Journal of Urban and Environmental Engineering, v.7, n.2 p. 247-252, 2013

ISSN 1982-3932
doi: 10.4090/juee.2013.v7n2.247252

Journal of Urban and Environmental Engineering

www.journal-uee.org

EVALUATION OF L-MOMENT AND PPCC METHOD TO DETERMINE THE BEST REGIONAL DISTRIBUTION OF MONTHLY RAINFALL DATA: CASE STUDY NORTHWEST OF IRAN

Babak Amirataee* and Majid Montaseri

Department of Water Engineering, University of Urmia, Urmia, Iran

Received 11 April 2013; received in revised form 6 September 2013; accepted 7 September 2013

Abstract: The analysis and use of hydrological data for decision making in water resources planning and management can only be meaningful if the data possess the appropriate characteristics. Whereas, rainfall stations are relation together in the studying area, so that choosing a best regionally probability distribution is necessary. In this paper, probability plot correlation coefficient (PPCC) test statistics and L-moment ratio diagrams are used to determine the goodness of fit the regional distribution of monthly rainfall data in 11 stations that located in Northwest of Iran. Two methods provide Pearson III as a best regional distribution of monthly rainfall data in our study area. As regards, PPCC test has been known as a powerful single-site test among many goodness of fit, but L-Moment approach is easy and can compare the fit of several distributions to many samples of data using a single graphical instrument.

Keywords: L-moment; PPCC method; best regional distribution

* Correspondence to: Babak Amirataee, Tel.: +98 441 2772407.
E-mail: Babak.amirataee@gmail.com

INTRODUCTION

Most statistical analyses of hydrological time series data at the usual time scale (e.g. monthly or annual) encountered in water resources planning studies are based on follows the appropriate probability distribution function (Adeloye & Montaseri, 2002). Therefore, the selection of an appropriate probability distribution is very important for the reasonable design. Generally, the selection of an appropriate probability distribution is based on the goodness of fit test which is the decision-making method to evaluate the fitness between sample data and its population for a given probability distribution (Sooyoung *et al.*, 2008). These include the Chi-squared test, the Kolmogorov-Smirnov test, the Probability Plot Correlation Coefficient (PPCC) test, and the Moment (L-moments and P-moments) ratio diagrams test (see Stedinger *et al.*, 1993).

The Probability Plot Correlation Coefficient (PPCC) test was first introduced by Filliben (1975) has been known as a powerful single-site test among many goodness of fit tests and now widely used (e.g., Stedinger *et al*., 1993; Kottegoda & Rosso, 1997; Beirlant *et al*., 2005; Heo *et al*., 2008). The probability plot correlation coefficient (PPCC) measures the linearity of the plot under an assumed distribution and provides a quantitative measure for comparing the relative goodness of fit of a fitted distribution (Vogel, 1986). PPCC test is easy to apply and yet has sufficient power to discriminate between different distribution hypotheses (Stedinger *et al*., 1993).

The test is based on the correlation coefficient between the ordered sample Y_i, i.e. such that $Y_1 \le Y_2 \le \ldots Y_n$ and their corresponding fitted quintiles, $W_i = G^{-1}(1 - P_i)$ i = 1, 2, …, n where P_i is the exceedence probability of Y_i and $G^{-1}(.)$ is the inverse of the cumulative distribution function (cdf) of the distribution being considered. Best estimates of P_i, can be obtained using (Cunnane, 1978):

$$P_i = \frac{R_{Yi} - 0.4}{n + 0.2} \tag{1}$$

where R_{ri} is the rank of y_i, in the ordered sample, i.e. R_{y1} = 1, R_{y2} = 2,..., R_{yn} = n. The estimated correlation coefficient between y_i, and w_i, is then compared with the critical points of the PPCC for the particular distribution. These critical points are provided by Vogel (1986) for the Normal, Log-Normal and Gumbel probability distribution functions. Vogel & McMartin (1991) also provide the critical values for the Gamma, Pearson type-3 (P3) and Log-Pearson type-3 (LP3) distributions. All these cover the range of distributions commonly used in hydrological studies (Adeloye & Montaseri, 2002).

