Hydrologic drought indices analysis by regionalization methods in southwest of Iran

Monireh Biabanaki, Seyed Esmalian, Jahangir Abedi-Koupai

Hydrologic drought indices analysis by regionalization methods in southwest of Iran

Monireh BIABANAKI [1], Seyed Saeid ESLAMIAN [2], Jahangir ABEDI-KOUPAI [2] and Alireza TABATABAEI NAEINI [3]

[1] PhD Student, Water Engineering Department, Isfahan University of Technology, Isfahan, 8415683111, Iran
(E-mail: m.biabanaki@yahoo.com)
[2] Associate Professor, Water Engineering Department, Isfahan University of Technology, Isfahan, 8415683111, Iran
[3] Graduate student, Water Engineering Department, Isfahan University of Technology, Isfahan, 8415683111, Iran

Abstract

Drought is a recurrent extreme climate event with tremendous hazard for every specter of natural environment and human lives. Drought analysis usually involves characterizing drought severity, duration and intensity. Usually, long-term datasets of hydrometric and hydrochemical information are needed to begin an evaluation of dominant low flow (as hydrologic drought indices) producing processes however, in many catchments, these data are not available. A major research challenge in ungauged basins is to quickly assess the dominant hydrological processes of watersheds. In this paper, for developing regional models, low flow analysis has been performed by 3 regression methods (multivariate regression, low flow index method, regionalization model of frequency formula parameters) and Hybrid low flow model in Karkheh basin (southwestern of Iran). Estimated error for four methods show although hybrid method can also use for low flow regionalization analysis but multivariate regression and low flow index methods are more suitable for this purpose.

Keywords

hybrid model; multivariate regression; low flow index; hydrologic drought; prediction in ungauged sites

INTRODUCTION

Droughts are sustained and regionally extensive occurrences of below-average natural water availability. They affect all components of the water cycle: from deficits in soil moisture through reduced groundwater recharge and groundwater levels to low streamflows or dried-up rivers. Droughts are reoccurring and worldwide phenomena, with spatial and temporal characteristics that vary significantly from one region to another and can have wide-ranging social, environmental and economic impacts.

So, estimates of the magnitude and frequency of annual low flow at ungauged sites on unregulated streams are necessary.

A 7-day annual low flow is the lowest flow recorded at a given site, over seven consecutive days in a single year. Annual values for the entire period of data are averaged to give the 7-day annual low flow for that site. However, this is only accurate for long term monitoring sites (primary sites) with continuous data record, so for those sites where there is only a limited number of spot gaugings available (tertiary sites) or there is not any gauging, statistical regressions are carried out to ascertain tertiary site flow characteristics from primary sites with a similar flow regime. These regression equations are then used to estimate 7-day annual low flows (Bayazit and Onoz, 2002; Riggs, 1985 and Zaidment et al., 2003).

It is necessary to determine the best-fitted distribution to studied data. The primary aim of frequency analysis is to relate the magnitude of extreme events to their frequency of occurrence using the probability of distributions (Chow et al., 1988). In this study seven well-known probable distributing models including two-parameter standard normal, two-parameter log normal, three-parameter log normal, two-parameter gamma, Pearson type III, log Pearson type III and Gumbel with moment and maximum likelihood parameters which are tested to determine the best fitted distributions as well as low flow in different return periods.

A lot of effort was made to establish low flow procedures and to evaluate methods to estimate low flow parameters at an ungauged site. The most popular tools in this type of regionalization are based on empirical methods and statistical approaches (Stall, 1962). Empirical methods, beside detailed knowledge of the basin physiography (e.g. morphology, geology, pedology, etc.) need additional information about the hydrological behaviour of the basin under investigation. The statistical approaches can be separated into correlation analysis and multiple regression analysis. The latter are some statistical methods to estimate flow parameters at an ungauged site and has to be considered the method most used in hydrology and water resources management.

Regional extreme value analysis involves three major steps:
1) Grouping of sites into homogeneous regions.
2) Estimation of the regional extreme value distribution for each homogeneous region (after dividing the random variable under study by the scale parameter). The data at different stations of the same homogeneous region can be combined for this task.
3) Estimation of at-site scale parameter and corresponding extreme value distribution.

