Fitting annual rainfall and annual streamflow by two transformed gamma distributions

HUYNH NGOC PHIEN AND THAMBIRAJAH JIVAJIRAJAH
Asian Institute of Technology, P.O. Box 2754, Bangkok 10501, Thailand

Abstract

The log Pearson type-3 (LP3) and the Grassia-Tadikamalla (GT) distributions which are both obtained from the gamma frequency curve are introduced to the statistical analysis of annual rainfall and annual streamflow data. The methods for estimating the parameters involved are described and suitable computational schemes suggested. Application of these distributions to actual data indicates that they may be used for the aforementioned purpose. Moreover, it is also found that the scale parameter of the LP3 distribution is frequently negative, leading to a bounded distribution like the GT distribution and both can be used as alternatives to the Johnson S_B curve.

Introduction

Although the log Pearson type-3 (LP3) distribution was recommended by the U.S.A. Water Resources Council (1967) for flood frequency analysis, its various forms as revealed by the work of Bobée (1975) indicate that it may be used for other purposes. From the studies of Nozdryn-Plotnicki and Watt (1979) and others reported in the literature, it was found that the scale parameter is commonly negative, leading to the existence of an upper bound for the random variable being represented. This may not be acceptable to hydrologists when they are concerned with flood estimation, but on the contrary, it may be acceptable for annual rainfall and annual streamflow analyses. With this conception in mind, the present study tries to explore the LP3 distribution in such context.

It should be noted that the LP3 distribution is obtained from the gamma distribution by means of an anti-logarithmic or exponential transformation. In addition to this form, Tadikamalla (1981) showed that the transformation proposed by Grassia (1977) leads to an even more versatile family of distributions because it covers a wider region in the moment-ratio -plane. Tadikamalla (1981) also suggested that the resulting distribution be used as alternative models to the Johnson S_B-curve which has been found to fit annual streamflows by Svanidze (1980) and Phien and Chow (1983). Evaluation of this suggestion is another purpose of the present study.

The material contained in this paper is organized as follows: After this introduction, the LP3 distribution and that obtained according to the transformation proposed by Grassia (1977) and Tadikamalla (1981), called the Grassia-Tadikamalla (GT) distribution, are described and methods for parameter estimation presented. Next, the computational schemes for the two commonly used goodness-of-fit tests, namely the chi-square and Kolmogorov-Smirnov tests, are discussed. The LP3 and GT distributions are then applied to annual rainfall and streamflow analyses, and finally conclusions are drawn.

The LP3 distribution

The density function of a LP3 variable x is written as

$$f(x) = [(\ln x - c)/a]^{b-1} \exp\{-(\ln x - c)/a\}/[|a|\,\Gamma(b)x] \quad (1)$$

where a, b and c are the scale, shape and location parameters, respectively; and $\Gamma(b)$ is the gamma function of b. Using the transformation $y = \ln x$, eq. 1 becomes:

$$g(y) = [(y - c)/a]^{b-1} \exp[-(y - c)/a]/[|a|\,\Gamma(b)] \quad (2)$$

which is a three-parameter gamma distribution, commonly known as a Pearson type 3 distribution.

When $a > 0$, y has a positive skewness. In this case, $c \leq y < +\infty$ and hence $\exp(c) \leq x < +\infty$. When $a < 0$, y has a negative skewness; $-\infty < y \leq c$ and $0 < x \leq \exp(c)$. Clearly, exp(c) is the upper bound of x in this case.

The mean, variance and skewness of y are as follows:

$$\left.\begin{aligned} m_1'(o) &= ab + c \\ m_1(o) &= a^2 b \\ g(o) &= 2a/(|a|\,b^{1/2}) \end{aligned}\right\} \quad (13)$$

in which the symbol (o) indicates a population value. The moment of order k of x can be shown to be given by the following equation:

$$M_k'(o) = e^{kc}(1 - ka)^{-b},\; 1 - ka > 0,\; k = 1, 2, \ldots \quad (4)$$

The mean, variance and skewness of x are then obtained as:

$$\left.\begin{aligned} M_1'(o) &= Be^c \\ M_2(o) &= A\,e^{2c} \\ G(o) &= C/A^{3/2} \end{aligned}\right\} \quad (5)$$

where

$$A = (1 - 2a)^{-b} - (1 - a)^{-2b}$$
$$B = (1 - a)^{-b}$$

and

$$C = (1 - 3a)^{-b} - 3[(1 - a)(1 - 2a)]^{-b} + 2(1 - a)^{-3b}$$

Even though the Water Resources Council (1967) recommended the use of the moments of $y = \ln x$ for the estimation of the parameters a, b and c, the simulation study by Phien and Hira (1983) showed that its performance is not satisfactory. In the following, a brief description of the four methods which were found to be acceptable by Phien and Hira (1983) is presented.

Method of Bobée (MBB)

Using eq. 4, Bobée (1975) arrived at the following system of equations:

$$\left.\begin{aligned} c - b\ln(1-a) &= \ln M'_1 \\ 2c - b\ln(1-2a) &= \ln M'_2 \\ 3c - b\ln(1-3a) &= \ln M'_3 \end{aligned}\right| \quad (6)$$

where M'_k denotes the sample moment of x of order k.