An L-moment diagram compares sample estimates of the L-moment ratios L-cv, L-skew and L-kurtosis with their population counterparts for a range of assumed distributions. An advantage of L-moment diagrams of other goodness of fit procedures is that one can compare the fit of several distributions to many samples of data using a single graphical instrument. Another advantage of L-moment diagrams over ordinary product moment diagrams is that L-moment ratio are approximately unbiased for all probability distributions, unlike ordinary product moment ratios, which are significantly biased (Vogel & Fennessey, 1993). Vogel and Fennessey show that L-moment diagrams are always preferred to ordinary product moment diagrams, regardless of the sample size, probability distributions, or skew involved.

L-moments were introduced by Hosking (1990). L-moments are linear combinations of order statistics that are less sensitive to outliers and virtually unbiased for small samples (Hosking & Wallis, 1997). The L-moments of any probability distribution are defined by:

$$L_1 = \beta_0 \tag{2}$$

$$L_2 = 2\beta_1 - \beta_0 \tag{3}$$

$$L_3 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{4}$$

$$L_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{5}$$

where $\beta_r (r = 0,1,2,3)$ is the probability weighted moment, which can be defined as:

$$\beta_r = n^{-1} \sum_{j=r+1}^{n} \binom{j-1}{r} \binom{n-1}{r}^{-1} X_{(j,n)}, \qquad r = 0, n-1 \tag{6}$$

L-moment ratios: L-coefficient of variation (L-cv, τ_2), L-coefficient of skewness (L-skewness, τ_3), and L-coefficient of kurtosis (L-kurtosis, τ_4).

L-moment ratio diagrams are the plots of L-cv (τ_2) versus L-skewness (τ_3) for 2-parameter distributions, and L-kurtosis (τ_4) versus L-skewness (τ_3) for 3-parameter distributions. Hosking (1990), Stedinger *et al.* (1993), Hosking & Wallis (1997), and others have summarized the theory of L-moments.

To construct an L-moment ratio diagram it is convenient to have simple explicit expressions for L-kurtosis in terms of L-skewness and for L-cv in terms of L-skewness for some commonly used 3-parameter and 2-parameter probability distributions (Vogel & Wilson, 1996). Polynomial approximations of the form **Eqs (7)** and **(8)** have been obtained, and the coefficients are given in the **Tables 1** and **2** for commonly distributions used in hydrological studies.

$$\tau_4 = A_0 + A_1\tau_3^1 + A_2\tau_3^2 + \ldots + A_8\tau_3^8 \tag{7}$$

$$\tau_2 = A_0 + A_1\tau_3^1 + A_2\tau_3^2 + \ldots + A_7\tau_3^7 \qquad (8)$$

METHODS

Study area

The rainfall data records used in this analysis consist of annual and monthly time series from eleven synoptic stations located in northwest of Iran as shown in **Fig. 1**. Statistical properties of rainfall and geographic location of synoptic stations are presented in **Table 3**. The data records are 50 years long (1961-2010) with the mean annual rainfall ranging from 235 to 1755 mm which covers two semi-arid and humid climates (McMahon *et al.*, 2007). It is note that the most of synoptic stations placed in semi-arid zone.

The rainfall data records were checked using a number of statistical tests to examine their suitabilities for using in the analysis (Adeloye & Montaseri, 2002). The tests were applied to investigate the statistical qualities of homogeneity, randomness and stationary in data records. The Double mass curve method was used to test the homogeneity of the data (McGhee, 1985).

Table 1. Coefficients of polynomial approximations of L-cv as a function of L-skewness

	2-parameter distributions	
Coefficient	Lognormal 2 (LN2)	Gamma (GAM)
A0	0	0
A1	1.16008	1.74139
A2	-0.05325	0
A3	0	-2.59736
A4	-0.10501	2.09911
A5	0	0
A6	-0.00103	-0.35948
A7	0	0

Note: The approximate are good for $-0.1<\tau_3<1$, except for GAM, in which case they are only good for $0<\tau_3<1$.

