Journal of Hydrology, 147 (1993) 61–81

0022-1694/93/$06.00

[4]

Likelihood-based confidence intervals for estimating floods with given return periods

Eduardo Sávio P.R. Martins*, Robin T. Clarke

Instituto de Pesquisas Hidraulicas, Porto Alegre, RS, Brazil

(Received 18 November 1992; accepted 22 December 1992)

Abstract

This paper discusses aspects of the calculation of likelihood-based confidence intervals for T-year floods, with particular reference to (1) the two-parameter gamma distribution; (2) the Gumbel distribution; (3) the two-parameter log-normal distribution, and other distributions related to the normal by Box–Cox transformations. Calculation of the confidence limits is straightforward using the Nelder–Mead algorithm with a constraint incorporated, although care is necessary to ensure convergence either of the Nelder–Mead algorithm, or of the Newton–Raphson calculation of maximum-likelihood estimates. Methods are illustrated using records from 18 gauging stations in the basin of the River Itajai-Acu, State of Santa Catarina, southern Brazil. A small and restricted simulation compared likelihood-based confidence limits with those given by use of the central limit theorem; for the same confidence probability, the confidence limits of the simulation were wider than those of the central limit theorem, which failed more frequently to contain the true quantile being estimated. The paper discusses possible applications of likelihood-based confidence intervals in other areas of hydrological analysis.

Introduction

The literature on the estimation of floods of given return period is vast and growing. Aspects discussed concern the following topics, among others: choices amongst probability distributions to describe records of annual floods at a given site (Bobee, 1975; Kite, 1977; Bobee and Robitaille, 1977; Houghton, 1978); comparison of procedures for estimating the parameters of the selected distributions, using the available flood record for the site (Greenwood et al., 1979; Landwehr et al., 1979; Hosking et al., 1985a,b); use of concomitant data from neighbouring sites with longer records of annual floods (Matalas and Jacobs, 1962; Hirsch, 1982; Grygier et al.,

* Corresponding author.

1989); possible information gain through the use of historic flood marks (Benson, 1950; Condie and Lee, 1982; Cohn and Stedinger, 1987); possible information gain through the use of more than one flood event per year (Shane and Lynn, 1964; Cunnane, 1973, 1979; Todorovic, 1978); estimation of low-order moments of appropriate probability distributions for floods at a site without records, by the use of regression analysis on basin characteristics (Thomas and Benson, 1970; Stedinger and Tasker, 1985, 1986; Hosking et al., 1985b); the use of a mathematical model of rainfall–runoff processes to transform long rainfall records into an estimated discharge sequence, from which extreme floods can be abstracted. All of the techniques discussed under these headings have as their principal objectives the estimation, for a site with or without flow records, of the flood with T-year return period, i.e. the flood which is observed once in T years in the long run.

An aspect which commonly receives less attention is the calculation of confidence intervals for estimates of T-year floods. In addition to the estimated T-year flood (here denoted by $\hat{X}_0$) at a particular site, it may also be useful to have an interval which includes the true value X_0 with given confidence probability. However, where the question of confidence intervals is discussed at all, the practice is commonly to assume that the estimate of the T-year flood is approximately normally distributed by virtue of the central limit theorem. Approximate confidence intervals are then quoted as $\hat{X}_0 \pm z\sqrt{(\mathrm{var}\ \hat{X}_0)}$ where z is the appropriate normal deviate, or perhaps as $\hat{X}_0 \pm 2\sqrt{(\mathrm{var}\ \hat{X}_0)}$ for the 95% interval; this was essentially the approach of Lu and Stedinger (1992) using the generalised extreme value distribution. However, both computational power and the sophistication of computational packages continue to increase rapidly as years pass, and must be used to develop the practice of hydrology. The use of methods known to be less than fully efficient – such as the method of moments – can now rarely be justified solely on computational grounds. This paper explores some of the possibilities for the use of likelihood-based confidence intervals as part of the procedure for estimating floods with return period T years; the calculation of such intervals requires more computational effort than the simple normal approximation set out above, but is well within the capabilities of current-generation desk-top computers. The theory of likelihood-based confidence intervals, like the normal approximation mentioned earlier, also requires that the number (n) of years of record be large; however, the approximations involved in their calculation are quite accurate even for quite small values of n, when the normal approximations give inaccurate confidence limits (McCullagh and Nelder, 1990).

This paper reports upon the computing aspects of the calculation of likelihood-based confidence intervals for floods with T-year return period. The value of $T = 100$ years is assumed; other values could also have been used. Larger values of T would tend to increase computing times, and conversely, but the longer computing times for larger T are not likely to be substantial. As a secondary objective, the paper explores the performance of the Box–Cox family of transformations for the estimation of T-year floods at 18 gauging stations in a flood-prone basin in southern Brazil.

The statistical basis of likelihood-based confidence intervals

It is assumed that the observed sequence of n annual floods at a gauging station is a sample of size n from a probability distribution of known form $f(x, \theta)$, where θ is the vector of model parameters for the distribution $f(.)$. The vector θ is taken to have p elements, commonly $p = 2$ or 3. It is also assumed that annual floods in different years are statistically independent. Provided that the probability distribution is 'correct', all information in the annual floods from the n years of historic record is contained in the likelihood function $L(\theta; \boldsymbol{x})$, or equivalently in its logarithm $l(\theta; \boldsymbol{x})$ defined by

$$l(\theta; \boldsymbol{x}) = \ln\ L(\theta; \boldsymbol{x}) = \sum_{i=1}^{n} \ln f(x_i, \theta) \qquad (1)$$

In practice, many different forms of probability distribution may be consistent with the data, each giving rise to its own likelihood function. The different log-likelihoods $l(.)$, corresponding to the different candidate probability distributions, can also be used to determine whether, for example, a log-normal distribution affords a better representation of the data than a gamma distribution (Atkinson, 1985). This aspect is not considered in this paper, which assumes that the appropriate distribution form $f(x, \theta)$ has been identified by likelihood methods or some other procedure.