A region can be considered homogeneous for low flow frequency analysis if sufficient evidence can be established that data at different sites in the region are drawn from the same distribution (except for the scale parameter). Hosking and Wallis (1993) developed several homogeneity tests for use in regional studies. The aim of these tests is to estimate the degree of heterogeneity in a group of sites and to assess whether they might reasonably be treated as a homogeneous region.

One of the low flow models is multivariate regression. In this model, the significant relationships between T-year low flows and characteristics of basin would be defined (Thomas and Benson (1970), Riggs (1973) and Tasker (1989), Dingman and Lawlor, 1995, Tasker, 1989, Vogel and Kroll, 1992). Another approach for analyzing low flows is low flow index method. The objective of this method is obtaining the regional frequency curves. Grover et al. (2002), Kumar et al. (2003) in India and Lim et al. (2003) in Sarawak have used the adjusted index method for regional flow frequency. They found, it is a suitable method for estimation of flow in the regions with insufficient records. Regionalizing the frequency formula parameters, mean and standard deviation, based on physiographic characteristics of basin is another method for low flow analysis (Tasker et al., 1989)

One of the methods for regional flood frequency that was developed for overcoming problems in arid regions is Hybrid model. This method has been originally used by Hjalmarsom (1992) for the arid regions in southwestern of United States (Hjalmarsom et al., 1992). Chavoshi and Eslamian (1998) have used this model for flood frequency analysis of Zayandehrud basin in Iran. They have found that for low and moderate return periods, Hybrid model has more accuracy in comparing with regression method and for the high return periods, both methods show similar results (Chavoshi et al., 1989).

This report presents the results of a study to characterize the magnitude and frequency of the 7-day annual low flow for selected streamflow-gaging stations and to develop regional regression equations to estimate the low-flow frequency at ungaged sites on in southwestern of Iran (Karkhek river basin). The low flow analysis is done using Hybrid method and the results were compared with: multi-variable regression, low flow index and regionalization of frequency formula parameters methods.

MATERIAL & METHODS

Homogeneity test

Andrews curves are examples of the space transformed visualization techniques for visualizing multivariate data, which represent k-dimensional data points by a profile line (or curve) in two- or three-dimensional space using orthogonal basis functions. Andrews curves are based on Fourier series where the coefficients are the observation's values. One advantage of the plot is based on the Parseval's identity (energy norm), which indicates that the information through transformation from the data space into the parameter space is preserved, and information that can be deduced in the hyperdimensional original space can be easily deduced in the two-dimensional parameter space. This duality empowers the discovery of correlated records, clusters and outliers based on the curve's intersections, gaps and isolations, respectively (Andrews, 1972 ; Nathan and McMahon 1990 ; Brian, 2000; Manciola and Casadei, 1991; Smakhtin, 2001).
For plotting Andrew's curves, the following relation has been used:

$$F(t) = \frac{X_1}{\sqrt{2}} + X_2 Sin(t) + X_3 Cos(t) + X_4 Sin(2\,t) + X_5 Cos(2\,t) + ... \quad (1)$$

where, X_1, X_2, ... are the characteristics of basins.

Low flow models

<u>Regression method</u>

Regression analysis is used to detect a relation between the values of two or more variables of which at least one is subject to random variation and to test whether such a relation, either assumed or calculated, is statistically significant. It is a tool for detecting relations between hydrologic parameters in different places, between the parameters of a hydrologic model and so on (Thompson, 1999). In this study the following relationship has been used :

$$Q_T = f[X_1^{b_1}, X_2^{b_2}, ...] \quad (2)$$

where, Q_T is low flow with T-year return period, X_i is characteristics of basin and b_i is the constant parameters that should be estimated from the multivariate regression technique.

Because a linear regression model is not always appropriate for the data, appropriateness of the model should assess by defining residuals and examining residual plots. The difference between the observed value of the dependent variable (Y_i) and the predicted value ($\hat{Y}_i$) is called the residual (e). Each data point has one residual:

$$e_i = Y_i - \hat{Y}_i \tag{3}$$

Both the sum and the mean of the residuals are equal to zero. That is, Σe=0 and e=0.

Variance inflation factors (VIF) are a measure of the multi-colinearity in a regression design matrix (i.e., the independent variables) (Thompson, 1999). Multi-colinearity results when the columns of X have significant interdependence (i.e., one or more columns of X is close to a linear combination of the other columns). Multi-colinearity can result in numerically unstable estimates of the regression coefficients (small changes in X can result in large changes to the estimated regression coefficients).

