ISSN 0097-8078, Water Resources, 2015, Vol. 42, No. 3, pp. 313–321. © Pleiades Publishing, Ltd., 2015.

WATER RESOURCES AND THE REGIME OF WATER BODIES

Comparison of Two Methods of Regional Flood Frequency Analysis by Using L-moments[1]

J. Mosaffaie

University of Yazd, Yazd, 89195-741 Iran

e-mail: jamalmosaffaie@gmail.com

Received Octoder 7, 2013

Abstract—A regionalized relationship to estimate flood magnitudes for ungauged and poorly gauged catchments can be established using regional flood frequency analysis (RFFA). Comparison of index-flood and multiple-regression analyses based on L-moments was the main objective of this study. Factor analysis was applied to determine main variables influencing flood magnitude. Results showed that the main variables are perimeter, equivalent diameter, time of concentration, length of main waterway and area. The study area was divided into two regions based on the Ward's method of clustering approach and site characteristics. The homogeneity test based on L-moments showed that two regions were acceptably homogeneous. Five distributions were fitted to the annual peak flood data of two homogeneous regions. Using the L-moment ratios and the Z-statistic criteria, generalized logistic (GLO) and generalized Pareto (GPA) distributions were identified for the first and second homogeneous regions respectively as the most robust distributions among five potential distributions. The relative root mean square error (RRMSE) measure was applied for evaluating the performance of the index-flood and the multiple-regression methods in comparison with the curve fitting (plotting position) method. In general, the index-flood method gives more reliable estimations for various flood magnitudes of different recurrence intervals.

Keywords: RFFA, multiple-regression, Qazvin province, index-flood, L-moments

DOI: 10.1134/S0097807815030112

INTRODUCTION

More than 75% of Iran area is located in arid and semi-arid regions and despite the low annual precipitation large floods often occur [12]. In developing countries annual flood series are too short to allow a reliable estimation of extreme events or there is no flow record available at the site of interest. An inadequate understanding of the probabilistic behavior of extreme flows may have significant economical impacts on the design of a hydraulic structure [12]. Reliable estimation of extreme flows with given return period is crucial at sites where dam constructions, reservoir management, low flow management, or any other hydraulic structure are needed. Underestimation of flood discharges will lead to increased flood risk, while overestimation of flood discharges will lead to increased construction costs. In practice, however, data are collected only at a limited number of sites and therefore it frequently happens that streamflow data are not available at sites where a hydraulic installation is to be constructed. In cases where at-site data are not available for flood assessment, one may use data from gauged catchments, or, in general, data from catchments with similar hydrologic regimes. Therefore, the estimation of a regional flood frequency distribution is popular and practical means of providing flood information at sites with little or no flow data available for the purposes of flood control and engineering economy [22].

RFFA usually involves two steps: (1) identification of homogeneous regions, and (2) selection of suitable regional frequency distributions and estimation of flood quantiles at sites of interest. Several methods are available to perform a regional analysis. The first step in a regional analysis is to define the region itself. The definition of a region depends on the quantities to be estimated. RFFA involves the identification of groups (or regions) of hydrologically homogeneous catchments and the application of a regional estimation method in the identified homogeneous regions.

Hosking [10] introduced the L-moments, which are linear functions of probability weighted moments (PWMs). The L-moments are more convenient than PWMs because they can be directly interpreted as measures of scale and shape of probability distributions. In this respect they are analogous to conventional moments. L-moment ratio diagrams have become popular tools for regional distribution identification and testing outlier stations. Hosking and Wallis [8] 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

[1] The article is published in the original.

a tool for identifying a regional distribution have been used in numerous studies [15, 19, 20, 21]. Chowdhury et al. [2] 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.

MATERIALS AND 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 regionalization 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 comparison of index-flood and multiple-regression as the two regional flood frequency methods.

Regionalization

The estimation of probability of extreme flood occurrence is an extrapolation based on limited data. Thus, the larger the database, the more accurate estimates will be. From a statistical point of view, estimation from small samples may give unreasonable or physically unrealistic parameter estimates, especially for distributions with a large number of parameters (three or more). Large variations associated with small sample sizes cause unrealistic estimates. In practice, however, data may be limited or in some cases not be available for a site. In this instance, regional analysis is the most useful.

Regionalization serves two purposes. For sites where data are not available, the analysis is based on regional data [3]. For sites with available data, the joint use of data measured at a site, called at-site data, and regional data from a number of stations in a region provides sufficient information to use probability distribution with greater reliability.