For the selected distribution $f(x, \theta)$, we denote by $\hat{\theta}$ the values of θ which maximise $l(\theta; \boldsymbol{x})$, satisfying $l(\hat{\theta}; x) \geq l(\theta; x)$ for all θ values lying within the space of feasible θ values. Clearly values of the unknown parameters θ that are close to $\hat{\theta}$ will be more consistent with the data (from which $\hat{\theta}$ was calculated) than values of θ which are remote from it; hence we can take as the confidence region all those values which lie 'sufficiently close' to $\hat{\theta}$. Statistical theory shows that, when n is large, the difference between the log-likelihoods at θ and at $\hat{\theta}$ is related to the χ^2 distribution. Formally, the difference

$$2\,[l(\hat{\theta}; \boldsymbol{x}) - l(\theta; \boldsymbol{x})] \qquad (2)$$

is approximately distributed as χ^2 with p degrees of freedom where, as indicated above, p is the number of elements in the vector of parameters θ. We can use this result to give a quantitative interpretation of 'sufficiently close'. In fact if α is a selected small probability, say 0.05 or 0.01, and if $\chi^2_{p,\alpha}$ is the tabulated value of χ^2 defining an upper-tail probability of α, then the set of θ values which satisfy the condition

$$2[l(\hat{\theta};\boldsymbol{x}) - l(\theta;\boldsymbol{x})] \leq \chi^2_{p,\alpha} \tag{3}$$

is an approximate $100(1-\alpha)\%$ confidence set for the parameters θ, and is usually more accurate in terms of coverage probability than intervals based on the normal approximation. The use of (3) is analogous to the use of the t-distribution for defining confidence intervals for the mean μ of a normal distribution; this statistic t is there used to define a region 'sufficiently close' to the sample mean, given by $|\bar{x} - \mu| \leq ts\sqrt{n}$.

Although the inequality (3) provides a confidence region for the set of parameters θ, of greater interest are the confidence intervals for the quantiles of the distribution $f(x,\theta)$: namely those values x_0 which satisfy the equation

$$\int_{\infty}^{X_0} f(x,\theta)\,\mathrm{d}x - P = 0 \tag{4}$$

where $P = 1 - 1/T$ in the case of the flood with return period T years. However, to every point θ in the region defined by (3) there corresponds a value X_0 which varies in one-dimensional (1-D) space as the parameter vector θ varies in the p-dimensional confidence region defined by (3). Thus by computing X_0 at points in this p-space, and identifying the largest and smallest X_0 values that result, we obtain a $100(1-\alpha)\%$ confidence interval for the T-year flood X_0.

The calculation of the confidence region for X_0 can be made more efficient if a sub-region of the p-space, within which θ varies, can be identified, that contains the maximum and minimum values of X_0. The following argument suggests how this can be done. Suppose that the distribution $f(x,\theta)$ is log-normal with two parameters μ and σ. Since the T-year flood for this distribution is the exponential of the T-year flood for a normal distribution with the same parameters, we can work with a normal distribution for which the p-space defined by (3) is a 2-D space, in which μ and σ vary. Intuitively, the larger μ becomes for fixed σ, the larger will X_0 become, unit increase in μ producing unit increase in X_0. If we increase σ with μ fixed, X_0 will also increase. Thus, as μ and σ vary over their 2-D confidence region given by (3), the function $X_0 = X_0(\mu,\sigma)$ increases with σ for fixed μ, and increases with μ for fixed σ. It follows that the maximum and minimum values of X_0 in the confidence region defined by (3) must correspond to points (μ,σ) lying on the boundary of that region. Hence, to obtain the maximum and minimum values of X_0 which define its $100(1-\alpha)\%$ confidence interval, it is sufficient to calculate $X_0 = X_0(\mu,\sigma)$ at points lying on the boundary of the $100(1-\alpha)\%$

confidence region for (μ, σ); it is not necessary to calculate $X_0 = X_0(\mu, \sigma)$ at points lying within that region.

The above intuitive argument can be formalised in the case of the log-normal distribution as follows. Denoting by G the expression on the left-hand side of (4), and differentiating both sides with respect to μ and σ, we have

$$(\partial G/\partial X_0)(\partial X_0/\partial \mu) + (\partial G/\partial u) = 0$$

$$(\partial G/\partial X_0)(\partial X_0/\partial \sigma) + (\partial G/\partial \sigma) = 0$$

and, on cancelling non-zero factors, we obtain

$$\partial X_0/\partial \mu = 1; \qquad \partial X_0/\partial \sigma = (X_0 - \mu)/\sigma \tag{5}$$

In practice we are concerned with upper-tail quantiles such that $X_0 > \mu$, so that both derivatives are positive. Hence the maximum and minimum values of X_0, as the parameters μ and σ assume different values throughout their joint confidence region, are given by

$$X_{\mathrm{U}} = \max X_0 \; : \; \partial X_0/\partial s = 0; \qquad X_{\mathrm{L}} = \min X_0 \; : \; \partial X_0/\partial s = 0$$

where s is measured along the curve defining the boundary of the confidence region in (μ, σ) space. Detransforming, the quantities $\exp(X_{\mathrm{U}})$ and $\exp(X_{\mathrm{L}})$ then define the confidence interval for the T-year flood in the original scale of measurement.

Similar arguments hold for the Gumbel distribution with cumulative distribution function $F(x, \mu, \sigma) = \exp\{-\exp[-(x - \mu)/\sigma]\}$, for which

$$X_0 = \mu - \sigma \ln(-\ln P) \tag{6}$$

whence $\partial X_0/\partial \mu = 1$ and $\partial X_0/\partial \sigma = -\ln(-\ln P)$, both positive quantities, and for the two-parameter gamma distribution.

Thus, for two-parameter flood frequency distributions $f(x, \theta)$, likelihood-based $100(1 - \alpha)\%$ confidence intervals for T-year floods X_0 can be derived by identifying the maximum and minimum values of the appropriate quantile of the distribution $f(x, \theta)$, as the point defined by θ moves along the curve defining the boundary of the joint confidence region for θ. When the distribution $f(.)$ has three parameters, a similar result holds, the confidence region for θ now being defined by a surface in 3-D space.

Computational considerations

Using the data to be described below, the calculation was restricted to the estimation of floods with return period 100 years, for which 95% confidence intervals were calculated. Three methods were explored. They are described for the case in which the distribution $f(.)$ of annual floods has two parameters, θ_1, θ_2; the methods generalise easily for three or more parameters.

