

Frequency Analysis of Low Flows

G. V. Loganathan, C. Y. Kuo, and T. C. McCormick

Dept. of Civ. Eng., Virginia Polytechnic Institute
and State University, Blacksburg, VA 24061, U.S.A.

The transformations (i) SMEMAX (ii) Modified SMEMAX (iii) Power and Probability Distributions (iv) Weibull (α, β, γ) or Extreme value type III (v) Weibull ($\alpha, \beta, 0$) (vi) Log Pearson Type III (vii) Log Boughton are considered for the low flow analysis. Also, different parameter estimating procedures are considered. Both the Weibull and log Pearson can have positive lower bounds and thus their use in fitting low flow probabilities may not be physically justifiable. A new derivation generalizing the SMEMAX transformation is proposed. A new estimator for the log Boughton distribution is presented. It is found that the Boughton distribution with Cunnane's plotting position provides a good fit to low flows for Virginia streams.

Introduction

Lows flows are critically important in the operation of run of the river power plants and water supply works. The diluting capacity of the river is highly reduced during low flow periods resulting in water quality problems. Also, a good knowledge of low flows is essential to support fish and wildlife, commercial and recreational uses. This paper compares the various low flow probability distributions in selecting the low flow of a given return period.

Frequency analysis is the procedure for estimating the probability of occurrence of an event. The sample for this type of analysis usually consists of the set of extreme values for each year. For floods, the maximum mean daily flow for each

year is used. For low flows, the duration of the extreme value for each year is often taken to be something other than one day. Longer durations are used because the harmful effects of low flows are often felt only after a prolonged period. This is especially true of instream flows where the harmful effects on fish and other wildlife become noticeable only after a few days. The durations commonly used for low flow analysis are 1, 3, 7, 14, 30, 60, 90, 120, 150, and 273 days (ASCE Task Committee Report 1980). As an example, if the seven day low flow is desired, the lowest mean flow for seven consecutive days in each year is taken as the sample.

For floods, a calendar year or water year is generally used. For low flows, some sort of climatic year, different from either of these years, is used. Critical periods of low flows usually occur in summer, fall, or winter. For long durations it is possible that the critical period could be separated into two incomplete parts by the end of a calendar or water year. This separation destroys the underlying premise of frequency analysis, that of the independence of the observations. To avoid this problem, a climatic year which begins April 1 of one year and ends March 31 of the next year is used. This climatic year is recommended by the American Society of Civil Engineers (ASCE) Task Committee on Low-Flow Evaluation, Methods, and Needs (1980) and is used in the Hydrologic Information Storage and Retrieval System (HISARS) data bank for the state of Virginia (Johnson et al. 1975).

A number of distributions have been proposed for the analysis of low flows. These distributions may be grouped into the following categories:

- 1) Transformation Methods
 - a) Power Transformation
 - b) SMEMAX (Small, MEDian, MAXimum) Transformation
 - c) Modified SMEMAX Transformation
- 2) Hypothetical Distribution Based Methods
 - a) Weibull (α, β, γ) or Extreme value type III
 - b) Weibull ($\alpha, \beta, 0$) Distribution
 - c) Log-Pearson Type III Distribution
- 3) Plotting-Position Based Methods
 - a) Log-Boughton Distribution

The three transformation methods considered, all transform the sample to a near normal distribution. It is called near normal distribution because the convergence of probabilities to the normal distribution may not be exact. It depends on the underlying distribution of the sample data. But the normal distribution provides a good approximation in the transformed range for the required probabilities. All of the distributions are three parameter distributions except the Weibull ($\alpha, \beta, 0$) and normal distributions. Based on previous research, Boughton (1980) concluded

that distributions with at least three parameters should be used for the frequency analysis of floods. At least three parameters are required to account for the variations in the shape of the underlying distributions at different stations.

The main objective of this study is to select the most suitable distribution for Virginia for the frequency analysis of low flows. The major criteria for selection are: 1) accuracy, and 2) reliability. Accuracy is the ability of the method to correctly find the low flow associated with a given return period. Since streamflow samples are from limited records, the accuracy obtainable is limited by the degree to which the sample is representative of the true underlying distribution. Thus, accuracy is the goodness of fit of a particular distribution to the sample. Sometimes, the theoretical probabilities based on the fitted distribution are compared with the plotting positions (Jennings and Benson 1969). But Cunnane (1978) points out the theoretical variations of plotting position formulas depending upon the true underlying distributions. Thus, theoretically it seems more reasonable to adopt goodness of fit test as the measure of accuracy. In this paper Kolmogorov-Smirnov test (K-S test) is adopted. The reliability of a given distribution is the ability to perform accurately over a wide range of basin characteristics. This then is the ability of the method to perform well for basins of varying size, topography, geology, etc.

Transformation Methods

In recent years a great deal of research has been undertaken to find a probability distribution that will adequately model the underlying distribution at any given station. The transformation methods attempt to find a unique distribution in the transformed range regardless of the distribution of the original sample.

Power Transformation – This transformation was first proposed by Box and Cox (1964), and Chander et al. (1978) used this transformation for flood frequency analysis. In a paper by Kumar and Devi (1982), the power transformation was applied to the frequency analysis of low flows for one station. The power transformation is form

$$z_i = \frac{[(y_i)^\lambda - 1]}{\lambda} \quad \text{for } \lambda \neq 0, y_i > 0 \quad (1)$$

$$z_i = \ln y_i \quad \text{for } \lambda = 0, y_i > 0 \quad (2)$$

where: z is normally distributed with mean M_z and standard deviation S_z ; y is the original sample value; λ is a parameter of transformation. Also, y can be replaced by $(y+k)$ for $y > -k$.

The proper value for λ is the value that produces a transformed sample with a coefficient of skewness (C_s) of zero. Alternatively, a maximum likelihood procedure can be used (Box and Cox 1964). An inverse transformation is required to convert the transformed cutoff points for each return period back to the low flows of interest. The inverse of the power transformation is as follows

$$Q_T = (\lambda z_T + 1)^{\frac{1}{\lambda}} \tag{3}$$

where: Q_T is the discharge having a return period of T ; z_T is given by $\phi(z_T) = 1/T = P(Z < z_T)$; ϕ is the normal cumulative probability distribution function.

If the transformed values are actually normally distributed, the skewness C_s and the kurtosis C_k will be equal to zero and three respectively.