Eliminating b and c in eq. 6 yields

$$\phi(a) = \ln[(1-a)^3/(1-3a)] - D\ln[(1-a)^2/(1-2a)] = 0 \quad (7)$$

where

$$D = (\ln M'_3 - 3\ln M'_1)/(\ln M'_2 - 2\ln M'_1) \quad (8)$$

Using the Newton-Raphson method (see Dahlquist and Bjorck, 1974), one can solve eq. 7 for a. More details can be found in Phien and Hira (1983). Once a has been determined, b and c can be computed according to the following equations:

$$b = (\ln M'_2 - 2\ln M'_1)/\ln[(1-a)^2/(1-2a)] \quad (9)$$

$$c = \ln M'_1 + b\ln(1-a) \quad (10)$$

Method of Maximum Likelihood (MML)

As suggested by Phien and Hira (1983), the parameter c should be obtained first, then a and b can be computed as follows:

$$a = (\bar{y} - c)/b \quad (11)$$

$$b = \beta/(\beta - N) \quad (12)$$

where

$$\beta = [\Sigma(\ln x_i - c)][\Sigma(\ln x_i - c)^{-1}]$$

and

$$\bar{y} = (1/N)\,\Sigma\, y_i = (1/N)\,\Sigma \ln x_i$$

for a sample $\{x_1, \ldots, x_N\}$ of size N. Although a trial-and-error procedure works well, it is better to devise a systematic method of solving c. Considering a and b as functions of c, one can differentiate eqs. 11 and 12 with respect to c to give

$$db/dc = -N^2(\beta - N^2)^{-2}\, d\beta/dc$$

$$da/dc = -(1 + a\,db/dc)/b$$

where

$$d\beta/dc = [\Sigma(\ln x_i - c)][\Sigma(\ln x_i - c)^{-2}] - N\,\Sigma(\ln x_i - c)^{-1}$$

The remaining equation obtained by the method of maximum likelihood

$$\phi(c) = \Sigma\ln[(\ln x_i - c)/a] - N\Psi(b) = 0 \quad (13)$$

may be treated as a function of c which, after differentiation with respect to c, gives

$$d\phi/dc = -\Sigma(\ln x_i - c)^{-1} - (N/a)\,da/dc - N\Psi'(b)\,db/dc \quad (14)$$

The functions $\Psi(b) = d[\ln\Gamma(b)]/db$ and $\Psi'(b) = d\Psi(b)/db$ are commonly referred to as the *digamma* and *trigamma* functions of b, which can be computed easily using the algorithms developed by Bernado (1976) and Schneider (1978), respectively. Using eqs. 13 and 14, c can be solved by using the Newton-Raphson method.

First Method of Mixed Moments (MM1)

This method, suggested by Phien and Hira (1983), uses the first two moments of x and mean of y for determining a, b and c. The equation containing only a is as follows:

$$\phi(a) = a + \ln(1-a) - [(\ln M'_1 - \bar{y})/(2\ln M'_1 - \ln M'_2)]\,[2\ln(1-a) - \ln(1-2a)] = 0 \quad (15)$$

which can be solved easily by the Newton-Raphson method. Having obtained a, one can compute b and c as:

$$b = -(2\ln M'_1 - \ln M'_2)/\ln[(1-a)/(1-2a)] \quad (9)$$

and

$$c = \bar{y} - ab \quad (16)$$

Second Method of Mixed Moments (MM2)

This method, also suggested by Phien and Hira (1983), uses the first moments of x and the variance of y to compute a, b and c. In this method, a should be obtained first from the equation

$$\phi(a) = a^2(2\ln M'_1 - \ln M'_2) - s^2\ln[(1-2a)/(1-a^2)] = 0 \quad (17)$$

where $s^2 = m_2$ is the sample variance of $y = \ln x$. Since

$$d\phi/da = 2a(2\ln M'_1 - \ln M'_2) + 2s^2a/[(1-a)(1-2a)] \quad (18)$$

is readily computed, the Newton-Raphson method can be used to solve for a. Then b and c are obtained from:

$$b = s^2/a^2 \quad (19)$$

$$c = \ln M' + (s^2/a^2)\ln(1-a) \quad (20)$$

The GT distribution

This distribution with four parameters was obtained by Tadikamalla (1981) who introduced two more parameters into the distribution obtained by Grassia (1977). The density function is written as follows:

$$f(x) = b^a(x-c)^{b-1}[-\ln\{(x-c)/d\}]^{a-1}/[d^b\,\Gamma(a)] \quad (21)$$

For eq. 21 to be defined in all cases, the random variable x must be bounded as follows:

$$c \leq x \leq c + d \quad (22)$$

The mean, variance, skewness and kurtosis of x can be shown to be given by

$$M'_1(o) = c + dB_1 \quad (23)$$

$$M_2(o) = d^2(B_2 - B_1^2) \quad (24)$$

$$G(o) = (B_3 - 3B_2B_1 + 2B_1^3)/(B_2 - b_1^2)^{3/2} \quad (25)$$

$$K(o) = (B_4 - 4B_3B_1 + 6B_2B_1^2 - 3B_1^4)/(B_2 - B_1^2)^2 \quad (26)$$

where

$$B_i = [b/(b+i)]^a, \qquad i = 1, \ldots, 4 \quad (27)$$

Tadikamalla (1981) employed the method of moments for parameter estimation. In this study, the method of maximum likelihood is also used. These two methods are described in the following.

Method of Moments (MM)

As the expressions of the skewness, G(o) and kurtosis, K(o) do not contain c and d, eqs. 25 and 26 can be used to solve a and b. These are rewritten with sample values G and K as:

$$F_1(a,b) = B_3 - 3B_2B_1 + 2B_1^3 - G(B_2 - B_1^2)^{3/2} = 0 \quad (28)$$

$$F_2(a,b) = B_4 - 4B_3B_1 + 6B_2B_1^2 - 3B_1^4 - K(B_2 - B_1^2)^2 = 0 \quad (29)$$

This system of nonlinear equations can be solved by some existing computer codes discussed by Moré and Cosnard (1979), and Hiebert (1982). However, since the partial derivatives of F_1 and F_2 can be obtained easily, Newton's method should be used instead. The iteration according to this method is written as:

$$\begin{pmatrix} a \\ b \end{pmatrix}_{k+1} = \begin{pmatrix} a \\ b \end{pmatrix}_k - \begin{pmatrix} \partial F_1/\partial a & \partial F_1/\partial b \\ \partial F_2/\partial a & \partial F_2/\partial b \end{pmatrix}_k^{-1} \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}_k \quad (30)$$

where the subscript k is used to denote the kth iteration. The partial derivatives are given in Appendix A. Having obtained a and b, one can compute c and d using eqs. 23 and 24 after setting $M_1'(o) = M_1'$ and $M_2(o) = M_2$, where M_1' and M_2 are the sample mean and variance of x, respectively.

