Turk J Agric For
29 (2005) 419-428

Comparison of Parameter Estimation Methods for the Three-Parameter Generalized Pareto Distribution

Tekin ÖZTEKİN
University of Gaziosmanpaşa, Faculty of Agriculture, Agriculture Technology, Tokat - TURKEY

Received: 07.09.2004

Abstract: The generalized Pareto distribution, which is a special case of both exponential and Wakeby distribution, has good potential for the analysis of flood peaks because of its inherent properties. In this paper, the parameter estimation methods of the moments, probability-weighted moments, maximum likelihood, principle of maximum entropy, and least squares to estimate the parameters in the three-parameter generalized Pareto distribution are compared. The usefulness and applicability of each method is discussed by application to observed annual discharge data for 50 different rivers, most of them in Turkey. The comparisons are based on the ability of each method to predict the elements of the sample series whose non-exceedence probabilities were determined by the Cunnane plotting position formula. Altogether the results demonstrate that for the annual discharge time series considered in this paper the moments method is superior to all the other parameter estimation methods employed.

Key Words: three-parameter generalized Pareto distribution, moments, probability-weighted moments, maximum likelihood, principle of maximum entropy, least squares

Üç Parametreli Genelleştirilmiş Pareto Dağılımı için Parametre Tahmin Yöntemlerinin Karşılaştırılması

Özet: Üstel ve Wakeby dağılımlarının özel bir durumu olan genelleştirilmiş Pareto dağılımı, yapısındaki özelliklerinden dolayı en yüksek debilerin analizi için iyi bir potansiyele sahiptir. Bu makalede, üç parametreli genelleştirilmiş Pareto dağılımı için parametre tahmin yöntemlerinden momentler, olasılık ağırlıklı momentler, maksimum olasılık, maksimum entropy prensibi ve en küçük kareler yöntemi karşılaştırılmıştır. Her yöntemin faydalılığı ve uygulanabilirliği, çoğunluğu Türkiye'den farklı elli adet nehrin yıllık en yüksek debi setleri kullanılarak tartışılmıştır. Karşılaştırmalar, her yöntemin aşılmama olasılıkları Cunnane pozisyon çizim formülü ile belirlenen örnek serilerin elemanlarını tahmin etme kabiliyetine dayandırılmıştır. Sonuçlar, dikkate alınan debi serileri için, momentler yönteminin diğer yöntemlere göre daha iyi olduğunu göstermiştir.

Anahtar Sözcükler: Üç parametreli genelleştirilmiş Pareto dağılımı, moment, olasılık ağırlıklı moment, maksimum olasılık, maksimum entropy prensibi, en küçük kareler

Introduction

The use of various distributions with a short period of recorded annual peak discharge data enables researchers to predict discharges corresponding to higher return periods than can be estimated with the existent lengths of data records. Furthermore, frequency analyses with probability distributions help designers to predict estimates as precisely as possible. The generalized Pareto (GP) distribution was introduced by Pickands (1975) as a two-parameter distribution and has been used widely by many scientists (van Montfort and Witter, 1985; Hosking and Wallis, 1987; Joe, 1987; Smith, 1991; Wang, 1991; Barrett, 1992; Rosbjerg et al., 1992; Moharram et al., 1993; Vogel et al., 1993; Prudhomme et al., 2003) in flood frequency analysis. A review of the literature shows that this distribution does not appear to have been employed for the analysis of any Turkish river data. The properties of the generalized Pareto distribution stated by Hosking and Wallis (1987) and Singh and Guo (1995) make the distribution a logical candidate for the analysis of extreme events. After determining the method for estimating the parameters of best fit, the generalized

Correspondence to: toztekin@gop.edu.tr

Pareto distribution can be employed and compared with previously employed distributions for the analysis of Turkish river data.

In the literature various investigations can be found that have discussed the moments (MOM), probability-weighted moments (PWM), L-moments, maximum likelihood (ML), principle of maximum entropy (POME) and least squares (LS) methods to determine the parameters in the GP3 distribution. Moharram et al. (1993) used the methods of MOM, PWM, ML, and LS to determine the parameters of the GP3 distribution. The performances of these methods were compared by Monte Carlo simulated data and the annual maximum flood series (Ahmad et al., 1988a, 1988b) of 2 rivers. In addition to MOM, PWM and ML, Singh and Guo (1995) employed the POME method for estimating the parameters of the GP3 distribution using Monte Carlo simulated data. They reported that the performance of the 4 estimation methods varied with the values of skewness (C_s) of samples and the values of the non-exceedence or cumulative probabilities (F).

Including Moharram et al. (1993) and Singh and Guo (1995), the review of the literature shows that MOM, PWM, ML, LS, and POME methods for estimating the parameters of the GP3 distribution using a wide range of observed annual peak discharge data of rivers were not compared with each other. Furthermore, the methods of LS and POME were not compared with each other either. The objective of the work presented in this paper is to determine the best parameter estimation methods among the MOM, PWM, ML, LS, and POME for the GP3 distribution using observed annual peak data series.

Generalized Pareto Distribution and Parameter Estimation Methods

The cumulative distribution function for the three-parameter form of the GP distribution (GP3) is

$$F(x) = 1 - \left[1 - \frac{a}{b}(x-c)\right]^{1/a} \quad \text{if} \quad a \neq 0 \tag{1}$$

$$F(x) = 1 - \exp\left(\frac{c - x}{b}\right) \quad \text{if} \quad a = 0 \tag{2}$$

where a is the shape parameter, b is the positive scale parameter, c is the position or location parameter, x is the random variable, and F(x) is the cumulative probability of x.