SMEMAX Transformation – This transformation was first proposed by Bethlahmy (1977) for the frequency analysis of floods. Prakash (1981) successfully applied the transformation to the analysis of low flows. A modified SMEMAX transformation was proposed by Rasheed et al. (1982). Calculation of the parameters used in the SMEMAX transformation equation is much easier than calculation of λ as required by the power transformation. For the SMEMAX transformation, the transformation equations simply ensure that, in the transformed sample, there are equal number of observations on either side of the mean value and thus the mean approximates the median. This relationship provides for near symmetry but is not a sufficient condition for normality. In the following a new general derivation of the SMEMAX transformation is proposed. Both the original SMEMAX and the modified SMEMAX are shown to be the special cases of this new derivation. The following notation is used: X original sample data point (known); X_s smallest (known); X_m median (known); X_1 largest (known); Z transformed value (computed); Z_s smallest (chosen arbitrarily); Z_m median = $(Z_1 + Z_s) / 2$; Z_1 largest (chosen arbitrarily). Considering Fig. 1

$$Z = Z_s + \frac{Z_m - Z_s}{(X_m - X_s)} (X - X_s) \quad \text{for } X_s \leq X \leq X_m \tag{4}$$

$$= Z_m + \frac{Z_1 - Z_m}{(X_1 - X_m)} (X - X_m) \quad \text{for } X_m \leq X \leq X_1 \tag{5}$$

It is noted that $Z_1 = 2Z_m$ for $Z_s = 0$. For this special case the transformation given by Eqs. (4) and (5) becomes

$$Z \equiv Z_m \frac{(X - X_s)}{(X_m - X_s)} \quad \text{for } X_s \leq X \leq X_m \tag{6}$$

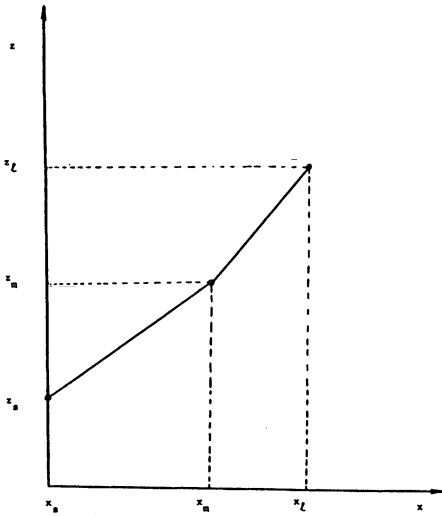


Fig. 1. Generalized SMEMAX.

$$\cong Z_m \left[1 + \frac{X - X_m}{(X_1 - X_m)} \right] \quad \text{for } X_m \leq X \leq X_1 \quad (7)$$

Eqs. (6) and (7) are precisely the same as those derived by Rasheed et al. (1982) as modified *SMEMAX*. Further by choosing

$$Z_1 = \|(X_1 - X_s)\| \quad \text{where } \|X_1 - X_s\| = [(X_1 - X_m)^2 + (X_m - X_s)^2]^{\frac{1}{2}}$$

it is seen $Z_m = \|(X_1 - X_s)\|/2$ for $Z_s = 0$. For this special case Eqs. (6) and (7) yield the original *SMEMAX* suggested by Bethlahmy (1977). These are given as Eqs. (8) and (9)

$$Z = (X - X_s) \frac{\|X_1 - X_s\|}{2(X_m - X_s)} \quad \text{for } X_s \leq X \leq X_m \quad (8)$$

$$= \left[(X_m - X_s) + (X - X_m) \frac{X_m - X_s}{(X_1 - X_m)} \right] \left[\frac{\|X_1 - X_s\|}{2(X_m - X_s)} \right] \quad \text{for } X_m \leq X \leq X_1 \quad (9)$$

It is noted that the median flow has a return period of 2 years. Thus, for higher return periods the low flows will be less than the median flow. Also, Z is normally distributed. From Eq. (8) the T year low flow Q_T for the original *SMEMAX* can be written as

$$Q_T = 2Z_T \frac{(X_m - X_s)}{\|X_1 - X_s\|} + X_s \quad (10)$$

Also from Eq. (6) Q_T for the modified SMEMAX can be written as

$$Q_T = Z_T \frac{(X_m - X_s)}{Z_m} + X_s \quad (11)$$

The new derivation has the following advantages over the original and the modified SMEMAX and the power transformations:

- 1) The new derivation is simple and direct.
- 2) The new derivation allows for the control of the range of values that the transformed variable Z can take. It is noted that the modified SMEMAX by Rasheed et al. (1982) assumes $Z_s = 0$, whereas in the new derivation Z_s is arbitrary.
- 3) In the power transformation, large negative λ values can lead to transformed values clustered around $1/\lambda$ whereas in the proposed method the user controls the range of the transformed values.

Hypothetical Distribution Based Methods

In this section, three well known distributions will be investigated. The extreme value type III or three parameter Weibull (α, β, γ) , two parameter Weibull $(\alpha, \beta, 0)$ and Log Pearson Type III are considered. The Weibull (α, β, γ) is based on the analytical study of distribution of minima whereas log Pearson Type III is chosen purely for the purpose of flexibility in curve fitting. The same procedure is followed in applying all three methods. First, the parameters of the selected distribution are calculated from the sample. The cumulative distribution function (*cdf*) is then used to find the low flows associated with the return periods of interest.

The difficulty in the application of some distributions to frequency analysis is the calculation of the ordinates of the *cdf*. The *cdf* of the Weibull distribution, has a tractable functional form. Like the normal distribution which is used in the transformation methods, the *cdf* of the log-Pearson distribution can only be stated as an integral of the probability density function (*pdf*). This greatly complicates calculation of the low flows associated with the return periods of interest.

Weibull (α, β, γ) Distribution – The *cdf* for this distribution is given as

$$F(y) = 1 - \exp \left[- \left(\frac{y - \gamma}{\beta - \gamma} \right)^\alpha \right] \quad \text{for } \gamma < y < \infty \quad (12)$$

where: $F(y)$ is the probability of a low flow less than or equal to y ; y is the cutoff value; β is the characteristic low flow (a location parameter); γ is the lower limit to y ; α is a scale parameter.