Method of Maximum Likelihood (MML)

From a sample $\{x_1, \ldots, x_N\}$ of size N, the log-likelihood function of the GT distribution is obtained as:

$$\ell(x,a,b,c,d) = Na\ln b - Nb\ln d + N\ln\Gamma(a) + (b-1)\Sigma\ln(x_i - c) + (a-1)\Sigma\ln[-\ln\{(x_i - c)/d\}]$$

where each Σ extends all over the N values of x. By setting the partial derivatives of ℓ with respect to a, b, c and d equal to zero, one obtains the following equations:

$$\ln b - \Psi(a) + (1/N)S_5 = 0 \quad (31)$$
$$Na - b(N\ln d - S_2) = 0 \quad (32)$$
$$(b-1)S_1 + (a-1)S_4 = 0 \quad (33)$$
$$Nb + (a-1)S_3 = 0 \quad (34)$$

where

$$S_1 = \Sigma(x_i - c)^{-1}$$
$$S_2 = \Sigma\ln(x_i - c)$$
$$S_3 = \Sigma[\ln\{(x - c)/d\}]^{-1}$$
$$S_4 = \Sigma[(x_i - c)\ln\{(x_i - c)/d\}]^{-1}$$
$$S_5 = \Sigma\ln[-\ln\{(x_i - c)/d\}]$$

From eq. 32,

$$b = -Na/S_6 \quad (35)$$

where

$$S_6 = \Sigma\ln\{(x_i - c)/d\} = S_2 - N\ln d$$

Under substitution of b into eqs. 31, 33 and 34, one arrives at the following system:

$$H_1(a,c,d) = Na + S_6S_7 = 0 \quad (36)$$
$$H_2(a,c,d) = NS_7 + (a-1)S_3 = 0 \quad (37)$$
$$H_3(a,c,d) = S_1(S_7 - 1) + (a-1)S_4 = 0 \quad (38)$$

where

$$S_7 = \exp[\Psi(a) - (1/N)S_5]$$

The system of nonlinear equations 36-38 can be solved by Newton's method because all the partial derivatives are readily available as given in Appendix B.

Goodness-of-fit tests

In evaluating the goodness-of-fit, the chi-square and Kolgomorov-Smirnov tests are most commonly used.

χ^2-test (see Yevjevich, 1972)

This test makes a comparison between the actual number of observations and the theoretical or expected number of observations that fall in the class intervals employed. The test statistic is calculated as follows:

$$\chi^2 = \sum_{i=1}^{h}(E_i - O_i)^2/E_i \quad (39)$$

where h is the number of class intervals (h≥5), while E_i and O_i are respectively the theoretical and observed numbers in interval i. This statistic is approximately distributed as a chi-square variable with h-p-1 degrees of freedom, where p is the number of parameters estimated. It would be desirable to have all the class intervals with the same theoretical number of observations. In this case,

$$E_i = N/h$$

where N is the sample size as denoted before.

Consequently, eq. 39 becomes:

$$\chi^2 = (h/N)\sum_{i=1}^{h}O_i^2 - N \quad (40)$$

The fitting is rejected if the computed value is greater than the critical value at the selected significance level.

Kolmogorov-Smirnov test (see Yevjevich, 1972)

This test is conducted as follows: Let F(x) be the theoretical distribution function to be fitted and let i/N denote the sample cumulative probability of the ith observation (after the sample has been arranged in ascending order). The KS-statistic is computed by

$$\Delta = \max_{1\le i\le N}|F(x_i) - i/N| \quad (41)$$

For a specified significance level, the fitting is rejected if the computed value of Δ is equal to or greater than the critical value.

Computation

For moderately large samples used (as seen later in Tables 1 and 2) the number of class intervals are chosen to be 7, each with probability 1/7. For the LP3 distribution, there are three parameters to be estimated, consequently, the degree of freedom for the chi-square test is 7 – 3 – 1 = 3. For the GT distribution, there are four parameters and the degree of freedom is 2.

Let

$$z = (\ln x - c)/a,$$

then the LP3 distribution is reduced to the one-parameter gamma distribution:

$$h(z) = z^{b-1} e^{-z}/\Gamma(b) \tag{42}$$

and

$$F(X) = \int_{\exp(c)}^{X} f(x)\,dx = \int_{0}^{Z} h(z)\,dz \tag{43}$$

where

$$Z = (\ln X - c)/a$$

The integral of h(z) involves the incomplete gamma function, which can be computed using the algorithm of Lau (1980). With this scheme, interpolation has to be made in order to obtain the class division points for the χ^2-test, while $F(x_i)$ in eq. 41 can be readily computed.

For the GT distribution, the transformation

$$z = -b\ln[(x-c)/d]$$

will simplify z to a one-parameter gamma variable. Consequently, the division points and $F(x_i)$ in eq. 41 can be determined in the same way as for the LP3 distribution.

Applications

Since the GT distribution is bounded, it should be applicable to the distribution of annual rainfall and streamflow sequences. The applicability of the LP3 distribution to these data should also be evaluated. For this purpose, 15 sets of annual rainfall in Portugal are selected from Santos (1981) and 20 sets of annual streamflows are taken from Yevjevich (1963), with the same criterion as mentioned by Phien and Chow (1983). These are listed in Tables 1 and 2, respectively. In order to see the range of the parameters of both distributions, their estimated values are given at five stations for annual rainfall sequences as well as for annual streamflows. For the case of the LP3 distribution (Tables 3 and 4) the values of a do not vary much. In all cases shown in Tables 3 and 4 as well as those which are not shown there a has magnitude less than 1. With all four methods, the values obtained for a are almost the same. For the parameter c, the range of possible values seems to be larger than that of a, but its values computed from the four methods are still similar. The range of possible values for parameter b is even larger, and its values estimated by these methods may be quite different.

