

UNICA UNIVERSITÀ DEGLI STUDI DI CAGLIARI



This is the Author's *accepted* manuscript version of the following contribution:

Muhammad Shafeeq ul Rehman Khan, Zamir Hussain, Ishfaq Ahmad, *Effects of L-Moments, Maximum Likelihood and Maximum Product of Spacing Estimation Methods in Using Pearson Type-3 Distribution for Modeling Extreme Values*, in *Water Resources Management*, vol. 35 (2021), pp. 1415–1431.

This version of the article has been accepted for publication, after peer review and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.

The Version of Record is available online at:

https://dx.doi.org/10.1007/s11269-021-02767-w

When citing, please refer to the published version.

This full text was downloaded from UNICA IRIS https://iris.unica.it/

Effects of L-Moments, Maximum Likelihood and Maximum Product of Spacing Estimation Methods in Using Pearson Type-3 Distribution for Modeling Extreme Values

Muhammad Shafeeq ul Rehman Khan • Zamir Hussain • Ishfaq Ahmad

Abstract
Modeling of extreme values like annual maxima's is important in many applications. Pearson Type-3 (PE3) distribution is an important probability distribution, widely used for modeling of extreme values with a variety of estimation methods. The focus of this study is to assess the effects of three methods of estimation of parameters for PE3 distribution namely L-moments (LM), maximum likelihood estimation (MLE) and maximum product of spacing (MPS). Assessment is based on a two-step approach. The first step uses simulation experiments while the second is based on empirical analyses, by varying size and shape characteristics of the sample. The study concluded that the estimates using LM method have low bias in case of small sample and when data exhibits small to moderate skewness and kurtosis. MPS is a reasonable alternative and provides efficient estimates, especially when the data shows large skewness and kurtosis with small to moderate size of sample. MLE method is useful in case of very large sample size with low values of shape characteristics of data. The results of this study provide useful guidelines for fitting PE3 distribution, especially to extreme values.

Keywords At-site analysis · Modeling of extreme values · Pearson Type-3 distribution · L-moments · Maximum product of spacing · T-years quantiles

Muhammad Shafeeq ul Rehman Khan
shafiqnaizi@gmail.com

Zamir Hussain
zamir@rcms.nust.edu.pk

Ishfaq Ahmad
ishfaq.ahmad@iiu.edu.pk

1 Introduction

Univariate modeling using probability distribution(s) also known as at-site frequency analysis is a popular area of research. Fitting of a probability distribution to a series of values generated from a random process can provide accurate and reliable estimates of quantiles, if modeled properly. Probabilistic modeling could be a challenge especially when dealing with extreme values because of their non-normal or skewed behavior and availability of limited span of data series at a site. There exists a variety of probability distributions for modeling a random variable with different methods of estimation. Success depends on the type, trends and tendencies, and shape of the variable under study along with the available size of the sample including many others. Details of dealing with extreme values in the area of statistical hydrology, meteorology and wind are available in Beirlant et al. (2004) and Naghettini (2017). A review of different methods of estimation of parameters in extreme values analyses (EVA) and few recommendations for best practices have been discussed in Palutikof et al. 1999 and Arns et al. 2013. In EVA, the choice of a method for parameters estimation is not the only source of variation for considering a distribution. Few stochastic uncertainties that belong to a specific method/model for the estimation of quantiles of T-year return period are a selection of threshold in a peak over threshold method, selection of block size (daily, monthly, quarterly, yearly) for maxima and minima analysis (Coles et al. 2001).

In EVA, the dealing variable usually consists of minima's or maxima's (hourly, weekly, monthly, quarterly or yearly) or peaks over a threshold (POT). Each series of values have certain advantages or disadvantages associated to it. The series of annual maxima's with a record length of at-least twenty years usually provide accurate estimates of T-year quantiles as compared to POT (Cook 1985; Palutikof et al. 1999; Ferreira and de Haan 2015). The variable in focus of this study will be annual maxima's as being a common variable in frequency analysis of extreme events. Major limitations attached with annual maxima's include availability of limited number of observations (span of the data series) and skewed shape of the distribution.

Pearson Type III (PE3) distribution is an important distribution in EVA as illustrated in a number of case studies round the world including IACWD 1982; Chang and Moore 1983; Hussain et al. 2017; Asquith et al. 2017; Li et al. 2017; Lei et al. 2018. Its inclusion in a set of five candidate distributions in the famous methodology of regional frequency analysis proposed by Hosking and Wallis (1997) is also advocating its significance and adequacy to models extreme values.

For fitting PE3 distribution, different methods of estimation have been proposed. For example, Song and Ding (1988) suggested to use probability weighted moments method while Hosking and Wallis (1997) provided L-moments (LM) estimators of its parameters. A comparison of maximum likelihood estimation (MLE) and the method of moments is illustrated in Bobee and Ashkar (1991). Koutrouvelis and Canavos (1999) proposed mixed moment estimators for PE3 distribution using exponentially transformed data. Jan and Shabri (2017) provided a comparison of L-moments and TL-moments estimators of PE3 distribution for river flow predictions in Johor, Malaysia.

The results of these mentioned studies reveal following important facts that there does not exist a universal method of estimation of parameters of PE3 distribution and the choice of estimation method is a critical factor, which depends heavily on size and shape characteristics of the sample. Therefore, this study is designed to compare the performance of three estimation methods (MLE, LM and Maximum product of spacing (MPS)) for fitting PE3 distribution

using simulations and empirical analyses by varying size and shape characteristics of the sample. A concise discussion on the use of these three methods is provided below:

In EVA, the efficiency of estimates using MLE (especially quantiles of T-year return period) is linked to the size of sample (Katz et al. 2002). To overcome this issue, LM derived by Hosking (1990) is a preferred choice. A large number of studies have used LM in regional frequency analysis while modeling annual maxima's, for instance (Lee and Kim 2019; Vivekanandan 2015; Drissia et al. 2019; Hussain 2017; Khan et al. 2019; Rutkowska et al. 2018).

Another, relatively less common choice, especially with small sample size is MPS. Few studies like Wong and Li 2006; Singh et al. 2014; Kumar Singh et al. 2016; Murage et al. 2019 and El-Sherpieny et al. 2020 showed that the method of MPS is a better choice of estimation than traditional methods. Asquith et al. (2017) reported that MPS and LM methods provide nearly identical parameter estimates of PE3 distribution. The study of Soukissian and Tsalis (2015) illustrated that estimates using MPS method are better in terms of bias, mean square error and variance of the estimated parameters for modeling extreme wind speed. But none of study so far has provided a comparison of LM, MLE and MPS methods for estimation of PE3 distribution.

