Theor Appl Climatol (2013) 113:83–94
DOI 10.1007/s00704-012-0763-2

ORIGINAL PAPER

Regional frequency analysis of extreme rainfalls using partial L moments method

Zahrahtul Amani Zakaria · Ani Shabri

Received: 12 April 2012 / Accepted: 31 August 2012 / Published online: 27 September 2012

Abstract An approach based on regional frequency analysis using L moments and LH moments are revisited in this study. Subsequently, an alternative regional frequency analysis using the partial L moments (PL moments) method is employed, and a new relationship for homogeneity analysis is developed. The results were then compared with those obtained using the method of L moments and LH moments of order two. The Selangor catchment, consisting of 37 sites and located on the west coast of Peninsular Malaysia, is chosen as a case study. PL moments for the generalized extreme value (GEV), generalized logistic (GLO), and generalized Pareto distributions were derived and used to develop the regional frequency analysis procedure. PL moment ratio diagram and Z test were employed in determining the best-fit distribution. Comparison between the three approaches showed that GLO and GEV distributions were identified as the suitable distributions for representing the statistical properties of extreme rainfall in Selangor. Monte Carlo simulation used for performance evaluation shows that the method of PL moments would outperform L and LH moments methods for estimation of large return period events.

Z. A. Zakaria (✉)
Faculty of Informatics, Universiti Sultan Zainal Abidin,
PO box 21300, Kuala Terengganu, Terengganu, Malaysia
e-mail: amanizakaria@gmail.com

A. Shabri
Department of Mathematics, Faculty of Science,
Universiti Teknologi Malaysia,
PO box 81310, Skudai,
Johor Bahru, Johor, Malaysia
e-mail: ani@utm.my

1 Introduction

Information regarding accurate estimation of extreme events such as flood magnitudes and their frequency of occurrence are of great importance in the planning, design, and management of hydraulic structures such as dams, spillways, culverts, and storm water management systems. A novel approach to the prediction of flood flows and also applicable to other hydrologic processes such as rainfall is the statistical method of regional frequency analysis. This approach promises a more reliable analysis by using information from several sites with identical behavior of flood, rather than only single-site information. With these, regional frequency analysis becomes a 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 economics Jingyi and Hall (2004).

The major developments in flood frequency analysis revolved the idea of probability weighted moments (PWM) introduced by Greenwood et al. (1979) and the theory of L moments proposed by Hosking (1990). The approach of L moments in regional frequency analysis has been applied successfully to model floods by a number of cases studied in Malaysia (Lim and Lye 1998; Zin et al. 2009), New Zealand (Pearson 1991), Southern Africa (Kjeldsen et al. 2002), Egypt (Atiem and Harmancioglu 2006), Turkey (Saf 2009), Iran (Rahnama and Rostami 2007), China (Chen et al. 2006), Italy (Noto and Logggia 2009; Cannarozzo et al. 2009), Pakistan (Hussain and Pasha 2009), Tunisia (Abida and Ellouze 2008), Canada (Glaves and Waylen 1997; Yue and Wang 2004), the UK (Fowler and Kilsby 2003), and India (Parida et al. 1998; Kumar et al. 1999, 2003).

Introducing partial probability weighted moments (PPWM), Wang (1990a) extended the definition of PWM for fitting distribution functions to censored samples. Partial L moments (PL moments) are variants of L moments and also analogous to the PPWM. In the case of flood estimation, the interest is focused mainly on the estimation of the right-hand tail of a distribution function. Because the concern of data on small flood events can sometimes be only of little relevance to the larger ones, the PL moments method are introduced for characterizing the larger events in data. Using PL moments may reduce undesirable influences that small sample events may have on the estimation of large return period events.

The PPWM and PL moments approach related to censor data sets has been employed in a number of studies. Previous researches have been done by Wang (1990a, b; 1996) and Bhattarai (2004) which utilized PL moments in fitting generalized extreme value (GEV) distribution to the censored flood samples. Bhattarai (2004) explored different censoring levels of PL moments and found that sampling properties of PL moments, with censoring flood samples up to 30 %, are similar those of simple L moments. Some other researches can be found in Kroll and Stedinger (1996), Koulouris et al. (1998), and Moisello (2007) and recently by Kochanek et al. (2008). However, literature review reveals limited usage of the proposed PL moments in regional frequency analysis.

Wang (1997) introduced another method which is also a generalization of the L moments, called LH moments for GEV. Since then, LH moments have been used by several authors in flood frequency analysis. Lee and Maeng (2003) analyzed design floods derived through LH moments using annual maximum floods in Korea watersheds. Meshgi and Khalili (2009a, b) developed regional flood frequency analysis based on LH moments of the Kharkhe watershed, located in Western Iran. They used GEV, generalized logistic (GLO), and generalized Pareto (GPA) distributions, and a comparative study had been made between LH moments and L moments method. Deka et al. (2011) studied the statistical modeling of annual maximum daily rainfall data in Northeast India fitted using LH moments. Bhuyan et al. (2010) used LH moments for regional flood frequency analysis of the north bank region of the river Brahmaputra, India.

In this study, the regional frequency analysis of the PL moments approach is developed, by first revisiting regional frequency analysis establishment based on the L moments by Hosking and Wallis (1997) and LH moments by Meshgi and Khalili (2009a, b). For this purpose, the previous developed relationship between PL moments and GEV distribution by Wang (1996) are revisited. Next, a new relationship for GLO and GPA distributions are developed. A total number of 37 stations within Selangor Malaysia are used for PL moments regional frequency analysis, and a comparative study has been made between L moments and LH moments of order two.

2 Methodologies

2.1 Method of L moments

The L moments, introduced by Hosking (1990), are another way of summarizing the statistical properties of hydrological data. L moments can be expressed as linear combinations of PWM. The PWM of order r was formally defined by Greenwood et al. (1979) as

$$\beta_r = \int_0^1 x(F)F^r\,dF \tag{1}$$

where $F=F(x)$ is a cumulative distribution function, $x(F)$ is an inverse distribution function or so-called quantile function of random variables x, and $r=0, 1, 2,\ldots$ is a nonnegative integer. The first four L moments, expressed as linear combinations of PWM, are

$$\begin{aligned}
\lambda_1 &= \beta_0\\
\lambda_2 &= 2\beta_1 - \beta_0\\
\lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0\\
\lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0
\end{aligned} \tag{2}$$