To estimate the parameters in equations (1) and (2) a variety of different methods have found widespread use in the field of hydrology. These include the MOM, PWM, L-moments, ML, POME and LS fitting methods.

The general log-likelihood function form of the GP3 distribution is

$$L(x_i,a,b,c) = -n\ln(b) + \left(\frac{1-a}{a}\right)\sum_{i=1}^{n} \ln\left[1 - \frac{a}{b}(x_i - c)\right] \tag{3}$$

where x_i is sample value and n is sample size. The procedures for estimating the parameters of the GP3 distribution using the ML method are based on maximizing equation (3). The procedures may employ a Golden Section Search (GSS) (Mathews, 1992; Press et al., 1992; Polak, 1997; Rheinboldt, 1998) for direct maximization of equation (3) or Newton-Raphson (NR) (Mathews, 1992; Press et al., 1992; Polak, 1997; Rheinboldt, 1998) by solving the partial derivative equations with respect to each unknown parameter. The parameters estimation equations for the ML method are

$$\frac{\partial L}{\partial a} = \frac{\sum_{i=1}^{n} \ln(1 - \frac{a(x_i-c)}{b}}{a^2} + \frac{(1-a)\sum_{i=1}^{n} \frac{x_i-c}{b\,(1 - \frac{a(x_i-c)}{b})}}{a} = 0 \tag{4}$$

$$\frac{\partial L}{\partial a} = \frac{n}{b} - \frac{(1-a)\sum_{i=1}^{n} \frac{a(x_i-c)}{b^2\,(1 - \frac{a(x_i-c)}{b})}}{a} = 0 \tag{5}$$

Since equation (3) is unbounded with respect to c (Singh and Guo, 1995; Singh, 1998), a ML estimator cannot be obtained for parameter c. Therefore, the lower sample value of x is used as an estimate of parameter c.

Following Hosking and Wallis (1987), the method of MOM equations for estimating parameters of the GP3 distribution are

$$\bar{x} = c+b/(1+a) \tag{6}$$

$$S^2 = b^2/[(1+2a)(1+a)^2] \tag{7}$$

$$C_s = 2(1-a)(1+2a)^{1/2}/(1+3a) \quad (8)$$

where $\bar{x}$, S^2 and C_s are the mean, variance and skewness of the sample, respectively.

The PWM method estimators of the GP3 distribution are

$$a = \frac{nl_1 + 2l_2(n-1)}{l_2(n-1) - l_1} \quad (9)$$

$$b = (1+a)(2+a)l_2 \quad (10)$$

$$c = x_1 - b/(n+a) \quad (11)$$

where $l_1 = M_0 - x_1$, with x_1 being the smallest element of the sample series, and $l_2 = M_0 - 2M_1$. The probability weighted moment values (M_0, M_1) of the sample are calculated using

$$M_k = \sum_{i=1}^{n} (1-P_i)^k x_i/n, \quad k = 0, 1 \quad (12)$$

where $P_i = (i-0.35)/n$ is the plotting position (Hosking and Wallis, 1987).

As given by Singh (1998) and Singh and Guo (1995), the entropy function of the GP3 distribution is

$$H(f) = -\ln a - \ln\frac{1}{a} + \ln b + n\left(1 - \frac{1}{a}\right)E\left[\ln\left(1 - \frac{a(x-c)}{b}\right)\right] \quad (13)$$

The parameter estimation equations for the POME method are

$$E\left[\ln\left(1 - a\frac{x-c}{b}\right)\right] = -a \quad (14)$$

$$E\left[\frac{1}{1-a(x-c)/b}\right] = \frac{1}{1-a} \quad (15)$$

$$\mathrm{Var}\left[\ln\left(1 - a\frac{x-c}{b}\right)\right] = a^2 \quad (16)$$

where E[.] is the mean of the bracketed quantity and Var[.] is the variance of the bracketed quantity.

Equations for estimating the parameters of the GP3 distribution using the LS method are

$$\begin{aligned} &a^2[\bar{x} z_1 \overline{zy} - \bar{x}\overline{z^2 y} - \overline{xz} z_1 \overline{zy} + \overline{xz}\,\overline{z^2 y} - \\ &\bar{z} x_1 \overline{z^2 y} + x_1 \overline{z^2 y} + \overline{z^2 x_1}\,\overline{zy} - \\ &\bar{z} x_1 \overline{zy}] - \overline{xyz}(\overline{z^2} - \bar{z} - z_1\bar{z} + z_1) = 0 \end{aligned} \quad (17)$$

$$b = \frac{a(\bar{x} - \overline{xz}) + a x_1(\bar{z}-1)}{(\overline{z^2} - \bar{z}) - z_1(\bar{z}-1)} \quad (18)$$

$$c = x_1 - (a/c)[1 - z_1] \quad (19)$$

where

$$\bar{x} = n^{-1}\sum_{i=1}^{n} x_i \quad \bar{z} = n^{-1}\sum_{i=1}^{n} z_i \quad \overline{z^2} = n^{-1}\sum_{i=1}^{n} z_i^2$$

$$\overline{xz} = n^{-1}\sum_{i=1}^{n} x_i z_i \quad \overline{zy} = n^{-1}\sum_{i=1}^{n} z_i y_i$$

$$\overline{z^2 y} = n^{-1}\sum_{i=1}^{n} z_i^2 y_i \quad \overline{xyz} = n^{-1}\sum_{i=1}^{n} x_t y_i z_i$$

$$z_i = (1-f_i)^a \quad y_i = \ln(1-f_i)$$

and fi is the Cunnane plotting position formula (Cunnane, 1978) given as (Moharram et al., 1993):

$$f_i = (i-0.4)/(n+0.2) \quad (20)$$

where i is the rank number of the i'th element in the sample series arranged in ascending order.

