Stoch Environ Res Risk Assess (2005) 19: 249–257
DOI 10.1007/s00477-004-0229-x

ORIGINAL PAPER

D. L. Fitzgerald

Analysis of extreme rainfall using the log logistic distribution

Published online: 3 February 2005

Abstract Computer-intensive methods are used to examine the fit of the log logistic distribution to annual maxima of Irish rainfall. The characteristics of the L-moment solutions are examined by using the conventional bootstrap on the data and by random sampling within the ellipse of concentration of the parameter estimates. A statistical method of examining uncertainty is provided by the maximum product of spacings solution. Factors derived from random division of an interval are proposed for estimation of short-duration falls for which no data are available.

Keywords Log logistic distribution · L-moment estimate · Maximum product of spacings method · Bootstrap · Ellipsoid of concentration · Geometric max-stable distributions

Introduction

In the Flood Estimation Handbook (1999) the generalised logistic distribution is preferred to the generalised extreme value distribution for flood estimation in the UK. The log logistic (LLG) distribution is a reparameterisation of the generalised logistic. Its three-parameter version (3pLLG) was employed in a study of river flows in Scotland (Ahmed et al. 1988) but, in spite of its theoretical attractiveness it has been little used in more recent flood or extreme rainfall studies. The L-moment solution of its three parameters is examined and then recast by fixing the location parameter; then the shape and scale parameters have a clear interpretation in terms of the data and, more importantly, their maximum likelihood (ML) and L-moment estimates coincide. Pooling L-moment solutions is treated in detail in Hosking and Wallis (1997). In this paper the focus is on examining the stability of the L-moment quantile estimates at individual stations under variation of the data and of the parameters. The statistical tools used are: (1) the conventional bootstrap to sample the data with replacement (2) the ellipsoid of concentration (Cramèr, 1946) which gives a method of randomly varying the parameters within bounds set by their covariance matrix. (3) The maximum product of spacings solution (Cheng and Amin 1983); for a sample of size n (x_1, x_2, ..., x_n) it is equivalent to the ML solution of a more extreme sample of size $(n+1)$ with $x'_1 < x_1$, $x'_2 < x_2$, ...$x'_n < x_n$, $x'_{n+1} > x_n$.

D. L. Fitzgerald
Met Éireann, Glasnevin Hill, Dublin 9, Ireland
E-mail: denis.fitzgerald@met.ie
Fax: +353-1-8064247

The log logistic distribution

This is of form

$$f(x) = \frac{b}{a}\frac{((x-g)/a)^{b-1}}{\left(1+(x-g/a)^b\right)^2}, \quad x > g\ a, b > 0 \tag{1}$$

and b ln $[(x-g)/a]$ has a logistic distribution of mean 0. Eq. 1 implies that the cumulative distribution function is:

$$F(x) = \frac{((x-g/a))^b}{\left(1+((x-g)/a)^b\right)} \tag{2}$$

This shows it as a special case of a Burr XII distribution (Johnson and Kotz 1970). Its inverse is given by

$$x = g + a\left[\frac{F(x)}{(1-F(x))}\right]^c = g + a(T-1)^c \quad c = 1/b > 0 \tag{3}$$

where $T = 1/[1 - F(x)]$ is the return period. It follows from Eq. 1 to Eq. 3 that

$$\text{Median} = g + a, \quad \text{mean} = g + \frac{ac\pi}{\sin(c\pi)},$$
$$\text{mode} = g + a\left(\frac{1-c}{1+c}\right)^{c} \tag{4}$$

To a good degree of approximation mode = 2(median)−mean for values of c less than 0.4; this covers all but the most extreme growth rates. Also from Eq. 2

$$c = 1 - 2F(\text{mode}) \tag{5}$$

The shape parameter c is thus the measure of skewness proposed in Bickel(2002) and using a plotting position formula we can get a first estimate of c. As Eq. 4 shows mode < median < mean and the skewness is positive.

Since $b\ \ln((x - g)/a)$ has a logistic distribution of mean 0 the scale parameter a can be estimated as the geometric mean of $(x{-}g)$ for the sample.

L-moment estimates of the parameters

The theory of L-moments and of the closely related probability weighted moments (PWM) is given in Hosking and Wallis (1997). With

$$l_1 = E(x),\ \ l_2 = 2E(xF) - E(x),$$
$$l_3 = 6E(xF^2) - 6E(xF) + E(x)$$

we get from Eq. 3

$$c = \frac{l_3}{l_2} = \text{ L - moment measure of skewness} \tag{6}$$

$$a = l_2 \frac{\sin(\pi c)}{\pi\ c^2} \tag{7}$$

$$g = l_1 - \frac{l_2^2}{l_3} \tag{8}$$

When g is known the two-parameter log logistic theory applies to $(x\text{-}g)$ giving

$$c = \frac{l_2}{l_1} = \frac{l_3}{l_2} \tag{9}$$

i.e. the shape parameter = the L-moment coefficient of variation = the L-moment coefficient of skewness (Fitzgerald 1996).

