Journal of Hydrology (N.Z.) Vol. 28, No. 1, 1989

ESTIMATING STATISTICAL PARAMETERS FOR AN N-SEGMENT DISCRETE HYDROLOGIC DATA SERIES

Wei Du

Graduate School of Geography
&
Center For Technology, Environment and Development,
Clark University, Worcester, Massachusets 01610

ABSTRACT

Extreme or rare floods, with probabilities in the range of a 10^{-2} to 10^{-7} chance of occurrence per year, are of interest to the hydrologic and engineering communities for project design and planning. However, streamflow records of greater than 100 years are meagre, and most records are considerably shorter. Paleoflood determinations can provide more information than just the magnitudes of ancient floods. Under favorable conditions, the paleohydrologic records may furnish information on all floods greater than some threshold. However, paleoflood records, or other nonsystematic records, usually add segments with large gaps to the hydrological data series. In the United States and other countries, the Pearson Type III and the Log-Pearson Type III probability distributions are recommended by governmental water resources agencies for flood frequency analysis. The accuracy of such analysis directly depends on the accuracy of estimating parameters of those distributions. However, extreme floods may significantly affect the values of parameters of those probability distributions (mean $\bar{Q}$ and Cv). This paper puts forward a new algorithm to estimate statistical parameters for an N-segment discrete hydrological data series and related conditional flood frequency analysis.

UNCERTAINTIES IN FLOOD FREQUENCY ANALYSIS

A major source of uncertainty in water resource engineering projects during planning, design, construction, operation, and management stages is the lack of knowledge of the stochastic nature of hydrologic events. Therefore, statistical tools are widely used in flood frequency analysis. The major sources of uncertainty in flood frequency analysis result from: (1) data error; (2) accuracy and imprecision of measurement and observation; (3) sampling uncertainty including the choice of samples and appropriate sample size; (4) selection of an appropriate probability distribution to describe these stochastic hydrologic events; (5) estimation of the hydrologic and statistical parameters in those models; (6) low probability flood extrapolation, for instance, problems in interpreting tails of hydrologic curves; (7) modelling assumptions; (8) the characterization of river basin parameters; (9) hydrologic forecasts, and (10) effects of human activities (Ratick and Du, 1988).

Flood frequency analysis is necessary for water resources engineering design and flood mitigation planning. However, there is no well-developed theoretical basis for flood frequency analysis. No single probability distribution has been proven theoretically to be uniformly superior to others. Standardized probability distributions can be regarded as "soft mathematical rulers" which can be shaped by their parameters, and the accuracy of flood frequency analysis depends on the accuracy of the estimation of statistical parameters.