The return period, T is defined as

$$F(x) = P(X \leq x) = \frac{1}{T} \tag{13}$$

Substituting Eq. (13) into Eq. (12) and rearranging the low flow Q_T with return period T is

$$Q_T \equiv \lambda + (\beta - \gamma) \left[-\ln\left(1 - \frac{1}{T}\right) \right]^{\frac{1}{\alpha}} \tag{14}$$

A number of techniques have been proposed to calculate the parameters α, β , and γ . Based on Deininger and Westfield (1969) and Kite (1977) the following methods are used: 1) the method of moments, 2) the maximum likelihood procedure, and 3) the method of smallest observed drought.

As previously noted, γ is the lower bound to the distribution. There are limits to the value that γ can assume. For a large river, the value of γ will in all likelihood be found to be greater than zero. The probability of a flow occurring below this lower bound is zero. From a purely physical viewpoint, there is a positive probability, of a low flow being less than any given positive value. Thus, any lower bound greater than zero cannot be completely justified. The lower bound is also limited by the fact that negative flows are impossible. Thus, actual lower bound should be zero. However, it may be practical to use a lower bound greater than zero if the fit can be improved.

By matching the sample and distribution moments, there is no limit on the value of γ . The value of γ need not be at all reasonable. This can take two forms: a value of γ that is less than zero, or a value of γ that is greater than one of the observed values. If γ is negative, there is a positive probability that the flow less than zero can occur. Thus, for small values of $F(x)$ (i.e., for long return periods), the computed low flows can become negative. A value of γ greater than the smallest observed value leads to similar problems. The smallest observed value clearly means that there is a positive probability associated with flows of this magnitude. The fitted distribution, however, excludes these probabilities.

In the maximum likelihood procedure, the objective is to select the distribution parameters such that the probability associated with the occurrence of the sample values is maximized. This procedure is preferred because of its ability to generate the minimum variance estimates but there is no guarantee that γ will be nonnegative. Another alternative estimating technique is the method of smallest observed drought. This method was proposed by Gumbel (1963) as a means of avoiding the situation where the smallest observed value is less than the calculated value of γ . The value of γ is estimated from the probability function of the smallest drought in a period of n years. While this method does guarantee a value of γ less than or equal to the smallest observed value, there is still no guarantee that γ is non-negative.

Weibull ($\alpha, \beta, 0$) *Distribution* – The *cdf* of this distribution is given as

$$F(y) \equiv 1 - \exp \left[- \left(\frac{y}{\beta} \right)^\alpha \right], \quad \text{for } 0 \leq y < \infty \quad (15)$$

In a manner analogous to that used for the Weibull (α, β, γ) distribution, Eq. (18) can be rearranged to give the low flows of interest as a function of the return period

$$Q_T = \beta \left[- \ln \left(1 - \frac{1}{T} \right) \right]^{\frac{1}{\alpha}} \quad (16)$$

The method of moments can again be used to find the parameters of this distribution (Tadikamalla 1978). At first glance this method appears to have significant advantages over the Weibull (α, β, γ) distribution. The distribution can assume a variety of shapes. It is easier to use because only two parameters must be estimated. Also, there are no problems associated with the lower bound. However, it is found that the ability of the Weibull ($\alpha, \beta, 0$) distribution to model low flows at different stations is not adequate as shown in the comparison of methods section.

Log-Pearson Type III Distribution – Bobee (1975) gives an excellent discussion of the properties of this distribution. The effect of variations of the parameters on the distribution shape is shown by a number of graphs. This shape flexibility is one of the major advantages of the log-Pearson distribution.

The probability density function is given as

$$f(y) \equiv \frac{1}{|\alpha| y \Gamma(\beta)} \left[\frac{\ln(y) - \gamma}{\alpha} \right]^{\beta-1} \exp \left[- \left(\frac{\ln(y) - \gamma}{\alpha} \right) \right] \quad (17)$$

for $\beta > 0$, $\alpha > 0$, $\exp(\gamma) \leq y < \infty$
 for $\beta > 0$, $\alpha < 0$, $0 < y \leq \exp(\gamma)$

where: $f(y)$ is the value of the *pdf* at y ; α is a scale parameter; β is a shape parameter (always positive); γ is a location parameter; $\Gamma(\beta)$ is the gamma function evaluated at β .

Since no closed form equation of the *cdf* is available, it is impossible to obtain a closed form relationship between the return period T and its associated low flow Q_T . A number of techniques have been proposed for calculating the parameters α , β , and γ . The method proposed by the Water Resources Council (WRC) has been called the indirect method of moments. Instead of calculating the three parameters such that the sample moments are reproduced by the fitted distribution, the sample is first transformed by a logarithmic transformation. A Pearson type III distribution is then fit to the log transformed values. The parameters are then

selected so that three moments of the distribution are matched with three moments of the new sample. Thus the moments are only matched indirectly. Q_T can be calculated as follows

$$Q_T = \exp(M_{1ny} + K_T S_{1ny}) \quad (18)$$

where K_T is called the frequency factor M_{1ny} and S_{1ny} are the mean and standard deviation of the logarithm of flow values. For this distribution, in addition to being a function of the return period T , K_T is also a function of C_s . To obtain an accurate value for K_T , the best approach is to use one of the extensive sets of tables of K_T values (see Bulletin 17A). For automatic computations, it would be preferable to use an equation for K_T . Kite (1977) and Bulletin 17A present an approximating equation that gives K_T as a function of C_s and the standard normal deviate z_T which is applicable for $-1.0 < C_s < 1.0$

$$K_T = \frac{2}{C_s} \left\{ \left[\frac{C_s}{6} \left(z_T - \frac{C_s}{6} \right) + 1 \right]^3 - 1 \right\} \quad (19)$$

It is interesting to note that if the skewness coefficient of the log transformed values is zero, the log-Pearson distribution approaches the log-normal distribution (Johnson and Kotz, p 170, 1970; Benson 1968). Thus, when the value of C_s is equal to zero, K_T is equal to the standard normal deviate. Similarly, a value of λ near zero reduces the power transformation to a log transformation and thus reduces to the log-normal distribution. Bobee (1975) suggested the direct method of moments. In this method, the moments of the distribution are matched to the moments of the actual sample (not the moments of the log transformed sample). By avoiding the distortion that accompanies the log transformation, this method gives each observation equal weight. Thus, this method is theoretically sound.

