Water Resour Manage (2011) 25:1299–1313
DOI 10.1007/s11269-010-9745-2

Estimation of the Parameters of Wakeby Distribution by a Numerical Least Squares Method and Applying it to the Annual Peak Flows of Turkish Rivers

Tekin Öztekin

Received: 28 October 2009 / Accepted: 18 November 2010 /
Published online: 6 January 2011

Abstract In this study, a numerical least squares (NLS) method for estimating the parameters of five-parameter Wakeby distribution was introduced. To asses the right tail estimate performances of the method, Monte Carlo simulated data and annual peak flows of 50 stations on Turkish rivers were used. Its results were compared to those by L-moments (LM) and curve fitting method of MATLAB. The biases from the LM for non-exceedence probability (F) of 0.999 mostly were less than those by the NLS. However, the values of relative root mean square error (*rrmse*) statistics from the NLS were better than those by the LM. In addition, the statistic of average deviation from the observed annual peak flows showed that NLS method exhibited mostly better results than those by LM for right tail predictions. Lastly, except the convergence problem of MATLAB, while both of the NLS and MATLAB produced the same determination coefficient (r^2) for the majority of data set, the NLS produced lower *rrmse* values than MATLAB.

Keywords Annual peak flow · Numerical least squares · Parameter estimation · Wakeby distribution

1 Introduction

The optimal combination of flood protection options is determined to minimize flood damages and construction cost of flood control options along the river (Karamouz et al. 2009). The needed design flood values for decided options especially when the lengths of recorded data are short, may require usage of various statistical distributions. These distributions enable us to predict values having return periods greater than the lengths of the recorded series. Therefore, choice of the distribution most

T. Öztekin (✉)
Agricultural Faculty, Department of Agricultural Engineering,
Gaziosmanpaşa University, Taşlıçiftlik Campus, 60240 Tokat, Turkey
e-mail: toztekin@gop.edu.tr

suitable to the recorded sample series is important from these aspects. Houghton (1978) stated that traditional distributions such as three-parameter log normal inadequately model flood flows for certain records. He introduced the Wakeby distribution to overcome this deficiency. Since the Wakeby distribution has five parameters, it can mimic the shapes of many commonly used skewed distributions. Furthermore, the distribution is very flexible, therefore it may well fit to observed data. Although a better performance is expected from the five-parameter Wakeby distribution (W5) than the three-parameter distributions, sometimes convincing evidence may not be found supporting this expectation as stated by Haktanir and Bozduman (1995). Houghton (1978) is the first researcher to practice this distribution for flood frequency analysis. Landwehr et al. (1979a, b), Kumar and Chander (1987), Rao and Arora (1987), Griffiths (1989), Haktanir (1992), Vogel et al. (1993), Haktanir and Horlacher (1993), Zalina et al. (2002), and Öztekin (2007) are some of the others who employed the W5 for flood frequency and rainfall analyses.

The W5 is defined only with the quantile estimation equation by

$$x(F) = \varepsilon + \frac{\alpha}{\beta}\left[1 - (1-F)^{\beta}\right] - \frac{\gamma}{\delta}\left[1 - (1-F)^{-\delta}\right] \tag{1}$$

where F is non-exceedence probability; $x(F)$ is F corresponding quantile value; ε is location, and α, β, γ, and δ are other parameters. More theoretical information about Wakeby distribution with an application to a design flood problem is given in Griffiths (1989). However, in Griffiths (1989), three-parameter case of Wakeby distribution ($\gamma = 0$ and $\delta = 0$) (equivalent to the three parameter generalized Pareto distribution) was employed.

In spite of its wide range of distributional shapes, several studies reported that the W5 could not fit to all data sets because there were some L-moment values that no W5 attains. Furthermore, although moments of x can be attained as functions of the parameters (Greenwood et al. 1979), the inverse relationship cannot be readily derived. As a result, either moments or maximum-likelihood estimates of parameters are not feasible or are not easily obtained (Rao and Hamed 2000). There is no study so far employing least squares (LS) method for estimating the parameters of W5. Therefore, estimating the parameters of W5 by the NLS is an original study.

The goals of this study are i) to present the procedures for estimating the parameters of W5 by a numerical least squares method, and ii) to compare the fitting potentials of this method to those of both LM and curve fitting method of MATLAB (The MathWorks Inc. 2005).

2 Materials and Methods

2.1 Parameter Estimation

In this study, the parameters of W5 were estimated using the LM method (Hosking and Wallis 1997; Hosking 2000), the curve fitting method of MATLAB, and a new alternative method based on NLS method. The LM method (which is equivalent to the probability weighted moment in principles) is the only one presently used to fit the W5 distribution. In this method, the values of the quantile ($x(F)$) and of the cumulative distribution function, cdf, ($F(x)$) are obtained by using "quawak" and "cdfwak" programs by Hosking (2000), respectively. The "quawak" program uses

Eq. 1. In the "cdfwak", F values are obtained solving Eq. 1 by employing the Halley's method which is the second-order analogue of Newton-Raphson method.

The LS method can be applicable for the W5 because the W5 has explicit quantile function (Eq. 1). In this study, the parameters by the NLS were obtained by numerically minimizing the sum of squares of residuals:

$$SSR = R = \sum_{i=1}^{n} [x_i - x(F)]^2$$
$$= \sum_{i=1}^{n} \left[x_i - \left\{ \varepsilon + \frac{\alpha}{\beta} \left\{ 1 - (1-F)^{\beta} \right\} - \frac{\gamma}{\delta} \left\{ 1 - (1-F)^{-\delta} \right\} \right\} \right]^2 \tag{2}$$

The NLS procedure for W5 starts by determining the non-exceeding cumulative probability values, F [0–1], with the plotting position formula, $(i - 0.35)/n$, for the sample series (sized n) arranged in ascending order. A small flowchart figure of the used Fortran subroutine for parameter estimation by NLS was given in Fig. 1. The downhill simplex method of Nelder and Mead for multi dimensions (Press et al. 1992) is employed to minimize SSR. This method requires only function evaluations, not derivatives. The program "amoeba" by Press et al. (1992) was modified and used

