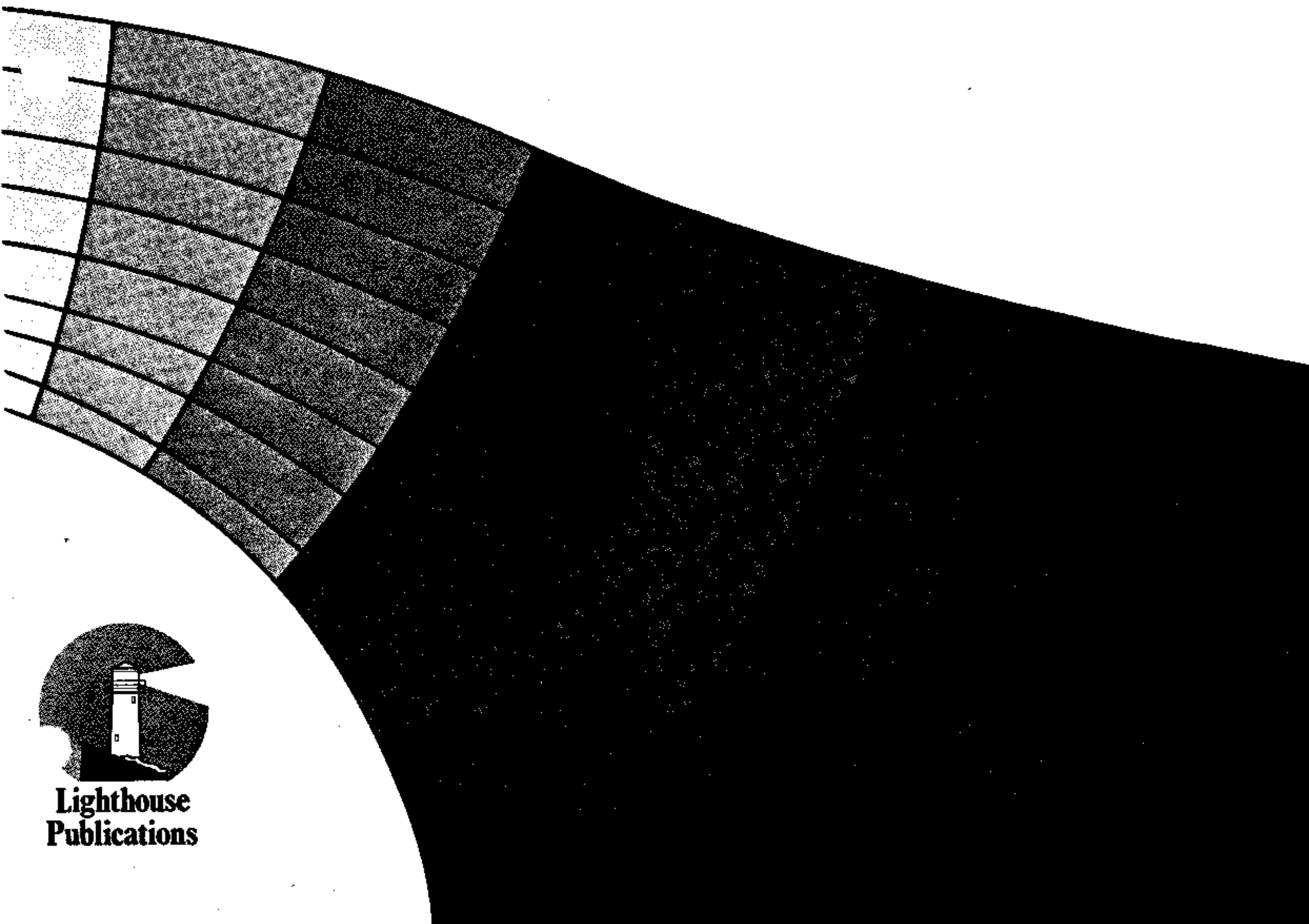
COMPUTATIONAL HYDROLOGY '87

Proceedings of the First International Conference
Anaheim, California, U.S.A., July – 1987

Editors
Theodore V. Hromadka
Richard H. McCuen



Lighthouse
Publications

Confidence Intervals For Floods

Robert Whitley
Department of Mathematics
University of California
Irvine, CA 92717

T.V. Hromadka II
Williamson and Schmid
Irvine, CA 92714

Abstract:

There is uncertainity in the estimation of the size of the T-year flood due to the uncertain estimation of the parameters in the Log Pearson III distribution which describes the occurance of the maximum annual discharge. Quantifying this uncertainty by means of computing confidence intervals is discussed.