Materials and Methods

In order to assess the performances of the parameter estimation methods such as MOM, PWM, ML, LS, and POME for the GP3 distribution, 50 sets of observed annual maximum discharge series for the unregulated rivers in Anatolia (Turkey) (Haktanir, 1992), the Spey at Kinrara and the Tay at Pitnacree (Scotland) (Ahmad et al., 1988a, 1988b), the Potomac at Rocks (USA) (Smith, 1987), the St. Marys at Stillwater (Canada) (Kite, 1978), and the Harricana at Amos (Canada) (Bobee and Ashkar, 1991) were used in this study. The data from the rivers in Anatolia were updated until the year 2000.

This study assumed the c parameter to be the smallest value as used by Moharram et al. (1993), Singh (1998), and Singh and Guo (1995). To determine the parameters of a and b, the NR procedure as suggested by van

Montfort and Witter (1985) was used to solve equations (4) and (5) with their partial derivatives. When $C_s < 1.0$, $a = 0.5$ and $b = \overline{x} + x_1$; otherwise $a = 1.0 \times 10^{-7}$ and $b = \overline{x} - x_1$ were used for the ML initial estimates. For convergence criteria, decreases in both absolute values of ($\partial L/\partial a$) plus ($\partial L/\partial b$) and increments/parameters were expected. As an alternative to the NR, the GSS procedure was also employed for the direct maximization of equation (3) with the $c \leq x_1$ constraint.

For the MOM estimate, equation (8) was solved for parameter a by employing the bisection procedure (Mathews, 1992; Press et al., 1992), and then equations (7) and (6) were used to determine parameters b and c.

After determining the probability weighted moment values with equation (12), estimating the parameters with the PWM method is straightforward just using equations (9), (10), and (11).

The parameters of GP3 by the POME were estimated by the NR procedure using equations (14), (15), and (16) with the first and second partial derivatives of those being used. When a non-convergence criterion was encountered in the NR procedure, the process was continued with a direct maximization of equation (13).

The parameters by the LS method were estimated by solving parameter a in equation (17) with the NR procedure first, and then parameters b and c were estimated by equations (18) and (19).

A fortran program was developed to determine the parameters by following all the steps mentioned above. In order to obtain accurate results, the loops employed in the program were terminated when the ratio of increment/parameter was less than 10^{-5}.

Quantiles for the non-exceedence cumulative probabilities of the Cunnane formula (equation (20)), were predicted by equations (1) and (2), after solving these equations for x.

The agreements between predicted quantiles and observed annual flood values for the non-exceedence cumulative probabilities of the Cunnane formula were quantified by computing the average deviation (AD), mean residual error (MRE), average relative percent error (ARPE), and coefficient of efficiency (CE) as follows:

$$AD = \frac{\sum_{i=1}^{n} |X_i - O_i|}{n} \tag{21}$$

$$MRE = \frac{\sum_{i=1}^{n} (X_i - O_i)}{n} \tag{22}$$

$$ARPE = 100 \times \frac{\sum_{i=1}^{n} (X_i - O_i)}{\sum_{i=1}^{n} (O_i)} \tag{23}$$

$$CE = \frac{\sum_{i=1}^{n} (O_i - \overline{O})^2 - \sum_{i=1}^{n} (X_i - O_i)^2}{\sum_{i=1}^{n} (O_i - \overline{O})^2} \tag{24}$$

where X_i is predicted quantile ($m^3\ s^{-1}$), O_i is observed annual peak discharge ($m^3\ s^{-1}$), and $\overline{O}$ is mean observed annual discharge ($m^3\ s^{-1}$). The MRE ($m^3\ s^{-1}$) gives information as to whether the method is over- or under-predicting; the ARPE (%) expresses this on a percentage basis, while the AD ($m^3\ s^{-1}$) is an indicator of quantitative dispersion between predicted and observed values. The CE evaluates the error relative to the natural variation in the observed values. A CE value of 1.0 represents a perfect prediction, while a value of zero represents a prediction that is not better than the random variation in the observed data. Increasingly negative values of CE indicate increasingly poorer predictions.

Results and Discussion

The predicted parameters and calculated AD, MRE, ARPE, and CE statistics for 8 selected station data sets and for each parameter estimation method are given in Table 1. Five of these stations are located outside Turkey and it is easy to obtain their annual peak discharges through the relevant references for further studies and comparison purposes. The other 3 stations, located in Turkey, are randomly selected to represent the other stations. Most of these parameters are obtained with an uncertainty less than 10^{-5}. The NR procedure used for the LS is sensitive to the initial estimate value of parameter c as stated by Moharram et al. (1993). The use of initial estimates $c = 0.3$ when $C_s > 1$, and $c = 0.6$ when $C_s < 1$, gave the best results for almost all the sets.

Table 1. Estimated values of GP3 distribution parameters by POME, ML, LS, PWM and MOM+ methods and statistical comparison for 8 selected rivers.