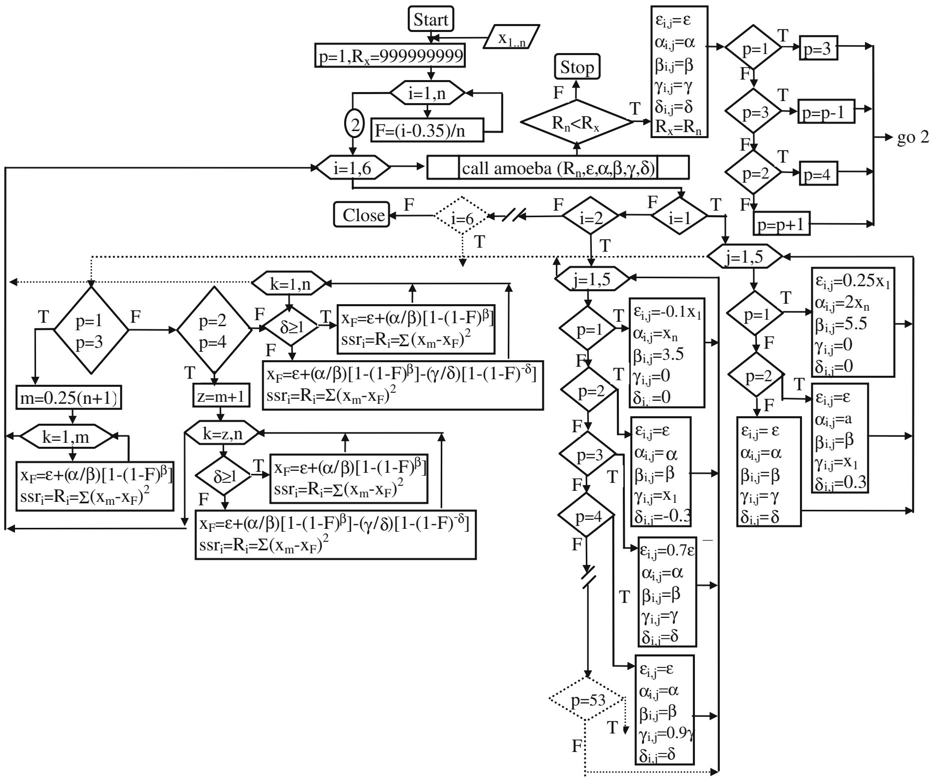

Fig. 1 A small flowchart of the used Fortran subroutine for parameter estimation by NLS

for this minimization. In the first phases of the algorithm, similar to Houghton (1978), the right-hand side of W5 (Eq. 1) was thought of in two parts: the left-hand side, $\varepsilon + \alpha/\beta\left\{1-(1-F)^{\beta}\right\}$ when $F \leq 0.25$; and the right-hand side, $-\gamma/\delta\left\{1-(1-F)^{-\delta}\right\}$ when $F > 0.25$. Furthermore, it is believed that the properties of both parts of W5 reflect the nature accurately since W5 can mimic the separation effect noted in nature (Houghton 1978) and when $\delta > 0$, the W5 has a heavy upper tail and can therefore give rise to data sets containing occasional high outliers, a phenomenon often observed in environmental sciences (Hosking and Wallis 1997). The procedure for estimating the parameters of W5 by the NLS consists of 53 phases.[1] In each phase, in five dimensions (parameters) forming the columns of a matrix $P(6,5)$, of $5+1$ points or locations forming rows of a matrix $P(6,5)$ having different initial values of parameters are searched until the algorithm encounters a (local, at least) minimum. The first five phases of the algorithm for estimating W5 parameters by the NLS are defined in Table 1. The phases 6 through 53 are similar to phase 5. In these phases, possible increases (5%), decreases (5%), or holding the same values of the parameters obtained in the previous phases are considered. At the end of phase 53, the algorithm goes back to phase 5, and this loop repeats until no further minimization is obtained by Eq. 2. The algorithm can be thought as a complex, however, it is just a description of the downhill simplex method of Nelder and Mead for multi dimensions and iterative process of phase 5 with different initial values. The random variables corresponding to non-exceedence cumulative probabilities (F) are obtained by solving Eq. 1 for the F term by the bisection method.

To measure the efficiency of developed NLS algorithm, the curve fitting efficiencies of W5 with its parameters estimated by NLS and MATLAB were also compared. For this aim, the curve fitting toolbox of MATLAB (Version 7.1) was employed to fit Eq. 1 to the annual peak flow series of Turkish rivers. The statistics of *rrmse* and r^2 between observed and predicted annual peak flows by the NLS and MATLAB were calculated. The NLS and curve fitting options with the method of nonlinear least squares in modern optimization softwares such as MATLAB are based on different theories. While the NLS is based on minimization of summed squares of residuals (Eq. 2) by modification of the downhill simplex method of Nelder and Mead, the MATLAB is based on minimization of summed squares of residuals by an iterative process involving initial estimates of parameters; production of the fitted curve for the initial estimates of parameters; calculation of the Jacobian matrix; adjustment of the parameters with one of the algorithms of Trust-region, Levenberg-Marquardt or Gauss-Newton (Press et al. 1992) and determine whether the fit improve; iterate the process by returning to the production of the fitted curve until the fit reaches the specified convergence criteria (The MathWorks Inc.).

2.2 Data

To assess the performance of the algorithm in terms of its efficiency in obtaining acceptable solutions and resulting estimates of the parameters and quantiles by the parameter estimation method of NLS for W5, Monte Carlo sampling experiments

[1]The Fortran source code of W5 by NLS can be obtained from the corresponding author.

Table 1 Algorithm for estimating the parameters of W5 by NLS method

Phase	Assume	Procedure
1	$P_1^a, \gamma = 0, \delta = 0$	Minimization only left-hand side of Eq. 1 $\left[\varepsilon + \frac{\alpha}{\beta}\left(1 - (1 - F)^{\beta}\right)\right]$ with respect to the sum of squares of residuals; as initial estimates of parameters employ values in matrix P_1; determine the elements of column vector R_1 (the values of sum of squares of residuals for the parameters given in the rows of P_1 matrix) for each set of parameters given in the rows of P_1 by considering the ordered observed values (x_i) located in the 1th quarter ($F \leq 0.25$); call "amoeba" which returns with minimum R (assign to M_1, $M_1 = R$) and parameters producing this M_1; go to phase 3 (see flowchart given in Fig. 1 when $p = 1$)
2	P_2^a	Minimize Eq. 1 by considering only x_i values in the 2nd, 3rd, and 4th quarters ($F > 0.25$); as initial estimates of parameters γ and δ employ values in the 4th and 5th columns of P_2 and holding ε, α, and β in the 1th, 2nd, and 3rd columns of P_2, respectively; determine the elements of column vector R_2 for each set of parameters given in the rows of P_2; call "amoeba" which returns with M_2 and the values of parameters γ and δ producing this M_2; go to phase 4
3	$P_3^a, \gamma = 0, \delta = 0$	Forming rows of initial matrix P_3 by holding-, increasing (50%)-, or decreasing (30%)-the obtained values of parameters β, α, and ε as shown in matrix P_3; determine the elements of column vector R_3 for each set of parameters given in the rows of P_3 by considering the x_i values located in the 1th quarter ($F \leq 0.25$); call "amoeba" which returns with min R_3 and parameters producing M_3; if determined $M_3 < M_1$ repeat this phase with new increased or decreased values of β, α, and ε until to reach the lowest M_3; go to phase 2
4	P_4^a	Forming rows of initial matrix P_4 by increasing (10%)- or decreasing (10%)-the obtained values of parameters γ and δ (also let parameter δ to take opposite sign) and holding β, α, and ε in P_4; determine the elements of column vector R_4 for each set of parameters given in the rows of P_4 by considering the x_i values located in the 2nd, 3rd and 4th quarters ($F > 0.25$); call "amoeba" which returns with M_4 and the values of parameters γ and δ producing this M_4; if determined $M_4 < M_2$ repeat this phase with new increased or decreased values of γ and δ until to reach the lowest M_4; go to phase 5
5	P_5^a	Minimize Eq. 1 by considering whole x_i values; forming the first row of P_5 with the values of β, α, and ε determined in phase 3 and the values of γ and δ determined in phase 4, and then forming other rows by increasing (5%)- or decreasing (5%)-the values in the first row; determine the elements of column vector R_5 for each set of parameters given in the rows of P_5; call "amoeba" which returns with M_5 and the values of parameters β, α, ε, γ and δ producing this M_5; if determined $M_5 < M_4$ repeat this phase with new increased or decreased values of parameters until to reach the lowest M_5; go to phase 6

