Nordic Hydrology, 17, 1986, 129-150

Frequency Analysis of Low Flows: Hypothetical Distribution Methods and a Physically Based Approach

G. V. Loganathan, P. Mattejat, C. Y. Kuo and M. H. Diskin

Dept. of Civil Eng., Virginia Tech.,
Blacksburg, VA-24061, U.S.A.

A mixed log Pearson type III distribution, a double bounded probability density function, partial duration series and a physically based approach are analyzed for frequency estimates of low flows. The mixed log Pearson III involves a point probability mass at zero for intermittent streams. The double bounded probability distribution has lower and upper bounds with a point mass at the lower bound. Two approaches are used in partial duration series i) truncation, and ii) censoring which represent curtailing of the population and the sample respectively. The parameters are estimated by maximum likelihood procedure. Considering low flows as part of the recession limb of stream flow hydrographs a physically based approach is formulated. By using the exponential decay of stream recessions and considering the initial recession flows, recession durations, and recharge due to incoming storms as statistically independent random variables, a first order random coefficient Markov model for initial recession flows is formed. The resulting steady state probability distribution for initial recession flows is combined with the probability distribution of the exponential decay to obtain the probabilities of low flow events. The methods are applied to both perennial and intermittent streams.

Introduction

Adequate streamflow is necessary to satisfy agricultural, human and wildlife needs. Accurate prediction of low flows is required for effective management of streamflow to meet all demands and water quality needs. Critical flows for specified return periods can be assessed by frequency analysis. A number of probability distributions are used in the frequency analysis of low flows. These distributions may be grouped into the following categories:

1) Transformation methods
2) Hypothetical distribution based methods
3) Plotting position based methods
4) Physically based probability methods

$\left.\begin{array}{l}1) \\ 2) \\ 3)\end{array}\right\}$ Black box methods

The first three categories are collectively called black box methods. The transformation methods attempt to find a unique probability distribution in the transformed range regardless of the distribution of the original sample. Power transformation proposed by Box and Cox (1964) transforms the original sample to a near normal distribution. This transformation has been widely used in hydrologic time series analysis. Kumar and Devi (1982) applied it to the frequency analysis of low flows. Bethlahmy (1977) suggested the SMEMAX (Small, MEdian and MAXimum) transformation to transform the original data to a near normal distribution. Prakash (1981) applied it to the analysis of low flows. Rasheed *et al.* (1982) suggested a modified SMEMAX transformation in which the range can be chosen arbitrarily. However, Westphal (1984) showed that the SMEMAX transformation could lead to a multi-modal probability density function when the original data were log Pearson or Gumbel distributed. McCormick (1984) indicated that the power transformation could result in clustering of transformed values if the parameter used in the transformation were a negative number. Also, neither method guarantees a kurtosis value of 3. In low flow analysis transformations to the normal distribution are widely used. There are also other transformations (Kruskal and Tanur 1978).

The hypothetical distribution method involves fitting a probability distribution directly to the sample data. The commonly used probability distributions for low flow analysis are the normal, log normal, gamma, log Pearson type III, Gumbel type III, and Weibull. McMahon and Arenas (1982), and Matalas (1963) presented extensive reviews of the application of various probability distributions to low flow data. Also, the ASCE Task Committee Report (1980) provided a review of the practical aspects and methods of analysis of low flows. Kite (1977) and Kottegoda (1980) present fitting techniques for various probability distributions to hydrologic data. Johnson and Kotz (1970) provide detailed accounts of continuous distributions. Many of these distributions involve a nonzero lower bound which is physically unsuitable for low flows. Intermittent streams have a positive probability of zero flows and so a mixed distribution should be used. Aitchison (1955) provided an excellent account of this approach. Jenings and Benson (1969) used such a mixed distribution for floods.

The plotting position based methods are probably the most straightforward of various methods. A plotting position formula is used to estimate the exceedance probability for each observed discharge and a curve is fit through these probability points. Cunnane (1978) emphasized unbiased plotting positions and dependence of the plotting position on the underlying probability distribution of the original data. Boughton (1980, 1983) suggested a general curve fitting procedure for the plotting position method. In this paper several black box methods namely the mixed log Pearson type III, a double bounded probability density function, and partial duration series are considered. The partial duration series applies both truncation and censoring.

The major limitation of the black box methods is that they do not incorporate the physics of the hydrologic process. To alleviate this problem, the physically based probability methods are suggested. These methods use hydrologic relationships and hypothesized probability distributions for independent random hydrologic variables to derive new distributions for the dependent variables. These may be quite involved compared to the black box methods. In this paper a physically based probability distribution for low flows is proposed. The method uses the well known exponential stream recession function with recession duration, initial recession flow, and recharge due to incoming storm as independent random variables to derive a probability distribution for the lowest flow on the recession limb. Application of the methods includes both perennial and intermittent streams.

Log-Pearson Type III Distribution

The log-Pearson type III (LPIII) distribution is widely applied for flood frequency analysis, as well as for low flow analysis. LPIII uses annual minimum low flows which represent an annual series. For annual low flows a water year different from the calendar year is used. Low flows generally occur in summer, fall, or winter. A long duration of low flows may overlap into another year and influence the succeeding minimum which contradicts the assumption of independence in frequency analysis. In this study the year begins in April and ends in March (ASCE 1980). The probability density function (pdf) for annual minimum flows is given by

$$f_Q(q) = \left(\frac{\ln q-\gamma}{\alpha} \right)^{\beta-1} \frac{\exp\left(-\frac{\ln q-\gamma}{\alpha} \right)}{|\alpha| \, q \, \Gamma(\beta)} \qquad (1)$$

where

α – scale parameter,
β – shape parameter,
γ – location parameter,
Γ(β) – gamma function at β.