Station (RL)[C_s]{K}*	Method	Parameters a	b	c	AD ($m^3 s^{-1}$)	MRE ($m^3 s^{-1}$)	ARPE (%)	CE#
Spey	POME	-0.0237	63.10	80.68	6.69	-1.51	-1.04	0.974
(31)	ML	-0.0265	62.89	80.70	6.64	-1.52	-1.05	0.975
[2.01]	LS	-0.00002	68.84	80.66	7.48	2.72	1.87	0.975
{4.75}	PWM	-0.0191	65.44	78.59	6.67	-1.53	-1.05	0.977
	MOM	-0.0019	67.00	78.18	6.97	-1.45	-1.00	0.975
Tay	POME	0.0863	126.22	225.21	22.21	-2.16	-0.63	0.932
(31)	ML#	0.4095	184.71	215.30	20.19	3.95	1.16	0.950
[0.73]	LS	0.0001	123.01	215.22	22.25	-6.14	-1.80	0.907
{-0.85}	PWM	0.0463	136.91	210.89	20.82	-2.37	-0.69	0.917
	MOM	0.4014	211.74	190.64	17.18	-0.78	-0.23	0.966
Potomac	POME	0.1063	2962.5	786.42	511.75	-57.89	-1.65	0.925
(92)	ML	0.1468	3101.1	787.19	513.50	-28.94	-0.83	0.919
[2.28]	LS	-0.53×10^{-5}	2499.0	787.01	505.00	-242.76	-6.92	0.935
{6.84}	PWM	0.6349	4527.9	738.31	570.74	-1.86	-0.05	0.783
	MOM	-0.0422	1999.5	1420.2	303.89	-20.99	-0.60	0.969
St. Marys	POME	0.1830	219.90	189.49	40.65	-35.09	-8.57	0.904
(59)	ML#	0.2969	278.72	189.61	32.07	-5.67	-1.38	0.926
[1.42]	LS	-0.2×10^{-4}	188.09	189.58	50.05	-33.98	-8.30	0.863
{3.20}	PWM	0.8620	422.25	182.56	34.76	-0.07	-0.02	0.857
	MOM	0.1267	186.48	243.82	19.08	-1.29	-0.30	0.971
Harricana	POME	0.1664	90.02	106.92	19.16	-7.65	-4.00	0.767
(69)	ML#	0.5463	136.25	98.80	16.65	-4.52	-2.36	0.836
[0.861]	LS	0.9964	179.51	98.78	14.93	-2.62	-1.37	0.852
{1.35}	PWM	1.6154	251.28	95.24	16.13	0.05	0.03	0.785
	MOM	0.3334	82.56	129.40	7.88	-0.18	-0.10	0.953
Dolluk	POME	0.0273	641.54	130.80	74.75	-8.62	-1.14	0.969
(61)	ML	0.0697	669.29	131.00	72.67	-6.22	-0.82	0.962
[2.22]	LS	-0.00001	616.76	130.90	105.57	46.16	4.79	0.966
{6.90}	PWM	0.2598	805.43	117.85	86.07	-1.94	-0.20	0.959
	MOM	-0.0338	537.70	200.65	52.53	-7.65	-1.01	0.977
Selçuk	POME	-0.0001	146.49	5.09	31.03	-3.02	-1.98	0.903
(44)	ML	0.1256	164.85	5.09	28.78	-2.22	-1.46	0.893
[2.54]	LS	-0.0001	137.97	5.05	30.34	-11.48	-7.54	0.913
{9.04}	PWM	0.6612	254.02	-0.60	31.61	-0.20	-0.13	0.749
	MOM	-0.0731	103.65	40.49	17.98	-2.35	-1.54	0.947
Kayırlı	POME	0.2168	212.25	36.54	41.58	-19.99	-8.69	0.816
(61)	ML	0.5755	322.67	16.40	35.87	-9.13	-3.97	0.867
[0.58]	LS	0.9978	419.59	16.33	29.61	-3.71	-1.61	0.896
{0.85}	PWM	1.5252	562.24	7.41	28.44	0.13	0.06	0.857
	MOM	0.4892	237.82	70.37	22.10	-0.30	-0.13	0.945

AD: absolute deviation; MRE: mean residual error; ARPE: average relative percent error; CE: coefficient of efficiency.

+ POME: principle of maximum entropy; ML: maximum likelihood; LS: least square; PWM: probability-weighted moments; MOM: moments.

* RL: record length; C_s: skewness; K: kurtosis.

The estimated parameters for the Spey and Tay rivers by the ML, MOM, PWM, and LS are almost the same as those obtained by Moharram et al. (1993). Minor improvements with the parameters in Table 1 for AD, MRE, ARPE, and CE values were obtained.

For 23 of the 50 sets, the NR procedure for the ML method did not converge, and therefore the GSS procedure was used for these sets. For all the 23 sets, parameter c obtained by the ML method was equal to the smallest value of the sets. This confirms the statement that the likelihood function (equation (3)) is maximum with respect to c when $c = x_1$ made by Singh and Guo (1995). For each data set, parameter c obtained by the PWM and LS methods is equal to or less than the smallest observed annual peak discharge. However, the same results were not found for parameter c estimated by the POME and MOM. For 42 of the 50 sets, the highest c values were obtained by the MOM such as for the last 5 stations listed in Table 1.

For 44 and 45 sets out of the 50, the minimum AD and maximum CE values were obtained with the MOM such as for the last 7 stations listed in Table 1. The MOM was followed by the ML method. Overall, the LS method produced 18 times the highest AD values in 50 sets such as for the stations on the Spey, Tay, St. Marys and Dolluk in Table 1. The POME and PWM parameter estimation methods yielded similar AD results. For all the sets, the MOM and POME methods produced negative MRE and ARPE values. This means that overall the POME and MOM under-predicted the non-exceedence frequencies' corresponding quantile values. For 44, 36, and 29 of the 50 sets, under-prediction was also observed by the ML, PWM, and LS methods, respectively.

To further evaluate the goodness of fit of each method, the observed and computed frequency curves such as Figures 1-4 for the Spey, Tay and Potomac rivers, and Dolluk station were plotted for each data set. Figures 1 and 2 are similar to Figures 1 and 2 in Moharram et al. (1993). In these figures, the observed and computed values were plotted against the corresponding EV1 reduced variates ($-\ln(-\ln F_i)$). From the figures, it was observed that the quantile estimation performances of the methods vary with F. As mentioned before, this result was also stated by Singh and Guo (1995). When $F \leq 0.90$ and $F > 0.90$, the AD, MRE, ARPE, and CE values for the estimation methods and for each station data set were also obtained and are given in Table 2 for the same stations listed in Table 1.