VIFs are a scaled version of the multiple correlation coefficients between variable j and the rest of the independent variables:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{4}$$

where R_j is the multiple correlation coefficient.

Variance inflation factors are often given as the reciprocal of the above formula. In this case, they are referred to as the tolerances.

Low flow index method

Another method for regional analysis is low flow index method, follows by these steps:
1) Homogeneity test for the stations would be done, 2) For each station, flow frequency curve draw, 3) Low flows with 2 years return period should be computed, 4) The ratio of $\frac{Q_T}{Q_2}$ for different return periods would be drawn (regional frequency curve), 5) Regression model for the low flows with 2 years return period be obtain (with basin physiographic characteristics) and Q_2 calculate, 6) Q_T (low flow for ungauged stations in different return periods) could be determined.

Hybrid method

In this method, at first the study area divide into some homogeneity classes and the annual discharge would be standardized for each class (Hjalmarsom et al., 1992). For the next step, the standardized discharge would be corrected and parameters of Hybrid equation have been calculated in composed of regression and regional analysis (using iteration method).
For each iteration, the regional relations for each class would be defined.
The model for this method is:

$$Q_T = aA^bB^cC^d \ldots \tag{5}$$

where, Q_T is discharge with T-year return period, A, B and C are independent hydrologic parameters and a, b, c and d are the constant components of the regression model.

Previous studies have shown that the basin area is the most significant independent variable (Hjalmarsom et al., 1992). So, the basin area enters as the first independent variable in relation (5) and other parameters (B, C,…) would be equivalent with one.
The maximum number of classes (j) for the area would be defined as follows:

$$j \le \frac{N_f}{100} \tag{6}$$

where, N_f is the summation of number of data.
The weighted mean of area for each class would be defined according to the following equation:

$$\overline{A}_i = \text{antilog}\left[\frac{\sum_{j=1}^{g}\sum_{k=1}^{h}\log A_{ijk}}{gh}\right] \quad (7)$$

where, A_i is the weighted mean of area for class i (i=1, 2,….,f), A_{ijk} is the basin area for station j in class i and for station–year k, j is the number of stations in class i (j=1, 2,…,g) and k is the years of station j in class i (k=1, 2,…,h).

The first step of iteration process for the area begins with standardization of annual discharges. Standardized discharge is :

$$S_{ijk} = \frac{Q_{ijk}}{\overline{A}_{ijk}^{\,b}} \quad (8)$$

where, S_{ijk} is k^{th} standardized discharge for station j in class i, Q_{ijk} is k^{th} annual discharge for station j in class i, A_{ijk} is weighted mean of k^{th} area for station j in class i, (the initial value of b is one) .

For obtaining the discharge values with a specific T-year return period, the computed value of S_{Ti} in each class should be standardized by:

$$Q_{Ti} = S_{Ti}(\overline{A}_i)^b \quad (9)$$

where, Q_{Ti} is T-year discharge in class i, S_{Ti} is the standard discharge in class i for T-year and A_i is weighted mean of area according to relation (7). Also b_T is:

$$b_T = \frac{\sum_{i=1}\overline{A}_i\, Q_{Ti} - \frac{\left[\sum_{i=1}^{f}\overline{A}_i \sum_{i=1}^{f} Q_{Ti}\right]}{f}}{\sum_{i=1}^{f}\overline{A}_i^{\,2} - \frac{\left[\sum_{i=1}^{f}\overline{A}_i\right]^2}{f}} \quad (10)$$

The next step of the process continues with replacing b in relation (8) and the process needs to be done until finding out the stable amount of b. The value of b will be fixed after one or two iterations otherwise, the linear relation does not exist between the parameter and discharge, so this parameter will be removed from the model.

Another parameters such as mean elevation and basin slope could be used in the same manner (Hjalmarsom et al., 1992).

Estimated error

In this paper, the differences between observed and estimated values calculated by relative error (RE) and root mean square of error (RMSE) (Chalise et al., 2003, Vogel and Kroll, 1990; Wiltshire, 1985).