Since $x > g$, the location parameter might, ideally, be expected to be a realistic estimate of the minimum of the annual rainfall maxima; in some cases it is but more often it is appears low and on occasion quite acceptable estimates of rainfall quantiles occur with a negative value of g.

The maximum likelihood (ML) parameter estimates

For a sample of size n the ML equations may be written

$$c = \frac{1}{n}\sum_{i=1}^{i=n}(2F(x_i) - 1)\ln(x_i - g) \tag{10}$$

$$0 = \sum_{i=1}^{i=n}(2F(x_i) - 1) \tag{11}$$

$$c = \frac{\sum_{i=1}^{i=n}(1 - 2F(x_i)/(x_i - g))}{\sum_{i=1}^{i=n}(1/x_i - g)} \tag{12}$$

When g is known Eqs. 10 and 11 are the two-parameter ML equations. In the three-parameter case a value of the location parameter g was taken, e.g. the L-moment estimate, and Eq. 10 and Eq. 11 used to estimate the shape and scale parameters; in turn the new values of c and a were used to get a new value of the location parameter. While often successful, this method did not always converge but occasional failure

Table 1 Parameter and quantile estimates for 2-day (0900-0900UCT) annual maximum rainfalls (mm)

Estimation method	Parameters			Return periods			
	c	a	g	RP 2	RP 100	RP 500	RP 1000
L-moment	0.225	32.4	17.0	49.3	108.1	148.3	170.5
ML	0.288	25.5	23.5	49.0	119.2	176.0	209.7

BOOTSTRAP

Estimation method	Statistic	Return Periods			
		RP 2	RP 100	RP 500	RP 1000
L-moment	Mean	49 ± 2	107 ± 11	146 ± 23	168 ± 31
	Median	49.4	106.1	144.1	165.1
ML	Mean	50 ± 2	105 ± 16	142 ± 28	163 ± 35
	Median	49.7	103.7	139.1	158.1

Station: Ffoulkesmills
Years of record = 62
Sample max = 105.2; Sample min = 35.5; Sample median = 48.2

is to be expected from Eq. 10 to Eq. 12; the most usual case of non-convergence was with g strongly negative and c low.

For quantile estimation beyond the range of the sample it is unwelcome that Eq. 10 gives high weight to low as well as to high values of $(x\text{-}g)$ in the estimation of the crucial shape parameter.

The information matrix

$$I = -\frac{1}{n}E\left((\partial^2 L)/\partial\theta_i, \partial\theta_j\right), \ \ \vartheta(c,a,g)$$

is:

$$\begin{bmatrix} \frac{\pi^2+3}{9c} & 0 & \frac{-\pi c(3c+(1-c^2)(\psi(2-c)-\psi(2+c)))}{3ac^2} \\ 0 & \frac{1}{3a^2c^2} & \frac{(1-c^2)\pi c}{3a^2c^2} \\ \frac{-\pi c(3c+(1-c^2)(\psi(2-c)-\psi(2+c)))}{3ac^2} & \frac{(1-c^2)\pi c}{3a^2c^2} & \frac{2\pi c(1-c^2)}{3a^2c^2} \end{bmatrix} \quad (13)$$

Here ψ is the digamma function.

Comparison of the L-moment and ML solutions for the three-parameter LLG

Table 1 presents a typical case of a good ML fit which generally has a higher value of the crucial shape parameter than the L-moment estimate. Bootstrapping the data to generate 5,000 samples gave much-reduced ML estimates at the higher return periods, closer to the L-moment estimates which were largely unchanged. However, nearly half the bootstrap samples did not converge in a fully satisfactory manner and so all that can safely be gleaned from this exercise is that the ML solution is not stable under bootstrapping. Only with 12 of the 10,000 L-moment solutions was it considered desirable to use the logistic instead of the LLG. Failures tended to occur in samples having a mean less than the median and hence unsuited to an LLG fit. Typically, for such samples the ML solution did not converge or produced estimates which were distinctly low; the L-moment solution produced low or even negative values of the location parameter but usually acceptable estimates of the rainfall quantiles.

Table 2 provides an example of an L-moment solution where the location parameter looked implausible and ML solution converged smoothly.

Note that the sample mean is less than the median and we may anticipate difficulty in fitting the LLG distribution. The three parameter L-moment solution has a very low value of the location parameter. The ML solution converged with a distinctly more realistic value of g. Bootstrapping provided L-moment estimates still well in accord with the quantile estimates and problems occurred with less than 12% of the 10,000 samples. As in Table 1 the ML bootstrap estimates reduce at high return periods and have a higher variance than the L-moment estimates. Again there is a strong element of subjectivity in the ML bootstrap results since there were convergence problems with nearly 30% of the 5,000 samples.