For the GT distribution (Tables 5 and 6), all four parameters have large ranges of possible values. The values of each of these parameters may be quite different when computed from the method of moments (MM) and method of maximum likelihood (MML). Fortunately, regardless of the differences in the estimated values of the individual parameters, the statistical descriptors (i.e. the mean, standard deviations, and skewness, etc.) of each annual rainfall or annual streamflow sequence are almost unchanged with the method for parameter estimation. This statement is applicable to all the data sets shown in Tables 3 through 6 as well as the others (not shown) for both distributions. For the two goodness-of-fit tests under consideration, the results obtained in Tables 3 through 6 indicate that their values seem to be less sensitive to the method used for parameter estimation. This observation is also valid with other data sets not shown in these tables.

Discussion

The maximum number of iterations used in the Newton-Raphson or Newton method was 100. Even with this relatively large number, there were some cases where the iterative scheme did not converge within this limit. For the LP3 distribution, the method of Bobée (MBB) did not converge for station R5 (annual rainfall), and the method of maximum likelihood (MML) did not coverge for station R11. Consequently, the total number of cases for both positive and negative values for the parameter a is 14 for these two methods (Table 7). All four methods were convergent for streamflow data. For the GT distribution, the MML did not converge at station R14 of rainfall data, and at stations S5, S6, S10 and S18 of streamflow data, while the method of moments (MM) did not converge only at stations S5 and S6 of streamflow data. Although the fitting cannot be evaluated in these situations, this study simply considers them to be rejected by the goodness-of-fit tests at any significance level. Although the iterative scheme can be made convergent by using other initial values for the parameters concerned, such intervention was not attempted in this study because this was considered to be less important.

As seen from Table 7, the scale parameter a of the LP3 distribution is negative for most data sets. Correspondingly, it seems that a bounded LP3 distribution would be more frequently encountered in practice. (See also Nozdryn-Plotnicki and Watt, 1979; and Jinajirajah, 1982 for additional information).

The evaluation of fitting the LP3 and GT distributions to annual rainfall and annual streamflow data is summarized in Tables 8 and 9. While the number of data sets fitted by the two distributions is almost similar for annual rainfall (Table 8), it seems that the LP3 distribution fits more frequently to annual streamflow data. However, since the number of fitted cases is influenced by the convergence of numerical schemes it is expected that both these distributions apply almost equally to the streamflow data as well.

The summarized results in Tables 8 and 9 indicate that the Kolmogorov-Smirnov test accepts the fit more frequently than the chi-square test. They also show that the number of cases accepted by the KS-test is less sensitive to the significance level employed. Although it is easier to compute the KS-statistic than the χ^2-statistic, the above observation indicates that one may easily commit the Type II error when adopting the KS-test, which was pointed out by Phien and Arbhabhirama (1980) in considering these tests for the Gumbel distribution. It should be noted that the KS-test is, in principle, applicable to distributions where no parameter estimation is required. However, this is seldom the case in practical situations.

TABLE 1
LIST OF ANNUAL RAINFALL DATA

Station	Area	Country	Length of record (years)	Period of record
R1	Serpa	Portugal	48	1932 – 1979
R2	Arronches	Portugal	48	1932 – 1979
R3	Azaruja	Portugal	48	1932 – 1979
R4	Redondo	Portugal	48	1932 – 1979
R5	Portel	Portugal	48	1932 – 1979
R6	Beja	Portugal	48	1932 – 1979
R7	Trindade	Portugal	48	1932 – 1979
R8	Algodor	Portugal	48	1932 – 1979
R9	Amareleja	Portugal	48	1932 – 1979
R10	Aldeia de Palheiros	Portugal	48	1932 – 1979
R11	Aljezur	Portugal	48	1932 – 1979
R12	Faro	Portugal	48	1932 – 1979
R13	Alfandega da Fe	Portugal	48	1932 – 1979
R14	Vilar Formoso	Portugal	48	1932 – 1979
R15	Alforigo	Portugal	48	1932 – 1979

TABLE 2
LIST OF ANNUAL STREAMFLOW DATA

Station	Name of river	Location	Length of record (years)	Period of record
S1	Kanawha	Kanawha Falls, U.S.A.	80	1877 – 1957
S2	St. Regis	Brasher Center, U.S.A.	46	1910 – 1956
S3	Osage	Near Bagnell, U.S.A.	50	1880 – 1930
S4	Petit Jean Creek	Danville, U.S.A.	41	1916 – 1957
S5	Brazos	Waco, U.S.A.	46	1898 – 1944
S6	Colorado	Bellinger, U.S.A.	50	1907 – 1957
S7	Pecos	Near Anton Chico, U.S.A.	46	1909 – 1957
S8	Little Beaver Creek	Near Pikes Peak, U.S.A.	48	1888 – 1938
S9	Verde	Below Burlett Dam, U.S.A.	50	1911 – 1957
S10	Carson	Near Fort Churchill, U.S.A.	46	1903 – 1956
S11	Kaweah	Near Three Rivers, U.S.A.	53	1895 – 1956
S12	Kings	Piedra, U.S.A.	61	1890 – 1957
S13	Drina	Zvornik, Yugoslavia	67	1890 – 1957
S14	Dal	Norslund, Sweden	70	1852 – 1922
S15	Lule	Trangfors, Sweden	46	1911 – 1957
S16	Don	Kalach, U.S.S.R.	48	1881 – 1929
S17	Broken	Goorambat, Victoria	62	1886 – 1948
S18	Glenelg	Balmoral, Victoria	44	1888 – 1932
S19	Owens	Wangaratta, Victoria	62	1886 – 1948
S20	Tama	Atsumi, Japan	37	1918 – 1955