From the definition of log - Pearson distribution given in Eq. (17), it is seen that depending on the sign of α , the distribution has either a lower bound or an upper bound. Thus, the comments with regard to the lower bound on the Weibull (α, β, γ) distribution apply to the log Pearson as well. As pointed out in the comparison of methods to follow, eight stations had a positive α implying a positive lower bound, out of the twenty stations considered. In such cases the use of the log Pearson is not physically justifiable.

Plotting Position Based Methods

Plotting position based methods are probably the most straightforward of the various types of methods. For each observed discharge, a plotting position formula is used to estimate an exceedence probability. A curve is then fit through the

data points. Since exceedence probability is one minus the probability of non-exceedence, the *cdf* can be generated directly. Thus, this approach directly fits the *cdf* to the probabilities as evidenced by the sample. In this study, one of these plotting position methods is used.

Log-Boughton Distribution – This distribution was proposed by W. C. Boughton for the frequency analysis of floods in 1980. After taking the logarithm to the base 10, of each sample data point, the mean and standard deviations are used to standardize the data. This standardized discharge is called the frequency factor (K_i). K_i is calculated as

$$K_i = \frac{z_i - M_z}{S_z} \tag{20}$$

where: K_i is the frequency factor; z_i is the common logarithm of the observed value; M_z is the mean of the transformed values; S_z is the standard deviation of the transformed values;

The return period is given in terms of a variable G_i which is defined as

$$G_i \equiv \ln \left[\ln \left(\frac{T^*}{T^*-1} \right) \right] \tag{21}$$

where T^* is the return period for floods. This equation can be rewritten in terms of the return period for low flows as

$$G_i = \ln [\ln(T)] \tag{22}$$

where T is the return period for low flows. Boughton observed that the function relating K_i and G_i very nearly fits a hyperbolic curve given by

$$C = (K_i - A)(G_i - A) \tag{23}$$

where A and C are constants to be determined. Boughton did not present a *cdf* for this distribution. From the equations given above, however, an estimator of the *cdf* can be derived (see Fig. 2). Combining Eq. (13) with Eq. (22) and solving for $F(z)$

$$F(z_i) = \exp[-\exp(G_i)] \tag{24}$$

Eq. (23) can be solved for G_i

$$G_i \equiv A + \frac{C}{K_i - A} \tag{25}$$

By substituting Eq. (20) for K_i into Eq. (25) and rearranging

Frequency Analysis of Low Flows

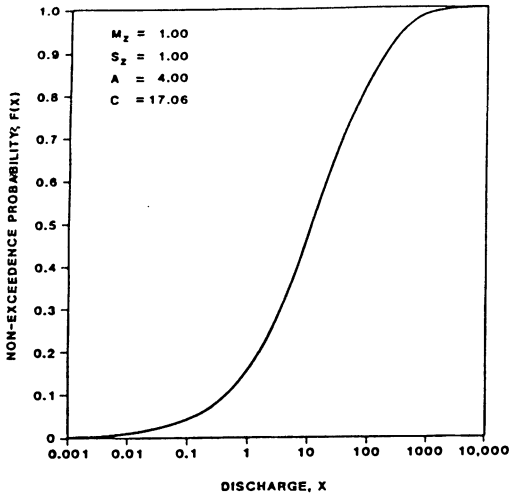


Fig. 2. Cumulative distribution function for the Log Boughton Distribution.

$$G \equiv A + \frac{CS_z}{z - M_z - AS_z} \quad (26)$$

Eq. (26) can then be substituted into Eq. (24) to obtain

$$\begin{aligned}
 F(z) &\equiv \exp[-\exp(A + \frac{CS_z}{z - M_z - AS_z})] \quad \text{for } -\infty < z < (M_z + AS_z) \\
 &= 1 \quad \text{for } z \geq M_z + AS_z
 \end{aligned} \quad (27)$$

Eq. (27) provides an estimator of the *cdf*. It can be seen from Eq. (27) for $z_1 > z_2$, $F(z_1) > F(z_2)$. As z goes to $-\infty$ Eq. (27) reduces to $\exp[-\exp(A)]$. If A is large (say around 3), this value becomes very close to zero. As z goes to $M_z + AS_z$ from below, $CS_z/(z - M_z - AS_z)$ goes to $-\infty$. Thus, $F(z)$ approaches one. As z exceeds $M_z + AS_z$ by a small amount, however, the term becomes positive infinity and thus $F(z)$ approaches zero. Thus, a discontinuity exists at $M_z + AS_z$ and $F(z)$ must be redefined to be one for $z > (M_z + AS_z)$ (see Fig. 2). The estimator $F(z)$ is used in the Kolmogorov-Smirnov goodness of fit test (Benjamin and Cornell 1970).

For methods similar to the log-Boughton distribution, the critical step is the selection of the plotting position formula to be used. Boughton and Shirley (1983) use the plotting position formula developed by Cunnane (1978) which is

$$P(X > x) = \frac{m - 0.4}{n + 0.2} \quad (28)$$

This formula is used to estimate the values of G_i in Eq. (21). Cunnane recom-

mends this formula as the best compromise plotting position for all distributions. Srikanthan and McMahon (1981) also found that the same formula gave good results when the underlying distribution was log-Pearson type III.

The parameter estimation technique developed by Boughton and Shirley (1983) used Eq. (23) to minimize the mean squared error of C . This procedure yields optimal A and C values. By using these optimal A and C values in Eq. (23) along with G_i for the data point Z_i the frequency factor for the fitted curve K_i^* may be written as

$$K_i^* = A + \frac{C}{G_i^{-A}} \quad (29)$$

Thus, for a data set of n values, there are n pairs of (K_i^*, Z_i) values. By using the form of Eq. (20) it can be written as

$$Z_i = M^* + K_i^* S^* \quad (30)$$

Eq. (30) is treated as a linear regression with coefficients M^* and S^* and these are estimated. Thus found M^* and S^* are used to estimate the low flow of return period T , Q_T as

$$Q_T = 10^{(M^* + K_T^* S^*)} \quad (31)$$

K_T^* is the frequency factor computed from Eq. (29) based on G_T which is a function of the return period of interest T .