When the figures of the frequency curves are examined for $F \leq 0.90$ ($EV1 \leq 2.25$), it appears that all the parameter estimation methods except for the LS and POME gave comparable results. The LS and POME methods gave evidently lower or higher quantiles than observed annual peak values.

When $F > 0.90$, the number of the calculated quantiles varied between 3 and 9 based on the sizes of the samples, which were between 31 and 92. These numbers may not be large enough to allow a comparison of the performance of the estimation methods. The frequency curves obtained by the estimation methods when $F > 0.90$ are not in as close agreement as when $F \leq 0.90$. This disagreement is especially evident in Figures 2 and 3. The values of the performance indices of AD, MRE, ARPE, and CE were also worse when $F > 0.90$ than when $F \leq 0.90$ (see Table 2). For 24 of the 50, both the least AD and the highest CE values were obtained with the MOM. The ML produced the second, the LS produced the third, the POME produced the fourth, and the PWM produced the fifth lowest AD values. From the best to the worst, CE values were obtained by the following order of the estimation methods: MOM, ML, POME, LS, and PWM. When MRE and ARPE values are considered, in 47 of the 50 sets, the PWM produced under-prediction, in contrast, when $F < 0.90$.

For the Turkish rivers, the order of the estimation methods from the best to the worse changed sometimes from region to region and a similarity in the order of the methods for the stations located in the same region of Turkey was noted.

When the kurtosis values of the samples are negative, the chances of yielding higher quantiles with the LS method are high (see Figure 2 for the river Tay). When both skewness and kurtosis values are less than 1.0 for $F \leq 0.90$, the chances of seeing lower quantiles with the POME are high. This can be seen from the values of MRE (-4.07 and -29.92) and ARPE (-1.28 and -14.52) for the river Tay and station Kayirli in Table 2.

Conclusions

In this paper, the parameter estimation methods of the moments, probability weighted moments, maximum likelihood, principle of maximum entropy, and least squares are compared to estimate the parameters in the

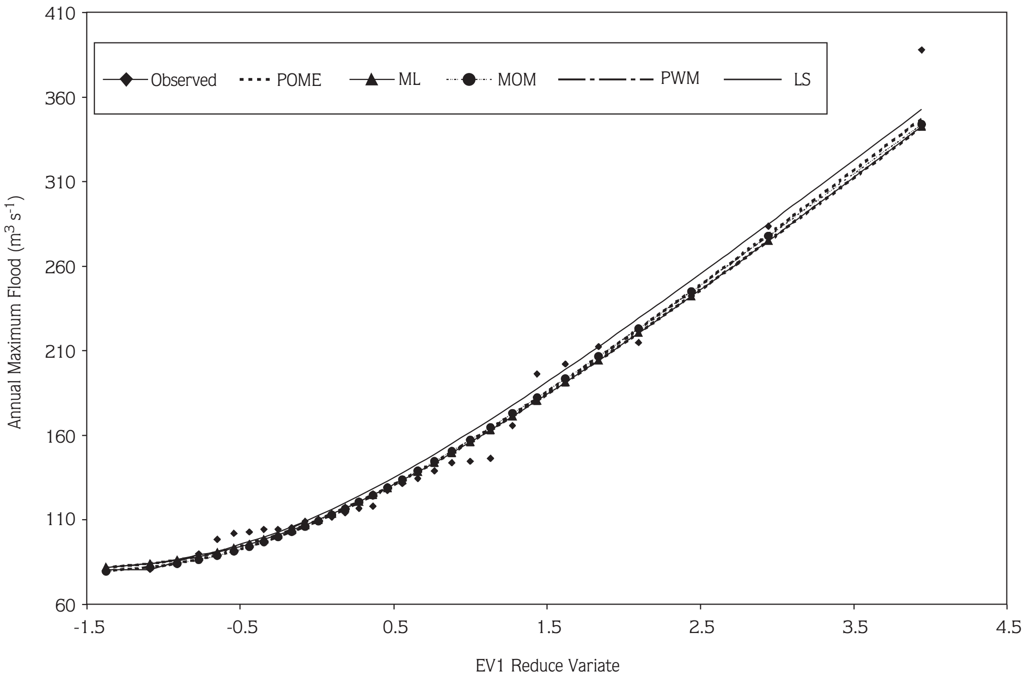

Figure 1. Comparison of the principle of maximum entropy (POME), maximum likelihood (ML), least squares (LS), probability-weighted moments (PWM) and moments (MOM) methods of fitting the GP3 distribution to annual peak discharges for the River Spey at Kinrara, UK (1952-1982).