[a]The matrices are given in Appendix

were conducted to generate data from W5. Generating data from a different distribution other than W5, for example the generalized extreme value distribution, but using the W5 to estimate the parameters and quantiles can be another research subject. Six flood population cases of W5 employed by Landwehr et al. (1979a, b) were used in this study. These cases are listed in Table 2. The relationships between the symbols of W5 quantile equations used by Landwehr et al. (1979a, b) and Eq. 1 are: $\varepsilon = m$, $\beta = b$, $\delta = d$, then $\alpha = \beta a$, and then $\gamma = \delta c$. Landwehr et al. (1979a, b) set m to zero. For each population case, 1,000 random samples of size 30, 60, and 90 were generated by using the pseudo random number (0–1) generator program "RAN3" (Press et al. 1992). These random numbers corresponding to the F values were inserted in Eq. 1 to determine quantile values. For accepted 1,000 samples with the sizes of 30, 60, and 90, the performances of the parameter estimators (LM and NLS) to estimate right tail quantiles for $F = 0.5$, 0.8, 0.9, 0.99, and 0.999 were also compared. These F values are corresponding to the 2, 5, 10, 100, and 1,000 year return period events, respectively.

To further evaluate the NLS for observed field data, 50 annual peak flow series measured by the General Directorate of Electrical Works Planning of the Government of Turkey (EIEI) (Table 3) at Turkish rivers were fit to the distribution of W5.

2.3 Performance Indices

For Monte Carlo generated data, the performances of the parameter estimation methods of LM and NLS for the W5 were measured using the performance indices of bias (*bias*) and relative root mean square error (*rrmse*):

$$bias = \frac{E(\hat{y}) - y}{y} \tag{3}$$

$$rrmse = \frac{E\left[(\hat{y} - y)^2\right]^{0.5}}{y} \tag{4}$$

where $\hat{y}$ is an estimate of y (parameter or quantile) and:

$$E(\hat{y}) = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i \tag{5}$$

where N is the number of Monte Carlo samples ($N = 1{,}000$ in this study).

Table 2 W5 population cases considered in the Monte Carlo sampling experiment

Case	Parameters				Population statistics				
	α	β	γ	δ	μ	σ	C_v	C_s	C_k
1	16.0	16.0	0.8	0.20	1.94	1.34	0.69	4.14	63.74
2	7.5	7.5	0.6	0.12	1.56	0.90	0.58	2.00	14.08
3	1.0	1.0	0.6	0.12	1.18	1.03	0.87	1.91	10.73
4	16.0	16.0	0.4	0.04	1.36	0.51	0.38	1.10	7.69
5	1.0	1.0	0.4	0.04	0.92	0.70	0.76	1.11	4.73
6	2.5	2.5	0.2	0.02	0.92	0.46	0.50	0.00	2.65

μ mean, σ standard deviation, C_v coefficient of variation, C_s coefficient of skew, C_k coefficient of kurtosis

Table 3 Some characteristics of gauging stations operated by EIEI and whose annual peak flows analyzed in this study