Method I

The first method consisted simply of searching along transects, starting at the point $(\hat{\theta}_1, \hat{\theta}_2)$ on the likelihood surface where the maximum occurred. Thus, starting at the point $(\hat{\theta}_1, \hat{\theta}_2)$, small steps were taken parallel to the axes $\theta_1 = 0$ and $\theta_2 = 0$, and along lines inclined at 45° to these axes; at each step, the log-likelihood was calculated to ascertain whether the point lay within the confidence region for (θ_1, θ_2) or outside it, this region being defined by (3) with $p = 2$, $\alpha = 0.05$ in the case where the vector θ of parameters has two components, θ_1 and θ_2. Eight points were therefore identified, lying on or very close to the boundary of the confidence region. The values of θ_1, θ_2 at these eight points were substituted in Eq. (4), which was solved for X_0 by a Newton–Raphson iteration. The largest and smallest of the eight X_0 values then approximately defined the $100(1-\alpha)\%$ confidence interval (X_L, X_U) for the T-year flood. When proceeding along each transect, two step sizes were explored: in one case the step sizes were 0.001 times $\hat{\theta}_1$ and 0.001 times $\hat{\theta}_2$, and in the other case the step sizes were 0.002 times these values.

Confidence limits for the T-year flood given by the two step sizes differed by very little, all intervals being strongly asymmetric about the estimated value. Since this estimation procedure was a first step in an attempt to obtain confidence limits by a more automatic method, results are not quoted in detail. However, the method is straightforward and its computational efficiency can probably be improved substantially: for example, by varying the step size along each transect, by starting a new transect from the point at which the previous transect terminated, and so on.

Method II

This method consisted of both maximising and minimising the solution X_0 of Eq. (4), with respect to the parameters θ, subject to the constraint $g(\theta) = 0$ where, from (3)

$$g(\theta) = 2[l(\hat{\theta}; \boldsymbol{x}) - l(\theta; \boldsymbol{x})] - \chi^2_{P,\alpha}$$

The constraint ensures that the search for the maximum and minimum values of X_0 is made along the boundary of the confidence region for the parameters θ. The search was made by means of the Nelder–Mead simplex algorithm; since this finds the minimum of a function, the value X_U (that is, the point on the confidence region boundary at which X_0 is a maximum with respect to the parameters θ) was found by calculating

$$\min_{\theta} \left(-X_0(\theta)\{1 + [1 + g^2(\theta)]^{-1}\}/2\right) \tag{7}$$

the negative sign ensuring that the function is maximised. Similarly, the value of X_L, the point on the confidence region boundary for which X_0 is a minimum, was found by calculating

$$\min_{\theta} \{X_0(\theta)[1 + g^2(\theta)]\} \tag{8}$$

The Nelder–Mead algorithm terminated when (1) the change in every parameter of the vector θ was less than 0.001, or (2) the number of iterations of the simplex exceeded a predetermined value, commonly 500. Although the use of (8) for the calculation of X_L always converged without difficulty, calculation of X_U using (7) sometimes failed to converge. Therefore Method II was modified as follows.

Method III

X_L was calculated using (8), but X_U was calculated by replacing (7) by

$$\min_{\theta} \{[1 + g^2(\theta)]/X_0(\theta)\} \tag{9}$$

With this modification, the Nelder–Mead algorithm converged for all 18 basins in Table 1.

Methods other than those described above are possible but have not been explored. For example, one parameter, say θ_1, could be fixed, and the roots of $g(\theta_1, \theta_2)$ could be calculated, giving θ_2; X_0 could then be calculated, conditional on θ_1, and its maximum and minimum values determined as θ_1 takes a series of values.

Data for stations on the Rio Itajai, SC Brazil

To explore the calculation of likelihood-based confidence intervals for floods with T-year return period, the data were used from 18 gauging stations, with annual flow records ranging from 15 to 51 years, on the Rio Itajai in the south of Brazil. Except for five stations with 6 years of record or less, which were omitted, the data constitute the whole of the published flood archive for this basin. The Itajai flows to the Atlantic Ocean, draining a region of approximately 15000 km^2. The rivers Itajai do Oeste and Itajai do Sul combine to form the Itajai-Acu, which is joined by the tributaries Hercilio and Benedito on its left bank. Subsequently it joins the Itajai-Mirim to form

Table 1
Gauging stations on the Rio Itajai and its tributaries, used in flood frequency analysis

River and gauging station	Basin area (km^2)	n	λ	Filliben correlation	
				Before	After
Taio	1575	51	0.463	0.9934	0.9956
Pouso Redondo	130	32	0.573	0.9779	0.9873
Trombudo	432	20	0.641	0.9851	0.9854
Adago	163	19	−0.306	0.9076	0.9794
Itajai, Barracao	364	18	0.789	0.9911	0.9900
Itajai do Sul, Jar.	720	24	−0.137	0.9630	0.9888
Itajai-Acu, R. do Sul	5100	36	0.864	0.9940	0.9940
Hercilio, Ibirama	3314	48	−0.085	0.9118	0.9918
Neisse Central	195	24	0.314	0.9621	0.9835
Itajai-Acu, Apiuna	9242	50	0.000	0.9408	0.9943
Benedito, B. Novo	692	49	0.600	0.9842	0.9900
Benedito, Timbo	1342	44	0.600	0.9832	0.9916
Itajai-Acu, Indaial	11151	46	−0.025	0.9440	0.9934
Testo, R. do Testo	105	33	−0.389	0.9458	0.9875
Itajai-Acu, Itoupava	11719	15	0.638	0.9925	0.9943
Garcia	127	32	0.425	0.9890	0.9943
Luis Alves	204	36	1.072	0.9946	0.9951
Itajai-Mirim, Brusque	1240	44	−0.270	0.9122	0.9930

n is the number of years of record; the columns headed 'λ' and 'Filliben' are explained in the text.

the Itajai proper, 7 km before it meets the Atlantic Ocean. The lower reaches of the Itajai are subject to extreme flooding. In the floods of 1983 and 1984, floods associated with El Niño events caused river levels to rise by more than 20 m, resulting in extensive damage in the city of Blumenau, SC. Table 1 lists the stations used in the analysis together with the number of years of record.

