The Log-Pearson III Distribution in Hydrology

R. J. Whitley * T. V. Hromadka II † M. J. Smith, ‡

Introduction

The procedures recommended for flood-frequency analysis by U. S. federal agencies and in general use in the United States are those given in [23], a reference which is commonly referred to as (Bulletin) 17B; also see [39], [31], [24], and [15]. In 17B the log-Pearson III distribution was adopted for yearly maximal discharges, this choice is discussed in appendix 14 of 17B, and see the comprehensive [15]. This chapter discusses the basic properties of this distribution which are needed for applications.

The Probability Density

The Pearson family of probability distributions is discussed in [25, Chap. 6]. The density for the Pearson III distribution has the form

$$f(t) = \frac{[(t-c)/a]^{b-1}}{|a|\Gamma(b)} e^{-[(t-c)/a]} \text{ for } (t-c)/a > 0, \tag{1}$$

$\Gamma(b)$ the gamma function, and f(t)=0 for $(t-c)/a < 0$ and with $a \neq 0$.

A random variable Y has a log-Pearson III distribution if logY has a Pearson III distribution. The same distribution results no matter which logarithm base is used and showing this will also establish other useful results. The formula for change of logarithm base from **b** to **a** is $\log_{\mathbf{a}}(t) = \log_{\mathbf{a}}(\mathbf{b}) \log_{\mathbf{b}}(t)$. The fact that the derivative of $\log_{\mathbf{a}}(t) = \log_{\mathbf{a}}(e)/t$ has the simplest form for $\mathbf{a} = e$ explains why base e is chosen for logarithms in mathematics, although base 10 is often used for engineering data. This common use of two different bases is why the stability of the log-Pearson distribution under a change of base in the logarithm is important.

In the following $f_Y(t)$ will denote the probability density of the random variable Y and $F_Y(t) = Prob(Y \leq t)$ the (cumulative) distribution function.

*P.O.Box 11133, Bainbridge Island, WA 98110 (rwhitley@math.uci.edu)

†Department of Mathematical Sciences, United States Military Academy, West Point, NY 10096, USA. (ted@phdphdphd.com)

‡Department of Mathematical Sciences, United States Military Academy, West Point, NY 10096, USA. (Mick.Smith@usma.edu)

Suppose that $\log_{\mathbf{b}}(Y) = X$ has a Pearson III distribution. Then

$$\log_{\mathbf{a}}(Y) = \log_{\mathbf{a}}(\mathbf{b}) \log_{\mathbf{b}}(Y).$$

A base for a logarithm is positive and greater than one, so in the above $\alpha = \log_{\mathbf{a}}(\mathbf{b}) > 0$, and $\log_{\mathbf{a}}(Y) = \alpha X$. Then

$$F_{\alpha X}(t) = Prob(\alpha X = t) = Prob(X \leq \frac{t}{\alpha}).$$

Differentiating,

$$f_{\alpha X}(t) = \frac{1}{\alpha} f_X(\frac{t}{\alpha}).$$

It will be useful to also determine the distribution of αX for $\alpha < 0$. In this case

$$F_{\alpha X}(t) = Prob(\alpha X \leq t) = Prob(X \geq t/\alpha) = 1 - Prob(X < t\alpha).$$

Differentiating,

$$f_{\alpha X}(t) = -\frac{1}{\alpha} f_X(\frac{t}{\alpha}).$$

In either case, $f_{\alpha X}(t) = \frac{1}{|\alpha|} f_X(\frac{t}{\alpha})$ and

$$f_{\alpha X}(t) = \frac{[(t - \alpha c)/\alpha a]^{b-1}}{|\alpha a|\Gamma(b)} exp\left(-[(t - \alpha c)/\alpha a]\right) \text{ for } (t - \alpha c)/\alpha a > 0 \ ,$$

which shows that αX has a Pearson III distribution with c and a multiplied by α and b unchanged.

Moments

The first three moments of a Pearson III distribution suffice to determine the density by determining the three parameters a,b, and c.