Station no	Station name	Stream name	Basin name	A (km^2)	H (m)	$Q_{ort.}$ (m^3/s)	T (year)
302	Döllük	M. Kemalpaşa	Susurluk	9,629	40	58.5	61 (1938–1998)
311	Küçükilet	Orhaneli	Susurluk	1,621	795	6.3	53 (1945–1998)[b]
314	Kayaca	Koca	Susurluk	2,308	25	19.5	45 (1953–1998)
316	Yahyaköy	Simav	Susurluk	6,454	32	45.3	46 (1953–1998)
321	Geçitköy	Nilüfer	Susurluk	1,290	63	15.6	45 (1954–1998)
324	Balıklı	Atnos	Susurluk	1,384	94	9.2	43 (1954–1998)
601	Selçuk	K. Menderes	K. Menderes	3,255	4	10.4	44 (1953–1998)[b]
701	Kayırlı	Çine	B. Menderes	948	262	6.6	61 (1938–1998)
706	Aydın Köp.	B. Menderes	B. Menderes	19,595	25	58.2	49 (1950–1998)
712	Burhaniye	B. Menderes	B. Menderes	12,798	120	38.0	48 (1951–1998)
713	Çitak Köp.	B. Menderes	B. Menderes	3,945	802	12.8	47 (1952–1998)
809	Kavaklıdere	Esen	Batı Akdeniz	546	1,115	3.7	42 (1957–1998)
901	Homa	Manavgat	Antalya	928	35	–[a]	44 (1941–1984)
902	Beşkonak	Köprüçay	Orta Akdeniz	1,942	116	84.5	59 (1940–1998)
1203	Beşdeğirmen	Porsuk	Sakarya	3,938	855	8.3	61 (1936–1998)
1219	Yağbasan	Dinsiz	Sakarya	410	25	7.0	44 (1953–1998)[b]
1221	Doğançay	Sakarya	Sakarya	52,531	41	–[a]	37 (1953–1990)
1222	Rüstümköy	Kocasu	Sakarya	2,021	198	18.2	46 (1953–1998)
1223	Hamidiye	Seydi	Sakarya	1,608	895	1.8	44 (1953–1997)[b]
1237	Dokurcun	Mudurnu	Sakarya	1,072	286	8.0	43 (1956–1998)
1302	Yakabaşı	B. Melen	B. Karadeniz	1,988	115	38.2	46 (1953–1998)
1401	Fatlı	Kelkit	Yeşilırmak	10,048	375	70.5	61 (1938–1998)
1402	Kale	Yeşilırmak	Yeşilırmak	33,904	190	158.0	57 (1939–1998)[b]
1413	Durucasu	Yeşilırmak	Yeşilırmak	21,667	301	64.5	44 (1955–1998)
1414	Sütlüce	Yeşilırmak	Yeşilırmak	5,409	510	26.3	44 (1955–1998)
1501	Yamula	Kızılırmak	Kızılırmak	15,581	995	–[a]	52 (1939–1990)
1503	Yahşihan	Kızılırmak	Kızılırmak	30,186	673	–[a]	47 (1939–1990)[b]
1508	Kaleboğazı	Kanak	Kızılırmak	2,918	960	6.6	38 (1952–1990)[b]
1517	Şefaatli	Karanlık	Kızılırmak	8,592	895	12.3	45 (1953–1998)[b]
1524	Kuyluş	Gökırmak	Kızılırmak	4,192	475	16.4	44 (1954–1998)[b]
1801	Himmetli	Göksu	Seyhan	2,596	665	30.2	63 (1936–1998)
1805	Gökdere	Göksu	Seyhan	4,242	312	59.8	58 (1939–1998)[b]
1806	Ergenuşağı	Zamanti	Seyhan	8,737	590	66.4	41 (1939–1998)[b]
1905	Torun Köp.	Karasu	Hatay	1,768	84	11.2	45 (1954–1998)
1906	Müşrüflü	Afrin	Hatay	2,764	98	9.0	44 (1954–1998)[b]
1907	Demirköprü	Asi	Hatay	16,170	85	33.2	49 (1949–1998)[b]
1908	Antakya	Asi	Hatay	22,624	73	63.2	47 (1949–1998)[b]
2001	Kılavuzlu	Ceyhan	Ceyhan	8,484	435	82.1	52 (1940–1991)
2005	Akçil	Ceyhan	Ceyhan	4,219	1,115	26.5	35 (1953–1990)[b]
2006	Karaahmet	Göksu	Ceyhan	739	1,324	9.1	45 (1954–1998)
2009	Poskoflu	Göksun	Ceyhan	1,387	1,040	12.3	44 (1954–1998)[b]
2015	Tanır	Hurman	Ceyhan	915	1,180	8.6	42 (1957–1998)
2119	Kemah boğ.	Fırat	Fırat	10,359	1,123	93.2	38 (1954–1998)[b]
2131	Kılayık	Beyderesi	Fırat	277	925	1.7	42 (1957–1998)
2201	Kürtün	Harşit	D. Karadeniz	2,750	480	–[a]	40 (1943–1988)[b]
2206	Kanlıpelit	Değirmendere	D. Karadeniz	708	263	11.2	40 (1951–1990)

Table 3 (continued)

Station no	Station name	Stream name	Basin name	A (km^2)	H (m)	$Q_{ort.}$ (m^3/s)	T (year)
2304	Bayburt	Çoruh	Çoruh	1,734	1,545	16.2	54 (1941–1998)[b]
2603	Beşiri	Garzan	Dicle	2,450	545	49.8	53 (1946–1998)
2605	Diyarbakır	Dicle	Dicle	5,655	570	71.1	50 (1946–1997)[b]
2610	Baykan	Bitlis	Dicle	640	910	19.8	44 (1955–1998)

A drainage area, *H* station altitude, $Q_{ort.}$ average discharge, *T* sample length and interval

[a]Missing data

[b]Stations with intermediate missing observations

For observed peak flows data, quantile–quantile plots were drawn and the statistics of modified Anderson-Darling (Sinclair et al. 1987; Ahmad et al. 1988) (AU_n^2) and average deviation (AD) were calculated by

$$AU_n^2 = \frac{n}{2} - 2\sum_{i=1}^{n} F(x_i) - \sum_{i=1}^{n}\left(2 - \frac{2i-1}{n}\right)\ln\left[1 - F(x_i)\right] \tag{6}$$

$$AD = \frac{\sum\limits_{i=n/2}^{n} |X_i - O_i|}{n/2} \tag{7}$$

where $F(x)$ and X_i are cumulative probability (0–1) and quantile ($m^3\ s^{-1}$) predicted by the W5, respectively, O_i is observed annual peak flow ($m^3\ s^{-1}$), i is sample order, and n is sample size. The statistic of AU_n^2 gives greater weight to the upper tail of the distribution. To further evaluate the upper right tail predictions, average deviations for the largest 7 observed data (AD_7), which had return periods of 30 $\left\{T = n - 1/(1-F) = 35 - 1/\left(1 - (28-0.35)/35\right)\right\}$ and greater years, were also computed. The lower values of these statistics show better fit.

3 Results and Discussions

Since naturally both parameter estimation methods of LM and probability weighted moments produce parameters that are more stable against possible outliers in the sample series (Haktanir and Bozduman 1995), the Grubbs and Beck test (Grubbs and Beck 1972) was employed for detection of outliers in the sample series generated by Monte Carlo sampling by both NLS and LM methods. In the analyses, the sample series were used with and without the outliers. However, the whole presented results are from the outliers discharged sample series. In general, employing this test supplied improvement in the results by NLS, while little by the LM.

While there is no theoretical reason to prefer any of the plotting position formula and the unbiased plotting formula is distribution specific, the effects of four common plotting position formulas (Rao and Hamed 2000), named Cunnane, Weibull, Gringerton, and Landwehr, on estimating the parameters and quantiles by NLS were also investigated with the Monte Carlo simulated data. Mostly the Landwehr formula $[(i-0.35)/n]$ produced better bias and relative root mean square error values than the others for all six population cases and three sample sizes. This formula also had been found to be the better combination of W5 with the parameter estimation method of probability weighted moment by Landwehr et al. (1979a, b), particularly in reference to the upper quantile values ($F \geq 0.50$) than any of the more common plotting position formulas.

Considering the results of biases of parameters and quantiles by two methods, for the parameters α, β, and ε, mostly the NLS produced smaller bias values than the LM, whereas, for the parameters γ and δ, the LM produced better results than the NLS. In addition, in the cases of 3 and 5, the LM produced less bias values than the NLS for the parameter α. When the bias values in the quantiles were considered, there was a significant reduction with increasing sample size, and there was significant increase with increasing non-exceedence probability for both methods. When $N = 30$, the NLS produced smaller biases than the LM for $F = 0.50$, 0.80, and 0.90. This is vice versa for $F = 0.99$ and 0.999. Both methods produced the same bias value (0.151) in the case of 4 for $F = 0.99$. The same results were also obtained when $N = 60$. In this case, the predictions by both methods produced comparable results for $F = 0.80$. However, the NLS produced smaller bias values than the LM in the case of 6 for $F = 0.99$. When $N = 90$, the NLS produced better results than the LM in the cases of 4, 5, and 6. Thus, in terms of *bias* indice, if the skewness value of data set (*Cs*) is less than 2.0, the NLS may be the preferred method for estimating the parameters and quantiles of W5 distribution when the sample size is high such as 60.