Table 2. Coefficients of polynomial approximations of L-kurtosis as a function of L-skewness

Coefficient	3-parameter distributions			
	Lognormal 3 (LN3)	Pearson III	Normal	Gumbel
A0	0.12282	0.12240	-	-
A1	0	0	-	-
A2	0.77518	0.30115	-	-
A3	0	0	-	-
A4	0.12279	0.95812	-	-
A5	0	0	-	-
A6	-0.13638	-0.57488	-	-
A7	0	0	-	-
A8	0.11368	0.19383	-	-
L-Skewness	-	-	0	0.16990
L-Kurtosis	-	-	0.12266	0.15004

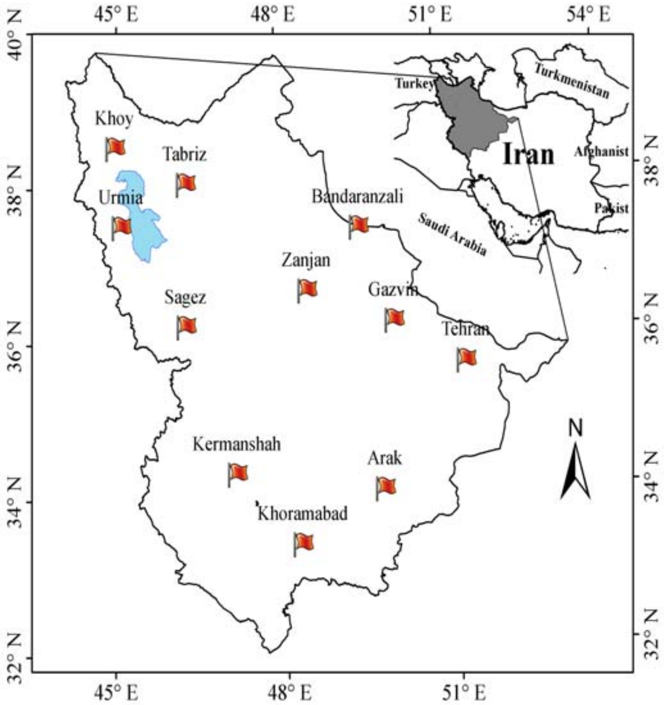

Fig. 1 Geographical location of stations in the study area.

Table 3. Statistical parameters of annual rainfall data

Station	Geographic coordinates		Statistical properties of rainfall series (1960-2010)		
	Lat.	Lon.	Mean	coefficient of variation (CV)	Skewness
Arak	34° 06′	49° 46′	332.3	0.30	0.32
Urmia	37° 32′	45° 05′	332.2	0.30	0.86
Bandaranzali	37° 28′	49° 28′	1755.5	0.19	0.76
Tabriz	38° 05′	46° 17′	283.2	0.30	0.95
Tehran	35° 41′	51° 19′	235.4	0.30	0.12
Khoramabad	33° 26′	48° 17′	503.1	0.24	0.07
Khoy	38° 33′	44° 58′	292.0	0.28	0.42
Zanjan	36° 41′	48° 29′	304.8	0.26	0.00
Sagez	36° 15′	46° 16′	487.6	0.27	0.55
Gazvin	36° 15′	50° 03′	317.0	0.27	0.33
Kermanshah	34° 21′	47° 09′	450.0	0.27	0.59

The Double mass curve procedure is based on the comparison of cumulative values of two data sets in a diagram form, one of the data sets being consistent, while the other is suspect. When plotted, the Double mass diagram should show a linear relationship when the suspect data set is consistent; otherwise, there will be a departure from Linearity.