This study is designed to check the performance of LM, MLE and MPS estimation methods for fitting PE3 distribution. A two-step approach is adopted. In the first step, simulation experiments have been performed by introducing linear variation in the shape parameter of the distribution (while selecting standard values of the location and scale parameters) for different sample sizes. These variations will help in assessment of the estimation methods with respect to different combinations of size and shape characteristics of the sample. The second step deals with empirical analysis, i.e. estimating flood quantiles considering annual maxima's of river discharges (AMRD) of four sites of Khyber Pakhtunkhwa (KPK), Pakistan. These sites have been selected keeping in view the sample size, scale and shape of the distribution associated to the observed data series. The findings of this study will provide useful information for the choice of PE3 distribution with adequate estimation method in EVA.

2 Pearson Type-3 Distribution

PE3 distribution also known as generalized gamma distribution is an important probability distribution in EVA. For a random variable Y having PE3 distribution, its density function is

$$f(Y) = \frac{1}{a\Gamma(b)} \left(\frac{y-o}{a}\right)^{b-1} e^{-\left(\frac{y-o}{a}\right)} \quad (1)$$

Where a, b and ε are scale, shape and location parameters respectively. If $y \geq \varepsilon$ and $a > 0$, the shape of PE3 distribution is positively skewed. If $y \leq \varepsilon$ and $a < 0$, its shape is negatively skewed.

The key standardizations of a random variable "Y" having PE3 distribution with (a, b, ε) are given as:

$$Z = \frac{Y-\varepsilon}{a}, \quad K = \frac{Y-\varepsilon-ab}{ab^{1/2}} \quad (2)$$

Here the random variable Z has Gamma distribution with one parameter which is equal to the PE3 (1, b, 0), and "K" is the frequency factor with mean "zero", variance "one" and skewness $\}2^{b^{1}\cdot 2}\}$.

3 Methods of Estimation

3.1 Method of L-Moments

First two L-moments and L-skewness of PE3 distribution are:

$$\lambda_1 = o + ab \tag{3}$$

$$\lambda_2 = r^{-1\cdot 2} a\Gamma\, b + \frac{1}{2}\ /\Gamma(b) \tag{4}$$

$$\tau_3 = 6I_{1/3}(b, 2b) - 3 \tag{5}$$

Here $I_x(p, q)$ represents the ratio of incomplete beta function. Further details are available in Hosking and Wallis (1997) (Appendix A-9).

3.2 Maximum Likelihood Estimates of PE3 Distribution

The log-likelihood function of PE3 distribution is:

$$lnL(a, b, o) = -nln(a) - nln\Gamma(b) + (b-1)\sum_{i=1}^{n} \ln(y_i - o) - n(b-1)\ln(a) - \frac{1}{a} \times \sum_{i=1}^{n} (y_i - o) \tag{6}$$

To obtain the estimates $a^\wedge$, $b^\wedge$ and $o^\wedge$, Eq. 6 is partially differentiated with respect to a, b, and ε and equating expressions equal to zero. These equations can be solved iteratively, as reported in Matalas and Wallis (1973). This iterative procedure cannot provide a solution when the sample size is too small and the value of $b < 1$. It is noted that the value of the coefficient of skewness must be less than 2. For a value of coefficient of skewness greater than 2, the conditional use of MLE for the parameter estimates of PE3 distribution are recommended by Bobee and Ashkar (1991).

3.3 Maximum Product of Spacing Estimates of PE3 Distribution

Cheng and Amin (1983) introduced order statistics based method of MPS for the estimation of any continuous probability distribution. General steps of the methodology of MPS are given below:

Suppose $y_{1:n} \leq y_{2:n} \leq \ldots \leq y_{n:n}$ be the ordered sequence of any continuous random variable "Y" having CDF $F(y, \theta)$ and pdf $f(y, \theta)$. The space between two CDF of the consecutive points

can be defined as.

$$Z_i(\theta) = F(y_{i:n}, \theta) - F(y_{i-1:n}, \theta) = \int_{y_{i-1:n}}^{y_{i:n}} f(y, \theta)dy, \qquad i = 1, 2, 3n + 1 \tag{7}$$

Where $y_{0:n} = -\infty$ and $y_{n+1:n} = +\infty$ then the sum of $Z_i(\theta)$ is equals to one.

$$\sum_{i=1}^{n} Z_i(\theta) = 1 \tag{8}$$

and

$$Z_1(\theta) = 1 - F(y_{n:n}, \theta) \tag{9}$$

Method of MPS provides optimum estimates for the value of θ by maximizing the product of probabilities between two adjacent sample points. The Z_i's are as near as possible to each other. We choose the value of the parameter θ as an estimate that maximizes the log of geometric mean (GM) of $Z_i(\theta)$.

$$K(\theta) = log \left(\prod_{i=1}^{n+1} Z_i(\theta) \right)^{1/(n+1)} = \frac{1}{n+1} \sum_{i=1}^{n+1} log(Z_i(\theta)) \tag{10}$$

The optimum MPS log estimator is given as.

$$K_{opt}(\theta) = log \frac{1}{n+1} \tag{11}$$

The relationship given in Eq. 11 shows that MPS has advantageous results relative to MLE because log-likelihood can reach to $+\infty$, whereas, MPS estimator is always bounded by $\log \frac{1}{n+1}$. It can easily be provided that the maximum and optimum value of GM is only obtained when all Z_i's are equal to $Z_i = i/(n+1)$. For more details, see Ranneby (1984).

The relationship of Eq. 7 can be used to obtain the estimates of parameters of PE3 distribution using MPS method. By using the pdf given in Eq. 1, following equation is obtained.

$$K_{opt}(a, b, \mu) = \frac{1}{n+1} \sum_{i=1}^{n+1} \log \int_{y_{i-1}}^{y} \frac{1}{a\Gamma(b)} \left(\frac{y - \mu}{a} \right)^{b-1} e^{-\left(\frac{y-\mu}{a}\right)} dy \tag{12}$$

Estimates of parameters a, b and ε can be obtained by maximizing the MPS log estimator given in Eq. 12. Closed form solution of Eq. 12 is not available. Therefore, non-linear optimization is used to obtain the numerical solutions of MPS estimators.