It seems clear that the L-moment estimates are the more stable. From the bootstrap means and standard errors it would also seem that they give fairly realistic estimates of the 1,000-year rainfall quantile for an individual station with about 50 years of record.

The 2p/3p solution

From (1) setting the location parameter g at some fixed value is an attractive option. For the two-parameter LLG the requirement that the L-moment coefficient of

Table 2 Parameter and Quantile estimates for 2-day (0900-0900UCT) Annual Maximum Rainfalls(mm)

Estimation Method	Parameters			Return Periods			
	c	a	g	RP 2	RP 100	RP 500	RP 1000
L-moment	0.151	58.9	04.5	63.9	122.2	154.7	171.2
ML	0.226	39.6	23.3	62.9	135.3	184.7	212.2

BOOTSTRAP

Estimation Method	Statistic	Return Periods			
		RP 2	RP 100	RP 500	RP 1000
L-moment	mean	64 ± 3	121 ± 12	153 ± 24	169 ± 32
	median	63.9	120.5	151.7	167.6
ML	mean	64 ± 3	125 ± 17	162 ± 37	182 ± 49
	median	63.6	122.9	157.5	175.6

Station: Brosna (Mt Eagle)
Years of record = 55
Sample max = 118.2; min = 38.4; median = 66.3; mean = 65.7

variation (*l-cv*) and the L-moment coefficient of skewness (*l-sk)* be equal is a key test of fit. A second test is that the scale parameter should equal the geometric mean of *(x-g)*. From a plot, the mode of a typical 30- to 60-year sample can usually be seen to fall within a certain range; the estimates of the mode provided by 2(median) − mean and Eq. 4 should agree closely and both should fall within the indicated range. For annual maxima of Irish rainfall the fit with $g=0$ is poor. However, a non-zero value of g can be found which gives a good fit. In the course of systematically varying g it was found that:

(a) The value of g obtained when *l-cv* = *l-sk* was never appreciably different from that of the three-parameter L-moment solution.
(b) The *l-sk* estimate of c was nearly constant as g varied.
(c) The *l-cv* and ML estimates of c were broadly similar and both varied systematically with g.
(d) A good fit with a common value of g could be found where the ML and the L-moment solutions for the shape and scale parameters essentially coincided; these values of g, a, c were in all cases so close to the three-parameter L-moment solution that there was no practical difference between their quantile estimates.

Based on (d) the model for annual maxima adopted was that the exceedances of a fixed, if initially unknown, threshold g follow a two-parameter LLG distribution; it is convenient that this amounts to the three-parameter L-moment solution with its location parameter g reinterpreted. An advantage is that the two-parameter ML theory now applies and Eq. 13 simplifies to give for a large sample of size n

$$\mathrm{var}(c) = \frac{9c^2}{n(\pi^2+3)}; \quad \mathrm{var}(a) = \frac{3a^2c^2}{n}; \quad \mathrm{covariance}(a,c) = 0 \tag{14}$$

Using Eqs. 3 and 14 the variance of the rainfall with return period T is:

$$\mathrm{var}(x_T) = (x_T - g)^2\left[\mathrm{var}(c)(\ln(T-1))^2 + \frac{\mathrm{var}(a)}{a^2}\right] \tag{15}$$

Ellipsoids of concentration and random variation

Cramèr (1946) shows that for a random variable ξ ($\xi_1, \xi_2, \ldots, \xi_n$) of known covariance matrix an ellipsoid of concentration can be defined with the property that randomly choosing a large number of points within the ellipsoid yields a random variable with the same mean and covariance matrix as ξ. The classical assumption is of a parent population distribution with the parameters estimated to a degree of accuracy determined by the sample size n; the ellipsoid of concentration re-expresses this in terms suited to computer simulation by defining a volume in parameter space within which all points can be taken as equally likely. Belonging to the population may then be taken as being within the ellipsoid of concentration and a random sample of n regarded as n random choices within the ellipsoid.

In the case of the preferred solution for the LLG the ellipsoid reduces to an ellipse having the parameter estimates *(a, c)* as the mean of the random variable and with covariance matrix given by Eq. 14. Since the random variation is of the parameters, the effect on the mean, median and variability of the quantile estimates

Table 3 Comparison of rainfall estimates (mm)