The double mass curve as a sample for Urmia station is shown in **Fig. 2**. This figure showed homogeneity with a linear correlation coefficient of 0.99. The Non-parametric Run Test and Rank Order Correlation Coefficient Test (McGhee, 1985) were also applied to check the randomness and trend of the rainfall data records. The results of tests indicate that the tests statistic of random and stationary. Randomness and stationary of annual (all stations) and monthly data (Bandaranzali station) are shown in **Table 4** and **5**.

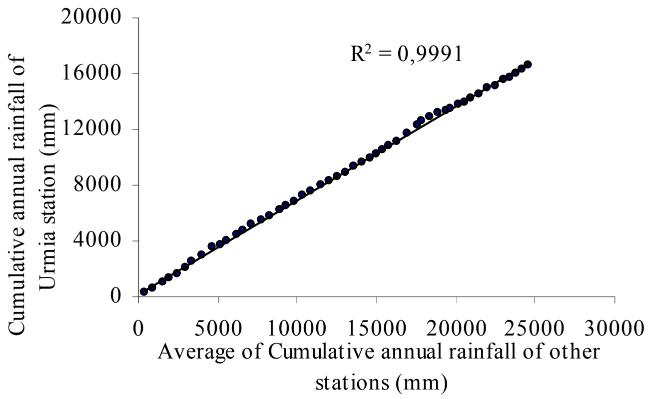

Fig. 2 Double mass curve for annual rainfall at Urmia station.

Table 4. Statistical tests of Randomness

Annual Data for all stations				
Stations	Randomness		Stationarity	
	Run test	Z (95%)	Trend	Z (95%)
Arak	-0.58	±1.96	0.05	±2.01
Urmia	-1.46	±1.96	1.67	±2.01
Bandaranzali	0.29	±1.96	-0.45	±2.01
Tabriz	-1.96	±1.96	2.00	±2.01
Tehran	-0.29	±1.96	-0.81	±2.01
Khoramabad	-1.46	±1.96	0.44	±2.01
Khoy	-0.87	±1.96	2.01	±2.01
Zanjan	-0.29	±1.96	-0.05	±2.01
Sagez	0	±1.96	0.76	±2.01
Gazvin	0.58	±1.96	-1.62	±2.01
Kermanshah	-1.46	±1.96	0.93	±2.01

Table 5. Statistical tests of Stationary

Monthly data for Bandaranzali station				
Month	Randomness		Stationarity	
	Run test	Z (95%)	Trend	Z (95%)
Oct.	0.58	±1.96	-0.04	±2.01
Nov.	0.29	±1.96	-0.30	±2.01
Dec.	-1.46	±1.96	0.59	±2.01
Jan.	0.29	±1.96	-0.79	±2.01
Feb.	1.17	±1.96	-2.01	±2.01
Mar.	-0.87	±1.96	0.45	±2.01
Apr.	1.46	±1.96	-0.60	±2.01
May	0.58	±1.96	0.22	±2.01
June	-0.29	±1.96	0.43	±2.01
Jul.	0.58	±1.96	-0.85	±2.01
Aug.	-1.75	±1.96	0.85	±2.01
Sep.	0.00	±1.96	-0.01	±2.01

Commonly probability distributions that used for annual and monthly hydrologic data are including Normal, Log-Normal, Pearson type three (Pearson III) and Log-Pearson type three (LP III) distributions (Yevjevich, 1972). So that, a variation of the Probability Plot Correlation Coefficient test (PPCC) was applied to the probability distributions of annual and monthly rainfalls. **Tables 6−11** present the PPCC correlation coefficients and RMSE for the annual and monthly rainfall data in all stations, respectively. Consequently, the Pearson type three distribution (Pearson III) is found to be the most appropriate distribution of annual and monthly rainfalls at all the synoptic stations. The Pearson III probability plots of the historical annual rainfall data for Urmia synoptic station are shown as an example in **Fig. 3**.

Based on these tables, it is observed that different months tended to have different "best" distribution on the basis of the maximum correlation coefficient and minimum RMSE of the PPCC test. Therefore, it was important to devise a criterion by which a single probability function could be selected for use across all eleven monthly rainfalls.