4 Simulation Experiments

The analyses include two steps for evaluating adequacy of three estimation methods for fitting PE3 distribution. The first step includes simulation experiments while the second step is based on empirical analysis using AMRD of four sites of KPK. In the first step, 1000 repetitions for different sample sizes, i.e. small, moderate and large (20, 40, 75 and 100) from PE3 distribution have been generated in each case of the estimation of parameters. Standardized form of PE3 distribution has been used as recommended by Wang 1990; NIST/SEMATECH 2012; Jan and Shabri 2017. Root mean square error (RMSE) and bias have been used as

accuracy measures of the estimates. A brief description of different steps in simulation experiments is as follows:

1) Random samples of size 20, 40, 75 and 100 have been generated, 1000 times, by setting the values of parameters of PE3 distribution as $\varepsilon = 0$, $a = 1$ and b. A linear variation in the shape parameter "b" has been introduced with a unit difference for the range of 1.5 to 6.5. This variation will help is assessing the performance of the three estimation methods at different levels of skewness.
2) For each case of 1000 samples, parameters of PE3 distribution have been estimated through LM, MLE and MPS. In few cases, MLE and MPS fail to produce the estimates of parameters at some repetitions during simulations due to the convergence issue of the optimization algorithm.
3) Vectors of estimated parameters through simulations have been obtained. These 1000 estimates have been used to calculate RMSE and bias associated to each parameter using the following expressions:

$$bias = E\hat{\theta} - \theta \tag{13}$$

$$RMSE = \sqrt{Var(\hat{\theta}) + (bias(\hat{\theta}))^2} \tag{14}$$

Where θ is the actual/assumed value of the parameter and $E\hat{\theta}$ is the expected value of the estimated values of the parameter calculated through simulations. The results of these simulation experiments for sample size 20 and 40 are illustrated in Table 1 while for sample size 75 and 100 are presented in Table 2.

For relatively small sample size, i.e. $n = 20$, LM estimation method provides estimates with low bias and RMSE relative to MLE and MPS for location and scale parameters. However, for the shape parameter, the estimates provided by LM method have more bias and are less efficient. The method of MPS provides more accurate and efficient estimates as the value of the shape parameter increases. Almost similar trends are obvious for a moderate sample size of $n = 40$. In few cases, while dealing with extreme values, especially on annual scale, a sample size of 40 may be considered as a large sample size.

For a sample size of $n = 75$ and $n = 100$, LM method provides estimates with less bias for location, scale and shape parameters in comparison to MLE and MPS. However, the estimates of shape parameter are less efficient. The method of MPS is a preferred choice in terms of RMSE, especially when the data exhibits large value of shape parameter. Another notable point is that the method of MLE provides low values of RMSE in comparison to LM method for the estimation of shape parameter when the sample size is quite large, i.e. 75 and 100.

In general, LM method provides estimates with low bias for location, scale and shape parameters for small, moderate and large sample sizes. Therefore, it can be concluded that LM estimates are relatively stable in terms of bias for estimating the parameters of PE3 distribution. The method of MPS is a preferred choice for estimating scale parameter of PE3 distribution for all the sample sizes and values of the shape parameters. The efficiency of MPS method, for estimating the shape parameter, increases with the increase in the value of the shape parameter.

Table 1 Values of bias and RMSE of the parameters estimated through LM, MLE and MPS for sample size 20 and 40

Values of Shape parameter "b"	n=20	LM			MLE			MPS		
		Location (ε)	Scale (a)	Shape (b)	Location (ε)	Scale(a)	Shape(b)	Location (ε)	Scale (a)	Shape (b)
1.5	RMSE	0.2267	0.2453	0.6401	0.2385	0.2759	0.5892	0.2348	0.2835	0.5676
	Bias	−0.0044	0.0053	−0.0951	−0.0319	0.0448	0.2573	0.0589	0.1351	0.0791
2.5	RMSE	0.2151	0.3591	0.8334	0.2136	0.3764	0.4987	0.2457	0.3863	0.5086
	Bias	0.0142	0.0253	−0.0442	0.0449	0.0865	0.0778	0.1090	0.1888	0.0916
3.5	RMSE	0.2033	0.4118	0.5938	0.2494	0.4725	0.5055	0.2699	0.5290	0.3036
	Bias	0.0127	−0.0363	−0.2787	0.0884	0.0634	−0.2925	0.1358	0.1918	−0.2008
4.5	RMSE	0.2065	0.5657	1.4380	0.2280	0.7245	1.3598	0.2238	0.6921	1.3129
	Bias	−0.0045	0.0286	0.1723	0.1184	0.2938	−0.0228	0.0431	0.1046	−0.1640
5.5	RMSE	0.2398	0.9972	2.2015	0.2795	1.1368	1.7119	0.2612	1.1228	1.6181
	Bias	0.0175	0.1727	0.4095	0.1158	0.3938	0.0228	0.0994	0.4063	0.3233
6.5	RMSE	0.2256	1.3475	4.2180	0.2708	1.5568	3.3284	0.2299	1.0058	2.5620
	Bias	0.0208	0.2618	0.9481	0.1144	0.5899	0.4289	0.0712	0.2534	0.0310
n=40										
1.5	RMSE	0.1493	0.1723	0.4968	0.1535	0.1904	0.4018	0.1550	0.1829	0.3600
	Bias	−0.0096	0.0127	0.0115	−0.0176	0.0498	0.1877	0.0302	0.0830	0.0720
2.5	RMSE	0.1623	0.2659	0.5269	0.1671	0.3098	0.3506	0.1918	0.3116	0.2827
	Bias	0.0001	0.0208	0.0244	0.0489	0.0838	0.0405	0.0708	0.1452	0.1191
3.5	RMSE	0.1443	0.2993	0.6944	0.1849	0.3818	0.5432	0.1874	0.3649	0.3550
	Bias	−0.0076	0.0052	0.0263	0.1067	0.1415	−0.1539	0.0444	0.0432	−0.1453
4.5	RMSE	0.1616	0.4659	1.0067	0.2190	0.6303	0.9439	0.1973	0.5771	0.9708
	Bias	0.0140	0.0748	0.1513	0.1364	0.2931	−0.1122	0.0838	0.2640	0.2248
5.5	RMSE	0.1635	0.5347	1.2858	0.2195	0.7056	1.1741	0.2080	0.6684	1.1048
	Bias	0.0093	0.0755	0.2138	0.1029	0.2904	−0.0550	0.0859	0.2516	0.0054
6.5	RMSE	0.1648	0.7520	1.7845	0.2089	1.0445	1.5147	0.1965	0.9100	1.4068
	Bias	0.0042	0.0579	0.1274	0.0788	0.3015	−0.1037	0.0629	0.1939	−0.2521