Considering the results of relative root mean square errors of parameter and quantile estimates by both LM and NLS methods, in 18 situations (6 cases $\times$ 3 sample sizes), overall the NLS produced 12, 16, 18, 6, and 6 times better results than the LM for the parameters α, β, ε, γ, and δ, respectively. Especially, for estimating the parameters of α, β, ε, the NLS became better than the LM in the cases of 1, 2, and 4 for three sample sizes. The performances of LM are fairly better than those of NLS for the parameters of γ and δ when the sample size is high. The NLS produced lower *rrmse* values than the LM in parameter estimates of δ for cases 3 and 5. When we consider the quantiles, the NLS produced smaller *rrmse* values than the LM for almost all cases and sample sizes. Thus, in terms of *rrmse* indice, the NLS may be the preferred method for estimating the parameters and quantiles of W5 distribution. As we compare parameter estimates by the methods in both performance indices (*bias* and *rrmse*), in general, α, β, and ε are better estimated by the NLS, while γ and δ by the LM.

In terms of computer time, the algorithm of NLS requires a much longer period of time than the LM. For example, in a personal notebook computer, the NLS required about 6 min to estimate the parameters and quantiles, and to write the results to a text file of 1,000 sample series each having 30 elements, for the 1st case. For this operation, the LM just required only 3–4 s. The percentages of successful estimates of parameters by the NLS algorithm were between 97–100%. This varied from case to case and sample sizes considered. In general, the success percentages increased with increased sample sizes. The low success rates in getting valid parameters defined in the parameter estimation section can be improved by reorientation of algorithm flow after setting each parameter to zero at the end of failed iteration, and restart to predict the values of rest of parameters. This kind of precaution was used by the LM, such as the routines turning back to the case of three parameter generalized Pareto distribution when both γ and δ are zero.

Table 4 The LM and NLS produced statistics of modified Anderson-Darling (AU_n^2) and average deviations for right tail (AD) and the highest seven values (AD_7)

Station and n	C_s	C_k	C_v	AU_n^2		AD		AD_7	
				LM	NLS	LM	NLS	LM	NLS
Döllük, 61	2.17	6.89	0.76	0.11	0.12	65.74	48.09	151.64	48.15
Küçükilet, 53	0.94	−0.14	0.67	0.29	0.26	10.32	7.71	17.35	7.87
Kayaca, 45	1.20	1.38	0.60	0.09	0.11	36.80	34.53	76.82	53.67
Yahyaköy, 46	1.68	3.65	0.57	0.07	0.15	50.29	41.44	114.14	85.86
Geçitköy, 45	1.77	3.73	0.56	0.16	0.36	17.21	17.81	50.86	43.08
Balıklı, 43	0.09	0.84	0.37	0.12	0.09	9.10	4.96	17.72	5.08
Selçuk, 44	2.54	8.99	0.80	0.10	0.13	11.70	11.41	23.50	19.29
Kayırlı, 61	0.58	0.94	0.49	0.10	0.11	12.55	11.31	34.59	30.79
Aydın Köp., 49	0.24	0.78	0.42	0.11	0.24	6.38	6.26	11.92	10.33
Burhaniye, 48	1.96	4.65	0.70	0.33	0.33	24.83	24.17	49.79	47.20
Çitak Köp., 47	2.40	6.92	0.43	0.09	0.13	1.97	1.76	5.20	4.62
Kavaklıdere, 42	1.25	1.14	0.65	0.24	0.34	8.31	8.94	18.50	15.39
Homa, 44	0.61	1.13	0.30	0.10	0.14	23.39	20.17	57.32	47.71
Beşkonak, 59	0.91	1.01	0.47	0.11	0.12	39.94	40.92	76.96	61.22
Beşdeğirmen, 61	2.61	10.90	0.76	0.07	0.16	2.46	2.47	6.59	4.05
Yağbasan, 44	0.78	0.02	0.55	0.17	0.17	7.33	6.49	10.76	8.37
Doğançay, 37	1.21	1.33	0.47	0.07	0.09	27.91	22.43	46.77	31.38
Rüstümköy, 46	0.59	0.11	0.52	0.08	0.07	7.85	5.21	13.45	3.08
Hamidiye, 44	1.02	0.49	0.80	0.16	0.19	3.29	2.96	5.30	3.59
Dokurcun, 43	1.49	1.65	0.77	0.24	0.36	14.29	10.48	26.01	11.26
Yakabaşı, 46	1.08	1.52	0.38	0.13	0.15	14.17	13.64	28.83	23.37
Fatlı, 61	−0.02	−0.42	0.34	0.17	0.11	14.02	10.17	16.36	7.67
Kale, 57	1.14	2.71	0.35	0.13	0.14	29.88	27.79	62.00	49.37
Durucasu, 44	0.67	0.53	0.45	0.10	0.11	14.06	12.33	26.61	19.94
Sütlüce, 44	0.74	0.55	0.37	0.14	0.18	7.50	7.02	13.97	10.20
Yamula, 52	1.09	2.16	0.40	0.07	0.09	14.90	14.22	26.24	20.12
Yahşihan, 47	1.07	0.47	0.62	0.24	0.34	45.71	41.40	91.26	61.28
Kaleboğazı, 38	0.98	1.85	0.52	0.40	0.16	4.50	2.22	9.27	2.27
Şefaatli, 45	0.49	0.16	0.50	0.10	0.12	2.82	2.65	5.11	4.28
Kuyluş, 44	1.71	3.20	0.55	0.07	0.12	11.82	11.50	27.03	20.46
Himmetli, 63	2.08	6.04	0.61	0.12	0.27	11.66	13.19	41.09	35.12
Gökdere, 58	1.84	4.54	0.56	0.07	0.37	32.75	29.36	80.82	41.97
Ergenuşağı, 41	2.32	8.31	0.43	0.09	0.09	9.73	7.37	17.15	10.44
Torun Köp., 45	1.15	1.78	0.46	0.16	0.18	9.39	8.47	19.19	17.22
Müşrüflü, 44	1.79	3.68	0.82	0.07	0.08	21.92	19.60	46.75	34.34
Demirköprü, 49	0.72	0.66	0.37	0.19	0.31	9.71	9.96	24.95	20.16
Antakya, 47	0.33	0.56	0.36	0.07	0.08	10.66	8.87	15.57	8.84
Kılavuzlu, 52	1.49	3.58	0.51	0.08	0.12	27.47	24.48	51.31	35.06
Akçil, 35	0.81	0.40	0.44	0.10	0.15	5.62	4.56	7.26	3.49
Karaahmet, 45	0.93	1.08	0.54	0.08	0.07	2.37	1.50	4.26	2.29
Poskoflu, 44	1.94	4.56	0.63	0.09	0.10	5.14	4.06	9.87	5.78
Tanır, 42	2.11	4.91	0.77	0.07	0.30	5.12	3.36	12.49	4.43
Kemah boğazı, 38	1.31	3.57	0.41	0.07	0.07	16.02	14.15	29.79	28.14
Kılayık, 42	2.55	7.71	1.07	0.09	0.12	4.39	3.43	8.99	5.63
Kürtün, 40	1.50	3.04	0.52	0.06	0.07	12.16	9.13	22.47	13.31
Kanlıpelit, 40	1.95	4.65	0.39	0.11	0.54	4.71	4.54	10.45	6.41