The criterion finally used was the highest score out of the total number of occasions in which a given distribution performed better than the others. Also, these tables contain the total score for each of the probability distribution functions; these show that the Pearson III produced the highest score for Urmia station.

Figures 4 and **5** shows the Relationships between L-cv, L-skewness and L-kurtosis for all monthly data in all stations. It is quite apparent that 2-parameter Gamma distribution and Pearson (III) as a 3-parameter distribution provides a much better fit to the observations than the other distributions.

Table 6. Correlation coefficients obtained from the PPCC test for annual rainfall data (maximum PPCC are underlined)

Month	Monthly (Urmia station)				
	Normal	Pearson III	LN(2)	LN(3)	LP III
Oct.	0.8570	0.9584	0.9695	**0.9864**	0.9841
Nov.	0.9421	**0.9950**	0.9288	0.9944	0.9843
Dec.	0.9034	0.9920	0.9915	**0.9961**	0.9951
Jan.	0.9719	**0.9894**	0.8793	0.9893	0.9810
Feb.	0.9827	**0.9936**	0.9721	0.9931	0.9907
Mar.	0.9614	**0.9958**	0.9756	0.9950	0.9951
Apr.	0.9848	0.9913	0.9776	0.9906	**0.9940**
May	0.9425	0.9839	0.9600	**0.9917**	0.9908
June	0.9301	**0.9815**	0.9449	0.9762	0.9773
Jul.	0.7291	**0.9714**	0.9536	0.9500	0.9573
Aug.	0.7165	**0.9724**	0.8622	0.8622	0.8866
Sep.	0.7881	**0.9695**	0.9364	0.9363	0.9394
Score	0	**8**	0	3	1

Table 7. Correlation coefficients obtained from the PPCC test for Monthly rainfall data (maximum PPCC are underlined)

Stations	Annual				
	Normal	Pearson III	LN(2)	LN(3)	LP III
Arak	0.9900	**0.9929**	0.9859	0.9927	0.9922
Urmia	0.9684	0.9894	0.9917	0.9918	**0.9923**
Bandaranzali	0.9798	**0.9956**	0.9940	0.9952	0.9951
Tabriz	0.9721	**0.9969**	0.9955	0.9965	0.9968
Tehran	0.9928	0.9933	0.9842	**0.9938**	0.9940
Khoramabad	0.9955	**0.9960**	0.9850	0.9850	0.9952
Khoy	0.9889	0.9935	0.9927	0.9929	**0.9949**
Zanjan	0.9938	**0.9939**	0.9746	0.9748	0.9906
Sagez	0.9792	0.9872	**0.9892**	0.9874	0.9895
Gazvin	0.9833	**0.9865**	0.9713	0.9851	0.9800
Kermanshah	0.9923	**0.9958**	0.9916	0.9941	0.9958
Score	0	**7**	1	1	2

Based on Correlation coefficients obtained from the PPCC test, according to **Table 8**, it is observed that, in the period of study (1961-2010), no month have a normal distribution, 95 month have a Pearson III distribution, no month have a LN(2) distribution, 29 month have a LN(3) distribution and 20 month have a LP(3) distribution.

Table 8. Number of months that have the highest correlation

Distribution	Normal	Pearson III	LN (2)	LN (3)	LP III
No. of month	0	**95**	0	29	20

Table 9. RMSE obtained from the PPCC test for annual rainfall data (minimum RMSE are underlined)

Stations	Annual				
	Normal	Pearson III	LN(2)	LN(3)	LP III
Arak	13.87	**11.50**	15.74	12.07	11.69
Urmia	24.71	**14.48**	14.72	14.69	14.72
Bandaranzali	66.24	33.63	38.22	35.91	**33.31**
Tabriz	19.84	7.07	7.63	6.98	**6.56**
Tehran	8.52	**8.23**	13.69	8.38	8.40
Khoramabad	11.15	**11.01**	19.43	19.43	11.26
Khoy	12.13	**9.36**	11.33	11.07	9.51
Zanjan	8.71	**8.73**	14.67	8.75	9.46
Sagez	26.58	**20.97**	21.84	21.47	21.30
Gazvin	10.73	**8.32**	11.59	8.50	8.47
Kermanshah	19.48	9.48	9.86	9.55	**9.08**
Score	0	8	0	0	**3**