The relations between the mean μ, standard deviation σ, and skew γ are

$$\mu = c + ab,$$

$$\sigma^2 = a^2 b$$

$$\gamma^2 = 4/b$$

where a and γ have the same sign. The parameters can be obtained by finding the first three moments of X. It is easier to work with

$$Z = \frac{X - c}{a}. \tag{2}$$

To obtain the density of Z, first consider the case $a > 0$: $F_Z(t) = Prob(Z \leq t) = Prob(X \leq c + at)$, and by differentiating $f_Z(t) = a f_X(c + at)$. If $a < 0$, then

$F_z(t) = Prob(X \geq c + at) = 1 - Prob(X < c + at)$, and $f_Z(t) = -af(X(c+at)$. For both $a > 0$ and $a < 0$,

$$f_Z(t) = |a| f_X(c + at) = \frac{1}{\Gamma(b)} t^{b-1} e^{-t} \text{ for } t > 0 \tag{3}$$

and $f_Z(t) = 0$ for $t < 0$. Thus Z has a gamma distribution with parameter b. [34]. (In general a gamma distribution has two parameters, but in the case here the other parameter is 1.)

The moment generating function for Z, $\phi_Z(\lambda)$ is [34]:

$$\phi_Z(\lambda) = E(e^{\lambda Z}) = \frac{1}{\Gamma(b)} \int_0^\infty e^{\lambda t} t^{b-1} e^{-t}\, dt = \frac{1}{(1-\lambda)^b}. \tag{4}$$

Note that, as is usually the case, the moment generating function is not defined for all λ but for "small" lambda, in this case for $\lambda < 1$. The first three moments of Z follow easily from the recursive formula $\Gamma(b+1) = b\Gamma(b)$, an equation which shows that for n an integer, $\Gamma(n+1) = n!$. The moments are: $E(Z) = \phi'(0) = b$, $E(Z^2) = \phi''(0) = b(b+1)$, and $E(Z^3) = \phi'''(0) = b(b+1)(b+2)$. Using equation (2),

$$\mu = ab + c.$$

Since $E(X^2) = a^2 b(b+1) + 2acb + c^2$,

$$\sigma^2 = a^2 b,$$

and X has standard deviation $\sigma = |a|\sqrt{b}$. A similar computation shows that $E[(Z-b)^3] = 2b$, from which $E[(X-\mu)^3] = E(a^3(Z-b)^3) = 2a^3 b$, which, using the formula for the skew, gives:

$$\gamma = \frac{E[(X-\mu)^3]}{\sigma^3} = \frac{2a^3 b}{[|a|\sqrt{b}]^3} = \frac{a}{|a|}\frac{2}{\sqrt{b}}.$$

So the parameter b is related to the skew γ by $b = 4/\gamma^2$, and since the factor $sgn(a) = \frac{a}{|a|}$ is 1 if $a > 0$ and is -1 if $a < 0$, the skew has the same sign as the parameter a.

Zero Skew

In the important case of zero skew the distribution, by a limiting argument, will be shown to be a normal distribution. The easiest proof is by the use of moment generating functions. Let W be the random variable Z scaled by subtracting its mean and dividing by its standard deviation, so that W has mean zero and standard deviation one:

$$W = (Z - b)/\sqrt{b}.$$

Consider the moment generating function for W:

$$E(e^{\lambda W}) = E(e^{\lambda X}) = e^{-\lambda\sqrt{(b)}} E(e^{(\lambda/\sqrt{b})Z}) = e^{-\lambda\sqrt{(b)}}(1 - \frac{\lambda}{\sqrt{b}})^{-b}.$$

Write

$$(1 - \frac{\lambda}{\sqrt{b}})^{-b} = e^{-b\log(1-\frac{\lambda}{\sqrt{b}})}$$

and use the three term Tayor's expansion for $\log(1+x)$

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3}(1+\tau)^{-3}, \text{ for some } \tau,\ 0 \leq |\tau| < |x| < 1$$

at $x = \frac{\lambda}{\sqrt{b}}$ to see that

$$\lim_{b\to\infty} \phi_X(\lambda) = e^{\lambda^2/2}. \tag{5}$$

The right-hand side of (5) is the moment generating function for a normal distribution with mean 0 and standard deviation one [34]. Applying the continuity theorem from probability [5] shows that X approaches N(0,1) in distribution as b tends to infinity. Since X=aZ+c, X tends to a normal distribution with mean c and standard deviation $|a|$, as b tends to infinity.