Table 4 (continued)

Station and n	C_s	C_k	C_v	AU_n^2		AD		AD_7	
				LM	NLS	LM	NLS	LM	NLS
Bayburt, 54	1.05	0.50	0.46	0.17	0.20	4.85	4.47	10.66	7.95
Beşiri, 53	0.73	0.23	0.44	0.06	0.07	17.39	14.95	41.37	28.00
Diyarbakır, 50	1.11	0.66	0.61	0.11	0.14	105.70	99.32	208.42	149.65
Baykan, 44	0.45	−0.50	0.44	0.07	0.06	6.26	5.34	10.77	8.05

n number of data in station after removal of outliers

For the annual river peak flow series, the statistics of AU_n^2, AD, and AD_7 for the stations and parameter estimation methods of W5 distribution are given in Table 4. The lowest statistic among two from the LM or NLS was underlined. For the AU_n^2 statistic, the LM produced the lowest values 43 times, and the NLS produced only 11 times. When the statistics of right tail average deviation (AD) are considered, the NLS produced 43 times better results than the LM. In terms of AD_7 values, the NLS yielded better results than the LM for all peak flow series. These kinds of AU_n^2 and AD statistics can be expected from the NLS and LM, since the NLS estimates are highly sensitive to the largest sample elements (Eq. 2), while the LM estimates occupy the medium positions. Furthermore, for unbounded distribution of residuum ($X_i - O_i$), the two dispersion measures, i.e. Eq. 2 (objective function for parameter optimization) and 7 (criterion of fit), are related. The 1:1 plots of the quantiles predicted by both LM and NLS methods versus the observed values for the recorded annual flood peaks series at gauging stations of Geçitköy and Fatlı are shown in Figs. 2 and 3, respectively. In general, these figures are representative of those for the other stations listed in Table 3. The closeness to the 1:1 line shows better fit. As seen in these figures, most of the quantiles produced by the methods

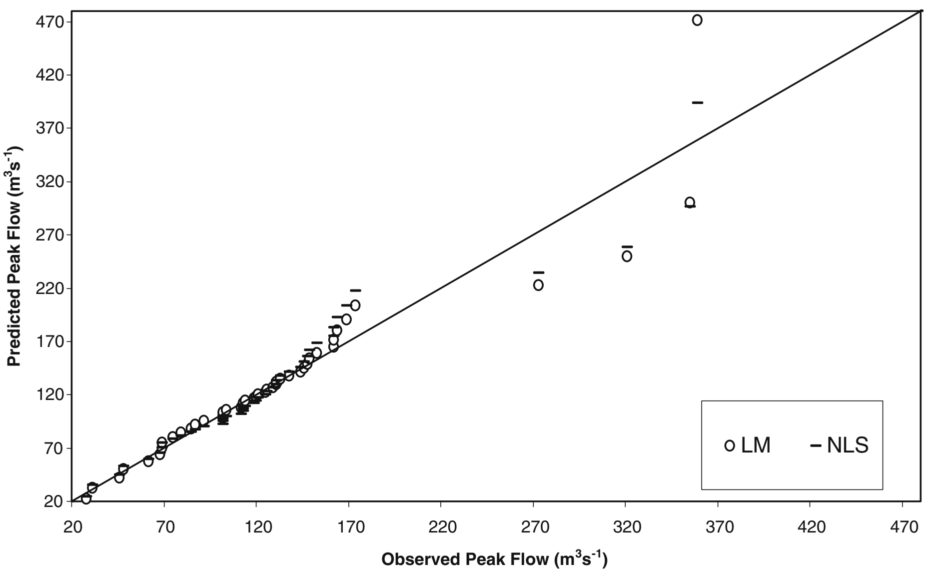

Fig. 2 Quantile–quantile plot by the LM and NLS parameter estimation methods for the annual peak flows of Geçitköy station

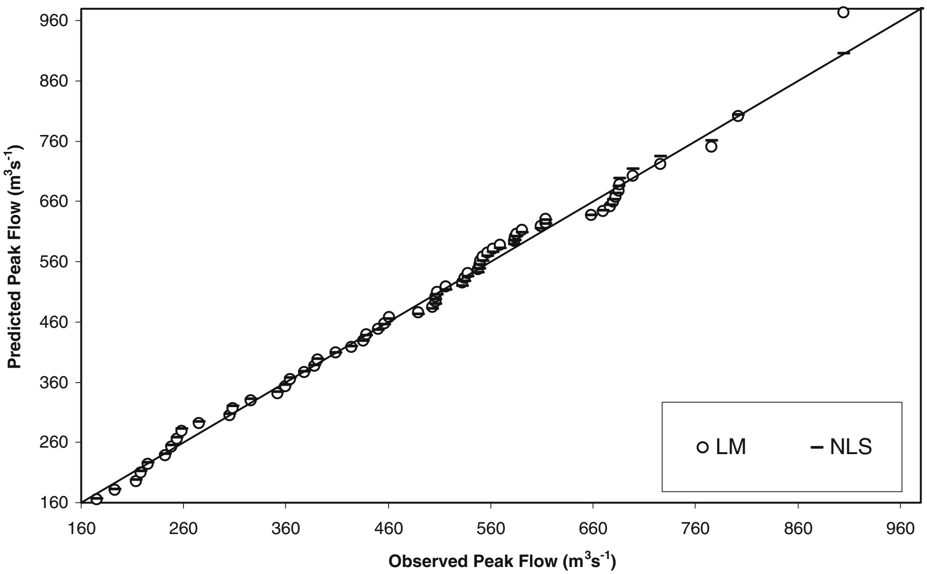

Fig. 3 Quantile–quantile plot by the LM and NLS parameter estimation methods for the annual peak flows of Fatlı station

are similar. In general, the differences are large for the highest last 5–10 quantiles. Similar to the statistics of AU_n^2 and AD in Table 4, the better predictions by LM for Geçitköy and by NLS for Fatlı can be seen in these Figs. 1 and 2, respectively. From both figures, the better prediction of NLS than that of LM for the largest events can also be seen.