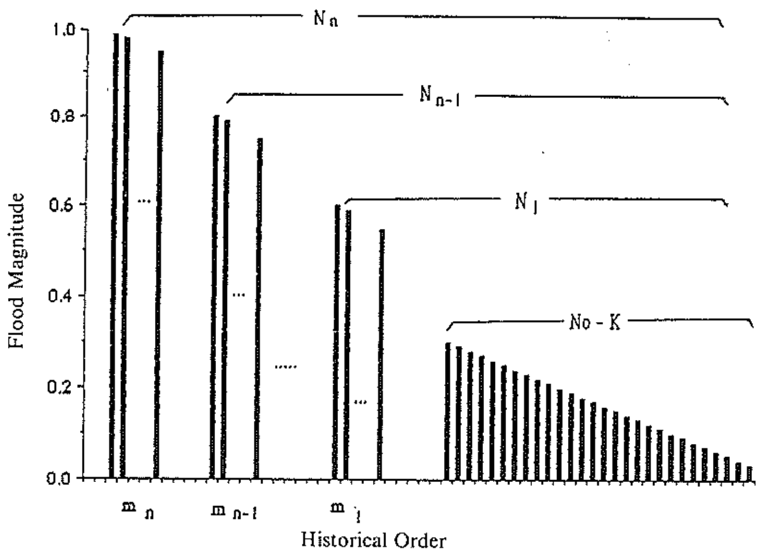
During the last decade, managing risk and uncertainty has become a major field in water resources research (Ashkar et al., 1980, Baecher, et al., 1980, Bao, et al., 1987, Byer, 1979, Cover, et al., 1986, Dandy, 1985, Hosking and Wallis, 1986, Karlesson, et al., 1988, Kitanidis, 1986, Leach et al., 1987, Lee, et al., 1986, Loaiciga et al., 1986, Reid, 1987, Stedinger, 1983, Tung, et al., 1981). Uncertainties in hydrologic parameters and hydrologic models have been studied primarily to improve flood frequency estimates (Bodo and Unny, 1976, Tung and Mays, 1980, Vicens et al., 1975, Wood and Rodriguez, 1975). Stedinger (1983) derived the sampling distributions of design discharge to construct confidence intervals for flood frequency analysis. Using various sources of flood data, different techniques have been developed, including streamflow sequence studies (Matalas and Jacobs, 1964, Crawford and Linsley, 1966), regional flood estimations (Greis and Wood, 1981; Wallis, 1983; Kucera, 1982), and the use of historic data (e.g., Benson, 1950; Leese, 1973). Some techniques have been incorporated in official procedures for flood frequency analysis (Natural Environment Research Council (NERC), 1975; U.S. Water Resources Council, 1982). An integrated sensitivity analysis for uncertainties in hydrologic, hydraulic, hydro-economic models, and other computerized resources evaluation systems has been developed by the Center for Technology, Environment, and Development at Clark University, Massachusetts (Ratick and Du, 1988).

TREATMENT OF EXTREME FLOOD VALUES AND BASIC MATHEMATICAL ASSUMPTIONS

Traditional methods for estimating extreme flood values for flood frequency analysis involve combining collected hydrological data, and then determining the historical position or historical orders of floods in terms of flood magnitude. Most rivers have less than one hundred years of continuously observed data available for flood frequency analysis. As was stated in the U.S. National Research Council report (NRC, 1988).

> "Floods of concern usually include those with annual probabilities in the broad range of from 10^{-2} to 'near 0'. However, in the United States streamflow records of greater than 100 years are meagre, and most records are considerably shorter. Consequently, any method for estimating probabilities of floods rarer than about the 100-year flood must include some form of extrapolation, a process that is likely to introduce errors and, at worst, strain credulity".

Paleoflood determinations can provide more information than just the magnitudes of ancient floods. Under favourable conditions, the paleohydrologic or nonsystematic record may indicate all occurrences of floods greater than some threshold. Auxiliary information is needed for statistical interpretation of the paleoflood or nonsystematic records, for instance, the discrete hydrological data set. Nonsystematic records may provide information about events of unusual magnitude, but often do not provide unequivocal evidence of the nonoccurrence of such events. Such information may be very useful for defining the potential for future occurrence of extreme events. Nonsystematic floods usually add discrete segments to hydrological data series to become a discrete form or pattern. For example, a nonsystematic record may be described as follows; "the following noteworthy floods on .... River were reported in the .... City newspaper: 1896, 30,000 cfs; 1907, 37,000 cfs; 1924, 45,000 cfs; 1939, 46,000 cfs. The continuous record began in 1943." No information is given as to the basis for observing or recording the peaks from before the continuous record period (NRC, 1988). In general, the basic structure of hydrological data series (Fig. 1) can be expressed as an N-segment discrete hydrological data series which we define.



N_i — number of years in the ith time period from the right to the left, $i = 0, 1, 2, \ldots, n$; and,

m_i — number of floods in the ith discrete segment from the right to the left, $i = 1, 2, \ldots, n$

k — number of the floods which were taken to a extreme flood segment.

FIG. 1—Structure of an N-Segment discrete Hydrologic Data Series.