Table 10. RMSE obtained from the PPCC test for monthly rainfall data (minimum RMSE are underlined)

Month	Monthly (Urmia station)				
	Normal	Pearson III	LN(2)	LN(3)	LP III
Oct.	18.00	**9.71**	12.36	11.94	11.70
Nov.	10.75	**3.78**	6.09	4.33	7.73
Dec.	10.27	**3.22**	3.34	3.25	3.23
Jan.	4.35	2.79	3.82	**2.76**	6.99
Feb.	3.00	1.85	3.15	1.96	**1.65**
Mar.	8.97	3.13	4.83	**3.05**	3.48
Apr.	5.35	**4.08**	7.59	4.13	4.73
May.	12.57	6.71	10.08	8.59	**5.00**
Jun.	4.77	**2.47**	3.93	3.21	6.69
Jul.	8.03	**2.80**	4.23	4.23	18.18
Aug.	4.04	**1.40**	2.09	2.09	19.59
Sep.	4.87	**1.89**	2.77	2.76	17.45
Score	0	**8**	0	2	2

Table 11. Number of months that have the lowest RMSE

Distribution	Normal	Pearson III	LN (2)	LN (3)	LP III
No. of month	0	**117**	7	7	13

Based on RMSE obtained from the PPCC test, (**Table 11**), no month have a normal distribution, 117 month have a Pearson III distribution, 7 month have a LN(2) distribution, 7 month have a LN(3) distribution and 13 month have a LP(III) distribution. Consequently, the Pearson type three distribution (Pearson III) is found to be the most appropriate distribution of annual and monthly rainfalls at all the synoptic stations of study area. Meanwhile, from the viewpoint of regional distribution, we used L-Moment approach to find a best regional distribution of rainfall data of study area. Also **Figs 4** and **5**, show that a 2-parameter Gamma distribution and Pearson III as a 3-parameter distribution provides a much better fit to the observations than the other probability distributions.

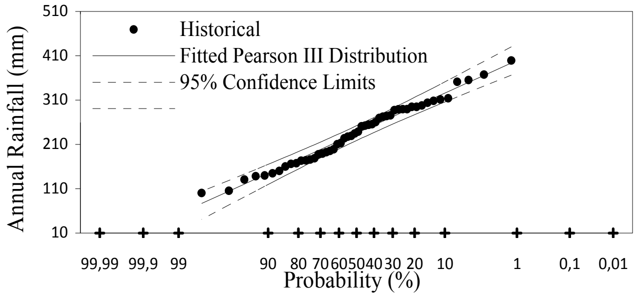

Fig. 3 Pearson III probability plots of annual data for Urmia station with 95% confidence limits.

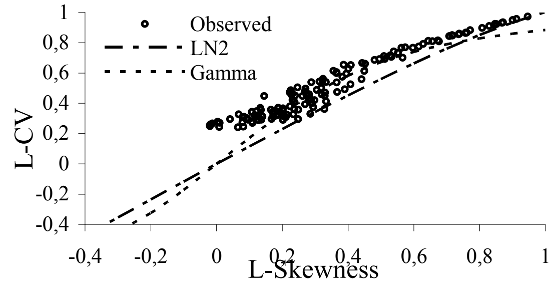

Fig. 4 Relationships between L-Cv, L-Skewness and L-Kurtosis: L-Cv versus L-Skewness.