RE in data is the discrepancy between an exact value and some approximation to it.

$$RE = \frac{\left|\hat{Q}_T - Q_T\right|}{Q_T} \times 100 \quad (11)$$

where, $\hat{Q}_T$ is the estimated low flow value and Q_T is the observed data.

The root-mean-square deviation (RMSD) or RMSE is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed. RMSE is a good measure of accuracy. These individual differences are also called residuals, and the RMSE serves to aggregate them into a single measure of predictive power.

$$\mathrm{RMSE} = \sqrt{\frac{\sum (\hat{Q}_T - Q_T)^2}{N}} \qquad (12)$$

where, N is the number of stations in each homogeneous area.
The models with more accuracy display the less RE and RMSE.

Study area

Karkheh basin is located in southwestern of Iran and south of Zagros Mountains. The basin area is about 51268 square kilometer (66 percent valley and 34 percent flat). The main slope of the basin is directed from north to the south. Karkheh River has been formed by the linkage of the following rivers: Gamasiab, Ghareso, Seimare and Kashkan. Figure 1 shows the location of study area and hydrometric stations.

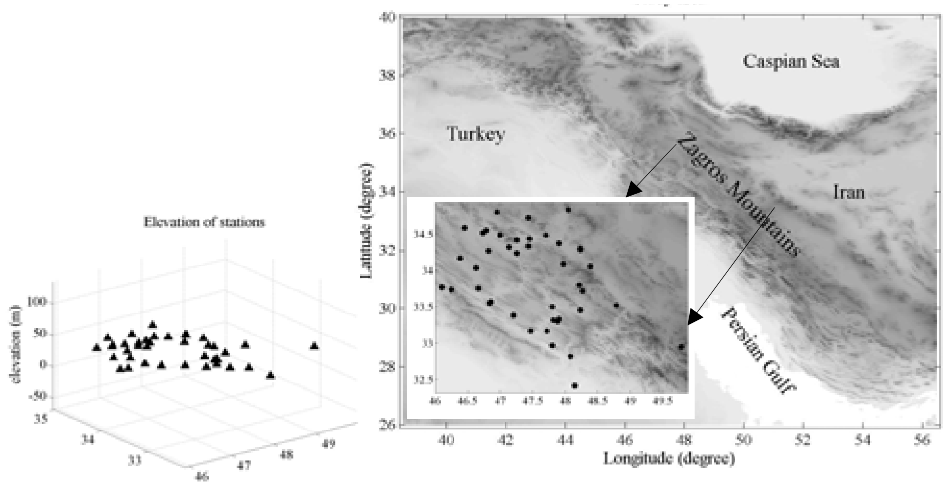

Fig. 1. Study area and hydrometric stations

RESULTS AND DISCUSSION

Data Sources

For this study, daily low flow data for 41 hydrometric stations have been used (Fig. 1). These data received from Iranian Water Resources Management Company. The run test and lag plot used for randomness analysis of the series.

7-day low flow values have been calculated by the time series. The gamma distribution has the best fitting on these data in study area. Using this distribution, low flows with different return periods have been determined.

Homogenity test

For investigation of homogenous area, Andrew's curves have been used. The results show that 35 stations are in a homogenous region.

Low flow models

Multivariate regression of low flow method

For this method, 7-day low flows with different return periods (5, 10, 20, 25, 50 and 100 years) used. The most important physiographic characteristics that are effective on the low flow estimates have been selected (by maximum correlation coefficient; minimum standard error and variance inflation factor lower than 5). Regression models are as:

$$Q_5 = 10^{-0.442BR - 5.01\times10^{-4}E + 1.38\times10^{-4}A + 0.535} \quad (13)$$

$$Q_{10} = 10^{-0.460BR - 5.44\times10^{-4}E + 1.405\times10^{-4}A + 0.452} \quad (14)$$

$$Q_{20} = 10^{-0.481BR - 5.94\times10^{-4}E + 1.411\times10^{-4}A + 0.386} \quad (15)$$

$$Q_{25} = 2.539\times10^{-4}A + 0.124S - 1.767 \quad (16)$$

$$Q_{50} = 2.211\times10^{-4}A + 0.111S - 1.589 \quad (17)$$

$$Q_{100} = 1.945\times10^{-4}A + 0.101S - 1.438 \quad (18)$$

where parameters are as: A : Area(km2), S : Slope (%), BR : Bifurcation Ratio (the ratio of the number of streams of any given order to the number of streams in next higher order), E : Elevation (m) and Q_T : 7-day low flow with different return periods for a specific basin.