The range of q is given by

$\beta > 0, \quad \alpha > 0, \quad \text{then } \exp(\gamma) < q < \infty$

$\beta > 0, \quad \alpha < 0, \quad \text{then } 0 < q < \exp(\gamma)$

LPIII has no closed form expression for the cumulative distribution function (cdf), however q_T, the flow occurring once in T years, can be calculated by

$$q_T = \exp(M_y + K_T S_y) \tag{2}$$

where

K_T – frequency factor,
M_y – mean of ln (q),
S_y – standard deviation of $\ln(q)$ [see also Notation].

Extensive tables for K_T are available for different values of return period, T, and skew coefficient, G (USWRC, 1977). Also Eq. (3) can be used for the range of skew $-1 < G < 1$ (Condie 1977; McGinnis and Sammons 1970).

$$K_T \equiv \frac{2}{G}\left[\left(\left(x_T - \frac{G}{6}\right)\frac{G}{6} + 1\right)^3 - 1\right] \tag{3}$$

where

K_T – frequency factor for T,
G – sample skew of $\ln(q)$,
x_T – standard normal deviate for $F(x_T) = \mathrm{P}(\mathrm{X}=\mathrm{x_T}) = 1/\mathrm{T}$.

To apply LPIII for data with zero flows, a positive constant may be added to each flow, and then for each calculated q_T the constant is subtracted. An alternative is to modify LPIII with a point mass probability for zero flow, called the conditional probability approach. The probability of zero flow is estimated by

$$P(Q=0) = \frac{k}{j} \tag{4}$$

where

k – number of zero flows in record,
j – number of flows in record.

Then the probability of flow equal to or less than q_T is

$$P(Q \leq q_T) \equiv P(Q=0) + P(Q \leq q_T | Q>0) \; [1 - P(Q=0)] \tag{5}$$

Recalling $P(Q \leqslant q_T) = 1/T$ it is obtained

$$P(Q \leq q_T | Q>0) \equiv \frac{(1/T) - (k/j)}{1 - (k/j)} \tag{6}$$

If there are no zero flows, for a prespecified return period, T, the standard normal deviate x_T in Eq. (3) can be chosen a priori from the standard normal tables. However for intermittent streams, Eq. (6) governs and x_T cannot be chosen a priori without knowing k and j which vary for different streams. Therefore it is preferable to use an expression relating $F(x_T)$ and x_T. Zelen and Severo (1972) give a series expansion for the cdf, $F(x)$, of the standard normal distribution which is

$$F(x) = 1 - \frac{1}{2}(1 + d_1x + d_2x^2 + d_3x^3 + d_4x^4 + d_5x^5 + d_6x^6)^{-16} \qquad (7)$$

where x = standard normal deviate

$d_1 = .0498673470$	$d_4 = .0000380036$
$d_2 = .0211410061$	$d_5 = .0000488906$
$d_3 = .0032776263$	$d_6 = .0000053830$

Since $F(x) = P(Q \leqslant q_T|Q > 0)$ is given Eq. (6), x can be found by trial error. The approximation deviates for return periods larger than 50 years, and thus care must be taken in applying the approximation. The estimated value of x is used in Eq. (3) to obtain K_T which is then substituted in Eq. (2) along with the standard deviation and mean of the logarithm of positive annual minimum flows to obtain q_T.

Double Bounded PDF with a Point Mass at Zero

In general low flows are limited by a lower bound, say U, and an upper bound, V. For intermittent streams the lower bound is zero and has a positive probability of occurrence. A comprehensive probability distribution for low flows requires a point mass at zero, bounded by an upper bound, and the ability to fit a variety of shapes for different stream data.

Based on the above characteristics an ideal probability density function (pdf) for low flows may have the form

$$\begin{aligned} f_X(x) &= F_0 && \text{for} \quad x = 0 \\ &= (1 - F_0) f_1(x) && \text{for} \quad 0 < x < 1 \end{aligned} \qquad (8)$$

where

x – $(q - U)/(V - U)$, and
q – the low flow variate.

The pdf given in Eq. (8) involves the parameters F_o, U, V, and the parameters of $f_1(x)$ for positive flows. In the following a maximum likelihood approach for the parameter estimation of f_X will be given. The likelihood function, L_1, for the distribution in Eq. (8) may be written as

$$L_1 = F_0^k (1-F_0)^{(j-k)} \prod_{i=1}^{j-k} f_1(x_i) \qquad \text{for} \qquad x_i > 0 \qquad (9)$$

where

j – total number of annual minima,
k – number of zero flows.

The derivatives of L_1 with respect to the parameters are set equal to zero to achieve the maximum likelihood. Kumaraswamy (1980) suggested the following form for $f_1(q)$

$$f_1(q) = \frac{ab}{V-U}\left(\frac{q-u}{V-U}\right)^{a-1}\left[1-\left(\frac{q-U}{V-U}\right)^{a}\right]^{b-1} \quad \text{for} \quad U < q < V \tag{10}$$

where

U – lower bound of q,
V – upper bound of q, taken to be the largest annual minimum,
F_o – point mass probability at U,
a, b – shape parameters positive.

The cumulative distribution function (cdf), $F(q)$, is given by

$$F(q) \equiv F_0+(1-F_0)\left[1-\left(1-\left(\frac{q-U}{V-U}\right)^{a}\right)^{b}\right] \tag{11}$$

The utility of the double bounded probability density function (DBPDF) is its flexibility to change shape with various combinations of a and b. Also the analytical form of Eq. (11) allows for straightforward application. However, estimating parameters a, b, F_o, and U (V is fixed as the largest annual minimum) requires an iterative technique. In this study a maximum likelihood method is used to estimate the parameters. Kumaraswamy (1980) presented a method of moments procedure for estimating the parameters.

Parameter Estimation

For intermittent streams $U = 0$; whereas for perennial streams U may assume any value according to the fitting of parameters to data. For perennial streams this is a shortcoming since U may be negative (station 03.1700.00) and thus giving a positive probability for negative flows which is unsatisfactory, or a positive U which physically seems unacceptable.