RP	Log logistic					Log logistic			Generalised extreme value				
Years	L M	ML	L-moment bootstrap			Ellipse of concentration			PWM	ML	PWM bootstrap		
			Mean	SE	Median	Mean	SE	Median			Mean	SE	Median
(Duration = 1 clock hour Sample max = 22.6 Sample min = 7.2)													
50	22.5	22.7	22.1	2.3	22.1	22.6	2.7	22.4	22.4	21.6	21.9	2.3	22.0
100	26.0	26.4	25.4	3.1	25.6	26.3	3.8	26.0	25.4	24.1	24.8	3.1	24.8
500	37.3	38.2	36.1	6.6	36.2	38.1	7.7	37.2	33.5	30.4	32.4	6.2	32.3
1000	43.9	45.2	42.4	9.0	42.3	45.1	10.2	43.7	37.7	33.4	36.4	8.2	36.1
(Duration = 24 clock hours Sample max = 94.0 Sample min = 34.2)													
50	94.1	101.6	92.8	7.3	92.9	94.4	8.2	94.0	92.4	91.4	91.0	7.4	91.2
100	104.5	116.1	102.9	9.8	103.0	104.9	10.5	104.4	99.7	98.7	98.2	9.7	98.3
500	133.0	159.6	130.7	18.9	130.0	133.9	17.3	132.8	116.0	115.0	114.5	16.8	113.6
1000	147.4	183.8	145.0	24.6	143.6	148.7	21.1	147.1	122.7	121.8	121.5	20.7	119.9
(Duration = 96 clock hours Sample max = 136.8 Sample min = 53.8)													
50	141.6	146.5	139.4	10.1	140.1	142.0	11.8	141.3	139.4	136.0	137.4	10.2	138.2
100	156.5	164.2	154.1	13.0	154.8	157.1	15.1	156.3	150.2	145.4	147.8	13.2	148.6
500	198.2	216.1	194.6	25.1	194.9	199.7	25.7	197.9	174.9	166.2	172.2	23.0	172.2
1000	219.7	244.2	215.8	32.8	215.2	221.8	31.7	219.2	185.3	174.8	182.9	28.4	181.9

Station: Cork Airport
Data: Annual maxima 1963–2002
L-M L-moment, *PWM* probability weighted moment, *ML* maximum likelihood, *SE* standard error, *RP* return period
Generalised extreme value: $x = \xi + \alpha(1 - (-\ln F)^k)/k$

can be compared with the preferred solution estimates, the three-parameter ML estimates and the values obtained from bootstrapping the data and using the L-moment solution. The results are presented in Table 3. For purposes of comparison Table 3 includes the probability weighted moment, the ML and the bootstrap PWM estimates for the generalised extreme value (GEV) distribution (Hosking and Wallis 1997). The ellipse of concentration statistics are based on 20,000 sample points.

Remarks on Table 3

- The GEV displays a lower growth rate but even at the 1,000-year RP the estimates are within 1.5 standard errors of the LLG estimates.
- Both the bootstrap and the ellipse of concentration give broadly similar quantile estimates for the LLG and both their medians and means agree well with the L-moment values.
- The very close agreement between the median of the ellipse of concentration estimates and the L-moment values at all return periods allows the identification of the L-moment solution with a central value of the well-defined population based on the sample data.

Fit of the estimates to the top sample values

For some rough guidance on the return period T_m associated with the higher order statistics of a sample of size n note that the distribution of $y = T_m - 1$ for the mth order statistic of any distribution is given by

$$f(y) = \frac{1}{B(m, n-m+1)} y^{m-1}(1+y)^{-(n+1)} \tag{16}$$

where B is the beta function and $f(y)$ an inverted beta distribution.

From Eq. 16$E(T_m) = (n/(n-m))$. The modal value $T_{\text{mode}} = (n+1/(n-m+2))$. The median can be obtained from Eq. 16 using incomplete beta functions. A restatement of Eq. 16 is that the CDF of the mth order statistic $F(x_m) = y/(1+y)$ has a beta distribution and Jenkinson (1977) exploited this to show that its median plotting position can be derived as

$$\tilde{F}(x_m) = \frac{m-0.307}{n+0.386}.$$

In turn this gives the good approximation:

$$T_m(\text{median})\tilde{T}_m = \frac{n+0.386}{n-m+0.693} \tag{17}$$

For the LLG the median value of x_m is given by the incomplete beta function expression

$$\frac{1}{2} = \frac{B_{y_m/1+y_m}(m, n-m+1)}{B(m, n-m+1)}$$
$$\text{where } y_m = \left(\frac{(x-g)}{a}\right)^b = T_m - 1 \tag{18}$$

and so, to the same approximation as in Eq. 17, we get

$$x_\text{m} = g + a(\tilde{T}_\text{m} - 1)^c, \quad c = \frac{1}{b} \tag{19}$$

As m approaches n Eq. 16 is a highly skewed distribution. Hence the median seems by far the best central value to select for the purpose of gauging the fit of any distribution by examining the quantile estimates for the top sample values.

Remarks on Table 4.