The L moments ratios (L coefficient variation, L skewness, L kurtosis, respectively) are defined as

$$\begin{aligned}
L-Cv &= \tau_2 = \tfrac{\lambda_2}{\lambda_1}\\
L-Cs &= \tau_3 = \tfrac{\lambda_3}{\lambda_2}\\
L-Ck &= \tau_4 = \tfrac{\lambda_4}{\lambda_2}
\end{aligned} \tag{3}$$

2.2 Method of LH moments

Wang (1997) introduced the concept of LH moments as a generalization of the L moments. The LH moments are the linear function of the expectation of the highest order statistic. The first four LH moments are given as

$$\begin{aligned}
\lambda_1^\eta &= E\left[X_{(\eta+1):(\eta+1)}\right]\\
\lambda_2^\eta &= \tfrac{1}{2}E\left[X_{(\eta+2):(\eta+2)} - X_{(\eta+1):(\eta+2)}\right]\\
\lambda_3^\eta &= \tfrac{1}{3}E\left[X_{(\eta+3):(\eta+3)} - 2X_{(\eta+2):(\eta+3)} + X_{(\eta+1):(\eta+3)}\right]\\
\lambda_4^\eta &= \tfrac{1}{4}E\left[X_{(\eta+4):(\eta+4)} - 3X_{(\eta+3):(\eta+4)} + 3X_{(\eta+2):(\eta+4)} - X_{(\eta+1):(\eta+4)}\right]
\end{aligned} \tag{4}$$

For $\eta=0$, LH moments will be Hosking (1990) L moments. The LH moments ratios (LH-Cv, LH-Cs, and

LH-Ck, respectively) are defined as

$$\tau_2^{\eta} = \frac{\lambda_2^{\eta}}{\lambda_1^{\eta}},\ \tau_3^{\eta} = \frac{\lambda_3^{\eta}}{\lambda_2^{\eta}}\ and\ \tau_4^{\eta} = \frac{\lambda_4^{\eta}}{\lambda_2^{\eta}} \tag{5}$$

The details on the estimation of parameters and regional flood frequency analysis can be found in Bhuyan et al. (2010), Meshgi and Khalili (2009b), and Deka et al. (2011).

2.3 Method of PL moments

There have been numbers of research discussing the definition of partial PWM by Wang (1990a, b; 1996), Hosking (1995), and Koulouris et al. (1998). In the present study, the definition of partial PWM by Wang (1996) is employed. Wang (1996) defined partial PWM as extended from the concept of PWM to be applied to a censored sample

$$\beta_r^{'} = \frac{1}{1-F_0^{r+1}} \int_{F_0}^{1} x(F)F^r dF \tag{6}$$

where $F_0=F(x_0)$, x_0 being the censoring threshold. When $F_0=0$, the partial PWM becomes the ordinary PWM. Given a complete sample $x_{(1)}\leq x_{(2)}\leq \ldots \leq x_{(n)}$, the following statistic is defined by Wang (1990a) as an unbiased estimator of $\beta_r^{'}$

$$b_r^{'} = \frac{1}{\left(1-F_0^{r+1}\right)n} \sum_{i=1}^{n} \frac{(i-1)(i-2)\ldots(i-r)}{(n-1)(n-2)\ldots(n-r)} x_{(i)}^{*} \tag{7}$$

where

$$x_{(i)}^{*} = 0 \text{ for } x_{(i)} \leq x_0$$
$$x_{(i)}^{*} = x_{(i)} \text{ for } x_{(i)} > x_0$$

The level of censoring, F_0, determines the number of the sample data points to be censored as

$$F_0 = \frac{n_0}{n} \tag{8}$$

where n is the length of the uncensored sample and n_0 is the number of occurrence of values which do not exceed x_0 in the sample (censored data points). The first four PL moments (ξ_1, ξ_2, ξ_3, and ξ_4) have the same definition and interpretations as the first four L moments.

The PL-Cv, PL-Cs, and PL-Ck are defined as

$$\varsigma_2 = \frac{\xi_2}{\xi_1}, \varsigma_3 = \frac{\xi_3}{\xi_2}\ and\ \varsigma_3 = \frac{\xi_4}{\xi_2} \tag{9}$$

2.4 Development of relationships between the PL moments and probability distribution

Many statistical distributions for regional frequency analysis have been investigated for extreme hydrologic variables. In this study, three probability distributions were considered: GEV, GLO, and GPA. The short-listed distributions were chosen based on previous studies such as those by Zin et al. (2009) and Zalina et al. (2002) of which these distributions were more prominent for tropical regions and by Kysely (2010) for modeling precipitation extremes. The details on the estimation of parameters of these distributions can be found by Hosking and Wallis (1993, 1997) for the L moments method and Bhuyan et al. (2010), Meshgi and Khalili (2009a, b), and Deka et al. (2011) for the LH moments method.

For the case of PL moments with the definition on Eq. (6), only the GEV distribution has been developed by Wang (1996). However, the development of relationships between PL moments and other distributions has not yet been available. In this section, the PL moments of the GEV distribution are revisited. Next, the PL moments for the GLO and GPA distributions are developed in this study. Upon the issues of censoring would improve the estimation of high return period, this study emphasizes on the censoring at 3 % of the complete data. This level of censoring would contribute to censoring level at F_0=0.03.

2.5 PL moments for GEV distribution

The cumulative distribution function of GEV is given by

$$F(x) = \exp\left\{-\left[1-k\left(\frac{x-b}{a}\right)\right]^{\frac{1}{k}}\right\} \quad k \neq 0 \tag{10}$$

and quantile function

$$Q(F) = b + \frac{a}{k}\left\{1-[-\ln(F)]^k\right\} \quad k \neq 0 \tag{11}$$

The expression of the first two PL moments and PL skewness of GEV are

$$\xi_1 = b + aH(0, F_0, k) \tag{12}$$

$$\xi_2 = a[H(1, F_0, k) - H(0, F_0, k)] \tag{13}$$

$$\varsigma_3 = \frac{2H(2, F_0, k) - 3H(1, F_0, k) + H(0, F_0, k)}{H(1, F_0, k) - H(0, F_0, k)} \tag{14}$$

where

$$H(r, F_0, k) = \frac{1}{k}\left\{1 - \frac{\mathrm{P}[1+k, -(r+1)\ln F_0]}{\left(1-F_0^{r+1}\right)(r+1)^k}\right\} \tag{15}$$

In Eq. (15), P(.,.)is an incomplete gamma function

$$P(1+k,-(r+1)\ln F_0)=\int_0^{-(r+1)\ln F_0}\theta^k e^{-\theta}d\theta \tag{16}$$