Introduction:

Flood control agencies generally select a flood protection standard, for example the 100-year flood, to be used in the design of local flood control facilities. To comply with this standard requires estimating the T-year peak flow rate Q_T, in the specific example Q_{100} for $T = 100$.

There are numerous sources of uncertainty in this estimation of Q_T which include:

a) The distribution choosen for Q; the usual choice being the Log Pearson III distribution recommended by the Water Resources Council Bulletins 17A and 17B [1].

b) The estimation of parameters from regional data; such as the skew of the Log Pearson III distribution.

c) Uncertainty in the data due to data measurement or changing catchment conditions.

d) The statistical variation in the estimation of parameters from local data; such as the estimation of the mean and standard deviation of the Log Pearson III distribution.

In spite of these sources of uncertainty, it is a common practice for flood control agencies to adopt a particular flood control goal, e.g., Q_{100} design flows, and then simply utilize a flood frequency curve to estimate Q_{100}. Even supposing that everything is exact except for the estimation of the unknown parameters of the distribution, with the usual estimation by expectations the true but unknown value of Q_{100} may be just as likely (in the sense of repeated estimations) to be more than the estimate as it is to be less than the estimate. Computing only the estimated value of Q_{100} does not indicate what confidence one can have that the true value actually is less than the estimated value and so leaves vague a crucial aspect of the protection provided.

In order to quantify the level of protection that the flood control agency receives, confidence intervals for the computed estimate of Q_{100} should be given. There are too many sources of uncertainty to give definitive confidence intervals, but it is possible, as will be discussed below, to give confidence intervals based on one major source of variability in the estimate.

Discussion:

Two basic assumptions underlie the following analysis:

a) The Q values come from a Log Pearson III distribution.

b) The skew parameter for this distribution is known. (Since the skew is estimated from regional data, following the guidelines of [1], the estimate of the skew is consequently based on much more data than the other parameters in the distribution, and so it is not unreasonable to assume that, relative to the other parameters, the skew is known.)

Thus only that variation in the estimate of Q_T which depends on the statistical variation in the local data is considered, and it is analyzed under the usual assumptions of statistical sampling: That the data form an independent sample from a population with a given distribution whose parameters one wishes to estimate.

It has been known for a long time, and is noted in bulletins 17A and 17B of the Water Resources Council [1], that in the case of a flood distribution whose logarithm is normally distributed, confidence intervals for the T-year flood, under the assumptions above, can be obtained by the use of the non-central student's t-distribution.

The more general case, following the guidelines of bulletins 17A and 17B, is when the logarithm of the flood distribution has a Pearson III distribution with a non-zero skew parameter. This case of non-zero skew is more complicated than the case of a lognormally distributed flood which is the case of zero skew. In an important paper [2] Stedinger shows that the confidence intervals, for quantiles, which are given in the U.S. Water Resources Council guidelines [1] are not satisfactory. He uses a variance formula due to Kite [3] and derives an expression for confidence intervals for the quantiles which he shows is satisfactory in several simulations.

A program described in [4] derives empirical confidence intervals for the T-year flood by means of a simulation. This simulation is similar to that done by Stedinger in [2] and similar to the a simulation done by Hardison [5], who used his results in order to assess the accuracy of an approximate formula for confidence limits for flood-frequency curves given in Bulletin 17A. (This is one of the formulas which Stedinger in [2] found to be unsatisfactory.) The results given in [4] are more accurate and cover a more extensive range of T, sample sizes, skews, and levels of significance than these other simulations.