- For both distributions the median plotting position is used to make the estimates. The fit to the sample values is good for 1 h and 24 h durations; for 96 h the fit is greatly improved by taking the fourth highest value (in brackets) into account. While it might be observed that they agree more closely with one another than with the data and that the GEV

Table 4 Comparison of top three sample values with median estimates

Duration	Distribution						
Clock hours	Log logistic			Generalised extreme value			Sample
	Parameter		Estimates	Parameter		Estimates	
01	c	0.272	23.3	k	−0.153	23.0	22.6
	a	5.9	19.6	α	2.2	19.5	21.3
	g	5.6	17.7	ξ	10.6	17.9	18.5
24	c	0.148	96.5	k	+0.035	94.2	94.0
	a	53.4	84.1	α	12.2	84.4	87.3
	g	−0.80	78.1	ξ	8.0	79.1	78.7
96	c	0.163	145.4	k	+0.011	141.8	136.8
	a	65.4	127.7	α	16.2	127.7	130.2
	g	18.8	119.4(113.9)	ξ	77.5	120.2(115.0)	129.3(115.7)

Station: Cork Airport
Data: annual maxima 1963–2002
Generalised extreme value $x = \xi + \alpha\,(1 - (-\ln F)^k)/k$

appears to fit the data a little better than the LLG, the notion of the LLG estimates as median values is supported.

– For 1 clock hour we have an instance of the location parameter g as a plausible minimum while the minimum for the GEV is −3.7. In general we may reasonably expect the growth rate to decrease or remain steady with increasing duration greater than 1 h or 2 h but, in practice contradictions between return periods arise e.g. here Table 3 gives 147.4 as the L-moment estimate of the 1,000-year rainfall quantile for 24 clock hours while the 18-h value is 149.3. Table 4 shows that this is caused by the negative value of the location parameter which forces a high value of the scale parameter and a low value of the shape parameter. The GEV estimates suffer from the same lack of consistency as they give a 1,000-year rainfall quantile of about 122 mm for 24 clock hours and an 18-h value of over 125 mm. This might fairly be taken as an indication of the inadvisability of estimating 1,000-year quantiles from forty data points but it is necessary to do so while imposing consistency over durations and return periods. The ML solution for the 3-parameter LLG converges to give 0.226, 35.1 and 17.1 as the estimates for c, a, g respectively and gives an 1,000-year quantile of 184 mm. However, its estimate of the 50-year quantile as 102 mm appears a little high as the median RP associated with the 40-year sample maximum of 94 mm is over 58 years. The location parameter may be considerably increased without much change to the estimates in Table 3 and in Table 4. The method, for a given value of the location parameter, is to first take the mean of the two estimates of the shape parameter in Eq. 9 and to use that in Eq. 7 to determine the scale parameter. Of itself this gives RP estimates quite close to the preferred solution estimates. Taking $g = 15$ we get estimates of 0.177 and 37.6 for the shape and scale parameters while the two-parameter ML solution gives 0.212 and 37.3. Giving the two solutions equal weight we get estimates of 95, 107, 141 and 159 for Table 3 and 97.6, 84.1 and 77.7 for Table 4, thus ironing out the contradiction between 18 h and 24 h estimates and retaining a decent fit to the top sample values. This, admittedly ad hoc, procedure can be used instead of or in conjunction with a depth-duration-frequency scheme [Flood Estimation Handbook (1999), vol 2, Chap. 10].

Maximum product of spacings (MPS) solution as a statistical safety factor

For parameter estimation of univariate distributions the MPS method of maximising the geometric mean of the sample spacings was developed by Cheng and Amin (1983) as an alternative to maximum likelihood. It was shown in Fitzgerald (1996) that the MPS method applied to a sample of size n is equivalent to using ML for a sample of size $(n+1)$ where $x_1' < x_1, x_2' < x_2, \ldots, x_n' < x_n, x_{n+1}' > x_n$. It was further shown for the two-parameter LLG that MPS gives a positively biased estimate of the shape parameter while the scale parameter is essentially equal to the ML or L-moment estimates. The location parameter of the preferred solution of Sect. 2.4 is accepted for the MPS; then, the median estimate of the two solutions is the same but for higher return periods MPS gives more extreme quantile estimates.

Maximum product of spacings chooses the sample spacings closest to a perfectly regular pattern and hence is a typical index of lack of clustering or ignorance. Compared with the estimates of the preferred solution it produces quantiles that are in line with climate change scenarios for mid latitudes, the amount of the increase being determined by what amounts to an application of the theorem of the mean to the sample values. Table 5 shows the results of using MPS estimates of the shape and scale parameters and their variances in an ellipse of concentration exercise as outlined in Sect. 3; the MPS results are very much along the lines of those found for the L-moment solution. The 100-year rainfall quantile increases from 90 to 107 ± 12 mm in what we may choose to regard as the new, median-conserving climate. The 1,000-year quantile increases from 157 ± 25 to about 210 ± 35; if the period 1941–2002 is used instead of the record for 1881–2002 the values become 173 ± 36 and 230 ± 54 respectively.