In practice, solving of the k parameter requires development of approximate methods based on Eq. (14). Equation (14) does not give an explicit solution for k and has to be solved numerically, solving using an iterative method. For this purpose, the polynomial function of computation of the following equation with good accuracy has been constructed based on ς_3 *as*

$$k=0.54-2.292\varsigma_3+1.773\varsigma_3^2-1.894\varsigma_3^3+0.954\varsigma_3^4 \tag{17}$$

Once the value of k is obtained, a and b can be estimated successively from Eqs. (12)–(13) as

$$a=\frac{\xi_2}{H(1,F_0,k)-H(0,F_0,k)} \tag{18}$$

$$b=\xi_1-aH(0,F_0,k) \tag{19}$$

2.6 PL moments for GLO distribution

The cumulative distribution function and quantile function of GLO are

$$F(x)=\left[1+\left[1+\left\{1-\frac{k}{a}(x-b)\right\}^{\frac{1}{k}}\right]\right]^{-1} \tag{20}$$

$$Q(F)=b+\frac{a}{k}\left\{1-\left(\frac{1-F}{F}\right)^k\right\} \tag{21}$$

The expression of the first two PL moments and PL skewness of GLO are

$$\xi_1=b+\frac{a}{k}\left[1-\frac{B_{1-F_0}(1+k,1-k)}{1-F_0}\right] \tag{22}$$

$$\xi_2=-\frac{a}{k}\times\left[\frac{2B_{1-F_0}(1+k,2-k)}{1-F_0^2}-\frac{B_{1-F_0}(1+k,1-k)}{1-F_0}\right] \tag{23}$$

$$\varsigma_3=\frac{\frac{6B_{1-F_0}(1+k,3-k)}{1-F_0^3}-\frac{6B_{1-F_0}(1+k,2-k)}{1-F_0^2}+\frac{B_{1-F_0}(1+k,1-k)}{1-F_0}}{\frac{2B_{1-F_0}(1+k,2-k)}{1-F_0^2}-\frac{B_{1-F_0}(1+k,1-k)}{1-F_0}} \tag{24}$$

where $B_{1-F_0}(.,.)$is an incomplete beta function

$$B_{1-F_0}(1+k,r-k+1)=\int_0^{1-F_0}\theta^k(1-\theta)^{r-k}d\theta \tag{25}$$

The parameter k of GLO can be computed using numerical solving of Eq. (24) in interval [−1, 1]. The estimate of parameter k is given by

$$k=0.184-1.333\varsigma_3+0.514\varsigma_3^2-0.575\varsigma_3^3+0.274\varsigma_3^4 \tag{26}$$

Once the value of k is obtained, a and b can be estimated successively and are then given by

$$a=\frac{-\xi_2 k}{\frac{2B_{1-F_0}(1+k,2-k)}{1-F_0^2}-\frac{B_{1-F_0}(1+k,1-k)}{1-F_0}} \tag{27}$$

$$b=\xi_1-\frac{a}{k}(1-\frac{B_{1-F_0}(1+k,1-k)}{1-F_0}) \tag{28}$$

2.7 PL moments for GPA distribution

The cumulative distribution function and quantile function of GPA are

$$F(x)=1-\left[1-\frac{k}{\alpha}(x-\xi)\right]^{\frac{1}{k}} \tag{29}$$

$$Q(F)=\xi+\frac{\alpha}{k}\left[1-(1-F)^k\right] \tag{30}$$

The expression of the first two PL moments and PL skewness of GPA are

$$\xi_1=\xi+\frac{\alpha}{k}\left(1-g_{1,1}\right) \tag{31}$$

$$\xi_2=-\frac{\alpha}{k}\left(2g_{1,2}-2g_{2,2}-g_{1,1}\right) \tag{32}$$

$$\zeta_3=\frac{6g_{1,3}-12g_{2,3}+6g_{3,3}-6g_{1,2}+6g_{2,2}+g_{1,1}}{2g_{1,2}-2g_{2,2}-g_{1,1}} \tag{33}$$

where $g_{r,s}=\frac{(1-F_0)^{k+r}}{(k+r)[1-(F_0)^s]}$

The parameter k of GPA can be computed using numerical solving of Eq. (33) in interval [−1, 1]. The estimates of parameter k is given by

$$k=-0.2705+0.1009\varsigma_3+1.1075\varsigma_3^2-0.2858\varsigma_3^3+0.0576\varsigma_3^4 \tag{34}$$

Once the value of k is obtained, a and b can be estimated successively and are then given by

$$\alpha = \frac{-\lambda_2 k}{2g_{1,2} - 2g_{2,2} - g_{1,1}} \quad (35)$$

$$\xi = \lambda_1 - \frac{\alpha}{k}\left(1 - g_{1,1}\right) \quad (36)$$

3 Regional frequency analysis based on L moments

Hosking and Wallis (1993, 1997) provided step-by-step guidelines for performing regional frequency analysis, using the L moments. The four steps involved in the regional frequency analysis are outlined as follow: (a) screening of the data using discordancy test, (b) identification of homogeneous regions, (c) choice of a regional distribution, and (d) estimation of the regional frequency distribution. A discussion of the first three steps is given next. The same procedure has been applied for LH moments (Bhuyan et al. 2010; Mesgi and Khalili 2009; and Deka et al. 2011).

4 Regional frequency analysis based on PL moments

The procedures discussed in Section 4 are similarly employed for the PL moments. PL-Cv, PL-Cs, and PL-Ck are equally replaced by L-Cv, L-Cs, and L-Ck for the discordancy and the homogeneity test. Selection of an adequately fitted distribution is carried out based on the PL ratio diagram and Z test using the regional PL-Cs and PL-Ck.

4.1 Discordance test

The main goal of the discordancy measure D test is to identify those sites for which point sample PL moments are markedly different from most of the other sites. Sites with great errors in data will stand out from the other sites and be flagged as discordant. The discordancy test, D_i, for site i is defined by Hosking and Wallis (1997) as

$$D_i = \frac{1}{3} N (u_i - \overline{u})^T S^{-1} (u_i - \overline{u}) \quad (37)$$

where $u_i = [\varsigma_2^i \ \varsigma_3^i \ \varsigma_4^i]^T$ is a vector containing the three sample PL moment ratios for site i, N is the number of sites in the region, and $\overline{u}$ represents the unweighted regional average of L moments ratio for each region

$$\overline{u} = \frac{1}{N} \sum_{i=1}^{N} u_i \quad (38)$$

and **S** is the sample covariance matrix expressed by

$$S = \sum_{i=1}^{N} (u_i - \overline{u})(u_i - \overline{u})^T \quad (39)$$

Generally, a site is declared as discordant from the group if the D_i value is greater than a critical value. Hosking and Wallis (1997) tentatively suggested $D_i \geq 3$ as the critical value for $N \geq 15$ sites. If the D statistic of a site exceeds 3, its data are considered to be discordant from the rest of the regional data.