TABLE 3
RESULTS OF FITTING ANNUAL RAINFALL AMOUNTS (MEASURED IN mm) AT FIVE STATIONS BY THE LP3 DISTRIBUTION

Station	Method (*)	Parameters			Goodness-of-fit		Descriptors of x (+)		
		a	b	c	χ^2	Δ	Mean	Std	Skewness
R3	MBB	– 0,04	45,35	8,25	1,875	0,054	624,4	167,9	0,499
	MML	– 0,04	47,48	8,29	1,875	0,054	626,3	167,9	0,509
	MM1	– 0,04	45,37	8,25	1,875	0,054	626,3	167,9	0,500
	MM2	– 0,04	46,37	8,29	1,875	0,051	626,4	169,9	0,512
R6	MBB	– 0,10	8,43	7,15	3,917	0,070	577,8	152,2	0,107
	MML	– 0,15	4,18	6,94	5,375	0,085	579,8	156,3	– 0,123
	MM1	– 0,11	7,17	7,09	6,833	0,073	577,8	152,5	0,056
	MM2	– 0,15	3,86	6,89	7,417	0,090	577,8	149,0	– 0,188
R9	MBB	– 0,10	8,41	7,13	8,583	0,138	553,7	151,9	0,135
	MML	– 0,04	49,62	8,26	8,875	0,122	553,5	153,0	0,537
	MM1	– 0,04	45,78	8,18	8,291	0,124	553,2	151,8	0,517
	MM2	– 0,04	47,95	8,26	8,875	0,120	553,7	155,6	0,543
R12	MBB	– 0,16	5,72	6,99	5,583	0,072	469,8	157,8	0,174
	MML	– 0,19	4,27	6,89	9,167	0,080	471,2	160,1	0,075
	MM1	– 0,17	5,21	6,95	5,583	0,074	469,8	157,8	0,141
	MM2	– 0,19	4,17	6,87	11,500	0,080	469,8	156,9	0,051
R15	MBB	– 0,11	8,55	7,40	5,667	0,073	664,2	199,0	0,210
	MML	– 0,11	8,33	7,38	5,667	0,073	664,2	198,2	0,199
	MM1	– 0,11	8,55	7,40	5,667	0,073	664,2	199,0	0,210
	MM2	– 0,11	8,61	7,41	5,667	0,073	664,2	199,7	0,215

(*) MMB: Method of Bobée; MML: Method of Maximum Likelihood; MM1: First Method of Mixed Moments, and MM2: Second Method of Mixed Moments.
(+) The mean, standard deviation (Std) and skewness are calculated directly on computer, not from the round-off values of the parameters shown in this table.

TABLE 4

RESULTS OF FITTING ANNUAL STREAMFLOWS (MEASURED IN m³/s) AT FIVE STATIONS BY THE LP3 DISTRIBUTION

Station	Method (*)	Parameters			Goodness-of-fit		Descriptors of x (+)		
		a	b	c	χ^2	Δ	Mean	Std	Skewness
S4	MBB	−0,22	8,26	4,79	4,244	0,073	24,0	13,2	0,860
	MML	−0,22	7,95	4,76	4,244	0,073	24,0	13,1	0,841
	MM1	−0,22	8,26	4,79	4,244	0,073	24,0	13,2	0,860
	MM2	−0,23	7,57	4,73	4,244	0,071	24,0	13,2	0,827
S8	MBB	−0,12	17,02	−2,18	4,208	0,079	0,0	0,0	0,716
	MML	−0,13	16,28	−2,19	4,208	0,078	0,0	0,0	0,873
	MM1	−0,13	15,97	−2,24	4,208	0,078	0,0	0,0	0,869
	MM2	−0,14	13,19	−2,41	4,208	0,077	0,0	0,0	0,810
S12	MBB	−0,12	15,00	5,90	5,443	0,053	65,2	28,7	0,757
	MML	−0,10	21,73	6,22	4,984	0,056	65,2	28,6	0,848
	MM1	−0,10	21,33	6,21	4,984	0,054	65,3	28,7	0,848
	MM2	−0,10	21,35	6,21	4,984	0,055	65,2	28,7	0,849
S16	MBB	−0,05	45,51	8,71	1,292	0,058	669,2	218,6	0,668
	MML	−0,04	72,59	9,26	1,292	0,058	669,1	217,6	0,729
	MM1	−0,02	245,00	11,55	1,583	0,059	668,8	218,5	0,868
	MM2	−0,02	253,60	11,72	1,583	0,060	669,2	222,6	0,889
S20	MBB	0,04	86,16	−1,03	6,703	0,098	11,5	4,5	1,568
	MML	0,12	8,23	1,37	8,216	0,080	11,5	4,8	2,638
	MM1	0,08	17,47	0,92	8,216	0,085	11,5	4,5	2,028
	MM2	0,10	13,15	1,09	7,459	0,079	11,5	4,7	2,253

(*) MMB: Method of Bobée; MML: Method of Maximum Likelihood; MM1: First Method of Mixed Moments, and MM2: Second Method of Mixed Moments.

(+) The mean, standard deviation (Std) and skewness are calculated directly on computer, not from the round-off values of the parameters shown in this table.

TABLE 5

RESULTS OF FITTING ANNUAL RAINFALL AMOUNTS (mm) AT FIVE STATIONS BY THE GT DISTRIBUTION

Station	Method (*)	Parameters				Goodness-of-fit		Descriptors of x (+)		
		a	b	c	d	χ^2	Δ	Mean	Std	Skewness
R3	MM	313,0	169,6	−1005,0	10280	2,458	0,061	624,4	169,7	0,391
	MML	9,9	9,0	112,4	1461	2,458	0,056	626,5	166,2	0,320
R6	MM	19,8	22,0	−210,3	1902	3,333	0,071	577,8	154,1	0,136
	MML	19,8	24,3	−278,3	1907	4,500	0,075	577,8	152,1	0,085
R9	MM	174,2	131,6	−984,9	5751	8,875	0,143	553,7	153,5	0,187
	MML	174,2	82,3	−398,7	7808	8,875	0,136	553,7	151,8	0,325
R12	MM	13,5	14,1	−178,2	1624	6,250	0,074	469,8	159,4	0,189
	MML	13,5	14,9	−203,5	1610	6,250	0,078	469,9	157,2	0,153
R15	MM	17,8	17,4	−201,9	2340	6,833	0,074	664,2	201,1	0,218
	MML	17,8	17,9	−212,6	2309	6,833	0,075	664,3	198,3	0,200

(*) MM: Method of Moments, MML: Method of Maximum Likelihood

(+) The mean, standard deviation (Std) and skewness are calculated directly on computer, not from the round-off values of the parameters shown in this table.