Calculation of likelihood-based confidence intervals for *T*-year floods: gamma distribution with two parameters

The two-parameter gamma distribution

$$f(x, \mu, \kappa) = (x/\mu)^{\kappa-1} \exp(-x/\mu)/[\mu\Gamma(\kappa)] \qquad 0 \leq x < \infty \tag{10}$$

was fitted by maximum likelihood, with κ estimated by solving

$$\ln \kappa - \partial \ln \Gamma(\kappa)/\partial\kappa = \ln A - \ln G$$

A and G being the arithmetic and geometric means of the data at each gauging station. The method used was that due to Greenwood and Durand (1960). The

Table 2
Two-parameter gamma distribution: comparison of upper 95% confidence limits, $X_{\mathrm{U}}^{\mathrm{II}}$ and $X_{\mathrm{U}}^{\mathrm{III}}$, obtained using Methods II and III

River, gauging station	$X_{\mathrm{U}}^{\mathrm{II}}$ ($\mathrm{m}^3\,\mathrm{s}^{-1}$)	$X_{\mathrm{U}}^{\mathrm{III}}$ ($\mathrm{m}^3\,\mathrm{s}^{-1}$)	$g(\theta)^{\mathrm{II}}$	$g(\theta)^{\mathrm{III}}$
Taio	574.43	575.39	0.0164	0.0081
Pouso Redondo	162.59	162.54	0.0354	0.0178
Trombudo	263.32	263.22	0.0396	0.0198
Itajai do Sul, Jar.	848.47	848.13	0.0396	0.0197
Testo, R. do Testo	66.04	66.03	0.0227	0.0113

The values of the constraint $g(\theta)$ are shown at the point where the criterion (9) was minimised.

estimate of μ was then $\bar{x}/\hat{\kappa}$. If the distribution $f(.)$ of annual floods were known to be gamma distributed, the maximum-likelihood estimates $\hat{\mu}$, $\hat{\kappa}$ are known to be sufficient statistics for these parameters, so that the estimate of the T-year flood would be distributed with minimum attainable variance, whatever the sample size. Of course, we never know if the distribution $f(.)$ is truly gamma, so benefits associated with statistical sufficiency are probably more apparent than real, as in the case considered below where the data are transformed to normality by Box–Cox transformation.

With the maximum of the log-likelihood surface given by $(\hat{\mu}, \hat{\kappa})$, the confidence region for the true values (μ, κ) was given by (3), with $\theta = [\mu, \kappa]$ and $\hat{\theta} = [\hat{\mu}, \hat{\kappa}]$. Searching along the boundary of this region gave $(X_{\mathrm{L}}, X_{\mathrm{U}})$, the confidence interval for X_0, the T-year flood.

In the discussion of Method II above, it was mentioned that the Nelder–Mead algorithm for calculating X_{U} by means of (7) did not always converge, although the calculation of X_{L} converged in all cases. In fact, when the two-parameter gamma distribution was used, the algorithm converged for only five of the 18 stations listed in Table 1. It was this poor performance which led to the substitution of (7) by (9). For the five stations for which both minimisation procedures, (7) and (9), converged, the values of X_{U} are compared in Table 2. The values $X_{\mathrm{U}}^{\mathrm{II}}$ and $X_{\mathrm{U}}^{\mathrm{III}}$ given by the two methods are close. Values of $g(\theta)$, the constraint which forces the minimisation procedure to search along the confidence region boundary, are also shown in the table; values of $g(\theta)$ should be near to zero if the constraint is satisfied. Table 2 does not show values of X_{L} as these were indistinguishable, and values of X_{L} given by Method III are shown in Table 3. Table 2 shows that this method appeared to search more closely to the curve $g(\theta) = 0$, the values for $g(\theta)$ being about half those given by Method II.

Table 3
Two-parameter gamma distribution: estimates of the 100-year flood, X_0, at each of the 18 basins in Table 1, together with 95% confidence intervals (X_L, X_U)

River, gauging station	X_L	X_0	X_U
Taio	433.8	488.5	574.4
Pouso Redondo	90.0	114.7	162.5
Trombudo	142.2	181.1	263.2
Adago	384.0	577.4	589.4
Itajai, Barracao	478.8	681.7	1168.1
Itajai, Jararaca	451.8	580.8	848.1
Itajai-Acu, Rio do Sul	1423.2	1770.4	2400.3
Hercilio, Ibirama	1565.5	1936.3	2570.9
Neisse Central	151.5	209.0	337.6
Itajai-Acu, Apiuna	3227.0	3970.9	5234.1
Benedito, Benedito Novo	334.7	400.8	510.8
Benedito, Timbo	802.7	978.9	1279.4
Itajai-Acu, Indaial	4155.1	5123.5	6770.7
Testo, Rio do Testo	45.7	53.0	66.0
Itajai-Acu, Itoupava Seca	2625.5	3644.8	6394.9
Garcia, Garcia	119.9	159.6	239.4
Luis Alves, Luis Alves	83.7	98.9	125.1
Itajai-Mirim, Brusque	413.3	499.8	645.0

Table 3 shows the 95% confidence intervals (X_L, X_U) for the 100-year flood X_0 at each of the 18 basins shown in Table 1, together with the estimates of X_0. The asymmetry of the 95% interval (X_L, X_U) is noticeable in all cases; this contrasts with the symmetry of the approximate confidence intervals obtained by assuming that $\hat{X}_0$ is normally distributed, by virtue of the Central Limit Theorem. When calculating X_L by Method III, the values of the constraint $g(\theta)$ ranged between 0.0043 and 0.0211; when calculating X_U, $g(\theta)$ ranged between 0.0081 and 0.0796.