Considering a sample set of j annual minima q_1, q_2, ..., q_j which contains k zero flows and j-k positive flows, the likelihood function L_1 can be written for the DBPDF as

$$L_1 = F_0^{k}\prod_{i=1}^{j-k}(1-F_0)\left(\frac{ab}{V-U}\right)\left(\frac{q_i-U}{V-U}\right)^{a-1}\left[1-\left(\frac{q_i-U}{V-U}\right)^{a}\right]^{b-1} \tag{12}$$

Taking the natural logarithm of Eq. (12) simplifies the differentiation of the maximum likelihood function, given by $L = \ln(L_1)$, and therefore

$$L \equiv k\ln F_0 + (j-k)\ln(1-F_0)+(j-k)(\ln a+\ln b) - ab(j-k)\ln(V-U) + (a-1)\sum_{i=1}^{j-k}\ln(q_i-U)+(b-1)\sum_{i=1}^{j-k}\ln[(V-U)^{a}-(q_i-U)^{a}] \tag{13}$$

Setting the derivatives of the log likelihood function to zero, gives the maximum likelihood estimates for *a, b,* F_0, and *U*. Differentiating *L* with respect to F_0 yields

$$\frac{\partial L}{\partial F_0} \equiv \frac{k}{F_0} - \frac{j-k}{1-F_0} \equiv 0 \tag{14}$$

and therefore

$$F_0 \equiv \frac{k}{j} \tag{15}$$

Also

$$\frac{\partial L}{\partial a} = \frac{j-k}{a} - b(j-k)\ln(V-U) + \sum_{i=1}^{j-k} \ln(q_i-U)$$

$$+ (b-1)\sum_{i=1}^{j-k} \left[\frac{(V-U)^a \ln(V-U) - (q_i-U)^a \ln(q_i-U)}{(V-U)^a - (q_i-U)^a}\right] \tag{16}$$

and

$$\frac{\partial L}{\partial b} = \frac{j-k}{b} - a(j-k)\ln(V-U) + \sum_{i=1}^{j-k} \ln[(V-U)^a - (q_i-U)^a] \tag{17}$$

by simplifying

$$b \equiv \frac{(j-k)}{\{a(j-k)\ln(V-U) - \sum_{i=1}^{j-k} \ln[(V-U)^a - (q_i-U)^a]\}} \tag{18}$$

$$\frac{\partial L}{\partial U} = \frac{ab(j-k)}{(V-U)} - (a-1)\sum_{i=1}^{j-k} \frac{1}{q_i-U} + a(b-1)\sum_{i=1}^{j-k} \frac{(q_i-U)^{a-1} - (V-U)^{a-1}}{(V-U)^a - (q_i-U)^a} \tag{19}$$

where $\frac{\partial L}{\partial a} = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial U} = 0$

It is interesting to note that the maximum likelihood estimator of F_0 in Eq. (15) is the same as the frequency estimator. Eq. (18) gives an explicit solution for *b*. The solution of *a* and *U* requires an iterative scheme and the following strategy is used for the solution.

STEP 1: Substitute Eq. (18) in Eq. (16) for *b*.
STEP 2: Assume an initial value of *U*.
STEP 3: For the assumed value of *U*, Eq. (16) can be solved for *a* (by method of bisection).
STEP 4: By using *a* from Step 3, compute *b* from Eq. (18).
STEP 5: Substitute computed *a, b* and the assumed *U* in (Eq. (19). If it is zero, the

parameters are the maximum likelihood estimates. If not, U is decreased (increased) by a small amount according as the sign of Eq. (19) is negative (positive). Proceed to Step 3.

Knowing the parameters and T, q_T is solved explicitly using Eq. (11)

$$q_T = U + (V-U)\left[1-\left\{1-\left(\frac{1/T-F_0}{1-F_0}\right)\right\}^{1/b}\right]^{1/a} \tag{20}$$

Partial Duration Series

The methods predicting low flows using annual minimum flows are called annual minima series methods, for example log-Pearson type III and Double Bounded pdf. Thus, for n years of record only n flows are used in the prediction. The main drawback in using only one low flow each year is that some years may have two or more low flows, which are severer than several annual minima in the record but only the severest is used. Partial duration series (PDS) is a series of data comprised of flows less than a pre-specified threshold, q_o. PDS has a larger record of flows than the annual minima series. The threshold value must be such that the selected low flows are independent of each other.

PDS can be viewed as two processes: i) Truncation, and ii) Censoring. For a chosen flow level q_0, let n flows be below q_0, in which n is a random variable and there be no information about flows which exceed q_0. This is called a truncation process. Traditional PDS is a form of truncation since any low flow occurring above q_0 is not considered. Suppose the sample size j is known. Then it is also known that in addition to the n flows below q_0, there exist (j-n) flows exceeding q_0. This is called Type I censoring in which flows below q_0 and the number of flows above q_0 but not their values (their values being censored) are known. There is also Type II censoring in which a fixed number, m, flows less than a threshold value, q_0 are observed and q_0 is a random variable. In censoring, observations are curtailed above or below a certain level, but in truncation the entire population is generated from a curtailed range. In hydrology Type I censoring is typically used. Leese (1971) has applied censoring for floods.

Partial Duration Series by Truncation

The specified q_0 sets the criterion for the magnitude of the highest low flow allowed. However, q_0 traditionally is based on judgement. A common decision is to set q_0 to the largest annual minimum in the record, so that each year will contribute at least one flow to the series. A probability distribution is empirically applied to the series data. Chow (Ch. 8, 1964) suggested the exponential distribution for maxima. Two types of variables can describe the truncated PDS. One using the flow values Q, the other the differences between the threshold flow, q_0 and the flow values, Q, called deviations. The deviation, X equals $q_0 - Q$, in which X and Q are random variables. The motivation for considering X is as follows: large low flows have a higher probability than small low flows, but the difference $X = q_0 - Q$ is small for the large low flows and reaches a maximum difference of q_0 (for zero flows). Thus large low flows have small differences and higher probability of occurrence which suits the exponential distribution. To allow for more flexibility the two parameter Weibull distribution is used. The Weibull distribution is given by

$$f_Y(y) = \frac{c}{b}\left(\frac{y}{b}\right)^{c-1} \exp\left[-\left(\frac{y}{b}\right)^{c}\right] \qquad 0 < y < \infty \tag{21}$$

where

b – scale parameter,
c – shape parameter

In the truncation process, low flow Q is defined to be equal to or less than q_0 and thus the range of low flows is $0 < Q < q_0$, but the range in Weibull distribution is $0 < y < \infty$. Therefore Eq. (21) must be modified so that positive probability of Q is obtained within the above range.