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 [5, 6].

Numerous techniques have been used to identify homogeneous regions for RFFA. Hosking and Wallis [9] recommended using methods that rely on site characteristics only when identifying homogeneous regions, and subsequently using the at-site characteristics to test independently 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. In this stage of study, firstly, preliminary determination was carried out by the cluster analysis, and then, the homogeneity of the region was tested by L-moments approach.

Study Area

The study area refers to Qazvin province of Iran, with the geographic coordinates from 35°24′ to 36°52′ N and 48°33′ to 50°51′ E (Fig. 1). The total area of Qazvin province is 15619 km^2, minimum elevation of 235 and maximum elevation of 4116 m above sea level in Sefidrud dam and Shahalborz heights, respectively.

The Alamutrud, Taleghanrud, Kharrud, and Hajiarab are the main rivers of Qazvin province. Annual flood series were firstly provided for the 32 gauging stations of Qazvin province from the Qazvin regional water company. A careful screening of data from 32 gauging sites was carried out. Most of gauging stations in this part was removed because of short data series with a large number of missing data or construction of storage dam at upstream or construction in recent years with the data record period less than 10 years. The selection of basin was made so that at least 15-year period of runoff data record was available. The average period of data record for the stations was 31 years with a range from 15 to 44 years. After the preliminary screening of the sites 15 gauging stations were selected for this study.

Main Basin Characteristics

After reviewing available resources along with the benefit of experts, 15 variables of selected stations were extracted as affecting flood magnitude (Table 1). Factor analysis is used for determination of the most important variables. Factor analysis seeks the patterns of relationship among many dependent variables to find out 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 independent variables directly observing.

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 not be single "true" distribution that applies to each site. Therefore, the aim is not to identify a "true" distribution but to find a distribution that will yield more accurate quantile estimates for each site. The chosen distribution does not need to be the distribution that gives the closest approximation to the observed data. Even when a distribution 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