Intermittent Streams

For the frequency analysis of low flows, intermittent streams are a major concern for two reasons: 1) many of the distributions require the calculation of the logarithms of the sample data and the logarithm of zero is negative infinity, and 2) continuous probability distributions lead to a zero probability that a specific flow will occur and thus, even if the stream is usually dry, the analysis produces a probability of a dry stream equal to zero. The first of these concerns can easily be rectified by transforming the data to a positive series by adding a constant to all of the data points. After application of the selected distribution, the calculated low flows are then found by subtracting the constant from the low flows found from the distribution. The fact the probability of the flow equal to the transformed zero flow is still zero is not addressed by this approach. Jennings and Benson (1969) proposed that a mixed distribution be used to alleviate the problem. A point mass is placed at the origin equal to the probability occurrence of a zero discharge. The

positive discharges are then represented by a truncated probability distribution. The point mass at zero is estimated as

$$P(x=0) = \frac{d}{n} \tag{32}$$

where: $P(X=0)$ is the probability that the random variable X equals zero; d is the number of zero observations in the sample; n is the sample size.

Eq. (32) assumes that n is large enough so that d/n approaches the true probability of zero flows. The conditional probability is given as $P(X < x | X > 0)$ which is the probability that X is less than or equal to x given that X is greater than zero. Thus,

$$P(0 < X \leq x) = (1 - \frac{d}{n}) F^*(x) \tag{33}$$

where $F^*(x)$ equals $P(X < x | X > 0)$. The *cdf* can be written as

$$F(x) = \frac{d}{n} + (1 - \frac{d}{n}) F^*(x) \tag{34}$$

To apply this approach, a selected distribution can be fit through the positive subset to obtain $F^*(x)$. Eq. (34) is then used to find $F(x)$.

A slightly different procedure can be used with the log-Boughton distribution. The exceedence probabilities associated with the sample can still be estimated from the sample with Eq. (28) using the total sample size in the denominator. A cutoff return period is defined at the exceedence probability associated with the zero flows. The flow for any return period greater than this cutoff return period is then set to zero. The low flow for shorter return periods is found directly from the distribution fitted through the positive values.

The conditional probability approach can be very difficult to apply. For methods with a tractable *cdf*, the method is very easy. But for methods such as those which use the log-Pearson and normal distributions, this method can be complex. Problems of interest usually require calculation of x for a given $F(x)$. In the case of non-intermittent streams, the problem typically reduces to that of finding the frequency factor associated with standard values of $F(x)$. With intermittent streams, however, the value of x associated with $F^*(x)$ must be found which in general requires a numerical integration of the *pdf* with varying upper limits by trial and error for intractable *cdf*'s. This numerical integration is in general tedious. Thus, normal and log Pearson are hard when intermittent streams are considered.

Also, positive skew for the log Pearson implies a positive lower bound. The Weibull (α, β, γ) also can have a positive lower bound. Thus, for the intermittent

streams, when the Weibull or the log Pearson is used along with the conditional probability approach, there is a positive mass at zero but any flow between the positive lower bound and zero has a probability zero. This aspect is physically unreasonable.

Comparison of Methods

A computer program has been written to implement each of the distributions discussed in this paper. Twenty stations in Virginia with continuous streamflow records exceeding 30 years were randomly selected from the HISARS data base. The summary in the Appendix contains the seven day low flows for return periods of 2, 5, 10, 25, 50, and 100 years. The results are given by individual stations and are identified by their USGS station numbers.

A comparison of the results from the SMEMAX and modified SMEMAX methods for the 20 stations leads to the conclusion that in all cases C_s is closer to zero for the transformed data than for the original data. The power transformation always produces a coefficient of skewness equal to zero. While the skewness using the SMEMAX transformation is near zero, it typically does not exactly equal to zero. The values of C_s in fact range from -0.153 to 0.176 .

The transformation methods had much less success in obtaining a value of C_k near three in the transformed range. The results from the 20 stations indicate a definite bias in the value of C_k . In all 20 cases using the power transformation and in 19 of the 20 cases using the SMEMAX transformation, the transformed value of C_k was found to be less than three. Thus, there is a bias towards platykurtic (flatter than normal) distributions. The bias is a much more difficult problem with the SMEMAX transformation since the data may or may not be symmetrical. Also, Westphal (1984) has shown that the SMEMAX transformation could lead to non-unimodal distributions.

Based on the ability of the power transformation to produce a symmetrical distribution, the power transformation is probably a better transformation method. Both methods, however, suffer because of the use of the normal distribution. Since a tractable form of the *cdf* is not available, the theoretically desirable approach developed by Jennings and Benson (1969) for intermittent streams cannot easily be applied on a computer. The transformation methods often produced negative flows (which show up as zeros) for large return periods. The negative flows are due to negative cutoff values (frequency factors) at low probabilities for normal distribution. Thus, the ability of these methods to adequately model the lower tails is in doubt.

Chander et al. found that for floods, the value of λ in the power transformation was usually between -1 and 1 . In this study, λ for the 20 stations ranged from -1.733 to 0.662 . The large negative value of λ causes the transformed values to

group around $1/\lambda$. Double precision arithmetic is required to differentiate between the transformed values.

The Weibull (α, β, γ) distribution when applied using the method of moments fitting procedure, was found to give very poor estimates of the lower bound. Physically reasonable values were generated in only 4 of 20 cases. In 14 cases, γ was found to be less than zero and in 2 cases γ was greater than the smallest observed value. The maximum likelihood fitting procedure was also found to be unacceptable. In 14 cases the technique failed to converge to an answer and thus produced no results. The method of smallest observed drought suffers from both of these shortcomings to a lesser degree. In 3 cases, the method did not converge to an answer. In 8 of the remaining 17 cases, a value of γ less than zero was found.

The Weibull $(\alpha, \beta, 0)$ distribution when there is a practical lower bound near zero (station 16150.00), compares favorably with the other methods. When there is a practical lower bound significantly different from zero (station 31730.00), however, the Weibull distribution tends to produce results that are less than those found by the other methods. Since reliability is an important characteristic for hydrologic applications this distribution is not acceptable for low flows even though it has shape flexibility (Tadikamalla 1978).

In fitting the log-Pearson type III distribution, the indirect method of moments tends to perform reliably for all stations. The method of moments, direct, while being more theoretically sound, is also more complex. This complexity can cause problems. When a value of α greater than zero is found, this method results in very small values of α and very large values of β (station 16695.00). Since the *pdf* of the distribution requires calculations of the gamma function of β , a computational overflow situation normally occurs. Because of these difficulties with the direct method of moments for low flows, the indirect method of moments is preferable. Positive α values were obtained for eight out of 20 cases. Positive α also implies a positive lower bound on the flow values. Even though a positive lower bound was imposed, the fit was generally good as shown in the Appendix.