Mathematical Discussion:

In the case of zero skew, X, the logarithm of the maximum annual discharge, has a normal distribution with mean μ and standard deviation σ. For the T-year flood, take $p=1-1/T$, and let y_p be the p-th quantile of X; i.e.,

$$P(X \leq y_p) = p \qquad (1)$$

It is y_p that we want to estimate. For the estimate

$$\hat{y}_p = \hat{\mu} + \hat{\sigma} z_p ,$$

where z_p is the p-th quantile for an $N(0,1)$ distribution, i.e., a normal distribution with mean 0 and standard deviation 1, and $\hat{\mu}$ and $\hat{\sigma}$ are the usual estimates for μ and σ based on m data points. It can be shown that

$$(y_p-\hat{\mu})/\hat{\sigma} = ([\mu-\hat{\mu}/\sigma]+z_p)/(\hat{\sigma}/\sigma) \qquad (2)$$

has the same distribution as

$$(1/\sqrt{m})[(Z+z_p\sqrt{m})/\sqrt{W}] \; . \qquad (3)$$

where the random variable in brackets in (3) has a non-central t-distribution, with non-centrality parameter $\delta = z_p\sqrt{m}$; the special case $\delta = 0$ is the student's t-distribution.

Since the distribution of (3) can be written in terms of the known non-central t-distribution, confidence intervals for y_p can be given. Because of the way in which the flood design value is used, the confidence interval of greatest interest is a one-sided interval which gives an upper bound. The choice of the value T, for the T-year flood, and the number m of data points determine the non-centrality parameter δ. The other quantity which must be specified is a probability p' for the one-sided confidence interval. If $t_{p'}$ is the p'-th quantile of the non-central t-distribution, then

$$P(y_p \leq \hat{\mu} + \hat{\sigma}(t_{p'}/\sqrt{m})) = p' \qquad (4)$$

gives the one-sided $100p'$ percent confidence interval for y_p.

In the case of non-zero skew, the logarithm X of the yearly peak discharge is assumed to have a Pearson type III distribution with density function

$$f(x) = (1/|a|\Gamma(b))[(x-c)/a]^{b-1}\exp-[(x-c)/a] \quad (5)$$

where, in the case of positive a, the density is given by (5) for $x > c$ and is zero for $x < c$, while in the case of negative a the density is given by (5) for $x < c$ and is zero for $x > c$. Computing the mean μ, standard deviation σ, and skew γ from (5) shows that

$$\sigma^2 = a^2 b$$
$$\gamma^2 = 4/b , \qquad (6)$$
$$\mu = c + ab$$

where a has the same sign as γ. As mentioned above, the skew coefficient γ is estimated either from a map of regional skews or from a large pool of data from that region. Consequently the error in estimating γ, which is ignored in this analysis, is of an entirely different kind than that which arises in estimating μ and σ by means of m data points for the specific area for which the T-year flood is being estimated.

The form of the density (5) shows that the random variable

$$Z = (X-c)/a \qquad (7)$$

has the one parameter density

$$g(x) = (1/\Gamma(b))x^{b-1}e^{-x} \qquad (8)$$

for $x > 0$ and 0 for $x < 0$; i.e., Z has a gamma distribution with shape parameter b and scale parameter 1.

In [4] it is noted that if t_q is the 100q-th percentile for the gamma distribution Z, $\hat{u}_Z$ and $\hat{\sigma}_Z^2$ are the usual estimators for the mean of Z and the variance of Z using m sample points, then

$$(y_p-\hat{\mu})/\hat{\sigma} = (t_p-\hat{\mu}_Z)/\hat{\sigma}_Z \quad \text{for} \quad a > 0 \qquad (9)$$
$$(y_p-\hat{\mu})/\hat{\sigma} = (\hat{\mu}_Z-t_{1-p})/\hat{\sigma}_Z \quad \text{for} \quad a < 0 .$$

Confidence intervals for y_p were obtained in [4], by simulating the distribution of

$$(\delta-\hat{\mu}_Z)/\hat{\sigma}_Z \qquad (10)$$

for Z having the gamma distribution (8).