4.2 Heterogeneity test

The next step in regional frequency analysis is the assignment of the sites to regions. Hosking and Wallis (1997) proposed a heterogeneity measure H_i that aims to estimate the degree of heterogeneity in a group of sites and to assess whether the sites might reasonably be treated as a homogeneous region. The heterogeneity test is then computed as

$$H = (V - \mu_V)/\sigma_V \quad (40)$$

where μ_V and σ_V represent the population mean and standard deviation of the simulated V value.

$$V = \sqrt{\sum_{i=1}^{N} n_i \left(\varsigma_2^{(i)} - \varsigma_2^R\right)^2 \Big/ \sum_{i=1}^{N} n_i} \quad (41)$$

Here, ς_2^R is the regional average PL moments ratio, calculated using the following formula

$$\varsigma_2^R = \sum_{i=1}^{N} n_i \varsigma_2^{(i)} \Big/ \sum_{i=1}^{N} n_i \quad (42)$$

where N is the number of sites and n_i is the record length at sites i. The four-parameter kappa distribution is used to generate a homogeneous region with population parameters equal to the regional average sample L moments ratios.

The criteria established by Hosking and Wallis (1997) for assessing heterogeneity of a region are

$H<1$—the region is acceptably homogeneous
$1 \leq H<2$—the region is possibly homogeneous
$H \geq 2$—the region is definitely heterogeneous

4.3 Selection of a regional frequency distribution

Hosking and Wallis (1997) suggested two approaches in testing whether the given distributions fit the data acceptably closely and hence choose the one that gives best fit to the data. The PL moments ratio diagram and Z test are employed for these purposes. The PL moments ratio diagram is a plot of PL-Cs and PL-Ck of the observed values and the calculated values from

Table 1 Equation coefficient of PL moment ratio for the GEV, GLO, and GPA distributions

Corresponding equation: $\tau_4^{DIS} = a_0 + a_1\tau_3 + a_2(\tau_3)^2 + a_3(\tau_3)^3 + a_4(\tau_3)^4$

Distribution	a_0	a_1	a_2	a_3	a_4
GEV	−0.0166	0.1578	0.8357	−0.1887	0.1453
GLO	0.0386	0.0925	0.6583	−0.0447	0.3587
GPA	−0.0993	0.1703	1.0107	−0.2091	0.0630

the distribution functions. Table 1 shows the coefficients for the newly developed relationships of PL-Cs and PL-Ck of the GEV, GLO, and GPA distributions based on PL moments for the range $-1 \leq {}_3 \leq 1$.

However, direct visual inspection of the PL moments ratio diagram is somewhat subjective. Hosking and Wallis (1997) preferred an alternative approach based on goodness-of-fit test, Z test, which works directly with the regional average of L moments statistics.

Table 2 Station name and statistics of maximum daily rainfalls for all the stations in Selangor

No.	Name of station	Station number	n	Mean (mm)	Stdv. (mm)	Kurtosis	Skewness	LM D_i	LHM D_i	PLM D_i
1	Ldg. Batu Untong	2615131	37	132.76	35.64	−0.54	0.28	0.61	0.41	0.63
2	Ldg. Telok Merbau	2616135	37	105.90	34.98	2.14	1.03	0.33	0.07	0.32
3	Ldg. Sepang	2617134	35	103.86	31.95	−0.07	0.88	1.08	0.30	0.59
4	Ldg. Bute	2717114	37	95.91	26.61	−0.23	0.28	0.76	0.28	0.71
5	P.Kwln P.ST. Gong	2913001	33	119.79	76.60	21.33	4.25	1.63	0.94	1.79
6	Ldg. West	2913121	37	108.33	39.88	1.29	0.82	0.51	0.01	0.03
7	Jps. Pulau Lumut	2913122	37	98.03	34.09	0.90	0.88	0.21	0.02	0.22
8	Pejabat Jps. Klang	3014084	36	86.81	26.33	2.31	1.51	1.37	0.02	1.78
9	Ldg. Sg. Kapar	3113087	37	105.11	30.08	−0.32	0.50	0.28	0.34	0.29
10	Ldg. Elmina	3115053	38	108.72	56.63	6.88	2.28	0.99	0.11	1.06
11	Sg. Buloh	3115079	38	94.10	29.12	0.18	0.41	0.39	0.28	0.63
12	Ldg. Edinburgh Stn 2	3116006	31	95.33	22.59	−0.66	0.43	0.87	0.53	*4.22*
13	Pemasokan Ampang	3118069	22	103.01	35.89	0.99	0.31	*2.13*	0.20	1.02
14	Sek.Keb.Kg.Lui	3118102	37	114.65	61.21	2.43	1.60	1.73	0.06	1.84
15	Ldg. Braunston	3213057	34	91.83	33.30	0.03	0.77	0.92	0.16	0.78
16	Ldg. Bkt. Ijok	3214055	35	106.97	46.24	1.70	1.49	0.63	0.03	0.66
17	Kg. Sg. Tua	3216001	36	98.23	29.78	0.81	1.26	1.04	0.09	0.81
18	Ibu Bekalan Km. 16	3217001	36	97.26	21.49	1.34	0.40	*3.12*	0.05	2.16
19	Empangan G. Klang	3217002	36	100.31	38.42	11.27	2.72	0.42	0.37	0.49
20	Stn. Jenaletrik Lln.	3218101	37	108.26	56.68	2.92	1.64	1.52	0.05	1.38
21	Ldg. Bkt. Belimbing	3312042	36	97.83	40.17	3.72	1.92	0.77	0.08	0.85
22	Jln. Kelang	3312045	37	96.56	32.38	7.14	2.20	1.06	0.15	*2.37*
23	Ldg. Bkt. Talang	3313040	35	98.65	48.73	6.76	2.43	0.60	0.26	0.62
24	Ldg. Kuala Selangor	3313043	37	100.60	40.46	0.56	0.84	0.58	0.05	0.28
25	Ldg. Sg. Buloh	3313060	38	93.24	31.41	1.44	0.99	0.27	0.01	0.32
26	Rmh Pam Jaya Setia	3314001	36	108.40	75.24	21.78	4.26	*3.50*	0.87	*3.10*
27	Ldg. Sg. Gapi	3316028	35	113.89	30.99	1.49	1.28	1.14	0.03	0.95
28	Parit 1 Sg. Burong	3411016	36	102.84	31.91	0.56	0.44	0.69	0.13	0.52
29	Bekalan Sg. Tengki	3412001	33	91.27	29.35	0.46	0.96	0.65	0.10	0.24
30	Ldg. Raja Musa	3412041	37	93.05	40.91	4.91	2.02	0.38	0.08	1.08
31	Ldg. Hopeful	3414030	35	109.25	36.91	0.15	0.94	1.11	0.20	0.77
32	Fdc. Sekichan	3510001	33	87.99	30.24	2.25	1.35	0.20	0.00	0.04
33	Parit 1 Sg. Besar	3609012	36	93.63	23.95	−0.11	0.56	0.61	0.51	0.82
34	Sg. Nipah	3610014	33	83.91	33.31	0.05	−0.23	*3.75*	0.56	*2.30*
35	Rumah Pam Jps Terap	3710006	37	86.53	22.68	0.04	0.36	0.51	0.27	0.72
36	Parit Sg. Air Tawar	3808001	33	89.92	39.67	2.67	1.54	0.43	0.01	0.46
37	Ldg Sg. Bernam	3809009	37	91.65	30.06	1.47	1.17	0.21	0.02	0.15