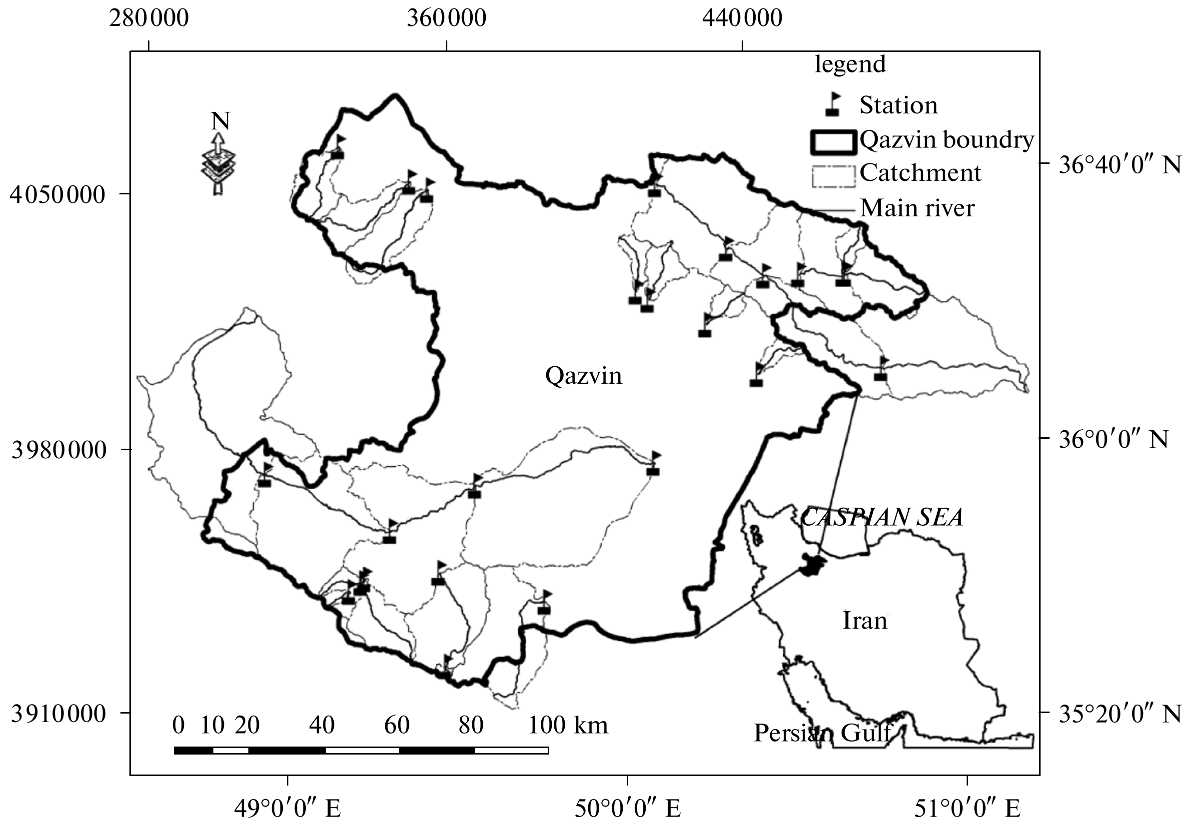

Fig. 1. Location map of study area in Iran.

can give rise to occasional outlying values far removed from the bulk data. In this study, five three-parameter distributions, i.e., generalized logistic (GLO), generalized extreme value (GEV), generalized Pareto (GPA), three-parameter generalized normal (GNO), and Pearson type III (PE3) were fitted to the flood data from the two identified homogeneous regions. The three parameters of these distributions were estimated by the L-moments approach.

Regional Flood Frequency Methods

Many types of regionalization procedures are available [3]. The index-flood and the multiple-regression methods were used in this study.

Index-Flood Method

The index-flood method which has been used for a long time was proposed by Dalrymple [4]. The key assumption in the index-flood method is that the distribution of floods at different sites in a region is the same except for a scale parameter, which reflects rainfall and runoff characteristics of each region. The scale factor is appointed as an index-flood and generally taken to be the mean annual flood [17].

There are two major parts of the 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 the mean annual flood is predicted at any point within the region. By combining the mean flood with the basic frequency curve, a regional frequency curve is produced. As reported by Hosking and Wallis [9] recent advances in regional frequency analysis include the use of L-moments together with the index-flood method. The methodology has been successfully applied in modeling floods in a number of case studies from USA [20], Australia [16], Southern Africa [13] and South Africa [11]. Heinz and Stedinger [7] generalized 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} ($j = 1, \ldots, n_i$). $Q_i(F)$, $0 < F < 1$, is 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, the frequency distributions of the

Table 1. Characteristics of selected hydrometric stations

Station	Average of annual water volume (CMC)	Average of annual peak flow (m^3/s)	Mean annual discharge (m^3/s)	Mean annual temperature (°C)	Annual precipitation (mm)	Drainage density (km/km^2)	Waterway slope (%)	Waterway length (km)	Average slope (%)	Height differences (m)	Time of concentration (h)	Compactness coefficient (Gr)	Equivalent diameter (km)	Perimeter (km)	Area (km^2)
Baghke-layeh	276.85	103.71	8.72	7.8	667	7.6	6.0	42.6	46.6	2900	3.36	1.37	28.0	121.1	615.3
Khuban	148.68	51.9	4.70	5.2	719	7.5	8.1	28.0	49.7	2616	2.15	1.31	17.7	73.3	245.6
Shotorak	7.41	16.1	0.25	11.9	436	7.5	7.0	15.5	24.2	1140	1.49	1.39	8.7	38.5	59.9
Behjat abad	7.42	19.1	0.23	12.0	418	8.2	6.3	18.9	30.5	1260	1.81	1.52	8.1	39.0	51.6
Barajin	15.08	24.48	0.45	11.9	430	7.8	6.2	19.9	22.6	1240	1.93	1.42	11.6	52.4	106.5
Amir abad	10.36	19.69	0.33	11.9	401	7.1	4.7	18.5	23.3	1030	1.91	1.52	9.3	44.4	67.3
Artesh abad	22.71	29.79	0.79	9.0	372	6.9	2.2	40.0	13.9	977	4.75	1.28	23.6	95.2	436.0
Shahab-basi	22.17	34.71	0.71	10.8	336	5.3	0.6	228.9	10.6	1714	28.68	1.78	83.1	467.4	5428.7
Abgarm	87.84	82.46	2.79	9.7	378	5.1	0.8	120.6	10.8	1339	15.04	1.54	55.9	272.6	2455.5
Haji arab	21.08	62.04	0.67	9.4	311	8.3	2.7	41.6	10.2	1275	4.49	1.48	27.0	126.5	571.4
Deh arvan	5.11	7.01	0.16	7.9	466	6.4	9.4	9.7	35	972	0.93	1.25	5.4	21.4	22.9
Pol arvan	16.01	24.57	0.51	9.0	423	6.7	6.8	14.8	23.9	1068	1.45	1.20	11.2	42.5	98.7
Tunel avaj	31.38	46.07	0.99	7.6	436	7.0	2.5	33.6	19.3	1136	3.66	1.42	20.0	90.0	314.2
Dashtak	17.24	54.28	0.55	9.5	386	4.3	1.1	76.2	8.1	1069	9.65	1.56	45.0	222.7	1590.9
Rahim abad	124.28	110.47	4.20	9.8	364	5.8	0.7	157.7	12.5	1508	19.59	1.62	72.4	371.4	4112.2