The Pearson Type III and the Log-Pearson Pearson Type III probability distributions are currently recommended for flood frequency analysis by governmental water resources agencies in many countries. For example, the U.S. Water Resources Council (1977) recommended the Log-Pearson Type III distribution for analysing flood frequencies; and the generalized extreme value distribution (GEV) was selected in the United Kingdom. Selecting a certain standard probability curve to approach natural hydrological phenomena means that the collected hydrologic data has been assumed to belong to the same statistical population and conform with the same probability distribution. The methods in this paper are to be used to estimate statistical parameters for the standard probability distributions which deal with mathematical treatments of extreme flood estimation for an N-segment discrete hydrological data set. Therefore it is assumed that:

(a) all hydrological data belong to the same statistical population;
(b) the statistical parameters ($\varnothing_i$) belong to same parametric family, that is:
   $\varnothing_i \subset \boldsymbol{\emptyset}$ (i = 1, 2, 3, ... N).
(c) Severe unusual K floods in the continuous hydrological data set should be taken out, and put into the extreme flood segment.

The following mathematical derivation is based on these three assumptions.

THE MATHEMATICAL DERIVATION OF THE SAMPLE MEAN OF AN N-SEGMENT DISCRETE HYDROLOGICAL DATA SERIES

For the continuous section (N_0) of hydrological data set, if k extreme values were removed from the continuous data set, and put in the extreme value segment according to their magnitudes, the statistical estimator for the mean of the continuous section of hydrological data can be written as:

$$\bar{Q}_0 = \frac{1}{N_0 - K} \sum_{i=K+1}^{N_0} Q_i \tag{1}$$

For a one-segment (N_1) discrete hydrological data set, accoring to the definition of the statistical mean, the mean estimator for the N_1 period can be written as:

$$\bar{Q}_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} Q_1 = \frac{1}{N_1} \left(\sum_{j=1}^{m_1} Q_j + \sum_{i=m_1+1}^{N_1} Q_i\right) \tag{2}$$

According to the basic model assumptions (a) and (b), let $\bar{Q}_0 = \bar{Q}_{N1-m1}$, so

$$\bar{Q}_0 = \frac{1}{N_0 - K} \sum_{i=K+1}^{N_0} Q_i = \frac{1}{N_1 - m_1} \sum_{i=m_1+1}^{N_1} Q_i \tag{3}$$

Therefore

$$\sum_{i=m_1+1}^{N_1} Q_1 = \frac{N1 - m1}{N_0 - K} \sum_{i=k-1}^{N_o} Q_i \tag{4}$$

By substituting (4) into (2), the mean for segment N_1 can be written as:

$$\bar{Q}_1 = \frac{1}{N_1}\left[\sum_{j=1}^{m_1} Q_j + \frac{N1 - m1}{N_o - K}\sum_{i=K+1}^{N_o} Q_i\right] \tag{5}$$

For a two-segment (N_2) discrete hydrological dat set, the statistical estimator for the mean ($\bar{Q}_2$) can be expressed as

$$\bar{Q}_2 = \frac{1}{N_2}\sum_{l=1}^{N_2} Q_2 = \frac{1}{N_2}\left(\sum_{l=1}^{m_2} Q_l + \sum_{l=m_2+1}^{N_2} Q_l\right) \tag{6}$$

According to the basic model assumptions (a) and (b), $\bar{Q}_1 = \bar{Q}_{N2-m2}$, so

$$\bar{Q}_i = \frac{1}{N_1}\left(\sum_{j=1}^{m_1} Q_j + \frac{N1 - m1}{N_o - K}\sum_{i=K+1}^{N_o} Q_i\right) = \frac{1}{N_2 - m_2}\sum_{l=m_2+1}^{N_2} Q_l \tag{7}$$

We can obtain from (7):

$$\sum_{l=m_2+1}^{N_o} Q_l = \frac{N2 - m1}{N_1}\left(\sum_{j=1}^{m_1} Q_j + \frac{N1 - m1}{N_o - K}\sum_{i=K+1}^{N_o} Q_i\right) \tag{8}$$