T-year Flood

With yearly maximal discharge for a catchment being modeled by a log-Pearson III distribution, the quantities of interest–in fact the main reason for modeling the discharge–are the values y_p of the T-year flood, $1 \leq T < \infty$. For a given T, y_p is the p-th percentile of Y, i.e. the value with

$$Prob(Y \leq y_p) = p = 1 - \frac{1}{T}.$$

A commonly used value of T is $T = 100$, with $y_{.99}$ the value for the 100-year flood. The estimation of y_p is done by estimating the T-year value x_p for $X = \log_{\mathbf{a}} Y$. Because $\log_{\mathbf{a}}(t)$ is an increasing function of t,

$$p = Prob(X \leq x_p) = Prob(log_{\mathbf{a}} Y \leq x_p) = Prob(Y \leq \mathbf{a}^{x_p})$$

showing that $y_p = \mathbf{a}^{x_p}$. Percentiles for the Pearson III distribution were computed in [18], [19], [20].

These percentiles can be readily obtained from percentiles for the Gamma distribution. Let z_p be the corresponding percentile for $Z = \frac{X-c}{a}$; $Prob(Z \leq z_p) = p$. The relation between x_p and z_p depends on the sign of a. First suppose $a > 0$, then

$$p = Prob(X \leq y_p) = Prob([(X-c)/a] \leq [(x_p - c)/a]) = Prob(Z \leq [(x_p - c)/a])$$

from which we see

$$z_p = (x_p - c)/a \text{ for } a > 0. \tag{6}$$

Next suppose that $a < 0$, then

$$p = Prob(X \leq x_p) = Prob([(X-c)/a \geq [(x_p - c)/a]) = Prob(Z \geq [(x_p - c)/a]) \tag{7}$$

so

$$Prob(Z \leq [(x_p - c)/a]) = 1 - p$$

and

$$z_{1-p} = [(x_p - c)/a] \text{ for } a < 0,$$

Estimation

Bulletin 17B [23] requires the use of the method of moments to estimate the parameters a, b, and c, which works in the following way: If the year maximal discharges are $\{y_1, y_2, \ldots, y_n\}$, take logarithms to get $\{x_1 = log_{\mathbf{a}} y_1, \ldots, x_n = log_{\mathbf{a}} y_n\}$. (This is a simplification as there is some data processing, for example for outliers, which is ignored here.) The usual estimators: $\hat{\mu}$ for the mean, $\hat{\sigma}$ for the standard deviation, and $\hat{\gamma}$ for the skew are:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2,$$

where the factor of $\frac{1}{n-1}$ is chosen so as to have an unbiased estimator, and

$$\hat{\gamma} = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu})^3}{\hat{\sigma}^3}.$$

Note that the use of an unbiased estimator for the standard deviation (e.g. the estimator $\hat{\sigma}$ having a factor of $n-1$, not n, in the denominator, which is done to have the expectation of the estimator equal the quantity being estimated) as in [23, pg. 10], is not generally considered to have particular importance [6, pg. 61], and is sometimes done and sometimes not done, which can lead to a certain amount of confusion. The estimator of [23] for the sample skew differs from the one given here for the same reason.

Using the relations between the moments of the Pearson III distribution of X and the parameters a, b, and c, these parameters are estimated by

$$\hat{b} = \frac{4}{\hat{\gamma}^2},$$

$\hat{a}$ has the same sign as $\hat{\gamma}$ and

$$\hat{a}^2 = \frac{\hat{\sigma}^2}{\hat{b}}$$

and

$$\hat{c} = \hat{\mu} - \hat{a}\hat{b}.$$

For the special case of zero skew γ ($\hat{\gamma}$ zero or "small"), the first two moments are the usual estimators for the parameters μ and σ^2 of a normal distribution.

An interesting property of the estimator for the skew is that it is bounded by a simple function of the number of data points n and independent of the size of the skew [26]. This result was derived in response to some simulations which had results he describes as "profoundly disturbing". To understand what was disturbing, consider the random generation of a Pearson III with positive skew, b=.01, and using n=100 point samples, and computing the sample skew each time. When done repeatedly an empirical distribution of the sample skew statistic is obtained. The skew for this distribution is $\sqrt{\frac{4}{b^2}} = 20$ whereas, using the bound below, the sample skew can be no larger than 10 which is far from the true skew. Such extreme skew values are not used for floods: the skew map in [23] has 97% of the skews in the range $[-1, 1]$ and all in the range $[-2, 2]$.