On 286 and 386 desk-top computers, the calculation of X_L and X_U was very slow, since to calculate them required solution of Eq. (4) and repeated calls on a sub-routine for the incomplete gamma function. Furthermore, this subroutine was recompiled at each call. To speed up the calculation, the method due to Bobee and Ashkar (1991) was used in which a frequency factor K_T is calculated by one of two formulae, depending on the value of $C_s = 2/\sqrt{\kappa}$. For $0 < C_s \leq 0.25$, K_T was calculated using the Wilson–Hilferty transformation as

$$K_T \cong (2/C_s)\left(\{(C_s/6)\,[u_P - (C_s/6)] + 1\}^3 - 1\right) \tag{11}$$

For $0.25 < C_s \leq 9.75$, the modified Wilson–Hilferty transformation was used to give K_T:

$$K_T \cong A\{\max[H, 1 - (G/6)^2 + (G/6)u_P]^3 - B\} \tag{12}$$

where in each case u_P is the pth quantile of the standard normal distribution, and where G, $1/A$, B and H^3 have polynomial approximations in C_s, as given in Bobee and Ashkar (1991, p. 38). Having calculated the frequency factor K_T by the appropriate method, the solution X_0 of Eq. (4) was estimated as

$$(K_T\sqrt{\hat{\kappa}} + \hat{\kappa})/\hat{\mu} \tag{13}$$

This modification reduced the time of calculation from hours to minutes; in the case of the Rio Taio, for example, the time of about 3 h on a 286 desk-top was reduced to 9 min. The changes in the values of X_L, $\hat{X}_0$ and X_U resulting from these modifications were negligible.

Calculation of likelihood-based confidence intervals for *T*-year floods: Gumbel distribution

For the Gumbel distribution

$$f(x, \alpha, \mu) = \alpha \exp\{-\alpha(x - \mu) - \exp[-\alpha(x - \mu)]\}, \qquad -\infty < x < \infty \tag{14}$$

the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\mu}$ are a solution of the non-linear equations

$$1/\alpha = \hat{x} - \sum x \exp(-\alpha x)/\sum \exp(-\alpha x) \tag{15}$$

$$\exp(-\alpha\mu) = \sum \exp(-\alpha x)/n \tag{16}$$

The first can be solved iteratively for α; substitution of α in the second then gives μ. The confidence region for $\theta = (\alpha, \mu)$, around the maximum-likelihood estimate $\hat{\theta} = (\hat{\alpha}, \hat{\mu})$, is then given by the inequality (3); Method III was used to search along this boundary for the values X_L, X_U defining the confidence limits for X_0. For the gauging stations listed in Table 1, values of $g(\theta)$, at convergence, ranged from 0.0052 to 0.0657. However, for one of the stations, on Rio do Testo, the calculation of the maximum-likelihood estimates $\hat{\theta} = (\hat{\alpha}, \hat{\mu})$ failed initially to converge. Graphical exploration of the likelihood surface – easily undertaken on a desk-top computer, and essential for thorough analysis – showed no reason why this should be so, and when Eq. (15) was modified to give

$$1 = \alpha\hat{x} - \alpha\sum x \exp(-\alpha x)/\sum \exp(-\alpha x)$$

Table 4
Gumbel distribution: estimates of the 100-year flood, X_0, at each of the 18 basins in Table 1, together with 95% confidence intervals (X_L, X_U)

River, gauging station	X_L	X_0	X_U
Taio	470.4	512.0	676.9
Pouso Redondo	96.1	113.3	165.7
Trombudo	211.4	245.5	783.0
Adago	506.5	636.9	1308.5
Itajai, Barracao	597.0	700.3	1997.4
Itajai, Jararaca	590.4	661.8	1909.9
Itajai-Acu, Rio do Sul	1697.2	1863.9	9161.3
Hercilio, Ibirama	2022.6	2192.2	4289.0
Neisse Central	169.4	207.8	340.3
Itajai-Acu, Apiuna	3924.8	4279.1	7074.7
Benedito, Benedito Novo	373.7	397.8	1237.6
Benedito, Timbo	990.3	1070.5	3978.3
Itajai-Acu, Indaial	5588.6	6062.3	14781.0
Testo, Rio do Testo	48.4	57.6	72.0
Itajai-Acu, Itoupava Seca	3139.5	3703.5	10096.0
Garcia, Garcia	130.2	154.4	239.1
Luis Alves, Luis Alves	115.1	125.4	472.8
Itajai-Mirim, Brusque	599.1	651.4	1737.2

the iterative solution of this equation by Newton–Raphson converged without difficulty. Failure to obtain a maximum-likelihood solution at the first attempt therefore does not necessarily mean that no solution exists. A further important point is that, even where the equations $\partial \ln L/\partial\theta = 0$ yield no solution within the parameter space, points on its boundary may be found which satisfy $\ln L(\hat{\theta}; x) \geq \ln L(\theta; x)$, so yielding maximum-likelihood estimates despite the fact that derivatives with respect to θ are non-zero.

Table 4 shows the values of the 95% confidence limits (X_L, X_U) for the maximum-likelihood estimate X_0 (also shown) of the 100-year floods at the 18 sites. The asymmetry of the limits, relative to $\hat{X}_0$, is again pronounced. Agreement with values of X_L, X_0 obtained using the gamma distribution was reasonable; agreement between values of X_U for the two distributions is less so. In both cases, the search leading to X_U is in a region of the likelihood surface where gradients are small.

Box–Cox transformations of the data

As an alternative means of modelling the skewness characterising the flood record at each site, the performance of the family of data transformations proposed by Box and Cox (1964) was explored. This well-known family is a transformation from the original scale of measurement, say x, to a new scale y in which the data are approximately normally distributed, where the transformation from x to y is given by

$$y = (x^{\lambda} - 1)/\lambda \qquad \text{for } \lambda \neq 0$$

$$y = \ln x \qquad \text{for } \lambda = 0.$$

The transformation can be generalised by writing $x + \alpha$ in place of x, requiring calculation of the additional parameter α; in this paper, the one-parameter form was used. The parameter λ of the transformation is estimated by maximum likelihood; for a sample of n annual floods x, the log likelihood $l(\theta, \lambda, y)$ of a normally distributed transformed variable y is written down (see Eq. (1) above) for a series of values of λ, commonly between -2 and $+2$. The value of λ for which $l(\theta, \lambda, y)$ is largest gives the maximum-likelihood estimate of this parameter. If λ is 'sufficiently close' to zero, a log-normal distribution is indicated for the data; a value of λ close to one-third would suggest a gamma distribution, this value corresponding to the well-known Wilson–Hilferty transformation (Kendall and Stuart, 1961) by which the gamma distribution is transformed to near-normality. The notion of 'sufficiently close' can be quantified by calculating likelihood-based confidence intervals for the estimate of λ; if this interval includes (say) zero, then the data can be transformed to near-normality by taking logarithms, so that a log-normal distribution describes the data. The confidence interval for the transfor-mation parameter λ is calculated as follows. In the plot of $l(\theta, \lambda, y)$ against λ, a horizontal line is drawn parallel to the λ-axis, at a distance $\chi^2_{1,\alpha}$ below the point at which the curve $l(\theta, \lambda, y)$ has its maximum. The two points at which this horizontal line cuts the curve $l(\theta, \lambda, y)$ give the $100(1 - \alpha)\%$ limits for λ. If $\alpha = 0.05$, the horizontal line is drawn at a distance $\chi^2_{1,0.05} = 3.841$ below the curve maximum. The calculation of the confidence limits for λ is again an application of the inequality in (3) above.