N sites are identical apart from a site-specific scaling factor, the index-flood.

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

where 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, $q(F)$ is the regional quantile of non-exceedance probability *F*, and q_T is the regional quantile of return period *T*.

In more general setting, the two steps in the modified version of the index-flood method are 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.

Multiple-Regression Method

This method is a traditional regression, required the use of generalized least-squares regression to define a set of predictive equations that relate peak discharges for the 2-, 5-, 10-, 25-, 50-, and 100-year recurrence intervals to selected basin characteristics for gauged sub-basins at the main basin. 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 [14]. The multiple-regression model can be expressed in the following form:

$$Q_T = \alpha A^a B^b C^c \ldots M^m, \qquad (2)$$

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 mean of determining flood peak magnitude for a given recurrence interval of the sites with little 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: $Q_2, Q_5, \ldots, Q_{100}$).

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

$$P(m) = m - a/N + b \quad \text{or} \quad T(m) = N + b/m - a. \qquad (3)$$

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 analyzed for all four data sets using the direct curve fitting method mentioned earlier. The observed probabilities were computed from the Gringorten plotting-position formula [18]:

$$P(m) = (m - 0.44)/N + 0.12, \qquad (4)$$

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 the multiple-regression methods.

Relative Root Mean Square Error (RRMSE)

RRMSE measure [23] was used to evaluate and compare the relative merits of the site estimation of different recurrence interval. This measure is defined as

$$\text{RRMSE} = \left[\frac{1}{NS}\frac{1}{j}\sum_{i=1}^{NS}\sum_{t=1}^{j}\left(\frac{\bar{Q}_t^i - Q_t^i}{Q_t^i}\right)^2\right]^{1/2}, \qquad (5)$$

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

RESULTS

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. Regionalization approaches such as cluster analysis require selection of variables that are used to define the similarity (or dissimilarity) for the catchments [1]. Hosking and Wallis [9] recommended using methods that rely on site characteristics only when identifying homogeneous regions, and subsequently using the site characteristics to test independently 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. 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 [8] is used for testing the heterogeneity of the proposed regions.

Total area of 15 selected sites is 16176 km^2. The identified homogeneous regions (1) and (2) include 2589 and 13587 km^2, respectively. Most of the sites with large area (1590–5428 km^2) are located in the region (2). Most of the smaller sites (area <314 km^2) are located in the region (1). 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.

Homogeneity of the Regions

L-moments and L-moment ratios are the bases for all stages of the L-moments approach, such as identification of unusual sites (discordancy test), homogeneity test, the goodness-of-fit measure for determina-