TABLE 6
RESULTS OF FITTING ANNUAL STREAMFLOWS (m^3/s) AT FIVE STATIONS BY THE GT DISTRIBUTION

Station	Method (*)	Parameters				Goodness-of-fit		Descriptors of x (+)		
		a	b	c	d	χ^2	Δ	Mean	Std	Skewness
S4	MM	10,0	5,3	−0,9	138,9	4,224	0,070	24,0	13,4	0,893
	MML	13,7	6,9	−2,3	168,7	4,224	0,070	24,0	13,1	0,894
S8	MM	27,8	11,4	−0,0	0,2	3,042	0,079	0,0	0,0	0,925
	MML	30,2	12,2	−0,0	0,2	3,042	0,076	0,0	0,0	0,918
S12	MM	44,3	17,0	−10,3	952,3	4,934	0,055	65,2	28,9	0,802
	MML	44,3	15,8	−4,4	1059,0	4,984	0,057	65,2	28,7	0,821
S16	MM	8,6	5,5	209,6	1949,0	3,333	0,062	669,2	220,9	0,696
	MML	8,6	5,7	198,4	1887,0	1,292	0,057	669,2	217,1	0,645
S20	MM	9,3	3,3	5,9	65,9	6,324	0,093	11,5	4,6	1,634
	MML	4,4	2,3	5,3	31,4	3,676	0,112	11,8	4,7	1,029

(*) MM: Method of Moments, MML: Method of Maximum Likelihood
(+) The mean, standard deviation (Std) and skewness are calculated directly on computer, not from the round-off values of the parameters shown in this table.

TABLE 7
NUMBER OF CASES WITH POSITIVE AND NEGATIVE VALUES FOR THE PARAMETER a OF THE LP3 DISTRIBUTION

Method (*)	Annual Rainfall		Annual Streamflow	
	Positive	Negative	Positive	Negative
MBB	1	13	5	15
MML	1	13	4	16
MM1	1	14	4	16
MM2	1	14	4	16

(*) MBB: Method of Bobée; MML: Method of Maximum Likelihood; MM1: First Method of Mixed Moments, and MM2: Second Method of Mixed Moments.

TABLE 8
NUMBER OF CASES ACCEPTED BY THE GOODNESS-OF-FIT TESTS FOR ANNUAL RAINFALL AMOUNTS

Method (*)	Significance Level 1% (1)	1% (2)	5% (1)	5% (2)	10% (1)	10% (2)
LP3 Distribution						
MMB	14	14	13	14	13	14
MML	14	14	11	14	10	14
MM1	15	15	12	15	11	15
MM2	14	15	10	15	9	15
GT Distribution						
MM	14	15	10	15	8	15
MML	13	14	8	14	7	14

(*) MBB: Method of Bobée; MML: Method of Maximum Likelihood; MM1: First Method of Mixed Moments, and MM2: Second Method of Mixed Moments.
(1) χ^2-test
(2) KS-test

TABLE 9
NUMBER OF CASES ACCEPTED BY THE GOODNESS-OF-FIT TESTS FOR ANNUAL STREAMFLOWS

Method (*)	Significance Level 1% (1)	1% (2)	5% (1)	5% (2)	10% (1)	10% (2)
LP3 Distribution						
MMB	20	20	17	20	15	20
MML	20	20	19	20	18	20
MM1	20	20	18	20	16	20
MM2	20	20	19	20	16	20
GT Distribution						
MM	15	18	12	18	9	18
MML	15	16	14	16	12	16

(*) MBB: Method of Bobée; MML: Method of Maximum Likelihood; MM1: First Method of Mixed Moments, and MM2: Second Method of Mixed Moments.
(1) χ^2-test
(2) KS-test

TABLE 10
PROBABILITY OF HAVING NEGATIVE VALUES CORRESPONDING TO A NEGATIVE LOWER BOUND OF THE GT DISTRIBUTION

Station	Annual rainfall		Station	Annual streamflow	
	Method (*)	Probability (in %)		Method (*)	Probability (in %)
R1	MM MML	0,008 0,010	S3	MM MML	0,004 0,004
R2	MM MML	0,013 0,012	S4	MM MML	0,009 0,041
R3	MM MML	0,000 —	S7	MM MML	0,384 0,002
R4	MM MML	0,016 —	S8	MM MML	0,002 0,002
R5	MM MML	0,012 —	S9	MM MML	0,023 —
R6	MM MML	0,007 0,006	S10	MM MML	0,000 —
R8	MM MML	0,008 0,007	S11	MM MML	0,024 —
R10	MM MML	0,019 0,035	S12	MM MML	0,007 0,006
R11	MM MML	0,045 0,037	S17	MM MML	0,002 0,082
R12	MM MML	0,004 0,004	S18	MM MML	0,396 nc
R13	MM MML	0,000 0,414	S19	MM MML	— 0,012
R15	MM MML	0,004 0,004			

(*) MM: Method of Moments, MML: Method of Maximum Likelihood
(–) c is positive (nc): iterative scheme not convergent

In practice, the parameter c of the GT distribution should be positive in rainfall and streamflow analyses. Normally, when c is negative, one automatically sets it equal to zero, reducing the GT distribution to a three-parameter frequency curve. However, such adjustment was not made in this work because the probability for x to be in the interval (c, 0) for $c < 0$ was found to be very small (Table 10). The maximum values of this probability were 0,414% and 0,396% for the rainfall and streamflow data, respectively. In other words, there is an extremely low probability for rainfall amounts or streamflows to be negative even if a negative value of their lower bound is accepted.