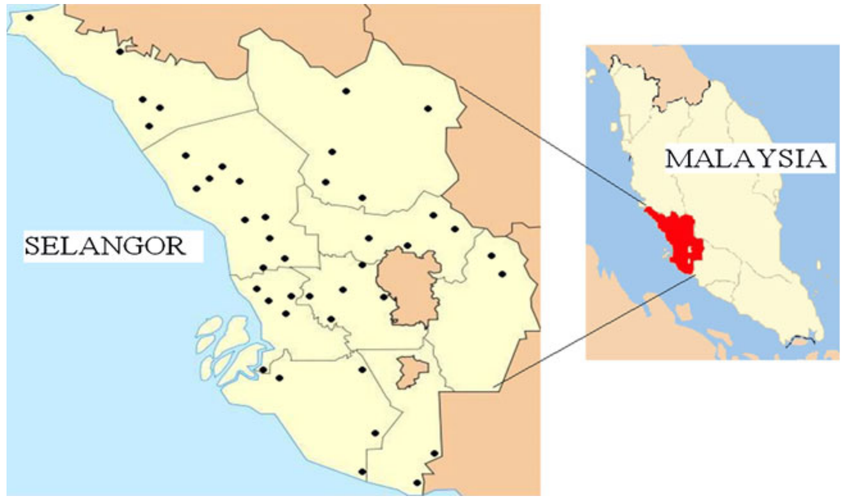

Fig. 1 Location of rain gauge sites used in the study

For each selected distribution, the Z test is calculated as follows:

$$Z^{DIS} = \left(\varsigma_4^{Dis} - \varsigma_4^{R}\right)/\sigma_4 \tag{43}$$

where ς_4^R and σ_4 are the simulated regional mean and standard deviation values obtained by kappa distribution, and ς_4^{Dis} is the regional value of the distribution function in interest.

Details computation is provided in Hosking and Wallis (1993, 1997). A calculated value of zero for $|Z^{\mathrm{DIS}}|$ indicates a perfect fit. The value of the Z statistic is considered to be acceptable at 90 % confidence level if $|Z^{\mathrm{DIS}}| \leq 1.64$. If more than one candidate distribution is acceptable, the one with lowest $|Z^{\mathrm{DIS}}|$ is regarded as the best-fit distribution.

5 Case study

Records of daily rainfalls from 37 stations in Selangor with record lengths of 22 to 38 years were acquired from the Department of Irrigation and Drainage, Malaysia. The statistics and basic information of the data are listed in Table 2. All the stations, numbered 1 to 37, are located in Selangor with latitudes ranging from 26° up to 38° and longitudes from 8° to 18°, as shown in Fig. 1.

As noted in Table 2, the means for the maximum daily rainfalls for the 37 sites in Selangor range from 83.91 mm (site 3610014) to 132.76 mm (site 2615131). Meanwhile, their standard deviations are from 21.49 mm (site 3217001) to 76.60 mm (site 2913001).

6 Results and discussions

This study would emphasize the regional frequency analysis of L moments, LH moments of order two, and PL moments at censoring level, F_0=0.03. When F_0=0, the PL moments become L moments as there are no data being censored.

Initially, the whole of Selangor was assumed as one homogeneous region, and the discordancy test was used for data verification and quality control. Results of the discordancy test, D_i, are given in Table 2. It is observed that for the L moments method, D_{critical}=3.0 is exceeded at three locations: stations 18, 26, and 34, with D statistic values of 3.12, 3.50, and 3.75, respectively. After the second round of discordancy test, station 13 is discarded for having a D statistic value greater than 3. Therefore, these four stations are excluded from the regional frequency analysis. In order to better illustrate the discarded stations, the D values in Table 2 were marked in italic. The values of heterogeneity

Table 3 Moment ratios and parameter values of the fitted kappa distribution

	Moment ratios			Parameters of the kappa distribution			
Method	t_2^R	t_3^R	t_4^R	b	a	h	k
L moments	0.1910	0.2280	0.1889	0.9710	0.2270	−2.3841	−0.2622
LH moments	0.1350	0.3165	0.1694	1.1690	0.1285	−6.9671	−0.2601
PL moments	0.1716	0.3272	0.1419	0.8580	0.1160	−2.0802	−0.2605

Table 4 Results of the homogeneity test and goodness-of-fit test

	Heterogeneity test	*Z* test		
Method	*H*	GEV	GLO	GPA
L moments	−1.1129	*−0.8349*	1.0029	−2.4475
LH moments	−0.0618	*0.3272*	0.9659	−1.6921
PL moments	0.6493	−1.4700	*0.0601*	−1.9431

measures computed by carrying out the 500 simulations based on the data of 33 stations are H=−1.1129.

For the LH moments method, the D statistic values for the 37 stations vary from 0.00 to 0.94. The largest D statistic value is 0.94 for station 5; hence, none of the stations have a D statistic exceeding the critical value. The heterogeneity measure H computed for all stations in this region was−0.0618, which suggested that the region was homogeneous.

For the PL moments method, it is observed that the D statistic values exceed at two stations of 12 and 26 with D statistic values of 4.22 and 3.10, respectively. After having several similar discordancy tests, another two stations are eliminated: stations 22 and 34. Thus, the value of heterogeneity measure based on the data of 33 stations is H=0.6493 which demonstrates acceptable homogeneity.