The log-Boughton method does not suffer from the shortcomings of the other methods. Intermittent streams are handled with ease. There is also no need to try and guess the underlying distribution. Finally, the method directly fits the probabilities to the observed values. The other methods use the observed sample only in estimating the moments, coefficients, etc. of the distribution. Also frequency factors are directly obtained without the need of tables.

Summary and Conclusions

The new general SMEMAX transformation enables the user to choose arbitrary lower and upper bounds; the method is also simpler than those found in literature (Bethlahmy 1977; Rasheed et al. 1982). It is also found that the power transforma-

tion leads to a more symmetrical distribution than the SMEMAX. The results of power transformation also compare favorably with the other distributions. There are at least three parameters needed in the distribution function to model the low flow probabilities. It is seen that Weibull ($\alpha, \beta, 0$) distribution does not provide a good fit. Even though Normal distribution has only two parameters, other parameters are introduced in transforming the data: λ in the case of the power transformation, Z_s and Z_1 in the case of SMEMAX. The Weibull (α, β, γ) and log Pearson suffer from the problem of positive lower bound which is physically hard to justify; but the three parameters lead to a better curve fit. The intermittent streams pose the special problem of prolonged zero flows. The point mass at zero must be considered. In the case of positive lower bounds, the probability mass being zero between the lower bound and zero is not physically justifiable. If the cumulative distribution does not have a closed form, extensive numerical schemes are required. In such cases log Boughton with Cunnane's plotting position may be preferred.

Acknowledgements

The work reported in this paper is based on the research supported by the Virginia Water Resources Research Center. The authors would like to thank Prof. W. R. Walker, Dr. Margaret Hrezo and Mr. Tom Johnson of the water center. The authors thank Prof. M. H. Diskin for his valuable comments on this study. The skillful typing by Ms. Ruby Shaw is appreciated.

References

- ASCE Task Committee on Low-Flow Evaluation, Methods, and Needs (1980) Characteristics of Low Flows, *Journal of the Hydraulics Division, ASCE, Vol. 105*, No. HY5, pp. 717-731.
- Benson, Manuel A. (1968) Uniform Flood-Frequency Estimating Methods for Federal Agencies, *Water Resources Research, Vol. 4*, No. 5, pp. 891-908.
- Benjamin, Jack R., and Cornell, C. Allin (1970) *Probability, Statistics, and Decision for Civil Engineers*, New York, McGraw-Hill.
- Bethlahmy, Nedavia (1977) Flood Analysis by SMEMAX Transformation. *Journal of the Hydraulics Division, ASCE, Vol. 103*, No HY1, pp. 69-78.
- Bobee, Bernard (1975) The Log Pearson Type 3 Distribution and its Application in Hydrology, *Water Resources Research, Vol. 11*, No. 5, pp. 681-689.

Frequency Analysis of Low Flows

- Boughton, Walter C. (1980) A Frequency Distribution for Annual Floods, *Water Resources Research*, Vol. 16, No. 2,1 pp. 347-354.
- Boughton, Walter C., and Shirley, Edward C. (1983) Fitting Procedures for the Log-Boughton Distribution, *Journal of the Hydraulics Division, ASCE*, Vol. 109, No. 4, pp. 579-589.
- Box, G. E. P., and Cox, D. R. (1964) An Analysis of Transformation, *Journal of the Royal Statistical Society*, Vol. B26, pp. 211-252.
- Chander, Subhash, Spolia, S. K., and Kumar, Arun (1978) Flood Frequency Analysis by Power Transformation, *Journal of the Hydraulics Division, ASCE*, Vol. 104, No. HY11, pp. 1495-1504.
- Cunnane, C. (1978) Unbiased Plotting Position – A Review. *Journal of Hydrology*, Vol. 37, No. 3, pp. 205-222.
- Deininger, Rolf A., and Westfield, James D. (1969) Estimation of the Parameters of Gumbel's Third Asymptotic distribution by Different Methods, *Water Resources Research*, Vol. 5, No. 6, pp. 1238-1243.
- Gumbel, E. J. (1963) Statistical Forecast of Droughts. *Bulletin of the International Association of Scientific Hydrology*, Vol. 8, No. 1, pp. 5-23.
- Jennings, M. E., and Benson, M. A. (1969) Frequency Curves for Annual Flood Series with Some Zero Events or Incomplete Data, *Water Resources Research*, Vol. 5, No. 1, pp. 276-280.
- Johnson, N. L., and Kotz, S. (1970) *Continuous Univariate Distributions-I Distributions in Statistics*, John Wiley, New York.
- Johanson, T. W., Wysor, J., Marchany, R., and Wallace, L. (1975) User's Guide for HISARS, Virginia Water Resources Research Center, Blacksburg.
- Kite, G. W. (1977) Frequency and Risk Analysis in Hydrology, Fort Collins: Water Resources Publications.
- Kumar, Arun, and Devi, Rema (1982) Statistical Determination of Design Low Flows – A Comment. *Journal of Hydrology*, Vol. 58, pp. 175-177.
- Prakash, Anand (1981) Statistical Determination of Design Low Flows, *Journal of Hydrology*, Vol. 51, pp. 109-118.
- Rasheed, Hameed R., Ramamoorthy, Marur V., and Al-Dabbagh., A. S. (1982) Modified SMEMAX Transformation for Frequency Analysis, *Water Resources Bulletin*, Vol. 18, No. 3, pp. 509-511.
- Srikanthan, R., and McMahon, T. A. (1981) Log Pearson III Distribution – An Empirically-Derived Plotting Position. *Journal of Hydrology*, Vol. 52, pp. 161-163.
- Tadikamalla, Pandu R. (1978) Applications of the Weibull Distribution in Inventory Control. *Journal of the Operational Research Society*, Vol. 29, No. 1, pp. 77-83.
- U.S. Water Resources Council (1977) Guidelines for Determining Flood Flow Frequency, Bulletin 17A, Washington, D.C.: GPO.
- Westphal, Jerome A. (1984) SMEMAX? Caution! *Journal of Hydraulic Engineering, ASCE*, Vol. 110, No. 3, pp. 209-218.