Summary and conclusions

Two distributions, namely the log-Pearson type 3 (LP3) and the Grassia-Tadikamalla (GT) distributions, which are obtained from the gamma frequency curves by the exponential transformations, are considered in this study. Their statistical descriptors such as the mean, variance and skewness are given and suitable methods for estimating the parameters involved are described along with appropriate computational schemes. While the LP3 distribution has been extensively applied to flood frequency analysis, it is applied, in this study, to the analysis of annual rainfall amounts and annual streamflows along with the GT distributions. From the analysis, the following conclusions may be drawn:

- There is no difficulty in using the presented methods for parameter estimation. Even though there were cases where computational schemes did not converge, it is expected that this minor difficulty can be overcome by different trials in the use of initial estimates.

- There is a very high portion of data sets fitted by the LP3 and GT distributions for both annual rainfall and annual streamflow data. This indicates their applicability to the frequency analysis of these data.

- The lower bound of the GT distribution may happen to be negative for some data sets, leading to the existence of negative values for annual rainfall and annual streamflow! A simple way to avoid this occurrence of negative values is to set the lower bound equal to zero when it is computed to be negative, and adjust the other parameters accordingly.

However, from the experience gained so far, there appears to be an extremely low probability for a negative value of rainfall or streamflow to occur, and hence such theoretical existence may be neglected in practical situations.

- The smallest sample size for the data used in this study is 37 (station S20), which is still relatively large. When the analysis is to be carried out with data having smaller sample sizes, it is envisaged that the values of sample skewness and sample kurtosis would fluctuate greatly, resulting in unreliable estimates for the parameters of the GT distribution in the method of moments. With the evidence provided by the results obtained in this study, one would think that the LP3 distribution may be used in preference of the GT distribution. Nevertheless, in comparison with the results obtained from the Johnson S_B curve for the same data sets (Phien and Chow, 1983; Phien and Jivajirajah, 1984), it is clear that both the LP3 and GT distributions fit them more frequently. Consequently, these two distributions can be used as alternative models to the S_B curve. Finally, as the lower bound parameter c of the GT is negative in many cases, it would be reasonable to set it equal to zero intentionally, thus reducing the GT distribution to a three-parameter frequency curve like the LP3 distribution. Such an attempt is being made and related results will soon be reported.

- It should be noted that the GT distribution is bounded, consequently one would not expect it to be applicable to the frequency analysis of extreme values.

- A set of computer programs have been developed for all the computations involved. These are obtainable from the authors upon request.

Acknowledgements

The authors wish to express their appreciation to the Asian Institute of Tehnology for use of its research facilities and to the Royal Government of the Netherlands for financing the study of the junior author.

References

BERNADO, J.M. (1976) Psi (digamma) function (Algorithm AS 103). *Applied Statistics* 25(3) 315-317.

BOBÉE, B. (1975) The log Pearson type 3 distribution and its applications in hydrology. *Water Resources Research* 11(5) 681-689.

DAHLQUIST, G. and BJORCK, A. (1974) *Numerical Methods*, p. 573 Prentice-Hall, Inc., New Jersey.

GRASSIA, A. (1977) On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma and derived compound distribution. *Australian J. Statistics* 19 108-114.

HIEBERT, K.L. (1982) An evaluation of mathematical software that solves systems of nonlinear equations. *ACM Trans. Mathematical Software* 8(1) 5-20

JIVAJIRAJAH, T. (1982) Applicability of Some Statistical Distributions in Hydrology. Master Thesis, Asian Institute of Technology, Bangkok.

LAU, C.L. (1980) A simple series for the incomplete gamma integral (Algorithm AS 147). *Applied Statistics* 29(1) 113-114.

MORÉ, J.J. and COSNARD, M.Y. (1979) Numerical solution of nonlinear equations. *ACM Trans. Mathematical Software* 5(1) 64-85.

NOZDRYN-PLOTNICKI, M.N. and WATT, W.E. (1979) Assessments of fitting techniques for the log Pearson type 3 distribution using Monte Carlo simulation. *Water Resources Research* 15(3) 714-718

PHIEN, H.N. and ARBHABHIRAMA, A. (1980) A comparison of statistical tests on the extreme-value type-1 distribution. *Water Resources Bulletin* 16(5) 856-861.

PHIEN, H.N. and CHOW, W.B. (1983) Applicability of the Johnson S_B curve to annual streamflow distribution. *J. Hydrol.* 64 15-24.

PHIEN, H.N. and HIRA, M.A. (1983) Log Pearson type-3 distribution: parameter estimation. *J. Hydrol.* 64 25-37.

PHIEN H.N. and JIVAJIRAJAH, T. (1984) Fitting the S_B curve by the method of maximum likelihood. *J. Hydrol.* 67 67-75.

SANTOS, M.A.B.A.D. (1981) On stochastic characterization of regional droughts. Thesis for Degree of Research Officer, Laboratorie Nacional de Engenharia Civil, Lisbon.

SCHNEIDER, B.E. (1978) Trigamma function (Algorithm AS 121). *Applied Statistics* 27(1) 97-99.

SVANIDZE, G.G. (1980) Mathematical modeling of hydrologic series. Water Resources Publications, Fort Collins, Colorado.

TADIKAMALLA, P.R. (1981) On a family of distributions obtained by the transformation of the gamma distribution. *J. Statist. Comput. Simul.*, 13 209-214.

WATER RESOURCES COUNCIL (1967) A uniform technique for determining flood flow frequencies. *Water Resources Research*, Washington, D.C., Bull. 15.

YEVJEVICH, V. (1963) Fluctuations of wet and dry years, Part I: Research data assembly and mathematical models. Colo. State Univ., Fort Collins, Colo., *Hydrology Paper No. 1.*

YEVJEVICH, V. (1972) Probability and statistics in hydrology, p. 302, *Water Resources Publications*, Fort Collins, Colorado.