Partial Duration Series by Truncation for Deviations – The pdf for deviation X is

$$f(x) \equiv \frac{c}{b}\left(\frac{x}{b}\right)^{c-1} \frac{\exp[-(x/b)^{c}]}{F(q_0)} \qquad 0 < x < q_0 \tag{22}$$

and from Eq. (22)

$$F(q_0) \equiv 1-\exp\left[-\left(\frac{q_0}{b}\right)^{c}\right] \tag{23}$$

The parameters are calculated by maximum likelihood estimation (MLE). The likelihood function is given as

$$L_1 \equiv \prod_{i=1}^{n} \frac{c}{b}\left(\frac{x_i}{b}\right)^{c-1} \frac{\exp[-(x_i/b)^{c}]}{F(q_0)} \tag{24}$$

Then L_1 is transformed by its natural logarithm, L and differentiated with respect to the parameters b and c and setting them equal to zero to obtain the maximum likelihood estimators.

The cumulative probability of Q is given by

$$F(q_T) = \frac{\exp\{-[(q_0-q_T)/b]^{c}\}-\exp[-(q_0/b)^{c}]}{F(q_0)} \tag{25}$$

where $F(q_T)=P(Q\leq q_T \mid Q<q_0)$.

For partial duration series, the return period of a given event, q_T, is expressed in terms of number of events ξT where ξ = mean number of low flows below q_o per year. On the average the flow q_T will occur once in T years which is the same as once in ξT events and thus (see Cunnane 1973; 1979)

$$P(Q\le q_T | Q<q_0) \equiv \frac{1}{\xi T} \qquad (26)$$

A closed form solution for q_T can be derived as

$$q_T = q_0 - b\{-\ln\{(\frac{1}{\xi T}) + [1-(\frac{1}{\xi T})]\exp[-(\frac{q_0}{b})^c]\}\}^{1/c} \qquad (27)$$

Partial Duration Series by Truncation for Flows – The pdf for flows is

$$f(q) = \frac{c}{b}(\frac{q}{b})^{c-1} \frac{\exp[-(q/b)^c]}{F(q_0)} \quad \text{for} \quad 0 < q < q_0 \qquad (28)$$

The solution for q_T is given as

$$q_T = b\{-\ln\{1-(\frac{1}{\xi T})\{1-\exp[-(\frac{q_0}{b})^c]\}\}\}^{1/c} \qquad (29)$$

The solution of parameters b and c for deviations and flows is obtained from a computer program WEIBUL by applying MLE (Zutter *et al.* 1982).

Partial Duration Series by Censorship

Censorship type I uses the data generated by truncation plus the number of low flow occurrences above the threshold (right censoring). Censorship below the threshold (left censoring) is also used, for example when stream level goes below the lowest gage mark which is not addressed in this study. As before, the Weibull distribution is applied to low flow data, but the pdf is not altered because of censorship. The influence of censoring occurs in parameter estimation. MLE is used for the parameter estimation. The probability of a flow being above q_o is given as probability of exceeding q_o which is

$$P(Q\ge q_0) = 1-F(q_0) = \exp[-(\frac{q_0}{b})^c] \qquad (30)$$

The likelihood function is given by

$$L_1 = [1-F(q_0)]^k \prod_{i=1}^{n} f(q_i) \quad \text{for} \quad 0< q_i < q_0 \qquad (31)$$

$$= \{\exp[-(\frac{q_0}{b})^c]\}^k \prod_{i=1}^{n} \frac{c}{b}(\frac{q_i}{b})^{c-1} \exp[-(\frac{q_i}{b})^c] \qquad (32)$$

n – number of flows less than q_o,
k – number of low flows greater than q_o,
j – total number of low flows, and $n = j - k$.

The log likelihood function L is differentiated with respect to b and c and the resulting equations are solved for b and c by equating them to zero. As the result a closed form solution is obtained for b as a function of c (Cohen 1965). An iterative scheme is required for c. The numer of low flows per year is denoted by ξ_1. Thus for period T, $\xi_1 T$ low flows occur. The T year low flow q_T is given as

$$q_T = b\{-\ln[1-(\frac{1}{\xi_1 T})]\}^{1/c} \tag{33}$$

Physically Based Method

It is well known that during low flow periods, streams derive their waters from the aquifer. The base flow is essentially the dry weather flow and is considered as the recession part of the hydrograph, which is given by

$$Q(D) = Q(0)\exp(-\frac{D}{K}) \tag{34}$$

where

$Q(0)$ – initial recession flow,
$Q(D)$ – Dth day flow,
D – length of the recession in days,
K – recession parameter (see also Notation).

The solution of the linearized Boussinesq equation with suitable boundary conditions (Brutsaert and Nieber 1977) for groundwater flow can also be expressed in the form of Eq. (34). The recession equation is a basic procedure in the separation of base flow from the runoff hydrograph, and can be used to predict low flows during dry periods. Many times the recession parameter is used to characterize the hydrogeology of the basin.

Fig. 1 illustrates a time series of discharge values. Because of long time intervals between rainfall events, the discharge time series contains peaks followed by recessions. The starting base flow discharge due to the n^{th} rainfall event, $Q_n(0)$, is given by (see Fig. 1)

$$Q_n(0) = Q_n(D) + R(n) \tag{35}$$

where

$R(n)$ – the recharge due to the n^{th} storm and
$Q_n(D)$– the lowest flow for the (n-1) event.