The regional average L moment, LH moment, and PL moment ratios of the respective study regions are calculated, and the corresponding parameter values of the fitted kappa distribution are found as presented in Table 3. Results of the H tests for the L moments, LH moments, and PL moments are given in Table 4.

After confirming the homogeneity of the study region, an appropriate distribution needs to be selected for the regional frequency analysis. Diagrams in Fig. 2 show a comparison of the observed and theoretical relations between the L moments, LH moments, and PL moments, respectively. In the L moment and LH moment ratio diagrams of Fig. 2, the point defined by the sample average values lies closest to the L moments and LH moments of the GEV distribution followed by GLO and GPA distributions. Analysis of the PL moment ratio diagram reveal that the sample average values of $t_3^R = 0.3272$ and $t_4^R = 0.1419$ in the diagram are better described by the GLO distribution rather than the GEV and GPA distributions.

Results of the Z test for the three distributions are given in Table 4. It has been observed that the values of Z test for both GEV and GLO distributions for all methods are less than the critical value of 1.64. On the other hand, GPA distribution failed the test with the Z test value exceeding the critical value of 1.64 for all methods. In general, the GEV and GLO distributions should be considered as the preferred distribution, as both distributions exhibit acceptable Z test values for the L moments, LH moments, and PL

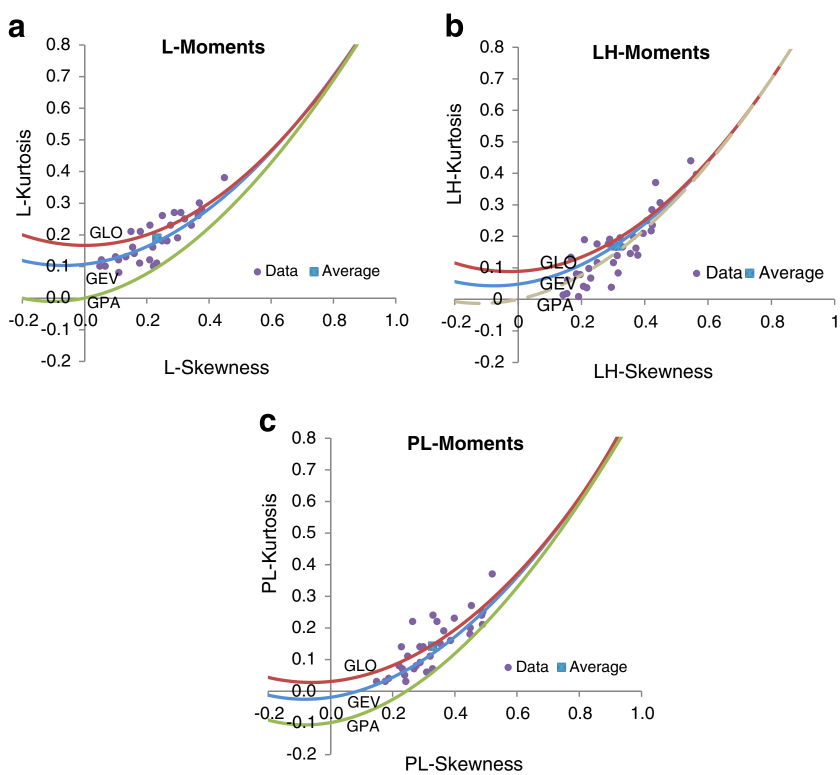

Fig. 2 **a** L moments, **b** LH moments, and **c** PL moments ratio diagram

Table 5 Regional parameter and quantile estimates of the GEV and GLO distributions for L moments, LH moments, and PL moments

Method		Parameters			Quantile estimates				
		b	*a*	*k*	Q_{10}	Q_{20}	Q_{50}	Q_{100}	Q_{200}
L moments	GEV	0.828	0.249	−0.083	1.446	1.669	1.978	2.226	2.488
	GLO	0.927	0.173	−0.228	1.419	1.651	2.009	2.329	2.702
LH moments	GEV	0.813	0.265	−0.088	0.911	1.237	1.471	1.711	2.045
	GLO	0.795	0.052	−0.627	1.041	1.237	1.664	2.191	3.004
PL moments	GEV	0.251	−0.059	0.838	1.442	1.653	1.939	2.164	2.397
	GLO	0.174	−0.204	0.935	1.418	1.638	1.970	2.261	2.595

moments methods. The regional parameters and the quantiles estimated for both selected distributions for T=2, 5, 10, 20, 50, and 100 years, using L moments, LH moments, and PL moments, are presented in Table 5.

6.1 Test for the robustness of the distribution

In regional frequency analysis, the final and imperative objective is to verify the robustness of the distribution in producing reasonably reliable estimates at all stations in the homogeneous region. The robustness of the selected regional frequency distribution is further investigated for estimation of proposed flood quantiles.

In this study, the accuracy of the estimates for the selected region is assessed using a Monte Carlo simulation procedure. In this simulation, flood quantile estimates for preferred distributions, in this case GEV and GLO distributions. In each simulation, 10,000 samples were generated from regional distributions for sample size n=20, 40, 60, and 100. Two of the common error measures of performance used in such cases

Table 6 RBIAS values for different quantiles of the best distribution for L moments, LH moments, and PL moments

Sample size (n)	Method	Distribution	Q_{10}	Q_{20}	Q_{50}	Q_{100}	Q_{200}
20	L moments	GEV	−0.0090	*−0.0090*	*−0.0010*	0.0110	0.0300
		GLO	−0.0110	−0.0110	−0.0060	*0.0050*	0.2300
	LH moments	GEV	−0.0098	−0.0179	−0.0203	−0.0127	*0.0057*
		GLO	−0.1609	−0.2904	−0.5078	−0.7202	−0.9971
	PL moments	GEV	0.0110	−0.0133	−0.0473	−0.0719	−0.0942
		GLO	*0.0080*	−0.0100	−0.0370	−0.0570	−0.0750
40	L moments	GEV	−0.0040	*−0.0030*	*0.0010*	0.0080	0.0180
		GLO	−0.0060	−0.0060	−0.0030	0.0030	0.0130
	LH moments	GEV	*−0.0029*	−0.0063	−0.0061	*−0.0009*	*0.0100*
		GLO	−0.1727	−0.2967	−0.4946	−0.6774	−0.9052
	PL moments	GEV	−0.0058	−0.0060	−0.0022	0.0044	0.0145
		GLO	−0.0060	−0.0060	−0.0030	0.0040	0.0110
60	L moments	GEV	−0.0020	*−0.0010*	*0.0020*	0.0060	0.0130
		GLO	−0.0030	−0.0030	0.0030	0.0040	0.0110
	LH moments	GEV	−0.0035	−0.0058	−0.0057	*−0.0019*	*0.0060*
		GLO	−0.1842	−0.3091	−0.5049	−0.6816	−0.8968
	PL moments	GEV	*0.0018*	−0.0047	−0.0138	−0.0201	−0.0255
		GLO	0.0020	−0.0020	−0.0060	−0.0090	−0.0100
100	L moments	GEV	−0.0020	*−0.0010*	*0.0010*	0.0040	0.0080
		GLO	−0.0030	−0.0030	−0.0020	*0.0010*	*0.0050*
	LH moments	GEV	−0.0015	−0.0024	−0.0014	0.0015	0.0070
		GLO	−0.1937	−0.3194	−0.5137	−0.6857	−0.8914
	PL moments	GEV	*−0.0004*	−0.0034	−0.0071	−0.0092	−0.0106
		GLO	−0.0090	−0.0040	0.0070	0.0200	0.0360