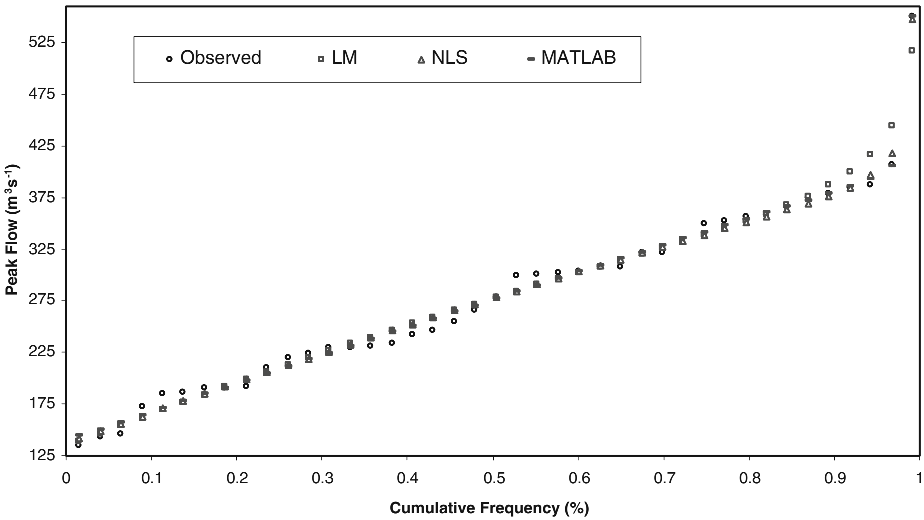

Fig. 4 The fitting efficiency of W5 by the parameter estimation methods of LM, NLS and MATLAB for Balıklı station

Since a preference is given to the estimation of upper quantiles ($F \geq 0.50$) of the W5 distribution in this study, omitting altogether the two parameters (γ and δ) associated with the lower tail of the distribution and keep only a three-parameter (as it is stated before which is equivalent to the three-parameter generalized Pareto distribution) was found a valuable research subject. To determine the effects of these two parameters on the performance of W5 with its parameters estimated by the developed NLS algorithm, instead of modification of the presented NLS algorithm, the performances of NLS (five-parameter) were compared with those of generalized Pareto distribution (three-parameter). In this analysis, the data series of Turkish rivers (Table 3) were employed with the statistics of AU_n^2, AD, and AD_7. The parameters of generalized Pareto distribution were estimated by LS (Moharram et al. 1993) with a FORTRAN subroutine. In the result, the generalized Pareto distribution produced better statistics for only about ten stations (20%) (10 stations with AU_n^2, 8 stations with AD, and 12 stations with AD_7). The C_s values of these five stations favored by all three statistics altogether are in the range between 1.07 and 1.44. The results may indicate also important effects of the parameters of γ and δ on the performance of W5 when $F \geq 0.50$.

The MATLAB could not produce any parameter estimates for 12 data sets of 50 (24%) because the fit computation did not converge. For the 31 of 38 stations, the NLS produced lower *rrmse* values than the MATLAB, and mostly both of NLS and MATLAB produced same values of r^2 (for 24 of 38 stations) and parameters (for 25 of 38 stations). For a representative for the other 38 stations, Figs. 4 and 5 for the stations of Balıklı and Tanır, respectively, show the fitting efficiencies of the predicted values by the NLS, MATLAB, and LM. Considering the deviations from observed values, similar fitting efficiencies by NLS and MATLAB and better estimates by NLS and MATLAB than those by LM especially for upper right tails are clear in Fig. 4. In addition, Fig. 5 shows better estimates by NLS than those by LM and MATLAB especially for upper right tail.

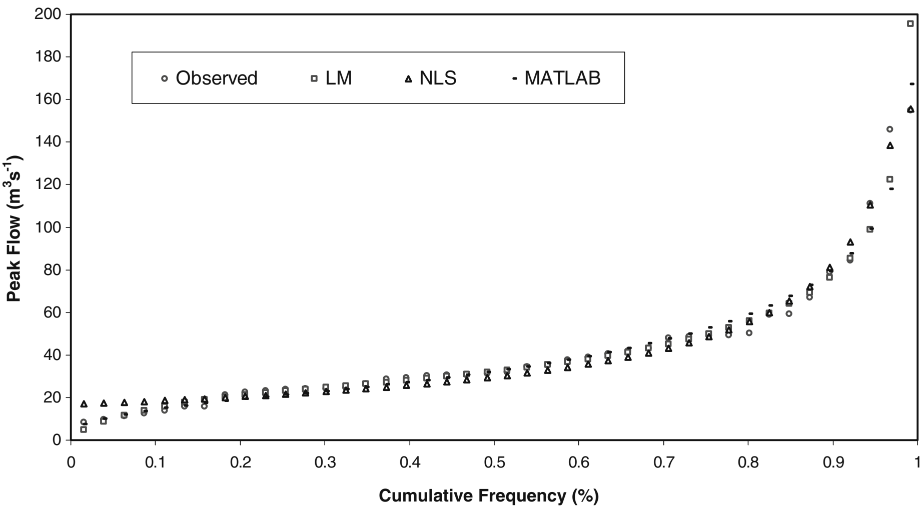

Fig. 5 The fitting efficiency of W5 by the parameter estimation methods of LM, NLS and MATLAB for Tanır station

4 Conclusions

The following conclusions can be drawn from this study: 1) an original numerical least squares parameter estimation method as an alternative to the L-moment (and probability weighted moment) was introduced for the five-parameter Wakeby distribution; 2) both performance indices of bias and relative root mean square error from the substantial Monte Carlo simulation experiment showed that the parameters of α, β, and ε of five-parameter Wakeby distribution are better estimated by the numerical least squares method, while γ and δ by the L-moment; 3) in terms of bias, the L-moment performance is better to predict high return period quantiles (for $F \geq 0.90$); 4) in terms of bias, the performances of numerical least squares are better than that of L-moment when both sample skewness and kurtosis are not too high ($C_s < 2.0$ and $C_k < 14.0$) and sample size is large ($n \geq 60$); 5) in terms of relative root mean square error, the performance of numerical least squares is better than that of L-moment; 6) for the observed river peak flows, while the modified Anderson-Darling statistics giving more weight to upper part of data series favored the L-moment, the average deviations giving weight to right part of data series, which is mostly preferable by engineers for flood frequency analysis to design a suitable flood value, favored the numerical least square; 7) in a limited success, the curve fitting algorithm of MATLAB can be used in place of the developed algorithm to predict the parameters of five-parameter Wakeby distribution.