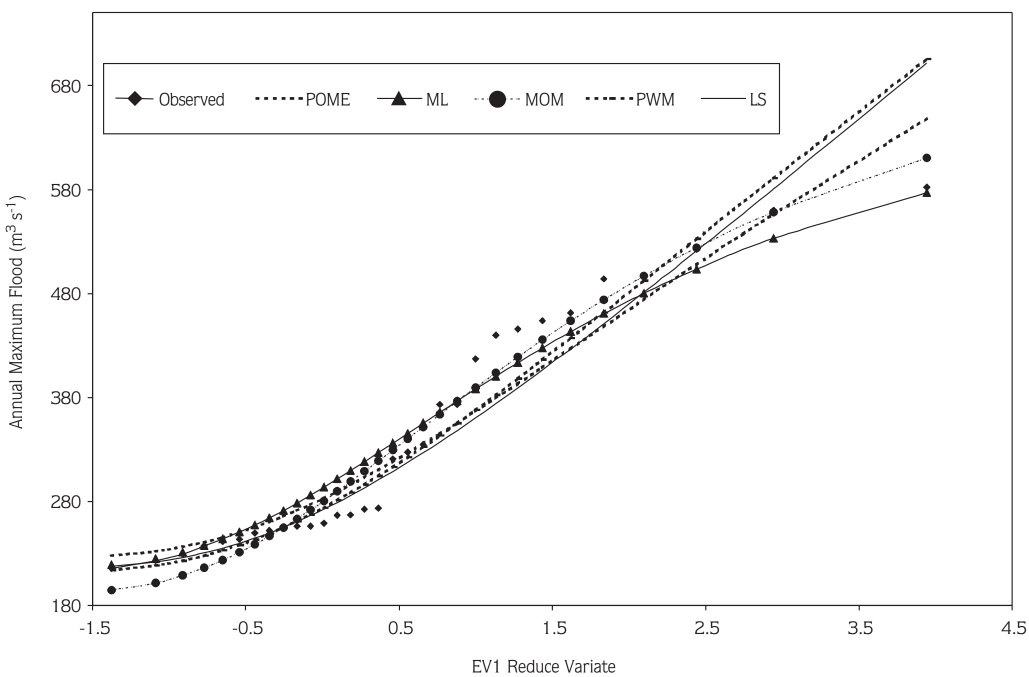

Figure 2. Comparison of the principle of maximum entropy (POME), maximum likelihood (ML), least squares (LS), probability-weighted moments (PWM) and moments (MOM) methods of fitting the GP3 distribution to annual peak discharges for the River Tay at Pitnacree, UK (1952-1982).

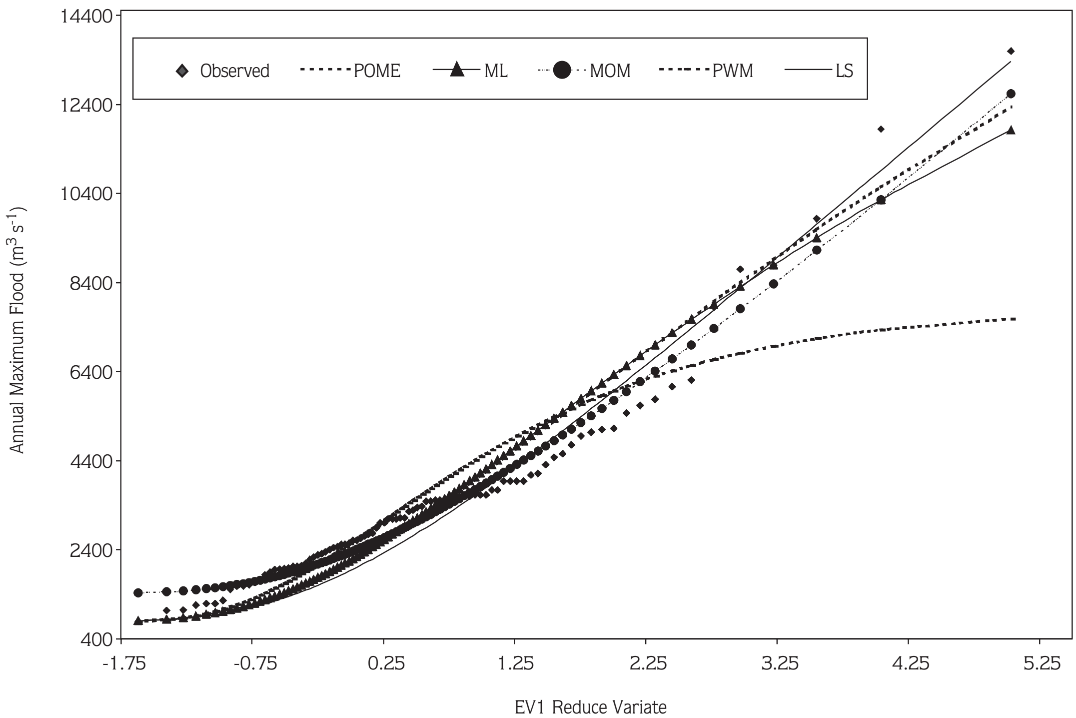

Figure 3. Comparison of the principle of maximum entropy (POME), maximum likelihood (ML), least squares (LS), probability-weighted moments (PWM) and moments (MOM) methods of fitting the GP3 distribution to annual peak discharges for the River Potomac at Rocks, USA (1895-1986).