Italicized numbers represent minimum RBIAS for the corresponding sample size

Table 7 RRMSE values for different quantiles of the best distribution for L moments, LH moments, and PL moments

Sample size (n)	Method	Distribution	Q_{10}	Q_{20}	Q_{50}	Q_{100}	Q_{200}
20	L moments	GEV	*0.1160*	*0.1500*	0.2150	0.2810	0.3630
		GLO	0.1220	0.1640	0.2420	0.3200	0.4180
	LH moments	GEV	0.1168	0.1469	*0.2116*	0.2843	0.3844
		GLO	0.2645	0.4425	0.8426	1.3608	2.1971
	PL moments	GEV	0.1206	0.1540	0.2123	*0.2633*	*0.3192*
		GLO	0.1240	0.1670	0.2390	0.3040	0.3800
40	L moments	GEV	0.0830	0.1080	0.1530	0.1960	0.2470
		GLO	0.0880	0.1200	0.1790	0.2350	0.3040
	LH moments	GEV	0.0826	0.1058	0.1550	0.2079	0.2760
		GLO	0.2531	0.4187	0.7718	1.218	1.9322
	PL moments	GEV	*0.0821*	*0.1051*	*0.1481*	*0.1905*	*0.2409*
		GLO	0.0850	0.1160	0.1720	0.2270	0.2960
60	L moments	GEV	0.0690	0.0890	0.1250	0.1590	0.1980
		GLO	0.0710	0.0980	0.1440	0.1890	0.2440
	LH moments	GEV	*0.0663*	*0.0858*	0.1274	0.1712	0.2261
		GLO	0.2521	0.4137	0.7451	1.1529	1.7968
	PL moments	GEV	0.0673	0.0867	*0.1213*	*0.1525*	*0.1874*
		GLO	0.0720	0.0980	0.1450	0.1880	0.2380
100	L moments	GEV	0.0520	0.0680	0.0960	0.1210	0.1510
		GLO	0.0570	0.0780	0.1150	0.1480	0.1870
	LH moments	GEV	0.0521	0.0679	0.1012	0.1353	0.1769
		GLO	0.2517	0.4117	0.7321	1.1211	1.7331
	PL moments	GEV	*0.0519*	*0.0670*	*0.0936*	*0.1177*	*0.1446*
		GLO	0.0540	0.0730	0.1100	0.1460	0.1860

Italicized numbers represent minimum RRMSE for the corresponding sample size

are the relative bias (RBIAS) and relative root mean square error (RRMSE) represented by

$$RBIAS = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Q_i^S - Q_i^C(F)}{Q_i^C(F)}\right) \tag{44}$$

$$RRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{Q_i^S - Q_i^C(F)}{Q_i^C(F)}\right)^2} \tag{45}$$

where Q_i^S and $Q_i^C(F)$ are the simulated and calculated quantiles of design flood, respectively. The robustness of the candidate distribution is evaluated by comparing the RBIAS and RRMSE of the estimated flood quantiles, whether the distribution is correctly determined or not.

Tables 6 and 7 present the RRBIAS and RRMSE values of quantiles computed using the L moments, LH moments, and PL moments methods, respectively. In order to better illustrate the results, the minimum achieved values are marked in bold. The results show that the RBIAS and RRMSE values generally increase with a reduction in the sample size and an increase in the return periods.

As shown by Table 6, PL moments of GLO and GEV distributions contribute to the smallest RBIAS values at low quantile, T=10 years. At T=20 and 50 years, L moments of GEV at corresponding sample sizes have produced minimum RBIAS values. At higher quantiles of T=100 and 200 years, the minimum RBIAS values are exhibited by LH moments of GEV distribution and L moments of GLO distribution.

As the results of Table 7, almost all PL moments of GEV distribution of corresponding sample sizes produced smaller RRMSE values compared with L and LH moments. The minimum RRMSE values of L moments appears at n=20 for return periods, T=10 and 20 years under GEV distribution. The minimum RRMSE values of LH moments appear also under GEV distribution at n=20 for T=50 years, and n=60 for T=10 and 20 years. In this case, the minimum RRMSE values are generally best described by GEV distribution for all the three methods of L, LH, and PL moments compared to GLO distribution.

It is interesting to note that from these results, the estimation of quantiles at higher return period is best estimated by PL moments of GEV distribution. These can be found at a high return period of T=100 and 200 years for all sample sizes which produce the lowest RRMSE values compared to L and LH moments. On the other hand, the minimum RRMSE values at lower return period are produced by L and LH moments of GEV distribution. This implies that L and LH moments appear to be more preferred in the estimation of low quantiles compared to the PL moments method.

7 Conclusions

The study provides a comprehensive evaluation of the L moments, LH moments, and PL moments, by first revisiting regional frequency analysis based on the L moments by Hosking and Wallis (1993, 1997) and LH moments by Meshgi and Khalili (2009a, b). Regional homogeneity was investigated by first assuming the entire study area as one homogeneous regional cluster. The corresponding relationships for regional homogeneity analysis by the PL moments are developed. PL moments for the GEV, GLO, and GPA distributions are also developed and used to provide the corresponding PL moments ratio diagrams and the goodness-of-fit test.

The results of this study have shown that from 37 stations in the study region, 33 stations based on L and PL moments and all of 37 stations based on LH moments are accepted statistically to be homogeneous. The Z test has shown that the GEV and GLO distributions of L, LH, and PL moments can be considered as the preferred distributions for modeling daily annual maximum rainfall in Selangor, Malaysia. Finally, Monte Carlo simulations used for performance evaluation show that the PL moments method is more efficient than L and LH moments methods of large return period events particularly at all tested sample sizes.