Using Eq. (34), Eq. (35) can be written as

$$Q_n(0) = A(n)Q_{n-1}(0) + R(n) \tag{36}$$

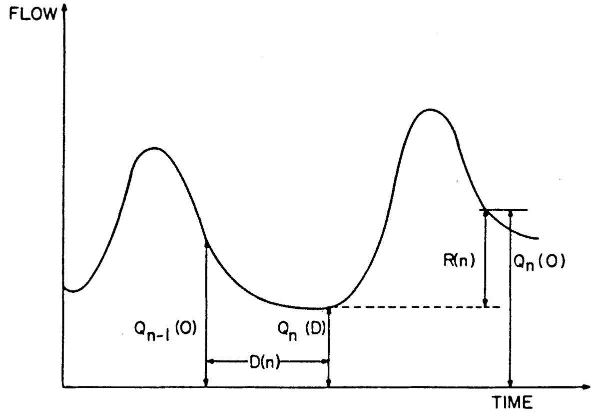

Fig. 1. Time series of flows.

which is a random coefficient Markov model and the recession factor, $A(n) = \exp[(-D(n)/K)]$, $0 < A(n) < 1$; and $D(n)$ = number of time units (days) to reach the lowest flow on the recession side of the (n-1) runoff event. Because the recharge amounts and recession periods are random variables, Eq. (36) represents a stochastic relationship. Also, since the rainfall events are physically independent, it can be assumed that $A(1)$, $A(2)$, ..., $A(n)$ and $R(1)$, $R(2)$, ..., $R(n)$ are independent identically distributed (i.i.d.) random variables. Thus, from Eq. (36), one can write

$$P[Q_n(0) \leq q \,|\, Q_{n-1}(0) \equiv c] \equiv P[Ac+R \leq q] \qquad (37)$$

which is the probability of $Q_n(0)$ being less than or equal to a specified discharge q given that, the previous initial base flow $Q_{n-1}(0)$ equals some constant c. Because of shape flexibility, and range (0,1), a beta distribution is fitted to the random variable A. Based on the available data Weibull distribution is fitted to R because of its shape flexibility and the closed form expression for its cumulative distribution function. For notational convenience hereafter,

$$Q_n(0) = Q(n) \quad \text{and} \quad Q_{n-1}(0) \equiv Q(n-1) \qquad (38)$$

The probability density function of A is given as

$$f_A(a) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} a^{\alpha-1}(1-a)^{\beta-1} \quad \text{for} \quad 0 < a < 1 \qquad (39)$$
$$\equiv 0 \qquad \text{elsewhere}$$

The probability density function of R is given as

$$f_R(r) \equiv \gamma\delta(r^{\delta-1})\exp(-\gamma r^{\delta}) \quad \text{for} \quad r > 0 \qquad (40)$$
$$\equiv 0 \qquad \text{elsewhere}$$

In order to compute the probability from Eq.(37)the distribution of S must first be derived, where

$$S \equiv Ac \qquad (41)$$

$$f_S(s) = \frac{1}{c} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} (\frac{s}{c})^{\alpha-1} (1-\frac{s}{c})^{\beta-1} \quad \text{for} \quad 0 < s < c \qquad (42)$$

$$= 0 \qquad \text{elsewhere}$$

Therefore Eq. (37) can be written as

$$P[Q(n) \leq q \,|\, Q(n-1)=c] = P[S+R \leq q] \qquad (43)$$

There are two ranges to be considered given by $q<c$ and $q>c$.

The cumulative probability distribution of R can be written as

$$F_R(r) = 1-\exp(-\gamma r^{\delta}) \qquad \text{for} \quad r > 0 \qquad (44)$$

$$\equiv 0 \qquad \text{otherwise}$$

and

$$P[S+R \leq q] = P[R \leq q-S] \qquad (45)$$

For $q \leqslant c$ Eq. (45) can be written as

$$P[S+R \leq q] \approx \int_0^q F_R(q-s) f_S(s)\, ds \qquad \text{for} \qquad 0 < q \leq c \qquad (46)$$

For $q > c$ Eq. (45) becomes

$$P[S+R \leq q] \approx \int_0^c F_R(q-s) f_s(s)\, ds \qquad \text{for} \qquad 0 < c < q \qquad (47)$$

Eqs. (46) and (47) can be used to generate the transition probabilities $P[Q(n) \leqslant q|Q(n\text{-}1) = c]$ in which the event $Q(n\text{-}1) = c$ is approximated as $c\text{-}\varepsilon < Q(n\text{-}1) < c+\varepsilon$ where ε is a positive number. By using the transition probabilities, the steady state probabilities can be obtained from

$$\lambda \equiv \lambda P \qquad (48)$$

where λ is the vector of steady state probabilities [i.e. $\lambda_i = P(Q(0) = t_i) = P(t_i - \varepsilon < Q(0) < t_i + \varepsilon), i = 1, \ldots, N$], P is the transition probability matrix of size $N \times N$ for N different states of $Q(0)$. Eqs. (34) and (48) can be used to find the frequencies of one day low flows. Eq. (34) can be written as

$$Z = Q(0)A \qquad (49)$$

where Z is $Q(D)$ and A is equal to exp $(\text{-}D/K)$. The return period T in years is defined as the time interval between occurrences of flow q_T. In T years μT recessions will occur in which q_T occurs once where μ is the mean number of recessions per year and therefore

$$\mu T \equiv \frac{1}{P[Z \leq z]} \tag{50}$$

in which

$$P[Z \leq z] = \int_q P[A \leq \frac{z}{q}] dF(q) \tag{51}$$

and therefore

$$P[Z \leq z] \equiv \sum_{i=1}^{N} g(t_i)\lambda_i \tag{52}$$

where

$$\lambda_i = P[Q(0) = t_i] \quad \text{and} \tag{53}$$

$$g(t_i) = \int_0^{z/t_i} f_A(a)\, da \tag{54}$$

It is noted that λ_i's are obtained from Eq. (48) and $g(t_i)$ are from direct integration of beta density. Therefore by using Eqs. (52) and (50) the T year low flow q_T (same as z) can be found by trial and error.