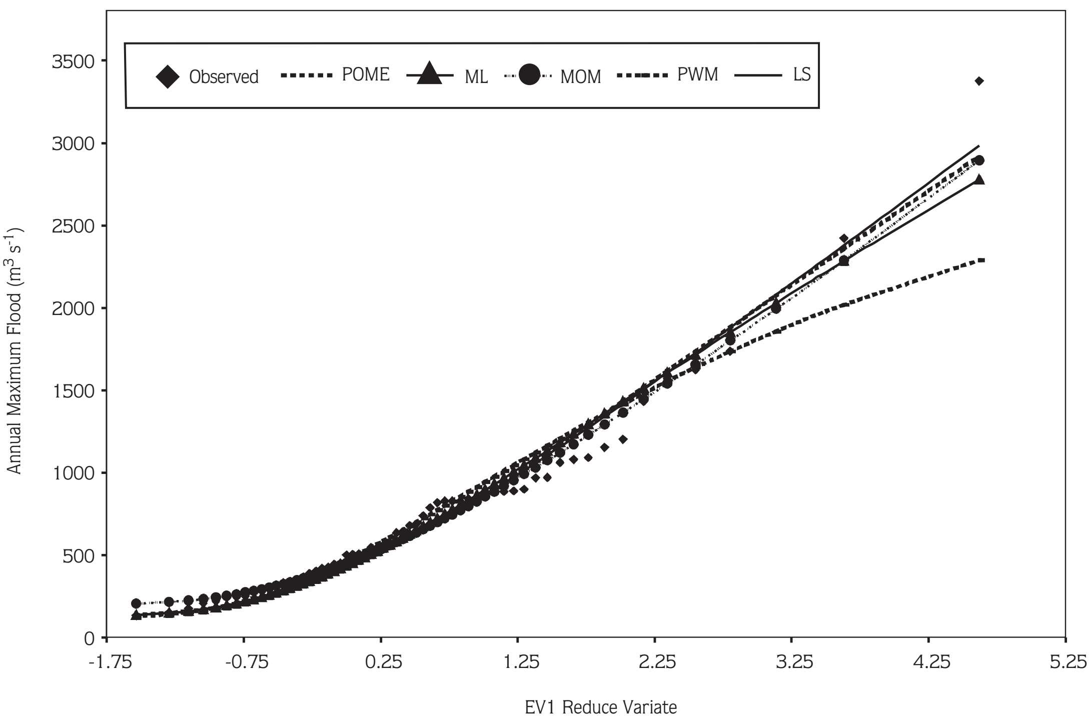

Figure 4. Comparison of the principle of maximum entropy (POME), maximum likelihood (ML), least squares (LS), probability-weighted moments (PWM) and moments (MOM) methods of fitting the GP3 distribution to annual peak discharges for the river M. Kemalpaşa at Dolluk, Turkey (1938-1998).

Table 2. Statistical comparison of quantile estimates of GP3 distribution by POME, ML, LS, PWM and MOM+ methods when $F \leq 0.90$ and $F > 0.90$ for 8 selected rivers.

		F ≤ 0.90				F > 0.90			
Station	Method	AD ($m^3 s^{-1}$)	MRE ($m^3 s^{-1}$)	ARPE (%)	CE	AD ($m^3 s^{-1}$)	MRE ($m^3 s^{-1}$)	ARPE (%)	CE#
Spey	POME	5.45	0.29	0.22	0.97	18.26	-18.26	-6	0.81
	ML	5.43	0.24	0.19	0.97	17.93	-17.93	-5.89	0.81
	LS	6.64	3.88	3.02	0.95	15.33	-8.1	-2.66	0.88
	PWM	5.66	-0.16	-0.13	0.97	16.05	-14.3	-4.7	0.85
	MOM	5.85	0.08	0.06	0.96	17.4	-15.69	-5.15	0.83
Tay	POME	21.57	-4.07	-1.28	0.91	28.22	15.74	2.84	-1.52
	ML	20.53	6.2	1.94	0.92	17.04	-17.04	-3.07	0.38
	LS	18.59	-10.14	-3.18	0.92	52.97	52.97	9.55	-8.41
	PWM	17.18	-8.49	-2.66	0.93	54.73	54.73	9.87	-7.95
	MOM	17.9	-1.86	-0.58	0.95	10.47	9.24	1.67	0.56
Potomac	POME	480.69	-76.33	-2.6	0.77	798.28	112.21	1.28	0.86
	ML	472.48	-28.7	-0.98	0.76	891.79	-31.14	-0.36	0.81
	LS	475.47	401.44	13.65	0.71	2499.7	-2442.1	-27.96	-0.8
	PWM	400.17	201.68	6.86	0.79	2143.7	-1878.9	-21.51	-0.26
	MOM	251.81	12.33	0.42	0.93	784.2	-328.3	-3.76	0.89
St. Marys	POME	39.91	-38.15	-10.2	0.79	47.18	-8.04	-1.12	0.85
	ML	28.11	-10.06	-2.69	0.88	67.06	33.11	4.59	0.73
	LS	48.27	-45.28	-12.1	0.71	65.81	65.81	9.13	0.78
	PWM	26.14	9.66	2.58	0.91	110.92	-86.02	-11.94	-0.13
	MOM	16.03	-2.07	-0.55	0.96	45.98	6.05	0.84	0.86
Harricana	POME	17	-12.84	-7.12	0.67	38.3	38.3	13.26	-0.4
	ML	16.11	-6.4	-3.55	0.7	21.38	12.13	4.2	0.54
	LS	13.62	-0.73	-0.41	0.81	26.57	-19.39	-6.71	-0.05
	PWM	13.49	4.53	2.51	0.8	39.58	-39.58	-13.7	-1.23
	MOM	6.89	-0.2	-0.11	0.93	16.63	0	0	0.74
Dolluk	POME	66.42	-7.15	-1.17	0.93	151.11	-22.14	-1.05	0.89
	ML	62.08	1.21	0.2	0.93	169.75	-74.36	-3.51	0.83
	LS	69.01	-17.82	-2.93	0.93	134.52	-9.8	-0.46	0.92
	PWM	54.61	26.58	4.37	0.93	276.73	-272.25	-12.85	0.43
	MOM	44.66	1.38	0.23	0.96	124.65	-90.41	-4.27	0.89
Selçuk	POME	28.12	-6.49	-5.29	0.7	60.22	31.73	7.09	0.83
	ML	24.97	-1.74	-1.42	0.74	66.89	-6.97	-1.56	0.71
	LS	28.21	-13.06	-10.64	0.74	51.67	4.33	0.97	0.84
	PWM	21.75	11.13	9.07	0.79	130.22	-113.52	-25.35	-0.31
	MOM	14.43	0.03	0.02	0.91	53.4	-26.12	-5.83	0.8
Kayırlı	POME	38.37	-29.92	-14.52	0.74	71.02	71.02	15.79	0.12
	ML	34.44	-13.52	-6.56	0.8	49.05	31.09	6.91	0.49
	LS	25.59	-0.36	-0.17	0.89	66.48	-34.43	-7.65	0.08
	PWM	22.31	8.67	4.21	0.91	84.63	-78.17	-17.38	-0.84
	MOM	19.28	-0.23	-0.11	0.93	47.92	-0.94	-0.21	0.62

AD: absolute deviation; MRE: mean residual error; ARPE: average relative percent error; CE: coefficient of efficiency.