Advantages of working on a normal scale, if a suitable transformation can be found, are that the mean and variance of the transformed y values are sufficient statistics for the normal distribution's mean and variance. Theory shows that any function of these statistics – such as, for example, a quantile of the normal distribution, or the inverse transform of this quantile, which gives an estimate of the T-year flood in the original scale of measurement – estimates its expected value with the minimum-attainable variance. Unfortunately, the advantage may be more apparent than real, since this expected value almost certainly will not coincide with the 'true' T-year flood. However, the same is true of whatever distribution is fitted to a record of annual floods.

We conclude that there is no a priori reason for the rejection of flood estimation methods based on transformation to a normal scale, and there may be advantages in its favour.

Calculation of likelihood-based confidence intervals for *T*-year floods: Box–Cox transformations to approximate normality

For each of the stations listed in Table 1, the Box–Cox parameter λ was calculated by maximising the log-likelihood of the data, conditional upon λ, as described above. Estimates of the λ value so obtained are shown in Table 1, together with the Filliben correlation coefficients before and after transformation. Confidence intervals for the estimates of λ, not presented here, were wide, sometimes including both zero (indicating the need for a log transform) and unity (indicating consistency with normality on the original scale). However, despite the width of these confidence intervals, flood records

Table 5
Estimates of the 100-year flood, X_0, at each of the 18 basins in Table 1, with upper and lower 95% confidence limits X_U, X_L

River, gauging station	X_L	X_0	X_U
Taio	430.5	481.7	561.1
Pouso Redondo	86.4	105.9	140.2
Trombudo	137.7	168.2	225.4
Adago	450.0	1228.4	3210.1
Itajai, Barracao	426.0	537.7	746.1
Itajai, Jararaca	476.3	685.8	1320.4
Itajai-Acu, Rio do Sul	1296.9	1506.5	1833.1
Hercilio, Ibirama	1684.2	2293.0	3640.6
Neisse Central	152.1	211.4	345.2
Itajai-Acu, Apiuna	3430.0	4530.0	6767.2
Benedito, Benedito Novo	320.0	371.9	452.1
Benedito, Timbo	761.5	895.8	1107.2
Itajai-Acu, Indaial	4450.6	5943.1	9052.9
Testo, Rio do Testo	47.1	58.0	84.1
Itajai-Acu, Itoupava Seca	2541.8	3317.4	5069.0
Garcia, Garcia	117.2	152.6	220.0
Luis Alves, Luis Alves	77.9	87.2	101.1
Itajai-Mirim, Brusque	446.2	606.8	996.5

Limits obtained by (a) Box–Cox transformations to approximate normality (Box–Cox parameter λ as shown in Table 1); (b) estimation of 99 percentile, with likelihood-based confidence limits, on the normal scale; (c) detransformation from normal to original scale of measurement.

were transformed according to the fitted Box–Cox parameter. For two of the 18 stations, Apiuna and Ibirama, simulation envelopes (Atkinson, 1985) were calculated, and the normal plots for these stations lay comfortably within the envelopes.

Having estimated λ, there were essentially three steps in the subsequent calculation to obtain confidence limits for the flood with return period $T = 1/(1 - P)$ years: (1) identification of the boundary of the joint confidence region for the parameters μ and σ on the transformed (normal) scale, (2) identification of the points (μ, σ) on this boundary for which the P-quantile $X_0^* \equiv X_0^*(\mu, \sigma)$ has maximum and minimum values, and, trivially, (3) detransformation by applying the inverse Box–Cox transformation to obtain the confidence interval (X_L, X_U) for the estimated T-year flood, X_0, on the original measurement scale.

Table 5 shows the 95% limits X_L and X_U for the 100-year flood $\hat{X}_0$ estimated by this procedure. Results are generally consisent with those given by the gamma and Gumbel distributions, with the notable exception of one station, Adago. The λ parameter for this station was negative and quite large, although not the largest; convergence of the Nelder–Mead algorithm was always more difficult where λ was negative, although convergence was achieved in all cases.

Calculation of likelihood-based confidence intervals for T-year floods: three-parameter distributions

For three-parameter distributions, the difficulty of calculating likelihood-based confidence limits is substantially increased. For a three-parameter gamma, with $(x - m)$ in place of x in (10), an obvious method is to fix m,

Table 6
Values of 95% confidence limits X_L and X_U for the 100-year flood, estimated by X_0, for the eight stations for which the Nelder–Mead algorithm converged, when a three-parameter gamma distribution was fitted

River, gauging station	X_L	X_0	X_U
Taio	427.4	487.9	589.2
Pouso Redondo	83.0	106.5	181.8
Trombudo	139.9	173.0	246.9
Itajai, Barracao	408.7	526.6	796.1
Itajai, Jararaca	465.0	782.6	1134.4
Benedito, Benedito Novo	311.1	374.8	512.5
Benedito, Timbo	737.2	902.5	1266.5
Itajai-Acu, Itoupava Seca	2453.4	3381.7	9758.6

use the Nelder–Mead algorithm to calculate the limits $(X^*_{\mathrm{L}}, X^*_{\mathrm{U}})$ conditional on m, and then to calculate

$$X_{\mathrm{L}} = \inf_m X^*_{\mathrm{L}}; \qquad X_{\mathrm{U}} = \sup_m X^*_{\mathrm{U}}$$

This is time-consuming on a small desk-top computer, but should not be difficult given greater computer power. Results were only obtained for the three-parameter gamma distribution; the Nelder–Mead algorithm converged for eight of the 18 basins. No attempt has been made to establish why convergence failed, and this needs further investigation. Quite possibly, a reparameterisation of the likelihood surface would secure convergence, as happened with the flood record from the Rio do Testo. For the eight gauging stations at which the Nelder–Mead algorithm converged, 95% confidence limits are shown in Table 6.