Stedinger derived an approximate expression for these confidence intervals in [2] using a variance formula due to Kite [3]. When restated in terms of the percentiles above Stedinger's formula is:

$$\Gamma_q = K_p + \lambda(\zeta_q(p)-z_p) \qquad (11)$$

The index p is related to the T-year flood by $p = 1 - 1/T$, and the constant z_p is the p-th percentile for a normal $N(0,1)$ distribution. The constant K_p is given by

$$K_p = (t_p-b)/\sqrt{b} \quad \text{for} \quad a > 0 , \qquad (12)$$

$$K_p = (b-t_{1-p})/\sqrt{b} \quad \text{for} \quad a < 0 .$$

in which t_p and t_{1-p} are the 100p-th and 100(1-p)-th percentiles of a gamma distribution with scale parameter 1 and shape parameter b, and which can be obtained by applying the Wilson-Hilferty transformation [6] to either z_p or z_{1-p}. The factor λ is a positive number given by Kite's variance formula

$$\lambda^2 = [1+\gamma K_p+(.5)(1+.75\gamma^2)K_p^2]/[1+.5z_p^2] \qquad (13)$$

To use formula (11) it is only necessary to find $\zeta_q(p)$, where $\zeta_q(p)\sqrt{m}$ is the q-th percentile for the non-central t-distribution with non-centrality parameter $\delta = z_p\sqrt{m}$. This non-central t-distribution is well-known and its q-th percentile can be found by means of a double numerical integration. The computation of this approximation is discussed in [7].

Approximate Confidence Intervals:

In [8] it was shown, by means of comparison with simulations, that for skew γ in a normal range, $-.75 \le \gamma \le .75$, Stedinger's approximate formula is satisfactory for computing not only the higher percentiles of the distribution but also for computing all of the percentiles 5%(5%)95%. This approximation is less accurate for larger values of skew, as will be seen in Table I; note that it was shown in [9] that the value $\gamma = 2$ corresponds to the common assumption of a linear relation between log Q and log T, T the return period. Table I compares simulation values to Stedinger's approximate values for the percentiles of the 100 year flood for skew in the range $-2 \le \gamma \le 2$.

Table I:

Values for $m = 10$ data points, $T = 100$ year flood

Tabulated values are relative percent error

100(A-B)/A for the indicated percentiles

A = values from simulation

(20,000 sets of 10 point samples)

B = Stedinger-Kite approximation

percentile	skew=	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
	5%	-22.2	-12.6	-5.1	-.9	.1	-1.4	-9.0	-21.1	-48.8
	25%	-13.1	-7.9	-3.2	-.4	.0	-1.0	-4.5	-10.2	-16.4
	75%	-1.3	-.9	-.2	-.2	-.2	-.3	-2.0	-4.4	-8.4
	95%	7.2	5.2	2.5	.1	.1	.3	-.2	-3.0	-8.2

More extensive simulation indicates that generally this relative error decreases as m increases ($5 \le m \le 30$) and decreases as T increases ($10 \le T \le 100$).

Conclusions:

Confidence intervals should be provided for design values of the T-year flood and they can be computed by simulation or, for the usual range of hydrologic parameters, by the Stedinger-Kite approximation.

References

[1] Guidelines for Determining Flood-Flow Frequency, Water Resources Council Hydrology Committee, Bulletin #17A, Washington, D.C., 1977. Guidelines for Determining Flood-Flow Frequency, Water Resources Council Hydrology Committee, Bulletin #17B, Washington, D.C., 1981.

[2] Stedinger, J. R., Confidence intervals for design events, J. of Hydraulic Engineering 109 (1983) 13-27.

[3] Kite, G. W., Confidence limits for design events, Water Res. Research 11 (1975) 48-53.

[4] Whitley, R. and Hromadka II, T., Computing confidence limits for floods I, Microsoftware for Engineers 2 (1986) 138-150.

[5] Hardison, C., Confidence limits for flood frequency curves computed from samples from Pearson type III populations, J. of Research U.S.G.S. 4 (1976) 545-547.

[6] Abramowitz, M. and Stegun, I., Handbook of Mathematical Functions, National Bureau of Standards, Washington, D.C., 1965.

[7] Whitley, R and Hromadka II, T., Computing confidence intervals for floods II, Microsoftware for Engineers 2 (1986) 151-158.

[8] Whitley, R. and Hromadka II, T., Estimating 100 year flood confidence intervals, to appear in Advances in Water Research.

[9] Whitley, R. and Hromadka II, T., Confidence intervals for flood control design, submitted.