Here is an easy derivation of a good bound, but not the best bound, on the sample skew. The estimators used are: Sample mean

$$\hat{\mu} = \frac{1}{n}\sum_{1}^{n} x_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum (x_i - \hat{\mu})^2$$

and skew

$$\hat{\gamma} = \frac{\frac{1}{n}\sum_1^n (x_i - \hat{\mu})^3}{\hat{\sigma}^3}.$$

The difference between these estimators from [24] and those used in [23] are various fudge factors; in [23], as noted above, for $\hat{\sigma}$ there is a factor of $\frac{1}{n-1}$ not $\frac{1}{n}$. For the skew, let $SS = \sum_1^n (x_i - \hat{\mu})^2$ and $SC = \sum_1^n (x_i - \hat{\mu})^3$, the estimator in [24] is

$$\sqrt{n}\frac{SC}{SS^{\frac{3}{2}}}$$

while [23] uses

$$\frac{n\sqrt{n-1}}{n-2}\frac{SC}{SS^{\frac{3}{2}}}.$$

The somewhat strange factor for the 17B skew is used to have an unbiased estimator for $E(X-\mu)^3$ and "reduces but does not eliminate the bias of $\hat{\gamma}$" [31, 18.5]. The example above, with a skew of 20 but a sample skew less than 10, shows that the expectation of the sample skew is certainly not the true skew, and that the bias can be considerable.

In $\frac{SC}{SS^{\frac{3}{2}}}$ if each $x_i - \hat{\mu}$ is multiplied by $\alpha > 0$ the fraction remains unchanged; do this with $\alpha = \sqrt{\frac{1}{SS}}$, so that

$$\frac{SC}{SS^{\frac{3}{2}}} = \frac{\sum u_i^3}{(\sum u_i^2)^{\frac{3}{2}}},$$

where $|u_i| \leq 1$ and $\sum u_i^2 = 1$, thus $|u_i|^3 \leq u_i^2$ and the above sum is bounded by 1. Thus with the estimators of [24]

$$|\hat{\gamma}| \leq \sqrt{n}.$$

This argument is quite a bit easier than [26] but he gets the bound

$$|\hat{\gamma}| \leq \frac{n-2}{\sqrt{n-1}}.$$

which is better than the bound above and in fact is best possible.

Uncertainty in Estimation

All statistical estimation involves uncertainty, but the estimation of the T-year flood is more uncertain than most estimations. The basic problem is that the estimate for the T-year flood is obtained by first estimating the parameters of the distribution, then, letting $\hat{X}$ denote the random variable obtained using these estimated parameters, finding the value $\hat{x}_p$ for which $Prob(\hat{X} \leq \hat{x}_p) = p = 1 - \frac{1}{T}$, a point which for larger values of T is far out on the tail of the estimated distribution. Then, a step which further amplifies any error, exponentiating to get $\hat{y}_p = \mathbf{a}^{\hat{x}_p}$, where **a** is the logarithm base used to transform the log-Pearson III random variable Y to the Pearson III random variable X by $X = \log_{\mathbf{a}} Y$.

Typically, statistical variation in estimation is described by the use of confidence intervals. A form of confidence interval that has great relevance for the estimation of T-year floods is a one-sided confidence interval.

The case of zero skew is well-known as it corresponds to X a normally distributed random variable. Let x_p be the true value for the T-year flood Pearson III random variable X and, as above, $\hat{x}_p$ the value obtained by the estimation procedure above. The relevant fact about a normal distribution with mean μ and standard deviation σ is that

$$\hat{\tau} = \frac{\mu - \hat{\mu}}{\hat{\sigma}} \tag{8}$$

has a known distribution, the t-distribution. In this case a $100q\%$ confidence level, $0 < q < 1$, for the mean can be obtained as follows. Think of sampling independent points from a normal X with unknown parameters μ and σ, and obtaining as above estimates $\hat{\mu}$ for the mean and $\hat{\sigma}$ for the standard deviation. For a given q, $0 < q < 1$, let c_q be the value of the t-distribution with the property that

$$Prob(\hat{\tau} \leq c_q) = q. \tag{9}$$

Consider an independent sample of n points from the normal distribution X. For each sample a value $\hat{\tau}$ is obtained. From equation (9), $100q\%$ of the time

$$\mu \leq \hat{\mu} + c_q\hat{\sigma}.$$

In hydrologic terms, when the procedure described–estimate the mean and standard deviation using the given formulas, choose q and thereby c_q– is followed, the estimate $\hat{\mu} + c_q\hat{\sigma}$ will give a value that is greater than or equal to the true (unknown) mean, $100q\%$ of the time, the mean being the T=2 year flood for the normal distribution. That is, this procedure provides protection from the

2-year flood $100q\%$ of the time when used repeatedly. The value q can be choose to provide the protection required or that which is affordable, 100% protection being impossible in this model.