Table 2 Values of bias and RMSE of the parameters estimated through LM, MLE and MPS for sample size 75 and 100

Values of Shape parameter "b"	n=75	LM			MLE			MPS		
		Location (ε)	Scale (a)	Shape (b)	Location (ε)	Scale (a)	Shape (b)	Location (ε)	Scale (a)	Shape (b)
1.5	RMSE	0.1139	0.1291	0.3259	0.1158	0.1325	0.2074	0.1164	0.1296	0.2042
	Bias	−0.0053	0.0067	−0.0018	−0.0059	0.0253	0.1089	0.0191	0.0507	0.0540
2.5	RMSE	0.1111	0.1755	0.3644	0.1367	0.2384	0.2488	0.1330	0.2083	0.1903
	Bias	0.0181	0.0359	0.0251	0.0753	0.1045	0.0146	0.0645	0.1208	0.0884
3.5	RMSE	0.1170	0.2341	0.5089	0.1716	0.3224	0.3743	0.2258	0.4333	0.3202
	Bias	−0.0031	0.0045	0.0136	0.1324	0.1516	−0.2380	0.0982	0.1404	−0.1370
4.5	RMSE	0.1250	0.3137	0.6595	0.1998	0.4653	0.5511	0.1892	0.4633	0.5289
	Bias	0.0026	0.0181	0.0520	0.1494	0.2512	−0.2922	0.1131	0.2386	−0.0667
5.5	RMSE	0.1138	0.3540	0.9044	0.1939	0.5447	0.7813	0.1594	0.4657	0.7790
	Bias	−0.0086	0.0004	0.1259	0.0959	0.2261	−0.1668	0.0632	0.1449	−0.1486
6.5	RMSE	0.1212	0.4422	1.0621	0.1706	0.6020	0.9172	0.1559	0.5477	0.8432
	Bias	0.0023	0.0131	0.0084	0.0878	0.2597	−0.1948	0.0567	0.1483	−0.2618
	n=100									
1.5	RMSE	0.0920	0.1008	0.2729	0.0920	0.1024	0.1666	0.0932	0.1006	0.1613
	Bias	0.0014	0.0067	0.0142	0.0015	0.0184	0.0732	0.0209	0.0384	0.0286
2.5	RMSE	0.1052	0.1648	0.3081	0.1297	0.2154	0.2189	0.1263	0.1888	0.1580
	Bias	0.0069	0.0186	0.0170	0.0655	0.0867	0.0045	0.0425	0.0829	0.0680
3.5	RMSE	0.1010	0.1952	0.4160	0.1545	0.2764	0.3180	0.1823	0.3018	0.2457
	Bias	0.0084	0.0204	0.0156	0.1315	0.1595	−0.1989	0.1393	0.2324	−0.0170
4.5	RMSE	0.1135	0.2831	0.5569	0.1781	0.4107	0.5016	0.1782	0.4240	0.4417
	Bias	0.0014	0.0273	0.1128	0.1344	0.2414	−0.2030	0.1149	0.2410	−0.0657
5.5	RMSE	0.0980	0.3139	0.7489	0.1852	0.5304	0.6599	0.1710	0.4980	0.6548
	Bias	−0.0013	0.0053	0.0270	0.1318	0.3077	−0.2466	0.0858	0.1899	−0.2300
6.5	RMSE	0.1050	0.4071	0.9570	0.1783	0.6251	0.8275	0.1592	0.5582	0.7654
	Bias	0.0029	0.0451	0.1807	0.0980	0.3142	−0.0788	0.0703	0.2156	−0.1191

The method of MLE provides comparable values of bias and RMSE for relatively large sample sizes, i.e. 75 and 100 and low values of shape parameter, i.e. for b = 1.5.

5 Empirical Analysis

In the second step, the performance of three estimation methods (LM, MLE and MPS) for fitting PE3 distribution has also been tested using a real life data set. AMRD in cubic feet per second (f^3/s) of four sites in KPK, Pakistan have been used. The secondary data is provided by the flood section of Irrigation Department of KPK. These sites are selected keeping in view the variations in the sample size, trends and tendencies of scale and shape characteristics (skewness and kurtosis) of the observed data series. None of the published studies so far (to the best of authors' knowledge) performed at-site frequency analysis of AMRD using PE3 distribution of these four sites. Geographical coordinates and record length of observed data series at each site is provided in Table 3. Few details of the fitting procedure are:

Time series plots of AMRD at each site are illustrated in Fig. 1. These graphs show that there exists random variation in the observed data series at each site. Moreover, the occasional occurrence of a large magnitude of flood is also obvious. To observe the general trends and tendencies of AMRD at each site, few descriptive measures are calculated and presented in Table 4. The information reveals that for the four sites, the sample size varies from 21 to 34. Data

Table 3 Geographical coordinates and record length of four sites

S. No.	Site Name	Latitude (North)	Longitude (East)	Record length in years
1	Wazir Ghari	33.9845	71.7749	1979–2010
2	Jundi Utmanzai	34.0094	71.8328	1987–2011
3	Bara Kohat Road	33.8638	71.5635	1982–2015
4	Shahban	34.0919	72.0391	1987–2007

exhibits variation as shown by standard deviation as a scale statistic. The shape of the data series at four sites is positively skewed with the values of skewness ranging from 1.45 to 3.19 and leptokurtic behavior as kurtosis values are showing more spread ranging from 1.48 to 10.67. These descriptive statistics show that the trends and tendencies of AMRD at four sites are different from each other, especially in terms of shape of the distribution. Therefore, the data is suitable to evaluate the effectiveness of different estimation methods for fitting PE3 distribution.