Appendix

Table 5 Matrices employed for determining the parameters of W5 by NLS[a]

$$P_1 = \begin{pmatrix} 0.25x_1 & 2x_n & 5.5 & 0 & 0 \\ -0.1x_1 & x_n & 3.5 & 0 & 0 \\ 0.35x_1 & 1.5x_n & 4.5 & 0 & 0 \\ 0.5x_1 & 2.5x_n & 4.0 & 0 & 0 \\ 0.5x_1 & x_n & 0.5 & 0 & 0 \\ 0.35x_1 & 1.4x_n & 1.0 & 0 & 0 \end{pmatrix} \qquad P_3 = \begin{pmatrix} \varepsilon_1 & \alpha_1 & \beta_1 & 0 & 0 \\ 0.7\varepsilon_1 & \alpha_1 & \beta_1 & 0 & 0 \\ \varepsilon_1 & 0.7\alpha_1 & \beta_1 & 0 & 0 \\ \varepsilon_1 & \alpha_1 & 1.5\beta_1 & 0 & 0 \\ 0.7\varepsilon_1 & 0.7\alpha_1 & \beta_1 & 0 & 0 \\ 0.7\varepsilon_1 & \alpha_1 & 1.5\beta_1 & 0 & 0 \end{pmatrix}$$

$$P_2 = \begin{pmatrix} \varepsilon_3 & \alpha_3 & \beta_3 & x_1 & 0.3 \\ \varepsilon_3 & \alpha_3 & \beta_3 & x_1 & -0.3 \\ \varepsilon_3 & \alpha_3 & \beta_3 & 2x_1 & 0.3 \\ \varepsilon_3 & \alpha_3 & \beta_3 & 2x_1 & -0.3 \\ \varepsilon_3 & \alpha_3 & \beta_3 & 0.5x_1 & 0.3 \\ \varepsilon_3 & \alpha_3 & \beta_3 & 0.5x_1 & -0.3 \end{pmatrix} \qquad P_4 = \begin{pmatrix} \varepsilon_3 & \alpha_3 & \beta_3 & \gamma_2 & \delta_2 \\ \varepsilon_3 & \alpha_3 & \beta_3 & 0.9\gamma_2 & \delta_2 \\ \varepsilon_3 & \alpha_3 & \beta_3 & \gamma_2 & 0.9\delta_2 \\ \varepsilon_3 & \alpha_3 & \beta_3 & 1.1\gamma_2 & \delta_2 \\ \varepsilon_3 & \alpha_3 & \beta_3 & \gamma_2 & 1.1\delta_2 \\ \varepsilon_3 & \alpha_3 & \beta_3 & \gamma_2 & -1.1\delta_2 \end{pmatrix}$$

$$P_5 = \begin{pmatrix} \varepsilon_3 & \alpha_3 & \beta_3 & \gamma_4 & \delta_4 \\ 1.05\varepsilon_3 & 1.05\alpha_3 & 1.05\beta_3 & 1.05\gamma_4 & 0.95\delta_4 \\ 1.05\varepsilon_3 & 1.05\alpha_3 & 1.05\beta_3 & 0.95\gamma_4 & 0.95\delta_4 \\ 1.05\varepsilon_3 & 1.05\alpha_3 & 1.05\beta_3 & 0.95\gamma_4 & 1.05\delta_4 \\ 1.05\varepsilon_3 & 1.05\alpha_3 & 0.95\beta_3 & 0.95\gamma_4 & 0.95\delta_4 \\ 1.05\varepsilon_3 & 1.05\alpha_3 & 0.95\beta_3 & 0.95\gamma_4 & 0.95\delta_4 \end{pmatrix}$$

[a] x_1 is the lowest, x_n is the highest observed values in sample, $\beta_{1,\ldots,53}$, $\alpha_{1,\ldots,53}$, $\varepsilon_{1,\ldots,53}$, $\gamma_{1,\ldots,53}$, and $\delta_{1,\ldots,53}$ are values of parameters obtained in steps 1,..., 53

References

Ahmad MI, Sinclair CD, Spurr BD (1988) Assessment of flood frequency models using empirical distribution function statistics. Water Resour Res 24(8):1323–1328

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

Griffiths GA (1989) A theoretically based Wakeby distribution for annual flood series. Hydrol Sci 34(3):231–248

Grubbs F, Beck G (1972) Extension of sample sizes and percentage points for significance tests of outlying observations. Technometrics 14(4):847–854

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

Haktanir T, Horlacher HB (1993) Evaluation of various distributions for flood frequency analysis. Hydrol Sci 38(1):15–32

Haktanir T, Bozduman A (1995) A study on sensitivity of the probability-weighted moments method on the choice of the plotting position formula. J Hydrol 168:265–281

Hosking JRM (2000) Research report Fortran routines for use with the method of L-moments. Version 3.03, IBM Research Division, New York, p 33

Hosking JRM, Wallis JR (1997) Regional frequency analysis. Cambridge University Press, New York, p 224

Houghton JC (1978) Birth of a parent: the Wakeby distribution for modeling flood flows. Water Resour Res 14(6):1105–1109

Karamouz M, Abesi O, Moridi A, Ahmadi A (2009) Development of optimization schemes for floodplain management; a case study. Water Resour Manage 23(9):1743–1761

Kumar A, Chander S (1987) Statistical flood frequency analysis: an overview. In: Proceedings of the international symposium on flood frequency and risk analyses, hydrologic frequency modelling. D. Reidel Publishing Company, Dordrecht, pp 19–35

Landwehr JM, Matalas NC, Wallis JR (1979a) Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resour Res 15(5):1055–1064

Landwehr JM, Matalas NC, Wallis JR (1979b) Estimation of parameters and quantiles of Wakeby distributions 1. Known lower bounds. Water Resour Res 15(6):1361–1372

Moharram SH, Gosain AK, Kapoor PN (1993) A comparative study for the estimators of the generalized Pareto distribution. J Hydrol 150:169–185

Öztekin T (2007) Wakeby distribution for representing annual extreme and partial duration rainfall series. Meteorol Appl 14(4):381–387

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipes in Fortran. Cambridge University Press, New York, p 963

Rao AR, Arora PS (1987) An empirical study of probability distributions of annual maximum floods. In: Proceedings of the international symposium on flood frequency and risk analyses, hydrologic frequency modelling. D. Reidel Publishing Company, Dordrecht, pp 449–465

Rao AR, Hamed KH (2000) Flood frequency analysis. CRC Press, New York, p 350

Sinclair CD, Spurr BD, Ahmad MI (1987) Modified Anderson-Darling test. Technical report, Department of Mathematics Science, St. Andrews University, St. Andrews, Fife, Scotland

The MathWorks Inc. (2005) MATLAB Release 7.1. Natick, MA

Vogel RM, McMahon TA, Chiew FHS (1993) Flood flow frequency model selection in Australia. J Hydrol 146:421–449

Zalina MD, Desa MNM, Nguyen VTV, Kassim AHM (2002) Selecting a probability distribution for extreme rainfall series in Malaysia. Water Sci Technol 45(2):63–68