Low flow index method

In this model, $\frac{Q_T}{Q_2}$ should be calculated for different return periods (Fig. 2).

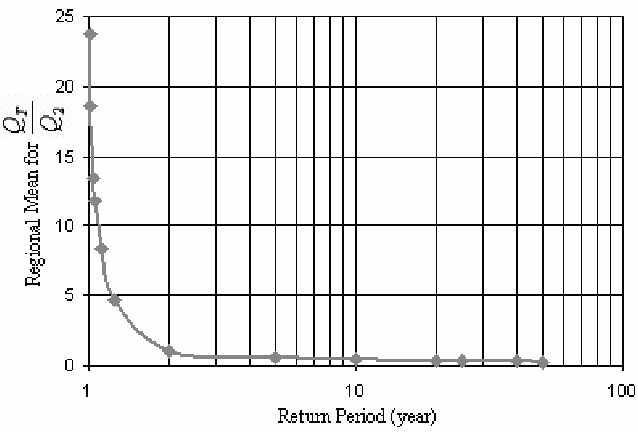

Fig. 2. Regional frequency curve

The final developed equation is:

$$Q_2 = 10^{-0.409BR - 3.65\times10^{-1}E + 1.38\times10^{-4}A + 0.634} \quad (19)$$

where, A : Area(km^2), BR : Bifurcation Ratio E : Elevation (m) and Q_2 : 7-day low flow with 2 years return period.

Regionalizing the frequency formula parameters method

For this method, the relation between average and standard deviation of 7-day low flows and characteristics of the basin has been studied:

$$\text{Mean}(Q) = 10^{1.435\times10^{-4}A - 2.678\,\text{Log}BR - 2.8\times10^{-4}E + 0.587} \quad (20)$$

$$\mathrm{Std}(Q) = 10^{6.147\times10^{-3} MsL - 0.263 \mathrm{Log} E - 0.203 BR + 4.618\times10^{-2}} \quad (21)$$

where, Mean(Q) is the average of 7-day low flows (m^3/s), A is the area of basin (km^2), BR is bifurcation ratio, E is basin elevation (m), MsL is the mainstream length (km) and Std(Q) is the standard deviation of 7-day low flows series.

Finally, the value of low flows with different return periods is calculated by cumulative distribution function of 2 parameters gamma distribution as follows (Smakhtin, 2001):

$$F_G(x : \alpha, \beta) = \int_0^\infty \frac{\alpha^\beta t^\beta e^{-\alpha} t}{t\Gamma(\beta)} dt \quad (22)$$

where, α and β are the scale and shape parameters respectively.

Hybrid Method

Classifying the area is the basis of Hybrid method. In this step, two factors (area and mean slope) selected and the study area divided into five different parts (by cluster method).
The Hybrid model is:

$$Q_T = aA^b S^c \quad (23)$$

where, Q_T is the 7-day low flow with T-year return period (m^3/s), A is the basin area (km^2), S is the mean slope (m/m) and a, b and c are the components of regression.
Each regression's component is determined separately. At the first step for determining component b, basin area is selected as the first factor and the others (mean slope of basin, S) are considered stable.
The values of low flow with different return periods (5, 10, 20, 25, 50 and 100 years) extracted for each region and the b values have been determined (for different return periods).

The regional models for estimating low flows by this method are:

$$Q_5 = 4.868 A^{-0.157} \quad (24)$$

$$Q_{10} = 14.102 A^{-0.031} S^{-0.808} \quad (25)$$

$$Q_{20} = 2.348 A^{-0.012} S^{-0.273} \quad (26)$$

$$Q_{25} = 1.770 A^{-0.008} S^{-0.197} \quad (27)$$

$$Q_{50} = 1.027 A^{-0.003} S^{-0.064} \quad (28)$$

$$Q_{100} = 0.780 A^{-0.001} S^{-0.016} \quad (29)$$

Estimated error

Estimated error of four mentioned methods have been considered by RE and RMSE (Fig. 3).