The key point here is that equation (8) provides a random variable τ which has a known distribution that does not depend on the parameters μ and σ of the underlying normal distribution X.

The same general approach will provide a $100q\%$ confidence level for the T-year flood value x_p for a normal distribution, the difference is that in the case $T \neq 2$ the random variable

$$\frac{x_p - \hat{\mu}}{\hat{\sigma}}$$

does not have a t-distribution, but does have a known distribution, the non-central t-distribution [33].

Note that the confidence level depends on the choice of the random variable τ, the estimators used in $\hat{\tau}$, as well as the distribution from which $x_1, \ldots, x_n$ are sampled.

Consider the case where X has a Pearson III distribution with non-zero skew.

For X with this distribution, let $Z = (X - c)/a$ Suppose that we have n independent samples of X, $X_1, X_2, \ldots, X_n$, with corresponding $Z_1, Z_2, \ldots, Z_n$. Consider the sample means:

$$\hat{\mu}_Z = \frac{1}{n}\sum_{1}^{n}((X_i - c)/a) = (\hat{\mu}_X - c)/a.$$

For the sample standard deviations

$$\hat{\sigma}_Z^2 = \frac{1}{n-1}\sum_{1}^{n}[(X_i - c)/a - (\hat{\mu}_X - c)/a]^2 = \frac{1}{a^2}\hat{\sigma}_X^2.$$

So then

$$\hat{\sigma}_Z = \frac{1}{a}\hat{\sigma}_X \text{ if } a > 0$$

and

$$\hat{\sigma}_Z = -\frac{1}{a}\hat{\sigma}_X \text{ if } a < 0$$

First from (6)

$$\frac{x_p - \hat{\mu}_X}{\hat{\sigma}_X} = \frac{z_p - \hat{\mu}_Z}{\hat{\sigma}_Z} \text{ for } a > 0. \tag{10}$$

while from (7)

$$\frac{x_p - \hat{\mu}_X}{\hat{\sigma}_X} = \frac{\hat{\mu}_Z - z_{1-p}}{\hat{\sigma}_Z} \text{ for } a < 0. \tag{11}$$

Any application of these equation will use the fact that a and the skew have the same sign.

As shown in (3) Z has a gamma distribution with parameter b. Thus, by considering τ the parameters a and c have been eliminated from consideration.

Known Skew

It has long been known that the procedures suggested in [23] for the calculation of confidence levels are not accurate [35]. These confidence levels depend on the procedure used and, if using the method of moments, on the variation in the estimators for the mean, standard deviation, and skew. This full problem, to be discussed below, is difficult but some insight can be obtained from the simpler problem which ignores the variability in the estimation of the sample skew.

The equations (6) and (7) allow the calculation of confidence levels under the simplifying assumption that value of the site skew is known exactly. Even in this case, the distribution involved is too complicated to be able to compute a density from which a confidence level could be calculated. An approximate formula, best used for small skew, has been derived in [35], and see [40]. More accurate values can be obtained from simulations and this was done in [41]. These simulations were applied in [21]. These calculations were redone recently [22] to incorporate the larger data sets now available, applying the same ideas but with faster simulations which give higher accuracy. The simulation approach to confidence levels in the case of known skew involves simulating the distribution of the right-hand sides of (6) and (7), which are too complex to have known densities. Using these equations, simulations can deliver accurate values for constants $K(N, \gamma, c)$ for which

$$Prob(x_p \leq \hat{\mu}_X + \hat{\sigma}_X K(N, \gamma, c) = c \tag{12}$$

is approximately true with error in the fourth decimal. This is to say that equation (12) displays a $100c\%$ confidence level for the unknown value x_p of the T-year flood, $p - \frac{1}{T}$. That this can be done is clear in principle but was impractical at the time 17B was written. To give an idea of the change in readily available computer power, a remark made in [40] can be used to estimate that a program which would have taken 40 hours using the personal computers and software of 20 years ago, would now take about 10 seconds. The computer programs for the simulations were compiled with a Lahey/Fujitsu Fortran 95 compiler using the IMSL software library. The simulations involved for a specific T-year flood take, in the case of N=50 data points, approximately three minutes. An accurate value for the constant $K(N, \gamma, c)$ is obtained by a simulation of the right-hand side of (6) or (7) in which an empirical histogram is used to obtain percentiles corresponding to the desired K values. After the values of K have been calculated, they are tested using a completely different random number generator from that used in the simulation.