Comparison of likelihood-based confidence intervals of 100-year flood with normal-approximation confidence intervals

Earlier sections of this work have explored the feasibility of calculating likelihood-based confidence intervals for floods X_0 with given return period. The question then arises of how do likelihood-based confidence intervals compare with those given by the central limit approximation confidence intervals, namely $\hat{X}_0 \pm 2\sqrt{(\mathrm{var}\,\hat{X}_0)}$? To answer this question for the distributions more commonly used in flood frequency analysis is beyond the scope of the present paper. However, as a first attempt at answering this question, we derived some preliminary results pertaining to the two-parameter log-normal distribution, or, more generally, to the case where a Box–Cox transformation to a normal scale is appropriate.

One hundred samples of size 20 were generated from the $N(0,1)$ distribution, and the upper quantile corresponding to a cumulative probability of $P = 0.99$ was estimated, this quantile corresponding to the '100-year flood.' For this estimated quantile, $\hat{X}_0$, we calculated (a) 95% confidence intervals using the central limit approximation, as above, and (b) likelihood-based (5%) confidence intervals. Despite the apparent remoteness from hydrological reality of the $N(0,1)$ distribution, it serves to explore the relative sizes of confidence intervals for X_0 since, after effecting the inverse Box–Cox transformation, confidence limits calculated using methods (a) and (b) would still bear the same relation to each other (in the sense that if the confidence interval using method (a) were wider than the confidence interval using method (b), it would remain so after detransformation from the normal to the original scale of measurement).

For the $N(0,1)$ distribution, the true value of the '100-year' quantile (corresponding to cumulative probability $P = 0.99$) is 2.326347. For each of the 100 generated samples, it was therefore possible to verify whether the confidence intervals given by (a) and (b) included this 'true' value, and also to see how far from the true quantile the maximum-likelihood estimate lay. Averaged over the 100 samples generated, the mean value of the estimates $\hat{X}_0$ of the $P = 0.99$ quantile was 2.234 ± 0.0423; the true value (2.326347) is slightly more than two standard deviations above this value, suggesting that small-sample bias may persist for samples of 20. Of more interest, however, are the mean widths of the confidence intervals: for the central limit approximation, the mean upper and mean lower 95% confidence limits were 3.405 ± 0.05877 and 1.171 ± 0.02744 respectively, giving a mean width of 2.234 ± 0.0367. For the likelihood-based confidence intervals, the mean upper and mean lower 95% confidence limits were 3.806 ± 0.04997 and 1.477 ± 0.02591, giving a mean difference of 2.329 ± 0.04532. The likelihood-based confidence intervals were therefore on average wider than those given by the central limit approximation, and were asymmetric relative to the true $X_0 = 2.326347$. Central limit confidence intervals, on the other hand, were more nearly symmetric relative to the true X_0, but failed more frequently to include it; when this happened, the upper confidence limit lay below X_0, the confidence interval being displaced to its left.

However, more intensive Monte Carlo study is required before firm conclusions can be drawn. This work will need to compare likelihood-based confidence intervals with central limit confidence intervals not only for the distributions used in flood frequency analysis, but also the effects of small sample sizes on each. The validity of the χ^2 approximation, used in (3) to define the confidence region for the parameters θ, also needs to be explored.

Discussion

The emphasis of this paper is on the estimation of confidence intervals for floods with T-year return period, using likelihood methods. However, it is possible that likelihood-based confidence intervals may rather wider hydrological application than in the case considered. We speculate on two possible applications.

First, we have considered the case where a good record of annual floods exists at a single site (say site A). Frequently, it will be useful, necessary even, to exploit the information in longer records at a nearby site B, when estimating the T-year flood at site A. At the expense of heavier calculation, confidence regions for parameters in the marginal distribution at site A can be calculated which exploit the record at sites B, the theory being a direct extension of that given earlier (e.g. McCullagh and Nelder, 1990). Having obtained such a confidence region, a search along its boundary will yield a confidence interval for the T-year flood at site A. Not only will this interval exploit information in the longer record at site B, but our speculation is that the procedure may provide a means of selecting which of the records available for sites B, C,..., really provide information on the T-year flood, X_0, at site A. For if site B provides information, the width of the confidence interval for X_0, resulting from its use, should be less than the width of the confidence interval for X_0 obtained using only the short record at site A. Thus it may be that a strategy for information transfer is to use information from nearby sites, whereas the use of such records results in smaller confidence intervals for X_0 at site A, and to stop using records from nearby sites, when their inclusion gives no reduction in the confidence interval for X_0.

Second, the calculation of confidence intervals for quantities predicted by rainfall–runoff models is receiving considerable attention (Binley and Beven, 1991; Binley et al., 1991; Beven and Binley, 1992). These authors proposed a 'generalised likelihood uncertainty estimation' (GLUE) procedure for calibrating rainfall–runoff models; this appears to work with multiple sets of parameter values and allows that, within the limitations of a given model structure and errors in boundary conditions and field observations, different sets of parameters may be 'equally likely as simulators of a catchment.'

It appears possible to formalise the arguments used by Binley and colleagues within the framework of likelihood-based confidence intervals. Suppose that a rainfall–runoff model describes runoff y_t in the tth time interval in terms of explanatory variables x_t (rainfall; potential evaporation; soil moisture; ... ;), the model being of the general form

$$y_t = f(x_t; \theta) + \epsilon_t$$

where θ are model parameters and ϵ_t is a 'residual.' Given (1) an appropriate probability distribution for ϵ_t and (2) observations on y_t and x_t, a likelihood function $L[\theta; (y_t), (x_t)]$ can be defined, specification of which will be simplified if the ϵ_t values are statistically independent. If all of the ϵ_t values are used in defining L, this simplification is inappropriate, since they will be serially correlated. However, it might be possible to formulate the likelihood function using only those ϵ_t values which are sufficiently separated to be considered effectively independent (perhaps by using only those ϵ_t values for points where y_t is a maximum, corresponding to flood peaks, or a minimum, corresponding to end-of-recession flows, or both). For example, fitting to flood peaks can be effected by defining a Boolean variable, w_t say, such that

$$w_t = 1 \quad \text{if } y_{t-1} < y_t \quad \text{and} \quad y_{t+1} < y_t$$
$$= 0 \quad \text{otherwise}$$

with model fitting based upon the criterion $\Sigma w_t[y_t - f(x_t; \theta)]^2$. By then allowing θ to vary over its joint confidence region, confidence limits can be calculated, in theory, for any quantity predicted by the model.