Table 5 Log logistic estimates of rainfall quantiles

RP (years)	L-moment	MPS	MPS ellipse of concentration: Mean	SE ±	Median
2	34.2	34.3	34.3	1.1	34.3
5	44.5	46.5	46.5	2.1	46.4
10	52.5	56.5	56.6	3.2	56.5
20	61.7	68.4	68.5	4.9	68.2
50	76.4	88.0	88.3	8.2	87.8
100	90.1	106.9	107.4	11.7	106.7
250	112.5	139.0	140.0	18.3	138.6
500	133.5	170.2	171.8	25.4	169.6
1,000	158.9	208.9	211.4	35.0	208.1

Station: Phoenix Park (Dublin)
Data: 1-day annual maxima 1881–2002
MPS Maximum product of spacings, *RP* return period, *SE* standard error

Table 6 Means and medians of maxima of the random division of the unit interval

	Number of intervals: 2	3	4	5	6
Mean	0.75	0.61	0.52	0.46	0.41
Median	0.75	0.59	0.50	0.44	0.39

Why use the log logistic distribution?

In the Flood Studies Report (1975) the equation for the rainfall growth curves is given by: $\ln \frac{M_T}{M_5} = c(y - 1.5)$, $M_T =$ T - year return period and $y = -\ln(-\ln(F(x)) \approx \ln(T - 0.5)$, $T > 5$

Here c= - k where k is the shape parameter of the GEV distribution and c ranges from 0.23 to 0.07.

For the LLG the corresponding growth curve is of strikingly similar form:

$\ln((x_T - g)/(x_5 - g)) = c(\ln(T - 1) - 1.4)$ where c is the LLG shape parameter.

The LLG distribution also exhibits geometric-extreme stability in the sense of Voorn (1987) i.e. if $F(x)$ is a continuous distribution independent of N, a distribution on the positive integers with generating function $Q(s)$, then F is maximum stable with respect to N if real numbers a and b exist such that

$$Q(F(x)) = \sum_{k=1}^{\infty} p_k F(x)^k = F\left(\frac{x-a}{b}\right), \quad b > 0 \tag{20}$$

If the number of rainfall events capable of producing the annual maximum is taken as geometric with generating function $Q(s) = (ps)/(1 - (1 - p)s)$i.e. $p_k = p(1 - p)^{k-1}$, k=1,2, ... and the total rainfall from the individual events is LLG(c, a, g) then

$$Q(F(x)) = F\left(\frac{x}{u}\right), \quad u = p^{-c}, \quad u > 1 \tag{21}$$

Hence over years with random numbers, $N(p)$, of events with an LLG distribution the maximum is an LLG distribution of a rescaled variable.

However, the geometric distribution used here has mean $1/p$ but its mode is 1 and this seems low for Irish conditions as most years are nondescript, having several events capable of yielding the maximum. If instead we consider a series of annual maxima having an LLG distribution and, after the manner of Mitnik and Rachev (1993), take the number of years into the future that the process is likely to remain free of radical change as a geometric variable $N(t)$, then the argument applied above gives the CDF of the maximum of a sample of size $N(t)$ as LLG, most plausibly if p is small. Here t is the rather ill-defined and subjective probability of 'major change in any year' and $N(t)$ the waiting time to the change. We might estimate t from our degree of belief in the length of time the ellipse of concentration generated from the LLG fitted to the annual series will adequately represent the process generating the maxima. A belief that there was only a probability of 0.5 that the current model would apply after 50 years would imply t=0.013 and for a shape parameter c=0.20 this yields a scaling factor of 2.38 to allow for the uncertainty in the length of the period. The MPS method deals with uncertainty by increasing the shape parameter while the method of (geometric) random sample size alters the scale parameter of the LLG distribution. However, the point of most interest is that under both these randomisations the LLG (and also the logistic distribution) is maximum stable. If the negative binomial is used instead of the geometric we get

$$Q(F(x)) = \left(\frac{pF(x)}{1-(1-p)F(x)}\right)^r = F\left(\frac{x}{u}\right)^r,$$

a Burr III distribution (Johnson and Kotz 1970); this may be viewed as LLG with a rescaling of the data and of the time unit, especially if r is an integer since in this case r bears the interpretation of the minimum number of events in unit time.

Experience dictates that when choosing a probability distribution for rainfall extremes it is wise to allow for much higher values than appear even in a long-term record. This leads to the choice of long-tailed distributions with shape and scale parameters. Among these the LLG has the advantage that, as Eqs. 4 and 5 show, it can be completely determined by three central values, leading to the hope that population characteristics may be accurately determined from a sample. Unfortunately the sample sizes in this study are rather small and the mode is not well-determined but may be inferred from the mean and median. The ellipse of concentration argument in Sect. 3 shows that the preferred solution closely mimics the median of a plausible representation of the population giving rise to the data. If, further, the parameter estimates of the preferred solution yield close approximations to the sample estimates of the median and mean then its fit to the LLG is regarded as good; this solution is computationally inexpensive and the simplicity of the underlying theory makes it easy to investigate its properties either directly or by computer-intensive methods.