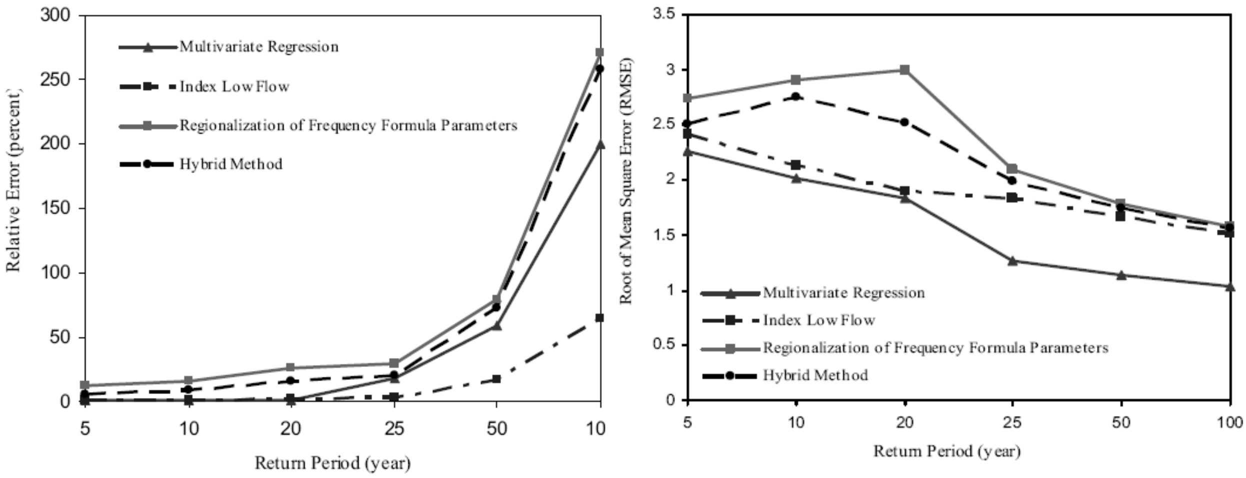

Fig. 3: Relative error (percent) and Root of mean square error (RMSE) for four methods

For all of return periods, the multivariate regression and low flow index methods have more accuracy in comparison with regionalizing the frequency formula parameters and Hybrid methods. In the basis of RE, low flow index and in the basis of RMSE multivariate regression methods are the suitable models for low flow analysis.

Also the results show multivariate regression and low flow index methods for the low return periods and regionalizing the frequency formula parameters and Hybrid models for the high return periods (higher than 50 years) do not have a significant difference.

Multivariate regression and low flow index methods are more suitable than Hybrid model.

Accuracy of Hybrid method (that usually used for flood regionalization) in comparison with other models has been indicated in Table 1.

Table 1: RE (percent) for three methods in comparing with Hybrid model

Low Flow Models	Return Periods					
	5	10	20	25	50	100
multivariate regression	84.44	87.65	75.13	103.60	99.68	98.91
low flow index	86.73	88.47	78.48	77.09	74.65	73.46
regionalizing the frequency formula parameters	128.30	171.43	134.78	114.53	104.99	90.33

CONCLUSIONS

Reliable estimation of low stream flows is necessary to investigate drought characteristics of the basin and to describe the capability of a stream to supply requirements for river navigation, municipal, industrial, liquid waste disposal, irrigation and maintenance of suitable conditions for aquatic life. The important characteristics in the study of low streamflow are its magnitude, frequency and duration. This study shows the evaluation of three regional low flow methods and also Hybrid method (a model for flood regionalization), for overcoming the problem of short record length of low flow data in an arid region (southwest of Iran). The results indicate, Hybrid method can be used for hydrologic drought prediction but multivariate regression and low flow index methods are more suitable than Hybrid model.

Using climatic data (for example Precipitation, Temperature and ..., that are important in low flows) instead of physiographic data (Area and Slope) in Hybrid method, maybe have more accuracy for applying this model (it can consider in future researches).