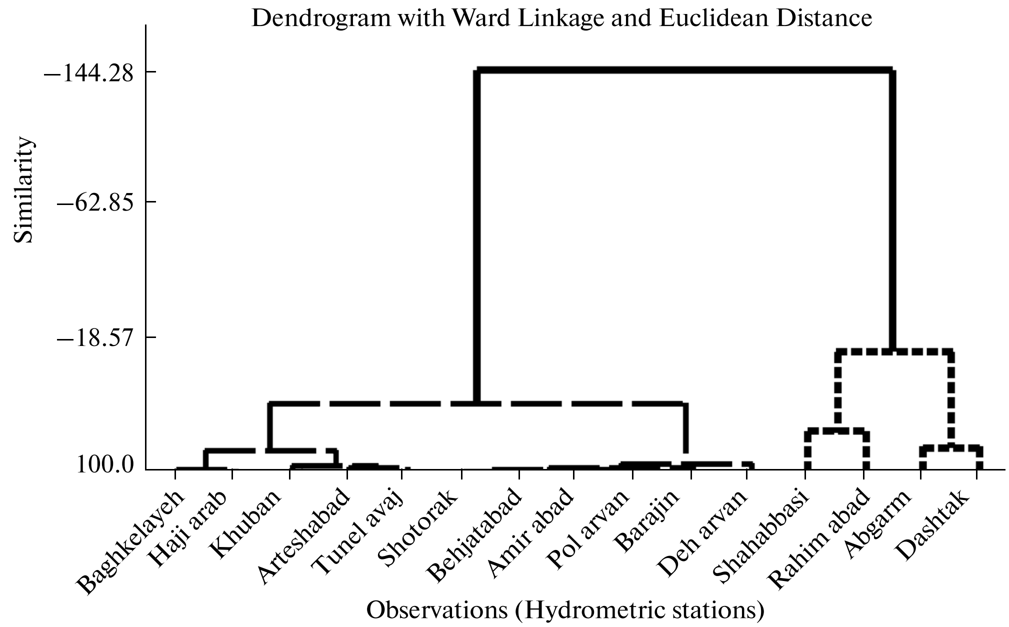

Fig. 2. Dendrogram of clustered basins by Ward's method.

tion of the best distribution for each homogeneous region, and parameter estimations (location (ξ), scale (δ), 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 PWMs (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 ($H < 1$), possibly heterogeneous ($1 < H < 2$) and definitely heterogeneous ($H > 3$).

The results of homogeneity test based on L-moments approach for whole study area show the selected sites should be divided into some homogeneous sub-regions. The results show that the absolute values of H-statistic for region (1) <2 and for region (2) <1 and these regions are acceptably homogeneous.

Identification of the Best-Fit Distribution

In this study, five three-parameter distributions were selected: GEV, GPA, GLO, three-parameter GNO, and 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 RFFA are obtained by the regional averages of the L-moments and L-moment ratios. The estimated regional parameters were obtained for GPA and GLO as the best-fit distributions. 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 two homogeneous regions.

As it is shown in Table 2 the GLO has the best goodness-of-fit with the data from region (1) and GEV, GNO and GPA are the other acceptable distributions. The Z value of these distributions is less than 1.64, and the GLO has the lowest value of Z-statistic. The GPA is the best-fit distribution for the flood analysis at the region (2); and other distributions are not acceptable.

Table 2. Goodness-of-fit analysis (Z^{DIST}) for five different frequency distributions

Region	PE3	GPA	GNO	GEV	GLO
1	–2.74	–1.45	–1.14	–0.2	0.09
2	2.46	1.56	2.12	4.76	3.68

Index-Flood Results

Two regional analyses were used to develop methods for estimating peak flood discharges for the basins in Qazvin province 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 generalized least-

Table 3. The values of regional growth curve (q^T) of different frequencies

Non-exceedance probability	0.99	0.98	0.96	0.90	0.80	0.50
Return period	100	50	25	10	5	2
Region 1	6.36	4.56	3.24	2.01	1.34	0.64
Region 2	4.14	3.67	3.15	2.38	1.72	0.73

Table 4. The predicted flood magnitudes (m^3/s) by best-fit distribution and direct use of the data

Homogeneous region	Station	Q_2	Q_5	Q_{10}	Q_{25}	Q_{50}	Q_{100}
Region 1	Artesh abad	19	39	59	96	135	189
	Amir abad	15	32	49	79	111	156
	Barajin	15	32	49	79	111	155
	Baghkelayeh	66	138	208	336	473	660
	Behjat abad	12	25	38	61	87	121
	Pol arvan	15	32	49	79	112	156
	Tunel avaj	29	61	92	149	210	293
	Haji arab	39	83	124	201	283	394
	Khuban	33	69	104	168	236	330
	Deh arvan	4	9	14	22	31	44
	Shotorak	10	21	32	52	73	102
Region 2	Abgarm	60	141	196	259	302	341
	Shah abbasi	25	59	82	109	127	143
	Dashtak	39	93	129	171	199	224
	Rahim abad	80	189	262	348	405	456

squares regression to define a set of predictive equations that relate peak discharges for the 2-, 5-, 10-, 25-, 50-, and 100-year recurrence intervals to the five obtained hydroclimatic and physical characteristics by factor analysis. The results of two methods were compared by RRMSE.