A point to emphasise is that the fit of the pure two-parameter LLG to the rainfall data was mostly poor and the move away from g=0 crucial to obtaining a good fit. Unfortunately for their applicability to this study, all the recent papers on the LLG in hydrology that have come to my attention have been on the pure two-parameter case. An example of the latter, remarkable for the depth of negative feeling expressed towards the 2pLLG, is the paper of Rowiński et al. (2002). They concentrate on the analysis of the tail region and for this study the most pertinent of their many reservations is the necessity to restrict the ranges of the parameters if population moments are to be kept finite; the practical implications of this are most easily discussed in terms of the following equation derived from Eq. 3:

$$E\left(\left(\frac{x-g}{a}\right)^r\right) = B(1 + rc, 1 - rc), \quad r = 1, 2, 3, \ldots, \quad 0<c<1,$$

B being the beta function.

In this form the fractional contribution to the population moment of the range 0–$F^{-1}(u)$ is given by an incomplete beta function. Also for a finite third moment we must have a shape parameter less than one-third and so on. This sort of restriction on the range of the shape parameter constitutes a difficulty for full tail analysis. For statistical practice, it is important to appreciate that with $c = 0.33$ about 90% of the value of the third moment is contributed by values with return periods of 10^5 years(time units) or more; with $c = 0.333$ the figure is 99%. Thus, the approach of the population moment of order three to infinity is mediated by values very unlikely to appear in samples. Clearly, in parameter estimation it would be unwise to employ third order moments. The restriction thus points up the advantages of the L-moment solution.

Values of the shape parameter c encountered in this study range from 0.1 to 0.33, with values of 0.33 to 0.4 occasionally appearing for short durations where growth rates are steeper. At $c = 0.4$ the population mean receives a 12% contribution from values with return periods of 50 years or more while for 100 years and 1,000 years the figures are 8% and 2% respectively. For $E(xF^r)$ these percentages rise only slowly with r . The indication is that, while in no way a drastic problem even at $c = 0.4$, low values of the shape parameter can be more accurately determined from a sample than high. In addition Eqs. 14 and 15 show how the variance of the quantile estimates increases with c. Values of $c > 0.33$ were treated with caution and, insofar as possible, checked for consistency against higher and lower durations, as in the next section.

Short-duration falls and random division of the unit interval

The sub-hourly falls available are predominantly from Dines rainfall recorders (HMSO 1981). Durations from 15 min to 24 h have been extracted from 42 stations. Fitting the LLG sometimes gives high growth rates, with values of the shape parameter in excess of 0.35, especially for durations ranging from 15 min to 3 h; checking for consistency of the quantile estimates raised the question of the possibility of finding scaling factors between durations. More importantly, urban drainage requires design values for durations less than15 min and for such short durations no reliable long-term data are available. When the sub-period can be considered part of a single event the theory of the random division of an interval can contribute plausible scaling factors. The distribution of the maximum of n intervals on (0, 1) (David 1981) gives the following.

That the factors are significant can be seen by checking them against the 5-year return period factors in Table II of Logue (1975) or Tables 3.6 and 3.7, vol 2, Flood Studies Report (1975), especially the mid-table values. However, their derivation as central values of the distribution of a maximum suggests that they apply more widely. Since, for example, it must be possible for more than 52% of the total to fall in one quarter of the time period the factors should apply better at low than at high return periods. Over Ireland the 500-year quantile for 4-h falls is usually less than 80 mm; it is possible to get that amount in 1 h (Rohan 1986) and so, in practice, high may mean about 1,000 years for some durations. In the particular case of the LLG using the factor to adjust the scale and location parameters and assuming the shape parameter is unchanged is equivalent to directly multiplying the estimates by the scaling factors; this may be attributed to the fact the parameters are determined by the median, mode and mean and these three may reasonably be taken to scale in the same way. If there is a difference Δc between the shape parameters of the two durations it appears as a multiplier $(T - 1)^{\Delta c}$ when estimating the T-year quantile; however, in the absence of data for the shorter duration the assumption that the shape parameter is constant seems a reasonable first approximation. Table 7 presents the estimates of 5-min quantiles from 30- and 15-min values.

Note that the 15-min estimates and 0.75 times the 30-min estimates agree to within one standard deviation even at the1,000-year return period, the scaled down 30-min values being systematically higher. The consistency between the 5-min estimates is heartening but their accuracy is difficult to gauge. It has been standard practice at Irish synoptic stations to extract daily maximum rates for 5 min but, as Logue (1975) points out, the Dines recorder is not accurate below about 15 min. Maximum values for 5 min from Dines recorders range from 12 mm to about 8 mm over a period of record mostly of 40–50 years. Fairly near Rosslare there was a remarkable fall of 38 mm in 12 min in August, 1949 (Rohan 1986); this strongly suggests that 5-min falls of over 20 mm occur in the general area but attaching a return period at a single point must be labelled a hopeful exercise and it seems a good idea to employ a regionalisation scheme to supplement single station analysis.

