R., Smithers, J.C., Schulze, R.E., 2002. Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method. Journal of Hydrology 255 (1–4), 194–211.
Kumar, R., Chatterjee, C., 2005. Regional flood frequency analysis using L-moments for North Brahmaputra region of India. Journal of Hydrologic Engineering 10 (1), 1–7.
Kumar, R., Singh, R.D., Seth, S.M., 1999. Regional flood formulas for seven subzones of zone 3 of India. Journal of Hydrologic Engineering 4 (3), 240–244.
Kumar, R., Chatterjee, C., Kumar, S., Lohani, A.K., Singh, R.D., 2003. Development of regional flood frequency relationships using L-moments for middle Ganga plains subzone 1(f) of India. Water Resource Management 17 (4), 243–257.
Malekinezhad, H., 2005. Comparative Study of Regional Flood Frequency Analyses in Some Basins in Central Iran. PhD thesis, Institute of Hydraulics and Rural Water Management, BOKU University, Vienna, Austria.
Mkhandi, S., Kachroo, S., 1997. Regional Flood Frequency Analysis for Southern Africa. Southern African Friend, Technical Documents in Hydrology No. 15 UNESCO, Paris, France.
Pandey, G.R., Nguyen, V.T.V., 1999. A comparative study of regression based methods in regional flood frequency analysis. Journal of Hydrology 225, 92–101.
Pearson, C.P., McKerchar, A.I., Woods, R.A., 1991. Regional flood frequency analysis of western Australian data using L-moments. International Hydrology and Water Resources Symposium, Perth, Australia, pp. 631–632.
Pilon, P.J., Adamowski, K., 1992. The value of regional information to flood frequency analysis using the method of L-moments. Canadian Journal of Civil Engineering 19, 137–147.
Saf, B., 2010. Assessment of the effects of discordant sites on regional flood frequency analysis. Journal of Hydrology 380, 362–375.
Singh, V.P., Wang, S.X., Zhang, L., 2005. Frequency analysis of nonidentically distributed hydrologic flood data. Journal of Hydrology 307, 175–195.
Vogel, Richard M., Fennessey, Neil M., 1993. L moments Diagrams should replace product moment diagrams. Water Resource Research 29, 1745–1752.
Vogel, R.M., Thomas, W.O., McMahon, T.A., 1993a. Flood-flow frequency model selection in south-western United States. Journal of Water Resources Planning and Management 119 (3), 353–366.
Vogel, R.M., McMahon, T.A., Chiew, F.H.S., 1993b. Flood-flow frequency model selection in Australia. Journal of Hydrology 146, 421–449.
Zrinji, Z., Burn, D.H., 1994. Flood frequency analysis for unguaged sites using a region of influence approach. Journal of Hydrology 153, 1–21.