Application

Log Pearson Type III (LP III)

LPIII with a point mass (mixed LPIII) gives consistent results in predicting low flows for intermittent streams. These results (Appendix) are compared with the double bounded pdf and LP III with an added constant for zero flows. No regular pattern of under or over estimation is found. LPIII also gives good results for perennial streams. The Kolmogorov-Smirnov (K-S) test at a significance level of 0.01 is passed by all stations.

Double Bounded PDF (DBPDF)

DBPDF is applied to perennial and intermittent streams. For streams with a derived lower bound, U equaling zero, the results compare favorably with LPIII. However, U can also be negative (station 03.1700.00), causing negative flows to have a positive probability of occurrence which is unacceptable. Generally for return periods 2 and 5 years the results are consistent, but for return periods of 10 years and above (especially 50 years) the results are extremely underestimated. The results are shown in the Appendix. Intermittent streams give much better and consistent results than perennial streams. All stations pass the K-S test at a significance level of 0.01.

Partial Duration Series

In the application of Partial Duration Series, q_o is set equal to the largest annual minimum in the flow record. This makes the data compatible with other methods, that is all flows from the minimum annual series will be included in the partial duration series data. Unfortunately, in truncated PDS for flows no convergence was obtained in the parameter estimation process (Zutter 1982). In contrast, for the deviations, $X = q_o - Q$, the parameter estimation process converged for all stations. In truncated PDS for deviations the majority of the stations have the scale parameter c close to 1.0 which indicates the suitability of the exponential distribution. Generally truncated PDS predicts lower values for q_T. Censored PDS compares favorably with LPIII and the recession model. Minor differences may be attributed to the length of data since PDS uses more number of low flows. Two intermittent streams (1.6685.00, 1.6700.00) failed the K-S test.

Physically Based Method

The Hydrologic Information Storage and Retrieval System (HISARS) is a good source for streamflow data for Virginia and some of its neighboring states. For this research, all data came from HISARS. For each year the number of recession periods are determined by separating the base flow from the runoff for a given storm. Base flow separation is done by plotting the natural logarithm of flow [ln(Q)] versus time (see Fig. 2). The plot of the natural logarithm of flows produces a straight line giving a distinct indication of the recession. The recession line begins at $Q = Q(0)$ and ends at $Q = Q(D)$. The time difference between the occurrences of $Q(0)$ and $Q(D)$ is the recession length, D.

At the end of a particular year attention must be paid to the possible overlap of a recession from one year to the next. The method of moments is used to calculate the distribution parameters. Since the beta distribution does not have a closed form

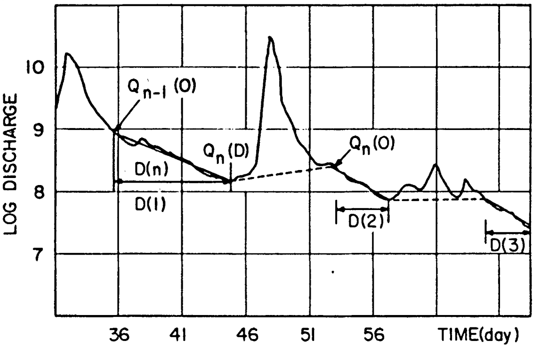

Fig. 2. Baseflow Separation.

for the cdf, it is numerically integrated using the computer programs from the International Mathematics Systems Library (IMSL). IMSL subroutine DCADRE was used to integrate Eqs. (46) and (47). The steady-state probabilities λ_i of $Q(0)$ in Eq. (48) were calculated using the IMSL subroutine EIGRF. The integration of the beta density in Eq. (54) was accomplished by the IMSL subroutine MDBETA.

A regression analysis was used for fitting the recessions. A subroutine calculates the optimum recession parameter using Lagrange multipliers, McCormick (1984) (also see Kelman 1980). For each recession the recession parameter K and the correlation coefficient between ln (Q) and time (for this study, days) are calculated. Any recession having a correlation coefficient less than 0.90 is ignored. Overall the computer application gave good results in determining the recessions and their respective recession lengths and recharge. For selected stations recessions done graphically were compared with the computer selection and almost all recessions were identical. The one day low flows for return periods of 2, 5, 10, 25, and 50 years are tabulated in the Appendix.

Summary

1) The traditional black box approach is to extract the minima at equal time intervals (typically a year) and fit the best curve for the probabilities of occurrences. A refinement of this, is to place a point mass at zero, (for dry streams) and an upper bound on flows so that all flows below this level are classified as low flows. The double bounded pdf fits these requirements aptly. The mixed log Pearson III is also considered because of its better curve fitting abilities.
2) The condition for minima from equal time intervals can be relaxed by imposing a threshold level and to consider flows below this level (partial duration series, PDS). There are two approaches possible with PDS: i) truncation and ii) censoring. In truncation the entire population is assumed to be generated from the truncated range. In censoring some observations are curtailed and the number of observations violating the censoring limit is known. Both the approaches are considered in the present study.
3) Low flows are governed by the recession part of the hydrograph and therefore a probability analysis of hydrograph recessions should lead to the frequency estimates of low flows. The recession model proposed in this paper demonstrates this hypothesis and performs well as compared to the other methods.

   The recession model injects physics among a plethora of black box methods. There are also some difficulties: i) closed form solutions are not available ii) lack of hourly data complicates the estimation procedure iii) the constant K value may be relaxed. The double bounded pdf performs well for intermittent streams and small perennial streams. However it provides low estimates for large perennial streams. This may be attributed to the lack of thick, long tail of the distribution. The log Pearson III results are consistent in that they agree well with the other methods. The truncation procedure generally provides low estimates for perennial streams and compares favorably with the recession model for intermittent streams. The censoring procedure yields results which are quite close to the log Pearson III and the recession model.