First received: 22 November, 1984

Revised version received: 28 February, 1985

2 Year Return Period – 7 Day Average Low Flows

Station Number	Power	SMEMAX & MOD SMEMAX (M of M)	Weibull (α, β, γ) (Max. L)	(M of SOD)	Weibull ($\alpha, \beta, 0$)	Log Pearson III (M of M,D) (M of M,I)	Bought-ton
16150	4.23	3.73	4.30	4.37	4.26	4.22	4.37
16320	3.40	3.46	3.59	3.47	3.49	3.43	3.53
16440	12.13	9.07	14.81	13.45	13.38	12.82	13.00
16460	11.23	10.81	11.35	11.37	10.98	11.07	11.89
16620	14.86	12.83	16.81	15.46	15.37	15.14	15.03
16650	2.14	1.88	2.91	2.14	2.13	2.13	2.30
16680	190.10	172.10	195.30	186.45	188.24	188.38	196.13
16685	2.16	1.89	2.70	2.36	2.32	3.00	2.10
16695	3.99	2.79	4.91	4.35	4.28	5.43	3.67
16725	38.35	37.56	39.76	37.59	37.60	38.31	38.84
20175	10.83	10.41	11.03	10.93	11.89	11.24	11.43
20225	6.69	6.65	6.78	6.69	7.10	6.82	6.88
20345	29.68	27.00	31.89	29.84	29.58	30.21	30.69
20425	17.53	16.21	18.15	17.41	17.18	17.37	17.82
20440	1.09	0.97	1.34	1.14	1.12	1.37	1.08
31730	43.76	41.63	44.61	44.60	47.65 ^F	45.23	46.30
31765	1515.54	1455.90	1497.96	—	1603.40	1519.42	1542.81
32090	10.85	9.28	14.77	12.24	12.06	11.26 ^F	11.70
34715	24.67	23.94	24.21 ^F	24.84	26.62	24.77	26.34
34880	33.71	32.54	34.26	33.83	36.74	34.82	35.24

F = Kolmogorov-Smirnov test fails.

— = did not converge.

M of M = Method of Moments; Max. L = Maximum Likelihood Procedure; M of SOD = Method of Smallest Drought M of M,D = Method of Moments Direct; M of M, I = Method of Moments Indirect.

Frequency Analysis of Low Flows

5 Year Return Period - 7 Day Average Low Flows

Station Number	Power	SMEMAX & MOD SMEMAX	(M of M)	Weibull (α,β,γ) (Max. L)	(M of SOD)	Weibull (α,β,0)	(M of M,D)	Log Pearson III (M of M,I)	Boughton
16150	2.22	2.40	2.16	2.17	2.15	2.14	2.16	2.25	2.22
16320	1.37	1.68	1.27	-	1.37	1.32	1.29	1.31	1.43
16440	4.34	5.04	3.03	-	4.42	4.29	3.20	4.27	4.59
16460	5.92	6.33	5.88	-	5.88	5.84	5.64	6.20 ^F	6.57
16620	5.19	6.32	4.71	-	5.73	5.55	4.56	4.84	5.50
16650	1.08	1.19	1.06	1.06	-	1.05	1.04	1.11 ^F	1.19
16680	90.73	101.12	89.00	88.06	-	90.17	85.93	85.58	94.26
16685	0.00	0.75	0.14	-	0.61	0.61	0.27	0.51	0.29
16695	0.00	1.24	0.29	-	1.12	1.12	0.46	0.96	0.82
16725	18.92	20.52	18.75	-	19.07	18.82	17.36	17.45	20.03
20175	8.75	8.87	8.44	-	8.51	8.31	8.59	8.69	8.76
20225	5.45	5.52	5.37	5.37	5.37	5.41	5.45	5.45	5.48
20345	13.64	14.99	13.41	-	13.86	13.72	12.03	12.88	14.57
20425	8.04	8.87	7.99	-	8.17	8.09	7.51	7.42	8.29
20440	0.00	0.39	0.27	-	0.39	0.39	0.34	0.33	0.27
31730	35.93	36.39	34.92	35.08	34.93	34.76 ^F	35.52	35.75	36.24
31765	1240.69	1257.51	1222.51	-	-	1219.98	1239.57	1238.09	1240.57
32090	2.98	3.94	2.30	-	3.95	3.94	1.46 ^F	2.69 ^F	3.85
34715	21.36	21.53	21.05 ^F	20.78	20.49	20.33	21.08	21.23	21.77
34880	26.87	27.31	26.06	26.13	26.15	25.80	26.54	26.77	26.94

F = K-S test fails.
 - = did not converge.

10 Year Return Period – 7 Day Average Low Flows

Station Number	Power	SMEMAX & MOD SMEMAX		Weibull (α,β,γ) (Max. L)	(M of SOD)	Weibull (α,β,0)	(M of M,D)	Log Pearson III (M of M,I)	Boughton
16150	1.45	1.71	1.36	1.40	1.27	1.35	1.43	1.39	1.38
16320	0.00	0.76	0.55	-	0.78	0.69	0.70	0.76	0.78
16440	2.38	2.93	0.00	-	2.16	2.02	1.28	2.40	2.36
16460	3.63	3.99	3.53	-	3.51	3.85	3.64	3.25 ^F	3.81
16620	2.51	2.92	0.59	-	3.03	2.83	2.06	2.55	2.65
16650	0.00	0.83	0.62	0.66	-	0.66	0.66	0.63 ^F	0.69
16680	53.10	63.75	47.28	53.61	-	55.39	51.82	48.62	53.79
16685	0.00	0.15	0.00	-	0.23	0.26	0.00	0.00	0.06
16695	0.00	0.43	0.00	-	0.44	0.46	0.00	0.00	0.31
16725	11.10	11.62	9.69	-	12.28	11.91	10.21	10.86	12.03
20175	7.96	8.07	7.60	-	7.77	6.56	7.50	7.86	7.60
20225	4.94	4.93	4.82	4.90	4.85	4.51	4.84	4.92	4.84
20345	7.82	8.71	5.64	-	8.23	8.26	6.43	7.67	8.26
20425	4.45	5.03	3.87	-	4.83	4.92	4.32	4.25	4.59
20440	0.00	0.09	0.00	-	0.18	0.19	0.00	0.00	0.08
31730	32.91	33.51	31.60	32.06	31.62	28.20 ^F	31.42	32.58	31.60
31765	1125.83	1153.81	1131.37	-	-	1018.05	1122.29	1126.55	1106.51
32090	1.18	1.14	0.00	-	1.78	1.88	0.31 ^F	1.22	1.56
34715	20.06	20.26	20.29 ^F	19.51	19.15	17.01	19.71	20.24	19.33
34880	24.23	24.58	23.29	23.73	23.56	20.42	23.09	23.99	23.36

F = K-S test fails.
 - = did not converge.