REFERENCES

- ANDREWS D. F. (1972). Plots of high-dimensional data. Biometrics, vol. 28, no. 1, pp. 125-136.
- BAYAZIT M. and Onoz B. (2002). L-moments for estimating low flow quantities. Hydrological Sciences Journal, vol. 74, 5rd edn, pp. 707-720.
- BRIAN S. C. (2000). Evaluation of low flow frequency analysis methods. Journal of Hydrology, vol. 39, 1rd edn, pp 19-47.
- CHALIE S. R., Kansakar S. R., Rees G., Croker K. and Zaydman, M. (2003). Management of water resources and low flow estimation for the Himalayan basins of Nepal. Journal of Hydrology, vol. 282, 1-4rd edn, pp. 25-35.
- CHAVOSHI S. and Eslamian, S. S. (1989). Estimation of flood discharge with different recurrence intervals in Zayandehrud basin by Hybrid method. Iranian Journal of Science and Technology of Agriculture and Natural Resources, vol. 3, 3rd edn, pp. 1-11.
- CHOW V. T., Maidment D. R. and Mays L. W. (1988). Applied hydrology. McGraw-Hill, New York, NY.
- DINGMAN S. L. and Lawlor S. C. (1995). Estimating low flow quantiles in New Hampshire and Vermont. Water Resources Bulletin, vol. 31, 2rd edn, pp. 243-256.
- GROVER P. L., Burn D. H. and Cunderlik J. M. (2002). A comparison of index flood estimation procedures for ungauged catchments. Canadian Journal of Civil Engineering, vol. 29, 5rd edn, pp. 734-741.
- HJALMARSOM H. W. and Thomas B. E. (1992). A new look at regional flood frequency relations for arid lands. ASCE, Journal of Hydraulic Engineering, vol. 118, 6rd edn, pp. 868-886.
  HOSKING J. R. M and Wallis J.R. (1993). Some statistics useful in regional frequency analysis. Water Resources Research, vol. 29, 2rd edn, pp. 271-281.
- KUMAR R., Chatterjee, C. and Kumar, S. (2003). Regional flood formulas using L-moments for small watersheds of stone subzone of India. Applied Engineering in Agriculture, vol. 19, 1rd edn, pp. 47-53.
- LIM Y., Howe, L. and Leonard M. (2003). Regional flood estimation for ungauged basins in Sarawak, Malaysia. Hydrological Sciences Journal, vol. 48, 1rd edn, pp. 79-94.
- MANCIOLA P. and Casadei S. (1991). Low flow Index and hydrogeological characteristics. National Conference on Hydraulic Engineering, pp. 930-936.
- NATHAN R. J. and McMahon T. A. (1990). Identification of homogeneous regions for the purpose of regionalization. Journal of Hydrology, vol. 121, pp. 217-238.
- RIGGS H. C. (1985). Streamflow characteristics. Elsevier, 249 pp.
- SMAKHTIN V. U. (2001). Low flow hydrology: a review. Journal of Hydrology, vol. 240, pp. 147-186.
- STALL J. B. (1962). Reservoir mass analysis by a low flow series. J. Sanit. Eng. Div. ASCE 88(SA5).
- THOMPSON S. A. (1999). Hydrology for water management. A.A. Balkema Publication Rotterdam, 362 pp.
- TASKER G. D. (1982). Comparing methods of hydrology regionalization. Water Resources Bulletin, vol. 18, 6rd edn, pp. 965-970.
- TASKER G. D. (1989). Regionalization of low flow characteristics using logistic and GLS regression. Proc. of New Direction for Surface Water Modeling, IAHS Publication, vol. 181, pp. 323-331.
- VOGEL R. M. and Kroll C. N. (1992). Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics. Water Resources Research, vol. 38, 9rd edn, pp. 2451-2458.
- VOGEL R. M. and Kroll C. N. (1990). Generalized low-flow frequency relationships for ungauged sites in Massachusetts. Water Resources Bulletin, vol. 26, 2rd edn, pp. 241-253.
- WILTSHIRE S. E. (1985). Grouping basins for regional flood frequency analysis. Hydrological Sciences Journal, vol. 30, pp. 151-159.
- YU P. S., Yang T. C. and Liu C. W. (2002). A regional model of low flow for southern Taiwan. Hydrological Processes, vol. 16, 10rd edn, pp. 2017-2034.

- ZAIDMENT M. D., Keller V., Young A. R. and Cadman D. (2003). Flow-duration-frequency behaviour of British rivers based on annual minima data. Journal of Hydrology, vol. 277, 3-4rd edn, pp. 195-213

17-WWW-YES-2012-Biabanaki-Paper-2012-10-04-DT.doc