This work can be extended by including all the rain gauge stations of Malaysia in order to identify the most suitable region distribution for the whole country.

References

Abida H, Ellouze M (2008) Probability distribution of flood flows in Tunisia. Hydrol Earth SystSci 12:703–714

Atiem IA, Harmancioglu NB (2006) Assessment of regional floods using L-moments approach: the case of the River Nile. Water Resour Manag 20:723–747

Bhattarai KP (2004) Partial L-moments for the analysis of censored flood samples. Hydrol Sci J 49(5):855–868

Bhuyan A, Borah M, Kumar R (2010) Regional flood frequency analysis of North-Bank of the River Brahmaputra by using LH-moments. Water Resour Manag 24:1779–1790

Cannarozzo M, Noto LV, La Loggia G (2009) Annual runoff regional frequency analysis in Sicily. Phys Chem Earth 34:679–687

Chen YD, Huang G, Shao Q, Xu C (2006) Regional analysis of low flow using L-moments for Dongjiang basin, South China. Hydrol Sci J 51(6):1051–1064

Deka S, Borah M, Kakaty SC (2011) Statistical analysis of annual maximum rainfall in North-East India: an application of LH-moments. Theor Appl Climatol 104:111–122

Fowler HJ, Kilsby CG (2003) A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. Int J Climatol 23:1313–1334

Glaves R, Waylen PR (1997) Regional flood frequency analysis in Southern Ontario using L-moments. Can Geogr 41(2):178–193

Greenwood JA, Landwehr JM, Matalas NC, Wallis JR (1979) Probability weighted moments: definition and relation to parameters of distribution expressible in inverse form. Water Resour Res 15 (5):1049–1054

Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. J R Stat Soc Ser B 52(1):105–124

Hosking JRM (1995) The use of L-moments in the analysis of censored data. In: Balakrishnan N (ed) Recent advances in life-testing and reliability. CRC Press, Boca Raton

Hosking JRM, Wallis JR (1993) Some statistics useful in regional frequency analysis. Water Resour Res 29(2):271–281

Hosking JRM, Wallis JR (1997) Regional frequency analysis: an approach based on L-moments. Cambridge University Press, UK

Hussain Z, Pasha GR (2009) Regional flood frequency analysis of seven stations of Punjab, Pakistan, using L-moments. Water Resour Manag 23(10):1917–1933

Jingyi Z, Hall MJ (2004) Regional flood frequency analysis for the Gang-Ming River basin in China. J Hydrol 296:98–117

Kjeldsen TR, Smithers JC, Schulze RE (2002) Regional flood frequency analysis in the KwaZulu-Natal province, South Africa using the index-flood method. J Hydrol 255:194–211

Kochanek K, Strupczewski WG, Singh VP, Weglarczyk S (2008) The PWM large quantile estimates of heavy tailed distributions from samples deprived of their largest element. Hydrol Sci J 53 (2):367–386

Koulouris et al (1998) L moment diagrams for censored observations. Water Resour Res 34(5):1241–1249

Kroll CN, Stedinger JR (1996) Estimation of moments and quantiles using censored samples. Water Resour Res 32(4):1005–1012

Kumar R, Singh RD, Seth SM (1999) Regional flood formulas for seven subzones of zone 3 of India. J Hydrol Eng 4(3):240–244

Kumar R, Chatterjee C, Panigrihy N, Patwary BC, Singh RD (2003) Development of regional flood formulae using L-moments for gauged and ungauged catchments of North Brahmaputra river system. J Inst Eng India Civ Eng Div 84(1):57–63

Kyselý J (2010) Coverage probability of bootstrap confidence intervals in heavy-tailed frequency models, with application to precipitation data. Theor Appl Climatol. doi:10.1007/s00704-009-0190-1

Lee SH, Maeng SJ (2003) Comparison and analysis of design floods by the change in the order of LH-moment methods. Irrig Drain 52:231–245

Lim YH, Lye LM (1998) Regional flood estimation for ungauged basins in Sarawak, Malaysia. Hydrol Sci J 48(1):79–94

Meshgi A, Khalili D (2009a) Comprehensive evaluation of regional flood frequency analysis by L- and LH-moments. I. A re-visit to regional homogeneity. Stoch Environ Res Risk Assess 23:119–135

Meshgi A, Khalili D (2009b) Comprehensive evaluation of regional flood frequency analysis by L- and LH-moments. II. Development of LH-moments parameters for the generalized pareto and generalized logistic distributions. Stoch Environ Res Risk Assess 23:137–152

Moisello U (2007) On the use of partial probability weighted moments in the analysis of hydrological extremes. Hydrol Process 21:1265–1279

Noto LV, Logggia GL (2009) Use of L-moments approach for regional frequency analysis in Sicily Italy. Water Resour Manag 23:2207–2229

Parida BP, Kachroo RK, Shrestha DB (1998) Regional flood frequency analysis of Mahi-Sabarmati Basin (Subzone 3-a) using index flood procedure with L-moments. Water Resour Manag 12:1–12

Pearson CP (1991) New Zealand regional flood frequency analysis using L-moments. J Hydrol 30(2):53–64

Rahnama MB, Rostami R (2007) Halil-Basin regional flood frequency analysis based on L-moment approach. Int J Agric Res 2(3):261–267

Saf B (2009) Regional flood frequency analysis using L-moments for the West Mediterranean region of Turkey. Water Resour Manag 23(3):531–551

Wang QJ (1990a) Estimation of the GEV distribution from censored samples by method of partial probability weighted moments. J Hydrol 120:103–110

Wang QJ (1990b) Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the GEV distribution. J Hydrol 120:115–124

Wang QJ (1996) Using partial probability weighted moments to fit the extreme value distributions to censored samples. Water Resour Res 32(6):1767–1771

Wang QJ (1997) LH-moments for statistical analysis of extreme events. Water Resour Res 33(2):2841–2848

Yue S, Wang C (2004) Determination of regional probability distributions of Canadian flood flows using L-moments. J Hydrol N Z 43 (1):59–73

Zalina MD, Nguyen VTV, Amir MK, Mohd Nor MD (2002) Selecting a probability distribution for extreme rainfall series in Malaysia. Water Sci Technol J 45(2):63–68

Zin WWZ, Jemain AA, Ibrahim K (2009) The best fitting distribution of annual maximum rainfalls in Peninsular Malaysia based on methods of L-moment and LQ-moment. Theor Appl Climatol 96:337–344