+ POME: principle of maximum entropy; ML: maximum likelihood; LS: least square; PWM: probability-weighted moments; MOM: moments.

generalized Pareto distribution. To do this, annual maximum discharge data for a set of 50 rivers, most of them in Turkey, were used. To measure the adequacy of the methods, the AD, MRE, ARPE, and CE statistics between observed and predicted quantiles corresponding to the non-exceedence cumulative probabilities of Cunnane were calculated.

The following conclusions can be drawn from this study: 1) to prevent the sensitivity of the LS method to the initial estimates of location parameter, the use of a location parameter equal to 0.3 when the skewness of the data set is greater than one, and the use of a location parameter equal to 0.6 when skewness is less than one may give better results; 2) to handle the non-convergence problem when employing the NR procedure for the ML method, the GSS procedure can be an alternative to the NR; 3) in the case of the ML method, the location parameter should be equal to the smallest value of the data set; 4) there is a tendency to estimate high location parameter values with the MOM method, such as 1420.2 for the Potomac river (Table 1); 5) quantile estimation performances of the estimation methods vary with the non-exceedence cumulative frequency values (F) (Table 2); 6) when $F \leq 0.90$, the PWM showed a tendency to over-predict and vice versa when $F > 0.90$ (see Figures 3 and 4 when EV1 < 2.25 and EV1 > 2.25); 7) the MOM method is superior to the other methods and the MOM was followed by the ML method (see Figures 1 and 2 and Table 1), and 8) the POME and LS produced comparable results for entire F and $F > 0.90$, and the POME yielded better results when $F \leq 0.90$.

References

Ahmad, M.I., C.D. Sinclair and A. Werritty. 1988a. Log-logistic flood frequency analysis. J. Hydrol. 98: 205-224.

Ahmad, M.I., C.D. Sinclair and A. Werritty. 1988b. Assessment of flood frequency models using empirical distribution function statistics. Water Resour. Res. 24: 1323-1328.

Barrett, J.H. 1992. An extreme value analysis of the flow of Burbage Brook. Stochastic Hydrol. Hydraul. 6: 151-165.

Bobee, B. and F. Ashkar. 1991. The Gamma Family and Derived Distributions Applied in Hydrology. Water Resources Publications, Colorado, USA.

Cunnane, C. 1978. Unbiased plotting position-a review. J. Hydrol. 37: 205-222.

Haktanir, T. 1992. Comparison of various flood frequency distributions using annual flood peaks data of rivers in Anatolia. J. Hydrol. 136: 1-31.

Hosking, J.R.M. and J.R. Wallis. 1987. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29: 339-349.

Joe, H. 1987. Estimation of quantiles of the maximum of N observations. Biometrika 74: 347-354.

Kite, G.W. 1978. Frequency and Risk Analyses in Hydrology. Water Resources Publications, Colorado, USA.

Mathews, J.H. 1992. Numerical Methods for Mathematics, Science, and Engineering. Prentice-Hall International, Inc., London, 646 p.

Moharram, S.H., A.K. Gosain and P.N. Kapoor. 1993. A comparative study for the estimators of the generalized Pareto distribution. J. Hydrol. 150: 169-185.

Pickands, J. 1975. Statistical inference using extreme order statistics. Ann. Stat. 3: 119-131.

Polak, E. 1997. Optimization Algorithms and Consistent Approximations. Springer, 767 p.

Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery. 1992. Numerical Recipes in Fortran, Cambridge University Press, Cambridge, New York, 963 p.

Prudhomme, C., D. Jakob and C. Svensson. 2003. Uncertainty and climate change impact on the flood regime of small UK catchments. J. Hydrol. 277: 1-23.

Rheinboldt, W.C. 1998. Methods for Solving Systems of Nonlinear Equations, SIAM, PA, USA. 148 p.

Rosbjerg, D., H. Madsen and P.F. Rasmussen. 1992. Prediction in partial duration series with generalized Pareto-distributed exceedances. Water Resour. Res. 28: 3001-3010.

Singh, V.P. 1998. Entropy-Based Parameter Estimation in Hydrology. Kluwer Academic Pub., London. 367 p.

Singh, V.P. and H. Guo. 1995. Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME). Hydrol. Sci. 40: 165-181.

Smith, J.A. 1987. Estimating tails of probability distributions. Ann. Statist. 15: 1174-1207.

Smith, J.A. 1991. Estimating the upper tail of flood frequency distributions. Water Resour. Res. 23: 1657-1666.

Van Montfort, M.A.J. and J.V. Witter. 1985. Testing exponentiality against generalized Pareto distribution. J. Hydrol. 78: 305-315.

Vogel, R.M., T.A. McMahon and F.H.S. Chiew. 1993. Flood flow frequency model selection in Australia. J. Hydrol. 146: 421-449.

Wang, O.J. 1991. The POT model described by the generalized Pareto distribution with Poisson arrival rate. J. Hydrol. 129: 263-280.