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 GLO and GPA distributions, as the best-fit distributions for both regions, were used for estimating the regional growth curve for different return periods. The results are given in Table 3 for two homogeneous regions.

Multiple-Regression Results

The aim is to develop the relationships between Q_T as dependent variable and main site and at-site characteristics as independent variables at each homogeneous region. The parameters of generalized extreme value distribution as the best-fit identified distribution, i.e., location (ξ), scale (δ), 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-, and 100-year intervals were separately obtained for each site. The results are shown in Table 4.

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

$$Q_T = \alpha P^p T^t D^d, \quad (6)$$

where Q_T is the peak flood magnitude (m^3/s) for T-year return period, P is the perimeter (km), T is time of concentration using kerpitch method (hr), D is diameter (km), α, p, t, d are the regression coefficients, R is the correlation coefficient defined by regression analysis. These coefficients were calculated

Table 5. The coefficients of regression equation for region (1)

Recurrence interval (T), year	α	p	t	R^2
2	−198.17	144.8	−97.67	0.842
5	−414.40	302.8	−204.2	0.859
10	−622.05	454.6	−306.6	0.882
25	−1002.87	732.9	−494.3	0.863
50	−1411.01	1031.1	−695.5	0.845
100	−1968.64	1438.6	−970.3	0.856

using multiple-regression for different flood probabilities at each homogeneous region. The coefficients of regression equation for each region are shown in Tables 5–6.

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, RRMSE measure was applied. The lower value of RRMSE for each method indicates that the method has better fitness to the data set.

As it has been shown in Table 7, RRMSE values indicate that the index-flood method gives better results for prediction of flood magnitude of different return periods for the region (1). For region (2) multiple-regression shows better performance than the index-flood method. In this region, correlation coefficient between Q_T and basin characteristics is high ($R >$ 0.946). In general, the large difference between RRMSE obtained by the index-flood and the multiple-regression methods shows that the index-flood method gives more reliable estimations for various flood magnitudes of different recurrence intervals. This method should be adopted as the regional flood frequency method for Qazvin province of Iran.

DISCUSSION

Firstly the study area was analyzed as a whole and then as two smaller sub-regions using cluster analysis technique. When data record period is short, the direct application of probability distributions for anticipating 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 RFFA. L-moments approach applies simultaneous use of all data from 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 15 independent variables could be referred to five factors. The area, perimeter, time of concentration, equivalent diameter, and water way length 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 regionalization objectives in hydrology. The dendrogram of clustered basins was cut from the distance of 20 to define the initial homogeneous regions in the study area. The results of homogeneity test based on L-moments approach at the whole study area show the selected sites should be divided into two homogeneous sub-regions. Using the L-moment ratios and the Z-statistic criteria, GLO and GPA distributions were identified as the most robust distributions among five potebtial distributions for regions (1) and (2), respectively. The estimated regional growth curves were significantly different for the two 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 identified distributions.

The RRMSE values indicate a considerable difference between the index-flood and the multiple-regression approaches. The RRMSE values of these two methods showed the multiple-regression gives acceptable estimations just for high values of correla-

Table 6. The coefficients of regression equation for region (2)

Recurrence interval (T), year	α	d	t	p	R^2
2	1826	1838	−302	−1142	0.918
5	4296	4325	−710	−2688	0.903
10	5948	5990	−984	−3722	0.922
25	7883	7937	−1304	−4933	0.896
50	9177	9241	−1519	−5743	0.912
100	10414	−1711	−6472	10343	0.914

Table 7. The RRMSE values of index-flood and multiple-regression methods

Region	Index-flood	Multiple-regression
1	3.674	5.692
2	0.995	1.002

tion coefficient. It can be concluded that the high difference between RRMSE obtained by the index-flood and the multiple-regression methods shows that the index-flood method gives more reliable estimations for various flood magnitudes of different return periods.

CONCLUSIONS

The study presented herein reports a regional analysis carried out in Qazvin province of Iran. Comparison of the index-flood and the multiple-regression analyses based on L-moments was the main objective of this study. Results showed the index-flood method gives more reliable estimations for various flood peaks of different recurrence intervals. Therefore, this method should be adopted as regional flood frequency method for the study area.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Qazvin regional water company (Iran) for providing data and financial support.