Acknowledgements

The work reported in this paper is based on the research supported in part by the Virginia Water Resources Research Center. The authors thank Prof. T. Ramalingam, Northern Illinois University for his valuable comments on this study. The skillful typing of Ms. Ruby Shaw is appreciated.

Notation

Log Pearson type III:

$F(x)$ – cdf of standard normal distribution
G – skew coefficient
K_T – frequency factor
M_y – mean of logarithm of annual low flow
Q – low flow random variable
S_y – standard deviation of logarithm of annual low flows
k – number of zero low flows
j – total numer of annual low flows
q – low flow cut off value
x_T – standard normal deviate at $F(x_T) = 1/T$
$\alpha, \beta \gamma$ – log Pearson distribution parameters

Double bounded probability density function (DBPDF):

F_o – probability of zero flows
L_1 – likelihood function
L – log likelihood function
U – lower bound of DBPDF
V – upper bound of DBPDF
a, b – parameters of DBPDF
f_1 – probability density function over positive range
j – total number of annual low flows
k – number of zero flows
q – low flow variate
x – $(q - V)/(V - U)$

Partial Duration Series (PDS):

$F(q_o)$ $\equiv 1 - \exp[-(q_o/b)^c]$
X – deviation, $(q_o - Q)$
Y – Weibull random variable
b, c – Weibull parameters
j – total number of low flows
k – number of flows above the threshold value q_o
m – given number of flows to be less than q_o
n – number of low flows below q_o
q_o – threshold value
ξ – mean number of flows below q_o per year
ξ_1 – mean nember of low flows per year

Physically based Method:

A $\equiv \exp(-D/K)$
D – recession length, days
K – recession parameter, days
P – transition probability matrix
N – total number of states
$Q(0)$ – initial recession flow, cfs
$Q(D)$ – Dth day recession flow, cfs
$Q_n(0)$ – initial recession flow for the n^{th} rainfall event, cfs
R – recharge, cfs
S – Ac
T – return period, years
Z – $Q(D)$
c – given value for $Q_{n-1}(0)$
$g(t)$ – cumulative distribution function (cdf) of A
n – integer counting variable
q_T – T year low flow, cfs
z – cut of value for cdf of Z
$\Gamma(\ )$ – gamma function
α, β – parameters of beta distribution of A
γ, δ – parameters of Weibull distribution of R
ε – range in which $Q(n\text{-}1)$ equals c whenever $c\text{-}\varepsilon < Q(\text{n-1}) < c + \varepsilon$
λ – vector of steady state probabilities with components λ_i
λ_i $= P[Q(0) = t_i]$
μ – number recessions per year

Appendix

Table 1 – 2 Year Return Period

	Recession		Mixed		PDS[3]	
Station	Model	LPIII[1]	LPIII[N]	DBPDF[2]	Truncated	Censored
		Perennial Streams				
3.1675.00	178.56	131.05		144.27	102.66	117.95
3.1680.00	1060.23	870.56		900.17	632.88	862.25
3.1700.00	98.23	97.92		111.35	65.46	88.41
3.2085.00	5.93	5.65		6.56	3.44	4.15
		Intermittent Streams				
1.6560.00	0.83	0.47	0.54	0.91	0.80	0.26
1.6685.00	–	1.42	1.59	1.98	1.33	1.86
1.6700.00	–	0.39	0.41	0.51	0.18	0.28
2.0205.00	2.89	5.79	4.57	5.19	3.02	3.65

1 LPIII ≡ Log Pearson Type III; for intermittent streams a positive value is added
2 DBPDF ≡ Double Bounded Probability Density Function
3 PDS ≡ Partial Duration Series
– ≡ no convergence for IMSL subroutine for numerical integration
N ≡ not applicable for perennial streams

Table 2 – 5 Year Return Period

	Recession		Mixed		PDS	
Station	Model	LPIII	LPIII[N]	DBPDF	Truncated	Censored
		Perennial Streams				
3.1675.00	141.54	102.94		92.91	79.71	98.67
3.1680.000	750.21	654.13		615.12	438.68	654.51
3.1700.00	73.96	72.38		63.85	41.12	70.98
3.2085.00	2.92	1.95		1.69	1.45	2.26
		Intermittent Streams				
1.6560.00	0.36	0.06	0.09	0.00	0.37	0.11
1.6685.00	–	0.18	0.16	0.13	0.54	1.05
1.6700.00	–	0.08	0.11	0.06	0.07	0.16
2.0205.00	1.20	1.86	1.86	1.61	1.28	2.19

Table 3 – 10 Year Return Period

Station	Recession Model	LPIII	Mixed LPIII$_N$	DBPDF	PDS	
					Truncated	Censored
Perennial Streams						
3.1675.00	121.45	91.23		69.51	63.38	86.26
3.1680.00	594.37	588.85		470.32	309.47	583.67
3.1700.00	60.80	62.35		42.20	26.59	60.17
3.2085.00	1.76	1.00		0.65	0.74	1.43
Intermittent Streams						
1.6560.00	0.19	0.01	0	0	0.20	0.05
1.6685.00	–	0.04	0	0	0.27	0.68
1.6700.00	–	0.02	0	0	0.06	0.11
2.0205.00	0.62	0.57	0.90	0.53	0.66	1.5

Table 4 – 25 Year Return Period

Station	Recession Model	LPIII	Mixed LPIII$_N$	DBPDF	PDS	
					Truncated	Censored
Perennial Streams						
3.1675.00	100.40	80.52		46.34	48.18	72.25
3.1680.00	548.72	529.36		332.59	173.80	501.85
3.1700.00	48.04	53.53		23.23	13.31	48.38
3.2085.00	0.91	0.44		0.19	0.30	0.78
Intermittent Streams						
1.6560.00	0.08	0	0	0	0.08	0.02
1.6685.00	–	0.01	0	0	0.11	0.39
1.6700.00	–	0.01	0	0	0.01	0.07
2.0205.00	0.27	0.09	0	0	0.27	0.90