The estimated parameters along with their RMSE and bias are given in Table 5. The two accuracy measures, i.e. RMSE and bias have been calculated using simulation experiments. For this purpose, estimated parameters of each site have been used to generate 1000 random samples from PE3 with sample size equal to its observed counterpart. Then for each site and generated sample, PE3 distribution is fitted using LM, MLE and MPS methods. These simulated values (of each parameter) have been used to calculate the RMSE and bias. The results of Table 5, for each site are discussed below:

i. For the site Wazir Ghari having relatively moderate sample size and high skewness and kurtosis values, the estimates of location parameter have nearly comparable RMSE for LM and MPS methods but LM method is showing less bias. However, for the estimates of scale and shape parameters, MPS method has obvious less values of RMSE while LM method is showing less bias.
ii. For the site Jundi Utmanzai, having relatively small sample size but high skewness and kurtosis values, estimates of MPS method for all the three parameters are showing less RMSE values. The estimates using LM method are showing less bias; but, interestingly, the estimates of MPS method are showing comparable bias for the estimation of shape parameter. Therefore, the performance of MPS method in case of small sample size with high skewness and kurtosis is comparable with LM method while estimating scale and shape parameters in terms of bias.
iii. For the site Bara Kohat Road having relatively large sample size with moderate shewness and high kurtosis, LM method provides estimates for location and scale parameters with low bias; however, the bias is comparable for LM and MPS methods for the estimation of shape parameter. Here MPS method provides estimates with low RMSE for all the three parameters of PE3 distribution.
iv. For the site Shahban having relatively small sample size and the distribution of observed data series is showing low skewness and kurtosis values, LM method provides estimates with low bias and RMSE for location and scale parameters. However, for the estimates of shape parameter, MPS method provides lowest RMSE value in comparison to LM and MLE.

The above discussion reveals that the method of LM generally provides estimates with low bias for all the parameters of PE3 distribution but MPS method provides efficient estimates,

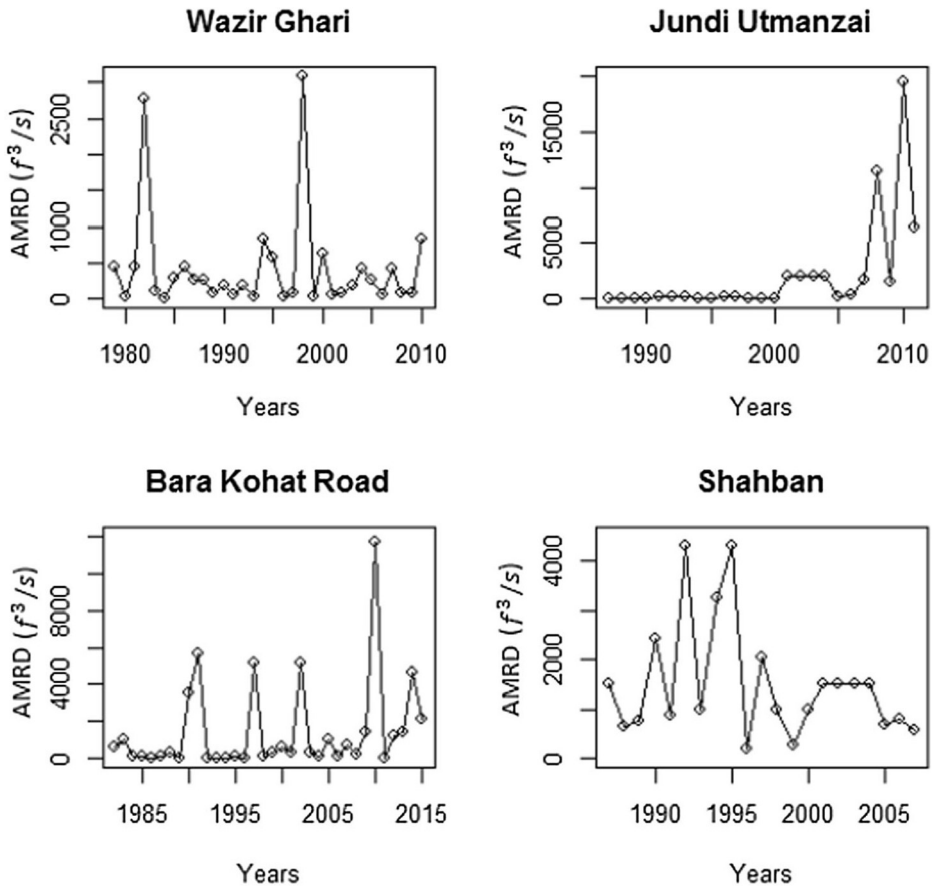

Fig. 1 Time series plots of the recorded data of annual maxima's of river discharge (ARMD) in cubic feet per second (f^3/s)

i.e. having low values of RMSE especially for the scale and shape parameters. The efficiency of LM method increases for relatively small sample size and small to moderate values of L-skewness and L-kurtosis, for instance like the site Shahban. However, as the severity of skewness and kurtosis increases, MPS method is more suitable to estimate the scale and shape parameters of PE3 distribution. The results of this application also show that MLE method is

Table 4 Descriptive statistics of AMRD of four sites. Here n is the number of observations, Min and Max are the minimum and maximum values in the data series, Skewness and Kurtosis are moments measures of skewness of kurtosis

S. No.	Site Name	n	Min	Max	Mean	Standard Deviation	Coefficient of Variation	Skewness	Kurtosis
1	Wazir Ghari	32	5	3080	427	697.73	163.61	3.16	10.03
2	Jundi Utmanzai	25	35	19,433	2053	4407.79	214.74	3.19	10.67
3	Bara Kohat Road	34	24	11,698	1453	2456	169.03	2.71	8.58
4	Shahban	21	218	4309	1516	1171.86	77.31	1.45	1.48

Table 5 Estimates of parameters of PE3 distribution along with their RMSE and Bias