Finally, the inequality given in (3) uses a χ^2 approximation for calculating the confidence region for θ. This is a large-sample approximation, although the literature reports that the approximation is a good one. With the rapidly increasing power of computers, it is becoming feasible to calculate the θ confidence region exactly, by identifying the smallest region, around the point $\hat{\theta}$ for which $L(\hat{\theta}) \geq L(\theta)$, such that the probability within that region is $100(1-\alpha)\%$.

Acknowledgements

It is a pleasure to acknowledge the support of our IPH colleagues, in particular Professors Marcos Leão and Carlos E.M. Tucci, whose support was essential in providing time and facilities for this work. The library staff at IPH, in particular Jussara Silva and Jussara Barbieri, were ever helpful and tolerant of persistent misuse of their library facilities.

References

Atkinson, A.C., 1985. Plots, Transformations and Regression: an Introduction to Graphical Methods of Diagnostic Regression Analysis. Oxford Science, Oxford.

Benson, M.A., 1950. Use of historical data in flood frequency analysis. EOS (Trans. Am. Geophys. Union), 31: 419–424.

Beven, K.J. and Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction. In: The Future of Distributed Models. Hydrol. Proc., 6(3): 279–298.

Binley, A.M. and Beven, K.J., 1991. Physically-based modelling of catchment hydrology: a likelihood approach to reducing predictive uncertainty. In: D.G. Farmer and M.J. Rycroft (Editors), Computer Modelling in the Environmental Sciences. IMA Series No. 28. Clarendon Press, Oxford, pp. 75–87.

Binley, A.M., Beven, K.J., Calver, A. and Watts, L.G., 1991. Changing responses in hydrology: assessing the uncertainty in physically based model predictions. Water Resour. Res., 27(6): 1253–1261.

Bobee, B., 1975. The log-Pearson type 3 distribution and its applications in hydrology. Water Resour. Res., 14(2): 365–369.

Bobee, B. and Ashkar, F., 1991. The Gamma Family and Derived Distributions Applied in Hydrology. Water Resources Publications, Fort Collins, CO.

Bobee, B. and Robitaille, R., 1977. The use of the Pearson type 3 distribution and log Pearson type 3 distribution revisited. Water Resour. Res., 13(2): 427–443.

Box, G.E.P. and Cox, D.R., 1964. An analysis of transformations. J. R. Statist. Soc., Ser. B, 26: 211–252.

Bras, R.L. and Rodriguez-Iturbe, I., 1985. Random Functions and Hydrology. Addison-Wesley, Reading, MA.

Camacho, F., McLeod, A.I. and Hipel, K.W., 1987. The use and abuse of multivariate time-series models in hydrology. In: I.B. McNeill and G.J. Humphrey (Editors), Stochastic Hydrology. Reidel, Dordrecht, pp. 27–43.

Cohn, T.A. and Stedinger, J.R., 1987. Use of historical information in a maximum likelihood framework. J. Hydrol., 96(1–4): 215–223.

Condie, R. and Lee, K., 1982. Flood frequency analysis with historical information. J. Hydrol., 57(1/2): 47–61.

Cunnane, C., 1973. A particular comparison of annual maximum and partial duration series methods of flood prediction. J. Hydrol., 18: 257–271.

Cunnane, C., 1979. A note on the Poisson assumption in partial duration series models. Water Resour. Res., 15(2): 489–494.

Greenwood, J.A. and Durand, D., 1960. Aids for fitting the gamma distribution by maximum likelihood. Technometrics, 2(1): 55–65.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R., 1979. Probability weighted moments: definitions and relation to parameters of several distributions expressed in inverse form. Water Resour. Res., 15(6): 1049–1054.

Grygier, J.C., Stedinger, J.R. and Byin, H., 1989. A generalised procedure for extending correlated series. Water Resour. Res., 25(3): 345–350.

Hirsch, R.M., 1982. A comparison of four record extension techniques. Water Resour. Res., 18(4): 1081–1088.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F., 1985a. Estimation of the generalised extreme value distribution by the method of probability weighted moments. Technometrics, 27(3): 251–261.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F., 1985b. An appraisal of the regional flood frequency procedure in the UK flood studies report. Hydrol. Sci. J., 30(1): 85–109.

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

Kendall, M.G. and Stuart, A., 1961. The Advanced Theory of Statistics, Vol. 2. Charles Griffin, London.

Kite, G.W., 1977. Frequency and Risk Analysis in Hydrology. Water Resources Publications, Fort Collins, CO.

Landwehr, J.M., Matalas, N.C. and Wallis, J.R., 1979. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resour. Res., 15(5): 1055–1064.

Lu, L.-H. and Stedinger, J.R., 1992. Variance of two- and three-parameter GEV/PWM quantile estimators: formulae, confidence intervals, and a comparison. J. Hydrol., 138(1–2): 247–267.

Matalas, N.C. and Jacobs, B., 1962. A correlation procedure for augmenting hydrologic data. J. Geophys. Res., 67(9): 3441–3448.

McCullagh, P. and Nelder, J.A., 1990. Generalised Linear Models, 2nd edn. Chapman and Hall, London.

Shane, R.M. and Lynn, W.R., 1964. Mathematical models of flood risk evaluation. J. Hydraul. Div. ASCE, 90(HY6): 1–20.

Stedinger, J.R. and Tasker, G.D., 1985. Regional hydrologic regression 1: ordinary, weighted and generalised least squares compared. Water Resour. Res., 21(9): 1421–1432.

Stedinger, J.R. and Tasker, G.D., 1986. Regional hydrologic analysis 2: model error estimates, estimates of sigma, and log-Pearson type 2 distributions. Water Resour. Res., 22(10): 1487–1499.

Todorovic, P., 1978. Stochastic models of floods. Water Resour. Res., 14(2): 345–356.

Thomas, D.M. and Benson, M.A., 1970. Generalization of streamflow characteristics from drainage basin characteristics. US Geol. Surv., Water-Supply Pap., 1975.