Table 5 – 50 Year Return Period

Station	Recession Model	LPIII	Mixed LPIII$_N$	DBPDF	PDS	
					Truncated	Censored
Perennial Streams						
3.1675.00	87.12	74.44		34.61	30.58	63.19
3.1680.000	467.65	495.66		256.59	102.77	447.73
3.1700.00	40.69	48.67		13.54	7.34	41.04
3.2085.00	0.56	0.25		0.08	0.15	0.50
Intermittent Streams						
1.6560.00	0.04	0	0	0	0.04	0.01
1.6685.00	–	0	0	0	0.05	0.25
1.6700.00	–	0	0	0	0.01	0.05
2.0205.00	0.14	0.02	0	0	0.13	0.62

References

Aitchison, J. (1955) On the Distribution of a Positive Random Variable Having a Discrete Probability Mass at the Origin, *Journal of the American Statistical Association, Vol 50,* pp. 901-908.

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.

Bethlahmy, Nedavia (1977) Flood Analysis by SMEMAX Transformation, *Journal of the Hydraulics Division, ASCE, Vol. 103,* No. HY1, pp. 60-78.

Boughton, W. (1980) A Frequency Distribution for Annual Floods, *Water Resources Research, Vol. 16,* No. 2, 1 pp. 347-354.

Boughton, W., and Shirley, E. (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 Transformations, *Journal of the Royal Statistical Society, Vol. B 26,* pp. 211-252.

Brutsaert, W., and Nieber, J. (1977) Regionalized Drought Flow Hydrographs from a Mature Glaciated Plateau, *Water Resources Research, Vol. 13,* No. 3, pp. 637-643, 1977.

Chow, Ven Te (1964) *Handbook of Applied Hydrology,* New York: McGraw-Hill.

Cohen, A.C. (1965) Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples, *Technometrics, Vol. 7,* Nor. 4, pp. 579-588.

Condie, R. (1977) The Log Pearson Type 3 Distribution: The T-Year Event and Its Asymptotic Standard Error by Maximum Likelihood Theory, *Water Resources Research, Vol. 13,* No. 6, 987-991.

Cunnane, C. (1973) A Particular Comparison of Annual Maxima and Partial Duration Series Methods of Flood Frequency Prediction, *Journal of Hydrology, Vol. 18,* pp. 257-271.

Cunnane, C. (1978) Unbiased Plotting Position – A Review. *Journal of Hydrology, Vol. 37,* pp. 205-222.

Cunnane, C. (1979) A Note on the Poisson Assumption in Partial Duration Series, *Water Resources Research, Vol. 15,* No. 2, pp. 489-493.

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,* New York: John Wiley.

Kelman, Jerson (1980) A Stochastic Model for Daily Streamflow, *Journal of Hydrology, Vol. 47,* pp. 235-249.

Kite, G.W. (1977) Frequency and Risk Analysis in Hydrology, Fort Colins: Water Resources Publications.

Kottegoda, N.T. (1980) *Stochastic Water Resources Technology,* London: Macmillan Press Ltd.

Kruskal, W.H., and Tanur, J.M. (1979) *International Encyclopedia of Statistics,* New York: The Free Press.

Kumar, A., and Devi, R. (1982) Statistical Determination of Design Low Flows – A Comment, *Journal of Hydrology, Vol 58,* pp. 175-177.

Kumaraswamy, P. (1980) A Generalized Density Function for Double-Bounded Randoms Processes, *Journal of Hydrology, Vol. 46,* pp. 79-88.

Leese, M. (1971) Use of Censored Data in the Estimation of Gumbel Distribution Parameters for Annual Flood Series, *Water Resources Research, Vol. 9,* No. 6, pp. 1534-1542.

Matalas, N. (1963) Probability Distribution of Low Flows U.S.G.S. Professional Paper 434-A, Washington, D.C.

McCormick, T (1984) Frequency Analysis of Low Flows, M.S. Thesis, Virginia Polytechnic Institute and State University.

McGinnis, D.G., and Sammons, W.H. (1970) Discussion of Paper: Daily Stream Flow Simulation, *Journal of the Hydraulics Division, ASCE, Vol. 96,* No. HY5, pp. 1201-1206.

McMahon, T.A., and Arenas, A.D. (1982) *Methods of Computation of Low Streamflow,* Paris: UNESCO.

Prakash, A. (1981) Statistical Determination of Design Low Flows, *Journal of Hydrology, Vol. 51,* pp. 109-118.

Rasheed, Hameed, Marur Ramamoorthy, and Al-Dabbagh, A.S. (1982) Modified SMEMAX Transformation for Frequency Analysis, *Water Resources Bulletin, Vol. 18,* No. 3, pp. 509-511.

U.S. Water Resources Council (1977) Guidelines for Determining Flood Flow Frequency Bulletin 17A, Washington, D.C.: GPO.

Westphal, J. (1984) SMEMAX? Caution! *Journal of Hydraulic Engineering, ASCE, Vol. 110,* No. 3, pp. 209-218.

Zelen, N., and Severo, N.C. (1972) Probability Functions, *Handbook of Mathematical Functions,* edited by M. Abramowitz and I.A. Stegun, New York: Dover Publications, Inc.

Zutter, B. *et al.* (1982) WEIBUL A Program to Estimate Parameters of the Weibull Distribution Using Complete, Censored, and Truncated Data. Pub. No. FWS-3-82, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University.

Received 3 July, 1986

Address:
Department of Civil Engineering,
Room 200 Patton Hall,
Virginia Tech,
Blacksburg, VA-24061, U.S.A.