Substituting (8) into (6), we get the mean of a hydrological data set with two segments of discrete hydrological data:

$$\bar{Q}_2 = \frac{1}{N_2}\left[\sum_{l=1}^{m_2} Q_l + \frac{N2 - m2}{N_1}\left(\sum_{j=1}^{m_1} Q_j + \frac{N1 - m1}{N_o - K}\sum_{i=1}^{N_o} Q_i\right)\right] \tag{9}$$

Using the same principle, the mean of a hydrological data set with three segments of discrete hydrological data can be derived as:

$$\bar{Q}_3 = \frac{1}{N_3}\left\{\sum_{m=1}^{m_3} Q_m + \frac{N3 - m3}{N_2}\left[\sum_{l=1}^{m_2} Q_l + \frac{N2 - m2}{N_1}\left(\sum_{j=1}^{m_1} Q_j + \frac{N1 - m1}{N_o - K}\sum_{j=K+1}^{N_1} Q_i\right)\right]\right\} \tag{10}$$

In general, for an N-segment descrete hydrological data set, a recurrence formula can be written as

$$\bar{Q}_n = \frac{1}{N_n}\left[\sum_{z=1}^{m_n} Q_z + (N_n - m_n)\bar{Q}_n - 1\right] \tag{11}$$

THE MATHEMATICAL DERIVATION OF COEFFICIENT OF VARIATION OF AN N-SEGMENT DISCRETE HYDROLOGICAL DATA SET

Using the above algorithm, the mean of an N-segments discrete hydrological data set ($\bar{Q}_n$, n = 1, 2, . . ., N) can be estimated. Therefore, we can apply $\bar{Q}_n$ to the calculation of the coefficient of variation of an N-segment discrete hydrological data. For a continuous section N_o of a hydrological data set, the coefficient of variation can be described as

$$Cv_o = \left[\frac{1}{N_o - K - 1} \sum_{i=K+1}^{N_1} \left(\frac{Qi}{\bar{Q}_n} - 1\right)^2\right]^{1/2} \tag{12}$$

For the two-segments (N_1) discrete hydrological data set (one continuous segment and one discrete segment), the coefficient of variation can be written as:

$$Cv_1 = \left[\frac{1}{N_1 - 1}\left(\sum_{j=1}^{N_1} \frac{Qj}{\bar{Q}_n} - 1\right)^2\right]^{1/2}$$

$$= \left\{\frac{1}{N_1 - 1}\left[\sum_{j=1}^{m_1}\left(\frac{Qj}{\bar{Q}_n} - 1\right)^2 + \sum_{j=m+1}^{N_1}\left(\frac{Qj}{\bar{Q}_n} - 1\right)^2\right]\right\}^{1/2} \tag{13}$$

According to the basic assumptions (a) and (b) of the model, let $Cv_o = Cv_{N1 - m1}$:

$$Cv_o = \left[\frac{1}{N_o - K - 1} \sum_{i=k+1}^{N_0} \left(\frac{Qi}{\bar{Q}_n} - 1\right)^2\right]^{1/2} = \left[\frac{1}{N_1 + m_1 + 1} \sum_{j=m_1+1}^{N_1} \left(\frac{Qi}{\bar{Q}_n} - 1\right)^2\right]^{1/2} \tag{14}$$

From (14), we can obtain:

$$\sum_{j=m_1+1}^{N_1} \left(\frac{Qj}{\bar{Q}_n} - 1\right)^2 = \frac{N1 - m1 - 1}{N_o - K - 1} \sum_{i=K+1}^{N_o} \left(\frac{Qi}{\bar{Q}_n} - 1\right)^2 \tag{15}$$

Substituting (15) into (13), the coefficient of variation of one discrete segment and one continuous segment hydrological data set (N_1) can expressed as:

$$Cv_2 = \left\{\frac{1}{N_2 - 1}\left[\sum_{j=1}^{m_2}\left(\frac{Qj}{\bar{Q}_n} - 1\right)^2 + \frac{N1 - m1 - 1}{N_o - K - 1} \sum_{i=K+1}^{N_o}\left(\frac{Qi}{\bar{Q}_n} - 1\right)^2\right]\right\}^{1/2} \tag{16}$$