Appendix A

Partial derivatives of F_1 and F_2
From eq. 27,

$$B_i = [b/(b+i)]^a$$

it follows that

$$\partial B_i / \partial a = B_i \ln a \quad (A.1)$$

$$\partial B_i / \partial b = ia\, B_i\, [b(b+i)]^{-1} \quad (A.2)$$

With G and K in eqs. 28 and 29 being treated as constants, one can easily obtain the following results:

$$\frac{\partial F_1}{\partial a} = [B_3 - 6B_1B_2 + 4B_1^2 - (3/2)G\,(B_2 - B_1^2)^{1/2}(B_2 - 2B_1^2)] \ln a \quad (A.3)$$

$$\frac{\partial F_1}{\partial b} = \frac{3aB_3}{b(b+3)} - \frac{3a(3b+4)B_1B_2}{b(b+1)(b+2)} - \frac{4aB_1^2}{b(b+1)} - 3aG(B_2 - B_1^2)^{1/2} \left[\frac{B_2}{b(b+2)} - \frac{B_1^2}{b(b+1)}\right] \quad (A.4)$$

$$\frac{\partial F_2}{\partial a} = [B_4 - 8B_3B_1 + 18B_2B_1^2 - 12B_1^4 - 2K(B_2 - B_1^2)(B_2 - 2B_1^2)] \ln a \quad (A.5)$$

$$\frac{\partial F_2}{\partial b} = \frac{4aB_4}{b(b+4)} - \frac{8a(2b+3)B_3B_1}{b(b+1)(b+3)} + \frac{12a(2b+3)B_2B_1^2}{b(b+1)(b+2)} - \frac{12aB_1^4}{b(b+1)} - 4aK(B_2 - B_1^2)\left[\frac{B_2}{b(b+2)} - \frac{B_1^2}{b(b+1)}\right] \quad (A.6)$$

Appendix B

Partial derivatives of H_1, H_2 and H_3

Partial differentiation of eq. 36 with respect to a gives:

$$\frac{\partial H_1}{\partial a} = N + S_6 \frac{\partial S_7}{\partial a} = N + S_6 S_7 \Psi'(a) \qquad (B.1)$$

since S_6 is independent of a. Similarly, S_2, S_3 and S_4 are independent of a and hence

$$\frac{\partial H_2}{\partial a} = NS_7 \Psi'(a) + S_3 \qquad (B.2)$$

$$\frac{\partial H_3}{\partial a} = S_1 S_7 \Psi'(a) + S_4 \qquad (B.3)$$

In order to compute the partial derivatives with respect to c, the following results are useful:

$$\frac{\partial S_6}{\partial c} = \frac{\partial S_2}{\partial c} = -S_1$$

$$\frac{\partial}{\partial c} \ln(-\ln \frac{x-c}{d}) = -[(x-c) \ln \frac{x-c}{d}]^{-1}$$

So

$$\frac{\partial H_1}{\partial c} = \frac{\partial S_6}{\partial c} S_7 + S_6 \frac{\partial S_7}{\partial c} = -S_1 S_7 + S_6 S_7 S_4 \qquad (B.4)$$

Now

$$\frac{\partial H_2}{\partial c} = N \frac{\partial S_7}{\partial c} + (a-1) \frac{\partial S_3}{\partial c}$$

$$\frac{\partial S_7}{\partial c} = S_7 (-\frac{1}{N}) \frac{\partial S_5}{\partial c} = (1/N) S_4 S_7$$

and

$$\frac{\partial S_3}{\partial c} = S_8$$

where

$$S_8 = \Sigma (x_i - c)^{-1} [\ln \frac{x_i - c}{d}]^{-2}$$

Upon substitution of these results into the expression for $\partial H_2 / \partial c$, one obtains:

$$\frac{\partial H_2}{\partial c} = S_4 S_7 + (a-1) S_8 \qquad (B.5)$$

Similarly, one can obtain the following expression:

$$\frac{\partial H_3}{\partial c} = S_9 (S_7 - 1) + S_1 S_4 S_7 / N + (a-1)(S_{10} + S_{11}) \qquad (B.6)$$

where

$$S_9 = \Sigma (x_i - c)^{-2}$$

$$S_{10} = \Sigma [(x_i - c) \ln \frac{x_i - c}{d}]^{-2}$$

and

$$S_{11} = \Sigma (x_i - c)^{-2} [\ln \frac{x_i - c}{d}]^{-1}$$

To compute the partial derivatives with respect to d, the following equation is used repeatedly:

$$\frac{\partial}{\partial d} \ln(-\ln \frac{x-c}{d}) = -[d \ln \frac{x-c}{d}]^{-1}$$

Now from eq. 36,

$$\frac{\partial H_1}{\partial d} = \frac{\partial S_6}{\partial d} S_7 + S_6 \frac{\partial S_7}{\partial d}$$

$$= -(N/d) S_7 + \frac{1}{N} S_6 S_7 \frac{\partial S_5}{\partial d}$$

Since

$$\frac{\partial S_5}{\partial d} = \Sigma \ d \ln \frac{x_i - c}{d}]^{-1} = \frac{1}{d} \Sigma [\ln \frac{x_i - c}{d}]^{-1} = \frac{S_3}{d}$$

So:

$$\frac{\partial S_7}{\partial d} = (Nd)^{-1} S_3 S_7$$

$$\frac{\partial H_1}{\partial d} = (Nd)^{-1} S_3 S_6 S_7 - (N/d) S_7 \qquad (B.7)$$

Similarly, from eq. 37

$$\frac{\partial H_2}{\partial d} = N \frac{\partial S_7}{\partial d} + (a-1) \frac{\partial S_3}{\partial d}$$

$$= (1/d) S_3 S_7 + (a-1) S_{12}/d \qquad (B.8)$$

where

$$S_{12} = \Sigma [\ln \frac{x_i - c}{d}]^{-2}$$

Finally

$$\frac{\partial H_3}{\partial d} = S_1 \frac{\partial S_7}{\partial d} + (a-1) \frac{\partial S_4}{\partial d}$$

After some simple manipulations, one can obtain the following equation:

$$\frac{\partial H_3}{\partial d} = (Nd)^{-1} S_1 S_3 S_7 + (a-1) S_8 / d \qquad (B.9)$$