Estimation Method		Wazir Ghari			Jundi Utmanzai			Bara Kohat Road			Shahban		
	Parameter	Estimate	RMSE	Bias	Estimate	RMSE	Bias	Estimate	RMSE	Bias	Estimate	RMSE	Bias
LM	Location (ε)	426.47	125.53	−0.35	2052.61	1072.06	−47.89	1453.00	486.94	1.23	1515.76	253.41	3.10
	Scale (a)	697.43	278.93	11.26	5135.75	3062.81	−36.94	2697.11	1128.06	60.30	1238.83	381.65	29.75
	Shape (b)	3.7573	0.8305	0.0110	5.2349	1.3625	0.0500	3.8403	0.8270	0.0204	2.0732	0.7478	−0.0087
MLE	Location (ε)	423.85	136.64	−6.99	2075.69	1259.12	124.23	1453.03	494.27	45.78	1554.06	313.49	28.20
	Scale (a)	773.01	330.79	45.69	5615.70	4079.00	899.56	2924.01	1184.45	284.63	1405.40	496.59	173.27
	Shape (b)	3.6911	0.8107	0.2839	5.5037	1.9993	0.6402	4.0923	0.9343	0.2982	2.1038	0.4968	0.2919
MPS	Location (ε)	517.10	126.73	51.46	2315.12	718.52	−12.35	1481.48	426.01	17.26	1573.92	299.97	96.71
	Scale (a)	678.62	196.89	101.77	4133.00	1529.32	−44.26	2457.26	791.08	44.97	1237.17	408.87	177.92
	Shape (b)	2.6355	0.2954	0.1215	3.6242	0.5394	−0.0319	3.3717	0.4810	0.0355	1.6276	0.4529	0.0548

not a reasonable choice of estimation in EVA for fitting PE3 distribution with small to moderate size of sample.

For a general numerical assessment of goodness-of-fit of the considered methods, two measures have been used namely standard error of fit (SEF) and Cramer Von-Mises (CVM) test.

The SEF is defined in Kite 1988 as:

$$SEF = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - n_p}} \tag{15}$$

Where y_i are the observed values of the data series at each site, $\hat{y}_i$ are the fitted values generated through fitted distribution, "n" is the sample size and n_p is the number of estimated parameters. CVM test is another goodness of fit test. In this test, F_0 (a cumulative distribution function (CDF)) is compared with a given F_n (empirical CDF). Suppose $X_1, X_2 \ldots X_n \sim F$, then a null hypothesis $H_0 : F = F_0$ is tested against the two-tailed alternative hypothesis. The test statistics of CVM test uses squared difference between F_n and F_0.

$$T = n \int_{-\infty}^{\infty} (F_n(x) - F_0(x))^2 dF_0(x) \tag{16}$$

After simplification of Eq. (16), following expression is used for the computations of T.

$$T = \frac{1}{12n} + \sum_{j=1}^{n} \left(F_0\left(X_{(j)}\right) - \frac{2j-1}{2n} \right)^2 \tag{17}$$

For a small value of T, the null hypothesis holds otherwise rejected. For more details of CVM test see Csörgő and Faraway (1996).

The values of these goodness-of-fit measures are given in Table 6. Comparable performances have been observed for LM and MPS methods. For the sites with relatively large sample size, skewness and kurtosis values, i.e. Wazir Ghari and Bara Kohat Road, method of MPS is a preferred choice; however, CVM method favors of LM method for the site Wazir Ghari. For the site Jundi Utmanzai having relatively small sample size and high skewness and kurtosis values, MPS method is the preferred choice. For the site Shahban having smallest sample size with low skewness and kurtosis values, LM method have least SEF value and highest corresponding p value of CVM. These results again show that method of MPS can be a preferred choice of estimation of PE3 distribution for sites having moderate to large sample size and high values of skewness and kurtosis. However, LM method provides better results for fitting PE3 distribution for small sample size with moderate skewness and kurtosis associated to the distribution of observed data series.

Flood quantiles for return periods of 20, 50 and 100 year have been estimated using the quantile function of PE3 distribution. These estimated quantiles along with their RMSE and bias are given in Table 7. Similar trends are obvious from the values of accuracy measures of the estimated quantiles. In general, the estimates using MPS method have low RMSE values while with LM method have low bias. The performance of LM method is superior for the site having small sample size and low values of skewness and kurtosis (for the site Shahban).

Another way of measuring the performance of estimation methods in fitting of a model is the calculation of 95% confidence intervals of the estimated quantiles. R-package "lmomco" provided by Asquith (2020) has been used for the development of 95% confidence intervals

Table 6 Values of goodness-of-fit measures for PE3 distribution. Here bold values indicate best fit method

Estimation Methods	Wazir Ghari			Jundi Utmanzai		
	SEF	CVM	P value	SEF	CVM	P value
LM	1009.51	0.1811	0.3082	6783.48	0.2960	0.1384
MLE	1060.71	0.6939	0.0126	7124.32	0.8168	0.0061
MPS	996.59	0.2212	0.2303	6208.14	0.2376	0.2053
		Bara Kohat Road			Shahban	
LM	3703.27	0.1009	0.5828	1793.43	0.0689	0.7638
MLE	3859.99	0.2815	0.1525	1917.02	0.1029	0.5748
MPS	3549.94	0.0696	0.7570	1795.14	0.0765	0.7177

for the estimated quantiles. These intervals along with growth curves of the estimated flood quantiles are illustrated in Fig. 2. This figure shows that the MPS estimation method provides shortest 95% confidence interval, especially a stable upper bound in the extreme upper tail of the distribution.

6 Summary and Conclusions

This study aims to assess the performance of three estimation methods MLE, LM and MPS for fitting PE3 distribution by varying size and shape characteristics of the sample. Two of the three methods (MLE and LM) are quite popular while scarce literature is available with respect to the application of MPS method. The assessment is based on simulations and empirical analyses. The fitting of PE3 distribution has been tested through two goodness-of-fit measures SEF and CVM. RMSE, bias and 95% confidence intervals of the estimated quantiles have been calculated for assessment analyses of estimation methods. Major findings of the study are:

i. The results of simulation experiments reveal that the estimates using LM method have low bias. The method of MPS is a preferred choice for estimating scale parameter of PE3 distribution for all the considered sample sizes and values of the shape parameters. The efficiency of MPS method, for estimating the shape parameter, increases with increase in the value of shape parameter. The method of MLE provides comparable values of bias and RMSE for relatively large sample sizes, i.e. 75 and 100 and low values of shape parameter, i.e. for b = 1.5.
ii. Similar tendencies of assessment measures have been observed through empirical analyses. These include that LM method is a preferred choice in case of small sample size and low or moderate skewness and kurtosis values of the observed data series. Alternatively, MPS method provides efficient estimates for moderate to large sample size and high values of skewness and kurtosis. The stability and efficiency of MPS method is obvious for the estimation of shape parameter of PE3 distribution. Moreover, MLE method is not a preferred method of estimation for fitting PE3 distribution for relatively small to moderate sample size.