Using the same principle, the coefficient of variation of two discrete segments and one continuous segment of a hydrological data set can be written as

$$Cv_2 = \{ \frac{1}{N_2 - 1} \{\sum_{I=1}^{m_2} (\frac{QI}{\bar{Q}_n} - 1)^2 + \frac{N2 - m2 - 1}{N_1 - 1} [\sum_{j=1}^{m_1} (\frac{Qj}{\bar{Q}_n} - 1)^2$$

$$+ \frac{N1 - m1 - 1}{N_o - K - 1} \sum_{i=K+1}^{N_o} (\frac{Qi}{\bar{Q}_n} - 1)^2]\}\}^{1/2} \qquad (17)$$

The coefficient of variation of three discrete segments and one continuous segment hydrological data set (Cv_3) can be described as:

$$Cv_3 = \{ \frac{1}{N_3 - 1} [\sum_{m=1}^{m_3} (\frac{Qm}{\bar{Q}_n} - 1)^2 + \frac{N3 - m3 - 1}{N_2 - 1} \{\sum_{I=1}^{m_2} (\frac{QI}{\bar{Q}_n} - 1)^2 + \frac{N2 - m2 - 1}{N_1 - 1}$$

$$[\sum_{j=1}^{m_1} (\frac{Qj}{\bar{Q}_n} - 1)^2 + \frac{N1 - m1 - 1}{N_o - K - 1} \sum_{i=K+1}^{N_o} (\frac{Qi}{\bar{Q}_n} - 1)^2]]\}\}^{1/2} \qquad (18)$$

In summary, for an N-segment discrete hydrological data set, a recurrence formula can be written as:

$$Cv_n = \{ \frac{1}{N_n - 1} [\sum_{z=1}^{m_n} (\frac{Qz}{\bar{Q}_n} - 1)^2 + (N_n - m_n - 1)\, Cv_{n-1}^2]\}^{1/2} \qquad (19)$$

THEORETICAL ASSESSMENT OF THE ALGORITHM

By derivation of algorithms of statistical estimators of the mean and the coefficient of variation of an N-segment discrete hydrological data set, we obtain two general recurrence formulae:

(i) $$\bar{Q}_n = \frac{1}{N_n} [\sum_{z=1}^{m_n} Q_z + (N_n - m_n)\, \bar{Q}_n - 1]$$

and

(ii) $$Cv_n = \{ \frac{1}{N_n - 1} [\sum_{z=1}^{m_n} (\frac{Qz}{\bar{Q}_n} - 1)^2 + (N_n - m_n - 1)\, Cv_{n-1}^{\ 2}]\}^{1/2}$$

From the equation (i) which estimates the mean of an N-segment discrete hydrological data series we can see that:

$$\lim_{N_n \to \infty} \bar{Q}_n = \lim_{N_n \to \infty} \left(\frac{1}{N_n} \sum_{z=1}^{m_n} Q_z + \frac{N_n - m_n}{N_n} \bar{Q}_{n-1} \right)$$

$$= \lim_{N_n \to \infty} \frac{1}{N_n} \sum_{z=1}^{m_n} Q_z + \lim_{N_n \to \infty} \left(\frac{Nn - m_n}{N_n} \right) (\bar{Q}_{n-1}) = \bar{Q}_{n-1} \tag{20}$$

Therefore,

$$\lim_{N_n \to \infty} | \bar{Q}_n - \bar{Q}_{n-1} | = | \bar{Q}_n - \bar{Q}_{n-1} | = 0 \tag{21}$$

Equation (21) proves that the algorithm is mathematically convergent for a period of length N_n. This conclusion is consistent with the model's assumption (a) that the statistical parameter $\bar{Q}_n$ belongs to the same statistical population. Using 'N + 1' to replace 'n' in the formula (20) derivation, yields:

$$\lim_{N_{n+1} \to \infty} | \bar{Q}_{n+1} - \bar{Q}_n | = | \bar{Q}_{n+1} - \bar{Q}_n | = 0 \tag{22}$$