Frequency Analysis of Low Flows

25 Year Return Period - 7 Day Average Low Flows

Station Number	Power	SMEMAX & MOD SMEMAX	Weibull (Max. L)	Weibull (α, β, γ) (M of SOD)	Weibull ($\alpha, \beta, 0$)	Log Pearson III (M of M,D)	Log Pearson III (M of M,I)	Boughton
16150	0.00	0.96	0.74	0.82	0.57	0.76	0.87	0.73
16320	0.00	0.00	0.08	-	0.42	0.31	0.34	0.35
16440	1.18	0.69	0.00	-	0.92	0.78	0.41	1.02
16460	1.65	1.50	1.50	-	1.45	2.27	2.12	1.38 ^F
16620	0.00	0.00	0.00	-	1.41	1.20	0.77	0.96
16650	0.00	0.45	0.26	0.36	-	0.37	0.38	0.27
16680	23.91	23.90	13.97	28.65	-	29.92	27.92	23.29
16685	0.00	0.00	0.00	-	0.05	0.08	0.00	0.00
16695	0.00	0.00	0.00	-	0.10	0.15	0.00	0.00
16725	4.76	2.12	1.85	-	7.16	6.67	5.26	5.59
20175	7.26	7.21	7.04	-	7.30	4.86	6.51	6.52
20225	4.47	4.30	4.38	4.55	4.45	3.59	4.27	4.16
20345	3.46	2.01	0.00	-	4.13	4.34	2.92	3.56
20425	1.72	0.94	0.50	-	2.36	2.62	2.18	1.87
20440	0.00	0.00	0.00	-	0.06	0.08	0.00	0.01
31730	30.22	30.53	29.24	30.00	29.27	21.66 ^F	27.64	27.09
31765	1020.06	1043.60	1068.45	-	-	809.99	1041.61	979.04
32090	0.00	0.00	0.00	-	0.56	0.74	0.04 ^F	0.37
34715	18.89	18.92	19.88 ^F	18.67	18.28	13.58	18.58	16.70
34880	21.90	21.67	21.35	22.15	21.79	15.19	19.94	20.02

F = K-S test fails.

- = did not converge.

50 Year Return Period – 7 Day Average Low Flows

Station Number	Power	SMEMAX & MOD SMEMAX (M of M)		Weibull (α,β,γ) (Max. L)	(M of SOD)	Weibull (α,β,0)	Log Pearson III (M of M,D)	Boughton
16150	0.00	0.49	0.46	0.56	0.24	0.49	0.61	0.48
16320	0.00	0.00	0.00	-	0.29	0.17	0.20	0.26
16440	0.00	0.00	0.00	-	0.53	0.38	0.18	0.81
16460	0.00	0.00	0.47	-	0.40	1.53	1.44	0.60 ^F
16620	0.00	0.00	0.00	-	0.85	0.64	0.37	0.71
16650	0.00	0.20	0.10	0.23	-	0.24	0.26	0.16 ^F
16680	11.11	0.00	0.00	18.02	-	18.95	17.88	14.00
16685	0.00	0.00	0.00	-	0.00	0.04	0.00	0.00
16695	0.00	0.00	0.00	-	0.01	0.07	0.00	0.00
16725	1.91	0.00	0.00	-	4.89	4.34	3.23	4.07
20175	6.87	6.66	6.81	-	7.12	3.89	5.95	6.79
20225	4.20	3.89	4.18	4.40	4.27	3.03	3.93	4.16
20345	1.61	0.00	0.00	-	2.38	2.70	1.63	2.61
20425	0.00	0.00	0.00	-	1.27	1.64	1.32	1.30
20440	0.00	0.00	0.00	-	0.01	0.04	0.00	0.00
31730	28.72	28.61	28.25	29.18	28.29	17.81 ^F	29.48	28.40
31765	959.47	971.79	1042.92	-	-	683.63	953.23	969.40
32090	0.00	0.00	0.00	-	0.15	0.37	0.01 ^F	0.25
34715	18.23	18.05	19.75	18.34	17.94	11.48	17.99	19.19
34880	20.60	19.79	20.56	21.54	21.08	12.20	18.15	20.28
								24.42
								904.27

F = K-S test fails.

- = did not converge.

Frequency Analysis of Low Flows

100 Year Return Period - 7 Day Average Low Flows

Station Number	Power	SMEMAX & MOD SMEMAX (M of M)	Weibull (α, β, γ) (Max. L) (M of SOD)	Weibull ($\alpha, \beta, 0$) ($\alpha, \beta, 0$)	Log Pearson III (M of M, D) (M of M, I)	Boughton
16150	0.00	0.06	0.39	0.02	0.44	0.24
16320	0.00	0.00	-	0.22	0.12	0.09
16440	0.00	0.00	-	0.33	0.08	0.29
16460	0.00	0.00	-	0.00	0.98	0.05
16620	0.00	0.00	-	0.55	0.18	0.16
16650	0.00	0.00	0.14	-	0.17	0.03
16680	3.70	0.00	11.38	-	11.60	4.41
16685	0.00	0.00	-	0.00	0.00	0.00
16695	0.00	0.00	-	0.00	0.00	0.00
16725	0.00	0.00	-	3.43	2.00	1.20
20175	6.56	6.16	-	7.02	5.47	5.39
20225	3.98	3.53	4.31	4.15	3.65	3.54
20345	0.00	0.00	-	1.28	0.91	0.68
20425	0.00	0.00	-	0.58	0.81	0.30
20440	0.00	0.00	-	0.00	0.00	0.00
31730	27.49	26.89	28.68	27.69	23.71	22.16
31765	909.45	907.53	-	-	902.83	841.62
32090	0.00	0.00	-	0.00	0.00 ^F	0.01
34715	17.69	17.26	18.15	17.74	17.54	13.49
34880	19.55	18.10	21.19	20.66	16.69	16.52

F = K-S test fails.
 - = did not converge.

G.V. Loganathan et al.

Address:

Department of Civil Engineering,
Room 200 Patton Hall,
Virginia Polytechnic Institute
and State University,
Blacksburg, VA-24061,
U.S.A.