Therefore, the study concluded that for fitting PE3 distribution, estimates using LM method have low bias in case of small sample and when data exhibits small to moderate skewness and kurtosis. MPS is a reasonable alternative and provides efficient estimates, especially when the data shows large skewness and kurtosis. MLE is useful in case of very large sample size with

Table 7 Predicted flood quantiles for various return periods (in years) along with their RMSE and Bias

Method	Site name	Wazir Ghari			Jundi Utmanzai			Bara Kohat Road			Shahban		
	Return period	20	50	100	20	50	100	20	50	100	20	50	100
LM	Quantile	1785	2696	3427	10,946	18,972	25,714	6681	10,250	13,119	3993	5150	6028
	RMSE	607.92	1067.47	1329.67	5061.44	10,055.32	16,705.13	2242.83	4036.10	5450.48	946.56	1438.33	2024.56
	Bias	−29.28	11.14	−17.06	1214.00	−366.20	−1014.00	−164.60	−541.90	−55.46	−29.99	−11.15	−369.40
MLE	Quantile	1934	2934	3735	11,522	20,516	28,136	7035	11,044	14,293	4367	5690	6695
	RMSE	661.05	1043.29	1647.86	6036.78	12,816.27	17,888.04	2677.70	4520.81	6561.35	1192.52	1713.46	2327.00
	Bias	−65.43	−49.62	−352.90	719.20	−844.20	−2752.50	−243.55	−785.70	−1062.00	−447.00	−425.84	−769.2
MPS	Quantile	1883	2607	3167	10,418	15,709	19,934	6351	9369	11,761	4005	5025	5785
	RMSE	494.41	790.46	993.25	4669.49	7590.64	7840.88	2068.11	3089.41	4018.14	1006.12	1452.33	1681.82
	Bias	−216.3	−277.4	−401.8	−543.9	168.90	−465.80	−370.60	−577.50	−354.80	−577.80	−606.30	−744.00

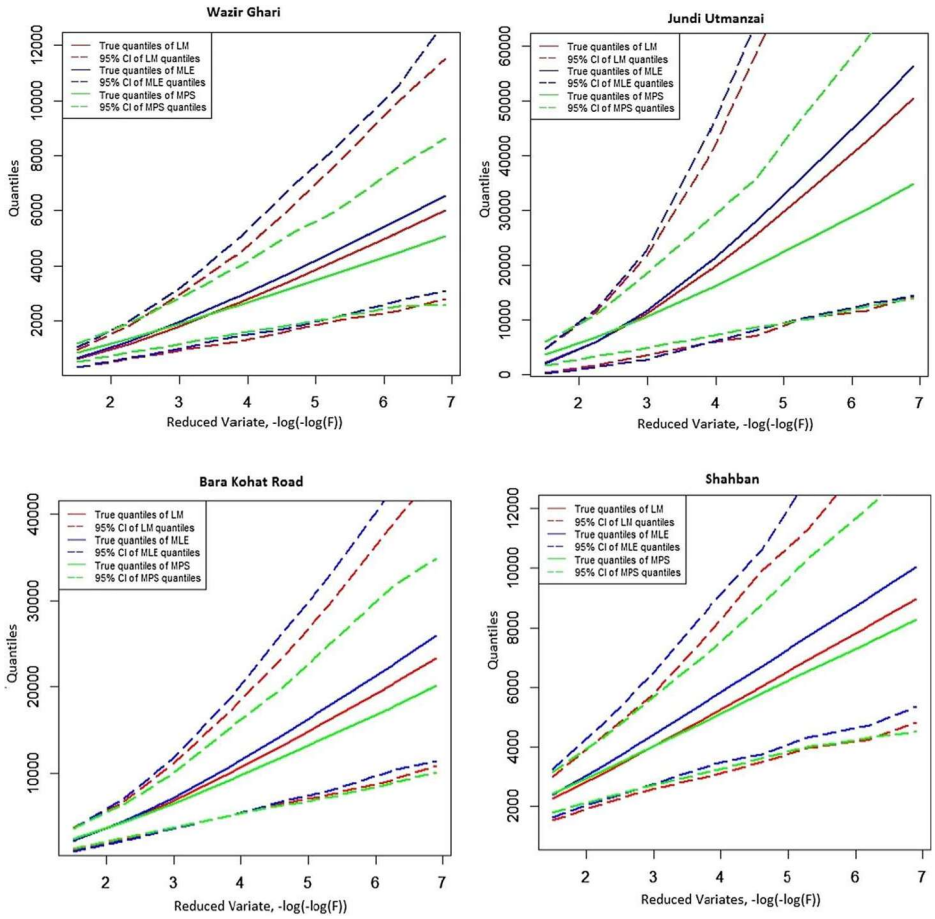

Fig. 2 Growth curves of predicted flood quantiles with 95% confidences intervals

low values of shape characteristics of data. The results of this study provide useful guidelines for fitting PE3 distribution to extreme values. These results can be improved in future by considering different probability distributions from the family of extreme value distributions, sample sizes and variations in terms of location, scale and shape parameters of the distributions.

Acknowledgements We acknowledge the support of Irrigation department of KPK, Pakistan for providing data for the study. We are also thankful to the handling editor, associate editor and anonymous reviewers for their constructive comments to improve the quality of the paper.

Authors Contributions Conceptual idea was initiated by [Muhammad Shafeeq ul Rehman Khan] and [Ishfaq Ahmad], Supervision [Ishfaq Ahmad], Data collection and analyses were performed by [Muhammad Shafeeq ul Rehman Khan] and [Zamir Hussain], original draft was written by [Muhammad Shafeeq ul Rehman Khan] and [Zamir Hussain], final draft was reviewed and edited by [Ishfaq Ahmad] and [Zamir Hussain].

Funding This study has not received any kind of funding form any other person or organization.

Data Availability Data of annual maxima's of river discharges of four sites of KPK Pakistan, used in this study and all the other relevant material will be available for the editorial office at any stage for non-commercial purpose.

Declarations

Ethical Approval All the authors declare that the current manuscript is original and complete, not a part of any other study, it is not submitted in any other place for publication and results are truly presented without fabrication.