Estimated Skew

The problem of finding confidence intervals which incorporate the uncertainty in the estimation of the mean, the standard deviation, <u>and</u> the skew is more difficult than the model problem of the last section in which the skew is taken as known. On difficulty is that [23, pg. 11] recommends the weighting of the site skew with other site values or using a map of skews [29] and [30]. It is impossible to test these methods without additional assumptions concerning the regional distribution of skews, a subject on which [23] is silent. The method of choice for most statisticians would be maximum likelihood [10], but at the time it was considered too difficult to use in a consistent way throughout the United States; this may no longer be the case if it were implemented in a program supplied by, say, the Army Corp of Engineers. Some proposed methods include [11], as discussed in [42], and a neural network approach [43]. A method that has been much studied recently [7],[8], [9], [12], [13], [16], [32], [37] is an iterative method which allows the incorporation of historical events, as does the method of maximum likelihood. This is an active area of research and the reader is directed to the quoted references; a convenient source for which is [36].

References

[1] Ashkar, F. and Bobée, B (1988), Confidence intervals for flood events under a Pearson 3 or log-Pearson 3 distribution, *Water Resour. Bull.*, *24* 639-650.

[2] Beard, L. R.(1962), Statistical Methods in Hydrology, U. S. Army Engineer District, Corp of Engineers, Sacramento, Calif., 74 pp.

[3] Bobée, B. (1975), The Log Pearson Type 3 Distribution and Its Application in Hydrology, *Water Resour. Res.*, *14* 365-369.

[4] Bobée, B. and Robitaille, R. (1977), The Use of the Pearson Type 3 Distribution and Log Pearson Type 3 Distribution Revisited, *Water Resour. Res.*, *14* 427-443.

[5] Breiman,L., Probability, *Addison-Wesley*, Reading, Mass., 1968.

[6] Breiman, L., Statistics: With a View Towards Applications, *Houghton Mifflin*, Boston, Mass., 1973.

[7] Cohen, T. A., Grifis, V. W., and Stedinger, J. R.(2004), Log Pearson type 3 quantile estimators with regional skew information and low outlier adjustment, *Water Resources Research*, *40*,W07503, doi:1029/2003WR002697.

[8] Cohen, T.A., Lane, W. L., and Gaier, W. G. (1997), An algorithm for computing moment-based flood quantile estimates when historical flood information is available, *Water Resour. Res.*, *33(9)*, 2089-2096.

[9] Cohen, T.A., Lane, W. L., and Stedinger, J. R. (2001), Confidence intervals for EMA flood quantile estimates, *Water Resour. Res.*, *37(6)* 1695-1706.

[10] Condie, R. (1977), The log Pearson type 3 distribution: the T-year event and its asymptotic standard error by maximum likelihood theory, *Water Res. Research*, *13* 987-991.

[11] Chowdhury, J. and Stedinger, J. D. (1991) Confidence intervals for design floods with estimated skew coefficients, *ASCE J. Hydraul. Eng.*, *11* 811-831.

[12] England, J. F., Jr., Jarrett, R. D., and Salas, J.D. (2003) Data-based comparisons of moments estimators using historical and paleoflood data, *J. of Hydrology*, *278(4)*, 172-196.

[13] England, J. F., Jr., Salas, J.D., and Jarrett, R. D. (2003) Comparisons of two moments-based estimators that utilize historical and paleoglood data for the log Pearson type III distribution, *Water Resour. Res.*, *39(9)*, 1243, doi:10.1029/2002WR001791.

[14] Foster, H.A. (1924), Theoretical frequency curves and their application to engineering problems, *Transactions of the American Society of Civil Engineers*, *87* 142-173; discussion 174-203.

[15] Griffis, V. W., and Stedinger, J. R, (submitted J. of Hydrology, July 2005), Evolution of flood frequency anaysis with bulletin 17,

[16] Griffis, V. W., Stedinger, J. R., and Cohn, T. A. (2004), LP3 quantile estimators with regional skew information and low outlier adjustments, *Water Resour. Res.*, *40*, W07503, doi:1029/2003WR002697.

[17] Hall, L. S. (1921), The probable variations in yearly run-off as determined from a study of California streams, *Transactions of the American Society of Civil Engineers,84* 191-231; discussion 215-257.