REFERENCES

1. Burn, D.H., Catchment similarity for regional flood frequency analysis using seasonality measures, *J. Hydrol.*, 1997, vol. 202, nos. 1–4, pp. 212–230.
2. Chowdhury, J.U., Stedinger, J.R., and Lu, L.H., Goodness-of-fit tests for regional GEV flood distributions, *Water Resour. Res.*, 1991, vol. 27, no. 7, pp. 1765–1776.
3. Cunnane, C., Methods and merits of regional flood frequency analysis, *J. Hydrol.*, 1988, vol. 100, pp. 269–290.
4. Dalrymple, T., Flood frequency analyses. Water Supply Paper 1543-A, *US Geological Survey*, 1960, pp. 11–51.
5. Groupe de recherche en hydrologie statistique (GREHYS), Presentation and review of some methods for regional flood frequency analysis, *J. Hydrol.*, 1996, vol. 186, nos. 1–4, pp. 63–84.
6. Groupe de recherche en hydrologie statistique (GREHYS), Inter-comparison of regional flood frequency procedures for Canadian rivers, *J. Hydrol.*, 1996, vol. 186, nos. 1–4, pp. 85–103.
7. Heinz, D.F. and Stedinger, J.R., Using regional regression within index flood procedures and an empirical Bayesian estimator, *J. Hydrol.*, 1998, vol. 210, pp. 128–145.
8. Hosking, J.R.M. and Wallis, J.R., Some statistics useful in regional frequency analysis, *Water Resour. Res.*, 1993, vol. 29, pp. 271–278.
9. Hosking, J.R.M. and Wallis, J.R., *Regional Frequency Analysis: An Approach Based on L-Moments*, London: Cambridge Univ. Press, 1997.
10. Hosking, J.R.M., L-Moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. R. Statis. Soc.*, 1990, vol. 52, no. 1, pp. 105–124.
11. Kjeldsen, T.R., Smithers, J.C., and Schulze, R.E., Regional flood frequency analysis in the KwaZulu–Natal province, South Africa, using the index-flood method, *J. Hydrol.*, 2002, vol. 255, nos. 1–4, pp. 194–211.
12. Malekinezhad, H., Nachtnebel, H.P., and Klik, A., Comparing the index-flood and multiple regression methods using L-moments, *Phys. Chem. Earth*, 2011, vol. 36, pp. 54–60.
13. Mkhandi, S. and Kachroo, S., *Regional flood frequency analysis for Southern Africa. Southern African friend, Technical Documents in Hydrology no. 15*, Paris: UNESCO, 1997.
14. Pandey, G.R. and Nguyen, V.T.V., A comparative study of regression based methods in regional flood frequency analysis, *J. Hydrol.*, 1999, vol. 225, pp. 92–101.
15. Pilon, P.J. and Adamowski, K., The value of regional information to flood frequency analysis using the method of L-moments, *Can. J. Civil Eng.*, 1992, vol. 19, pp. 137–147.
16. Saf, B., Regional flood frequency analysis using L-moments for the West Mediterranean region of Turkey, *Water Resour. Manag.,* 2009, vol. 23, pp. 531–551.
17. Saf, B., Assessment of the effects of discordant sites on regional flood frequency analysis, *J. Hydrol.*, 2010, vol. 380, pp. 362–375.
18. Singh, V.P., Wang, S.X., and Zhang, L., Frequency analysis of nonidentically distributed hydrologic flood data, *J. Hydrol.*, 2005, vol. 307, pp. 175–195.
19. Vogel, R.M. and Fennessey, N.M., L-moments diagrams should replace product moment diagrams, *Water Resour. Res.*, 1993, vol. 29, pp. 1745–1752.
20. Vogel, R.M., Thomas, W.O., and McMahon, T.A., Flood-flow frequency model selection in south-western United States, *J. Water Resour. Plann. Manage.*, 1993, vol. 119, no. 3, pp. 353–366.
21. Vogel, R.M., McMahon, T.A., and Chiew, F.H.S., Floodflow frequency model selection in Australia, *J. Hydrol.*, 1993, vol. 146, pp. 421–449.
22. Zhang, J. and Hall, M.J., Regional flood frequency analysis for the Gan-Ming River basin in China, *J. Hydrol.*, 2004, vol. 296, pp. 98–117.
23. Zrinji, Z. and Burn, D.H., Flood frequency analysis for unguaged sites using a region of influence approach, *J. Hydrol.*, 1994, vol. 153, pp. 1–21.