References

Arns A, Wahl T, Haigh ID, Jensen J, Pattiaratchi C (2013) Estimating extreme water level probabilities: A comparison of the direct methods and recommendations for best practise. Coast Eng 81:51–66

Asquith WH (2020) L-moments, censored L-moments, trimmed L-loments, L-comoments, and many distributions. R package version 2.3.6, Texas Tech University, Lubbock, Texas. https://cran.r-project.org/web/packages/lmomco/lmomco.pdf. Accessed 18 May 2000

Asquith WH, Kiang JE, Cohn TA (2017) Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities (no. 2017–5038). US geological survey

Beirlant J, Goegebeur Y, Teugels J, Segers J (2004) Statistics of extremes. Wiley series in probability and statistics

Bobee B, Ashkar F (1991) The Gamma family and derived distributions applied in hydrology (no. GB656. 2. M34. B63 1991)

Chang SK, Moore SM (1983) Flood frequency analysis for mall watersheds in Southern Illinois. Water Resour Res 19(2):277–282

Cheng RCH, Amin NAK (1983) Estimating parameters in continuous univariate distributions with a shifted origin. J R Stat Soc Ser B Methodol 45(3):394–403

Coles S, Bawa J, Trenner L, Dorazio P (2001) An introduction to statistical modeling of extreme values, vol 208. Springer, London, p 208

Cook NJ (1985) The designer's guide to wind loading of building structures part 1: background. Damage survey, wind data and structural classification building research establishment, Garston and Butterworths London

Csörgő S, Faraway JJ (1996) The exact and asymptotic distributions of cramér-von mises statistics. J R Stat Soc Ser B Methodol 58(1):221–234

Drissia TK, Jothiprakash V, Anitha AB (2019) Flood frequency analysis using L moments: a comparison between at-site and regional approach. Water Resour Manag 33(3):1013–1037

El-Sherpieny ESA, Almetwally EM, Muhammed HZ (2020) Progressive type-II hybrid censored schemes based on maximum product spacing with application to power lomax distribution. Phys A Stat Mech Appl 553: 124251

Ferreira A, De Haan L (2015) On the block Maxima method in extreme value theory: PWM estimators. Ann Stat 43(1):276–298

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

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

Hussain Z (2017) Estimation of flood quantiles at gauged and ungauged sites of the four major rivers of Punjab, Pakistan. Nat Hazards 86(1):107–123

Hussain Z, Shahzad MN, Abbas K (2017) Application of regional rainfall frequency analysis on seven sites of Sindh, Pakistan. KSCE J Civ Eng 21(5):1812–1819

Interagency Advisory Committee on Water Data [IACWD] (1982) Guidelines for determining flood flow frequency: Bulletin 17b of the hydrology subcommittee, office of water data coordination, U.S. geological survey, Reston, Va., 183 p
Jan NAM, Shabri A (2017) Estimating distribution parameters of annual maximum streamflows in Johor, Malaysia using TL-smoments approach. Theor Appl Climatol 127(1–2):213–227
Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. Adv Water Resour 25(8–12): 1287–1304
Khan MSR, Hussain Z, Ahmad I (2019) A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in regional frequency analysis. Appl Ecol Environ Res 17(3): 6937–6959
Kite GW (1988) Frequency and risk analyses in hydrology. Water Resources Publications, Littleton
Koutrouvelis IA, Canavos GC (1999) Estimation in the Pearson type 3 distribution. Water Resour Res 35(9): 2693–2704
Kumar Singh R, Kumar Singh S, Singh U (2016) Maximum product spacings method for the estimation of parameters of generalized inverted exponential distribution under progressive type II censoring. J Stat Manag Syst 19(2):219–245
Lee DH, Kim NW (2019) Regional flood frequency analysis for a poorly gauged basin using the simulated flood data and L-moment method. Water 11(8):1717
Lei GJ, Yin JX, Wang WC, Wang H (2018) The analysis and improvement of the fuzzy weighted optimum curve-fitting method of Pearson–type III distribution. Water Resour Manag 32(14):4511–4526
Li W, Zhou J, Sun H, Feng K, Zhang H, Tayyab M (2017) Impact of distribution type in bayes probability flood forecasting. Water Resour Manag 31(3):961–977
Matalas NC, Wallis JR (1973) Eureka! It fits a Pearson type: 3 distribution. Water Resour Res 9(2):281–289
Murage P, Mung'atu J, Odero E (2019) Optimal threshold determination for the maximum product of spacing methodology with ties for extreme events. Open J Model Simul 7(03):149
Naghettini M (ed) (2017) FFundamentals of statistical hydrology. Springer International Publishing, Cham
NIST/SEMATECH (2012) E-handbook of statistical methods. NIST/SEMATECH 2012. http://www.itl.nist.gov/div898/handbook/. Accessed in Dec 2019
Palutikof JP, Brabson BB, Lister DH, Adcock ST (1999) A review of methods to calculate extreme wind speeds. Meteorol Appl 6(2):119–132
Ranneby B (1984) The maximum spacing method. An estimation method related to the maximum likelihood method. Scand J Stat 1(2):93–112
Rutkowska A, Żelazny M, Kohnová S, Łyp M, Banasik K (2018) Regional L-moment-based flood frequency analysis in the upper Vistula River basin, Poland. In: Geoinformatics and atmospheric science. Birkhäuser, Cham, pp 243–263
Singh U, Singh SK, Singh RK (2014) A comparative study of traditional estimation methods and maximum product spacings method in generalized inverted exponential distribution. J Stat Appl Probab 3(2):153
Song D, Ding J (1988) The application of probability weighted moments in estimating the parameters of the Pearson type three distribution. J Hydrol 101(1-4):47–61
Soukissian TH, Tsalis C (2015) The effect of the generalized extreme value distribution parameter estimation methods in extreme wind speed prediction. Nat Hazards 78(3):1777–1809
Vivekanandan N (2015) Flood frequency analysis using method of moments and L-moments of probability distributions. Cogent Eng 2(1):1018704
Wang QJ (1990) Estimation of the GEV distribution from censored samples by method of partial probability weighted moments. J Hydrol 120(1–4):103–114
Wong TST, Li WK (2006) A note on the estimation of extreme value distributions using maximum product of spacings. In: Time series and related topics. Institute of mathematical statistics 52:272–283

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.