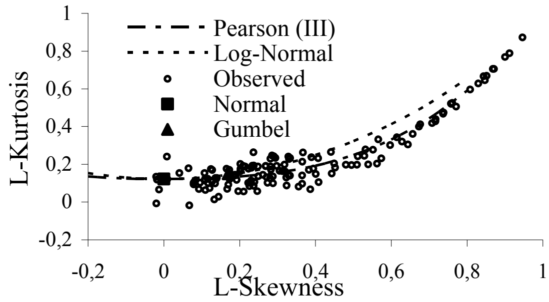

Fig. 5 Relationships between L-Cv, L-Skewness and L-Kurtosis: L-Kurtosis versus L-Skewness.

CONCLUSION

In this paper, PPCC and L-Moment approach were performed to determine the regional best distributions of rainfall data in northwest of Iran. It is observed that, two methods provide Pearson III as a best distribution of monthly rainfall data in our study area. As regards, PPCC test has been known as a powerful single-site test among many goodness of fit tests (e.g., Stedinger *et al.*, 1993; Kottegoda & Rosso, 1997; Beirlant *et al.*, 2005; Heo *et al.*, 2008), but L-Moment approach is easy and can compare the fit of several distributions to many samples of data using a single graphical instrument.

REFERENCES

Adeloye, A.J. & Montaseri, M. (2002) Preliminary streamflow data analyses prior to water resources planning study. *Hydrolog. Sci. j.* **47**(5), 679–692.

Antevs, E. (1954) Climate of New Mexico during the last glacio-pluvial. *J. Geology*. **62**, 182–191.

Beirlant, J., Goegebeur, Y., Segars, J. & Teugels, J., (2005) *Statistics of Extremes*, Theory and Applications. Wiley, Hoboken.

Cunnane, C. (1978) Unbiased plotting positions-a review, *J. Hydrol*. **37**, 205–222.

Filliben, J.J. (1975) The Probability Plot Correlation Coefficient Test for Normality, *Technometrics*. **17**(1), 111–117.

Heo, J.-H., Kho, Y.W., Shin, H., Kim, S. & Kim, T. (2008) Regression equations of probability plot correlation coefficient test statistics from several probability distributions. *J. Hydrol*. **335**(1–4), 1–15.

Hosking, J.R.M. (1990) L-moments: Analysis and estimation of distributions using linear combinations or order statistics, *J. R. Stat. Soc. B Met.* **52**, 105–124.

Hosking, J.R.M. & Wallis, J.R. (1997) *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge Univ. Press, New York.

Kottegoda, N.T. & Rosso, R. (1997) *Statistics, Probability, and Environmental Engineers*. McGraw–Hill, New York.

McGhee, J. W. (1985) *Introductory Statistics*. West Publishing Co., New York, USA.

McMahon, T. A., Peel, M.C., Vogel, R.M. & Pegram, G.G.S. (2007) Global streamflows - Part 3: Country and climate zone characteristics, *J. Hydrol*. **347**, 272–291.

Sooyoung, K., Younwoo, K., Hongjoon, Sh., & Jun-Haeng, H. (2008) Derivation of the Probability Plot Correlation Coefficient Test Statistics for the Generalized Logistic and the Generalized Pareto Distributions, *World Environmental and Water Resources Congress*, Ahupua'a, ASCE.

Stedinger, J.R., Vogel, R.M. & Foufoula-Georgiou, E. (1993) *Frequency analysis of extreme events*. In: Maidment, D.R. (Ed.), Handbook of Hydrology, McGraw–Hill, New York.

Vogel, R. M. (1986) The probability plot correlation coefficient test for the normal, log-normal, and Gumbel distribution hypotheses, *Water Resour. Res*. **22**(4), 587–590.

Vogel, R. M. & McMartin, D. E. (1991) Probability plot goodness-of-fit and skewness estimation procedures for the Pearson Type III distribution. *Water Resour. Res*. **27**(12), 3149–3158.

Vogel, R.M. & Fennessey, N.M. (1993) L-moment diagrams should replace product-moment diagrams, *Water Resour. Res*. **29**, 1745–1752.

Vogel, R.M. & Wilson, I. (1996) Probability distribution of annual maxi-mum, mean and minimum streamflows in the United States, *J. Hydrol. Eng*, **1**(2), 69–76.