[18] Harter, R. H. (1969), A new table of percentage points of the Pearson type III distribution, *Technometrica,11* 177-186.

[19] Harter, R. H. (1971), More percentage points of the Pearson type III distribution, *Technometrica,13* 203-204.

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

[21] Hromadka, T. et. al. (1987) Derivation of a rainfall-runoff model to compute N-year floods for Orange County watersheds, *Publication, U. S. Army Corp of Engineers, Los Angeles Division*

[22] Hromadka, T. et. al. (2006) The analysis and update of Los Drainage Area (LACDA) stream gauge data using a Log-Pearson III simulation computer program. *Publication, U. S. Army Corp of Engineers, Los Angeles Division*

[23] Interagency Committee on Water Data (1982), Guidelines for determining flood flow frequency, *Bull 17B*, 28pp. Hydrol. Subcomm., Washington, D.C., March.

[24] Committee on Techniques for Estimating Probabilities of Extreme Floods, Water Science and Technology Board, Commission on Physical Sciences, Mathematics, and Resources, Nation Research Council (1988), Estimating Probabilities of Extreme Floods, *National Academy Press*, Washington, D.C.

[25] Stuart, A. and Ord, J. K., Kendall's Advanced Theory of Statistics 5^{th} Ed., Vol 1, *Griffin & Co*, London, 1987.

[26] Kirby, W. (1974), Algebraic boundedness of sample statistics, *Water Res. Research, 10* 220-222.

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

[28] Maidment, D. (1992), Handbook of Hydrology, *McGraw-Hill*, New York.

[29] McCuen, R. H. (1979), Map Skew???, *J. Water Res. Plann. Manage. Div. Am. Soc. Civ. Eng. (now) J. Water Res. lLann. Manage., 105* 265-277 [with Closure, *107*, 582 (1981)].

[30] McCuen, R. H. (2001), Generalized flood skew: map versus watershed skew, *J. of Hydrologic Eng.,64* 293-299.

[31] Maidment, D. R., Editor in Chief (1993), Handbook of Hydrology, *McGraw-Hill*, New York.

[32] Reis, D. S., Stedinger, J.R., and Martins, E. S. (2005) Bayesian generalized least squares regression with application to log Pearson 3 regional skew estimation, *Water Resour. Res.,41, W10419, doi:10.1029/2004WR003445.*

[33] Resnikoff, G. and Lieberman, G. (1957) Tables of the Non-Central t-Distribution, *Standford University Press*, California.

[34] Ross, S. M. (2003), Introduction to Probability Models, 8^{th} ed., *Academic Press*, New York.

[35] Stedinger, J. R. (1983), Confidence Intervals for Design Events, *J. of Hydraulic Engeering, 109* 13-27.

[36] Stedinger, J. R. and England, J. R., Jr., Key documents for discussion of evolution of Bulletin 17B and proposed changes (2005), *ftp://ftp.usbr.gov/jengland/HFAWG/.*

[37] Tasker, G. D. and Stedinger, J. R. (1986), Estimating Generalized Skew with Weighted Least Squares Regression, *J. Water Resour. Plann. Manage, 112* 225-237.

[38] Thomas, W. O. (1985), A Uniform Technique for Flood-Frequency Analysis, *J. Water Resour. Plann. Management, 111* 321-337.

[39] Uncertainty of Discharge-Probability Functions (1996). Chapter 4 in Manual No. 1110-2-1619, *Department of the Army, U.S. Army Corps of Engineers*, Washington, D.C. (Available online as EM 1110-2-1619 from www.usace.army.mil/inet/usace-docs/eng-manuals.)

[40] Whitley, R. J. and Hromadka II, T. V. (1986), Computing Confidence Intervals for Floods I, *Microsoftware for Engineers*, *2*,138-150.

[41] Whitley, R. J. and Hromadka II, T. V. (1986), Computing Confidence Intervals for Floods II, *Microsoftware for Engineers*, *2*,151-158.

[42] Whitley, R. J. and Hromadka II, T. V. (1997), Chowdhury and Stedinger's approximate confidence intervals for design floods for a single site, *Stochastic Hydrl. Hydraul.*, *11* 51-63.

[43] Whitley, R. J. and Hromadka II, T. V. (1999), Approximate confidence intervals for design floods for a single site using a neural network, *Water Res. Research* , *35* 203-209.