Table 7 Five-minute quantile estimates from 15 min and 30 min values

LLG estimates			Table 6 estimates	
RP Years	15 min mm	30 min mm	5 from 15 mm	5 from 30 mm
2	7.2 ± 0.4	9.8 ± 0.6	4.4	4.0
5	9.4 ± 0.6	12.9 ± 0.9	5.7	5.3
10	11.1 ± 0.8	15.5 ± 1.2	6.8	6.4
20	13.1 ± 1.1	18.5 ± 1.5	8.0	7.6
50	16.1 ± 1.7	23.3 ± 2.2	9.8	9.6
100	18.9 ± 2.3	27.9 ± 2.9	11.5	11.4
250	23.4 ± 3.5	35.3 ± 4.5	14.3	14.5
500	27.5 ± 4.8	42.4 ± 6.4	16.8	17.4
1000	32.5 ± 6.6	51.1 ± 9.0	19.8	20.9

Station: Rosslare
Data: 45 annual maxima for 15 min and 30 min (millimetres)
Distribution fitted: log-logistic with shape, scale and location parameters (c, a, g)
15-Min: $c = 0.254$, $a = 5.3$, $g = 1.9$; 30-min: $c = 0.285$, $a = 6.7$, $g = 3.0$

Discussion and conclusions

The parameters of the log logistic distribution are readily interpretable in terms of the data. Its form and L-moment solution are such that it is well suited to computer-intensive investigations. Bootstrap and ellipse of concentration exercises prove very informative about the stability of the quantile estimates. Curiously, having gone to some pains to obtain a solution with attractive statistical properties, it was found that the parameters so obtained can be varied fairly considerably without much change to the quantile estimates; this is a useful property that can obviate the need for depth-duration schemes to iron out the occasional inconsistencies that arise at high return periods when a single station record is analysed over a range of durations.

Some idea of more extreme rainfall regimes may be gleaned from the maximum product of spacings solution or, even more vaguely, from the property of geometric-maximum stability.

An interesting and important digression was that in the examination of the fit of the log logistic distribution to falls of short duration D, scaling factors k based on the theory of the random division of an interval were found to give acceptable estimates for durations such as $D/2$, $D/3$, $D/4$, $D/5$, ... In terms of the log logistic distribution the degree of approximation implied is that an LLG(c, a, g) at duration D scales into an LLG(c, ka, kg) at the shorter duration.

Over a wide range of durations fitting the log logistic distribution to annual maxima of Irish rainfalls using the L-moment method of parameter estimation proved successful. In comparisons with estimates based on the generalised extreme value distribution the log-logistic values agree closely for return periods up to about 100 years and tend to be higher at longer return periods.

References

Ahmad MH, Sinclair CD, Werrity A (1988) Log-logistic flood frequency analysis. J Hydrol 98:205–224

Bickel DR (2002) Robust estimators of the mode and skewness of continuous data. Comput Stat Data Anal 39:153–163

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

Cramér H (1946) Mathematical methods of statistics. Princeton University Press

David HA (1981) Order statistics. Wiley, New York

Fitzgerald DL (1996) Maximum product of spacings estimators for the generalised Pareto and log-logistic distributions, Stochastic hydrology and hydraulics, vol 10. Springer, Berlin Heidelberg New York, pp 1–15

Flood Estimation Handbook (1999) Institute of hydrology, vols 1 and 2. Wallingford, UK

Flood Studies Report (1975) Meteorological studies, vol 2. NERC, London

HMSO (1981) Handbook of meteorological instruments. Her Majesty's Stationery Office, London

Hosking JRM, Wallis J R (1997) Regional frequency analysis (An approach based on L-moments), CUP

Jenkinson AF (1977) The analysis of meteorological and other geophysical extremes, MET O 13 Technical Note No 58, British Meteorological Office

Johnson NL, Kotz S (1970) Continuous univariate distributions-1. Wiley, New York

Logue JJ (1975) Extreme rainfalls in Ireland. Met Éireann, Dublin

Mitnik S, Rachev ST (1993) Modelling asset returns with alternative stable distributions. Econometric Rev 12(3):261–230

Rohan PK (1986) The climate of Ireland. Met Éireann, Dublin

Rowsiński PM, Strupczewski WG, Singh VP (2002) A note on the applicability of log-Gumbel and log-logistic probability distributions in hydrological analyses: I known pdf. Hydrol Sci J 47(1):107–122

Voorn WJ (1987) Characterization of the logistic and log-logistic distributions by extreme value related stability with random sample size. J Appl Prob 24:838–851