Therefore, when N_n is very large (->∞), $\bar{Q}_n \to \bar{Q}$, $\bar{Q}_{n+1} \to \bar{Q}$, . . ., $\bar{Q}_m \to \bar{Q}$, that is, $\bar{Q}$ is the mathematical limit. This mathematical property satisfies natural hydrological phenomena, which have natural limitations. For example, the flood discharge (Q) in every river has its quantitative limit, the maximum flood peak or PMF. On the other hand, from equation (11), we can derive the same conclusion for the algorithm of the coefficient of variation for an N-segment discrete hydrological data set.

$$\lim_{N_n \to \infty} Cv_n^2 = \lim_{N_n \to \infty} \left[\frac{1}{N_n - 1} \sum_{z=1}^{m_n} \left(\frac{Qz}{\bar{Q}_n} - 1 \right)_2 \right] + \lim_{N_n \to \infty} \left[\left(\frac{N_n - m_n - 1}{N_n - 1} \right) Cv_{n-1}^2 \right]$$

$$= Cv_{n-1}^2 \tag{23}$$

Therefore

$$\lim_{N_n \to \infty} | Cv_n^2 - Cv_{n-1}^2 | = | Cv_n^2 - Cv_{n-1}^2 | = 0 \tag{24}$$

Equation (24) shows that the algorithm of estimating the coefficient of variation of an N-segments discrete hydrological data set is mathematically convergent. By the same principle, using 'n + 1' to replace 'n' in the formula (23) derivation, the following relationship results:

$$\lim_{N_n \to \infty} | Cv_{n+1}^2 - Cv_n^2 | = | Cv_{n+1}^2 - Cv_n^2 | = 0 \tag{25}$$

Therefore, when

$$Nn \to \infty, Cv_{n+1} \to Cv, Cv_{n+2} \to Cv, \ldots, Cv_M \to Cv.$$

Finally, we will prove the new estimators are unbiased. First, taking statistical expectation for both sides of equation (5), we can obtain:

$$E(\hat{\bar{Q}}_1) = E\{ \frac{1}{N_1} [\sum_{j=1}^{m_1} Qj + \frac{N1 - m1}{N_o - K} \sum_{i=K+1}^{N_o} Q_i]\}$$

$$= \frac{1}{N_1} [\sum_{j=1}^{m_1} E(Qj) + \frac{N1 - m1}{N_o - K} \sum_{i=K+1}^{N^o} EQ_i]\}$$

$$= \frac{1}{N_1} [m_1 E(Qj) + \frac{N1 - m1}{N_o - K} (No - K)E(Q_i)] \tag{26}$$

According to the model's assumption (a) and (b), E(Qj) = E(Qi), so that equation (26) can be written as:

$$E(\hat{\bar{Q}}_1) = \frac{1}{N_1} [m_1 E(Qj) + (N1 - m1) E(Qj)] = \bar{Q}_1 \tag{27}$$

Using equation (21), the above relationship can apply to any number of discrete hydrologic data series. The same principle can apply to estimator C_{vn}, and its unbiased character also can be proven.

THE ESTIMATION OF CONDITIONAL FLOOD FREQUENCY AND THE APPLICATION OF THE ALGORITHM FOR ESTIMATING OF STATISTICAL PARAMETERS

Because the structure of hydrological data series was treated as an N-segment discrete data series, the traditional method of calculating flood exceedence probabilities:

$$P_i = m/(N + 1), \qquad M = 1, 2, 3, \ldots, N \tag{28}$$

can not be used. For example, consider a three-segment discrete hydrological data series. When we consider the exceedence probability of a flood in the second discrete segment, we should regard the flood exceedence probability as a conditional probability on the extreme, but less severe than those in the third segment. According to statistical theory, if there is a group of events B_1, B_2, ..., which satisfy the following conditions:

(i) $B_i \cap B_j = \varnothing$ (i = j);
(ii) $P(B_i) > 0$ (i = 1, 2, ...)
(iii) $P(B_1 \cup B_2 \cup .....) = 1$ (29)

for any event A, then

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + ..., \quad (30)$$

We apply this to the above problem. Assume that an event which severe floods in the second discrete segment have occurred in F_2, and severe floods do not occur as F_2'. Obviously, F_2 and F_2' can consist of an entire probability space ($\varnothing$), that is $F_2 \cap F_2' = \varnothing$, and $F_2 \cup F_2' = 1$, and $P(F_2) > 0$ and $P(F_2') > 0$. Let F_1 be a flood in the first discrete segment (See Figure 1, both number of segments and magnitude are increased from right to left). The sample condition satisfies equation (29). Therefore, the probability of F_1 can be described as:

$$P(F_1) = P(F_2)\, P(F_1/F_2) + P(F_2')\, P(F_1/F_2') \quad (31)$$

In fact, $P(F_1/F_2) = 1$, because, in segment 2, the magnitudes of all floods are greater than in segment 1, that is $F_2 > F_1$, therefore F_2 dominates Flood F_1, so the probability of occurrence of the flood in segment 1 under the condition of occurrence of the flood in the second segment is equal to 1. Therefore, the equation (31) can be simplified as:

$$P(F_1) = P(F_2) + P(F_2')\, P(F_1/F_2') \quad (32)$$

This equation provides the basic principle for calculating exceedence probability for an N-segment hydrologic data set:

Exceedence probability of a lower magnitude flood in certain position of the discrete hydrologic data series
= (the exceedence probability of the higher magnitude flood in the higher segment)
+ (the probability that the high magnitude flood does not occur) × (exceedence probability of the lower magnitude flood that the higher magnitude flood has been taken out of the hydrologic data set)

According to this principle, the conditional flood frequency can be given by the following equations:

$$P_n = \frac{m}{N_n - 1}, \quad m = 1, 2, ..., m_n \quad (33)$$

$$P_{n-1} = P_n \,|\, m_n + \frac{Nn - mn}{N_n} \times \frac{m}{N_{n-1} + 1}, \quad m = 1, 2, ..., m_{n-1} \quad (34)$$

$$P_1 = P_2 | m_2 + \frac{N2 - M2}{N_2} \times \frac{m}{N_1 + 1}, \quad m = 1, 2, \ldots, m_1 \tag{35}$$

$$P_0 = P_1 | m_1 + \frac{N1 - M1}{N_1} \times \frac{m}{N_0 - K + 1}, \quad m = 1, 2, \ldots, m_0 \tag{36}$$

We use the above equations to estimate exceedence probability for a hydrological data set in a river basin in the Sichuan Province of China. The results of the analysis are shown in Table 1.

The Pearson type III distribution is used traditionally in flood frequency analysis. We apply the above algorithm to estimate statistical parameters for the Pearson type III probability distribution in this sample calculation. In this case, the hydrological data form a two-segment discrete structure. According to paleohydrologic study and hydrological analysis of historical data in the river basin, within the last 400 years, no flood with a magnitude higher than 432,000,000 cubic meters per day occurred. Therefore, N_2 = 400 years was determined. According to historical data, for the 74 year period from 1903 to 1977, a flood in 1903 was the largest one in the river basin, so N_1 = 74 years. In the river basin, there were 24 years of continuous hydrologic gauged record, so N_0 = 24 years. The extreme flood in 1956 was much more severe than others in the continuous data segment, so the flood was taken out and was put in segment 1. Therefore, K was assigned as 1, that is K = 1. According to equation (9) and equation (17), the mean of Q (annual maximum daily volume of flood) is 81.9 million cubic meters, and C_v = 0.632. Using a regional skew coefficient and applying the above specified parameters to the Pearson type III distribution, the results gave a good fit to the hydrological data series.

CONCLUSION

This new algorithm for estimating statistical parameters for flood frequency analysis differs from traditional methods in two ways: (1) the method concerns an N-segement discrete hydrological data series, and its mathematical treatment of extreme floods more closely approximates real hydrological situations than traditional methods; and (2) the method deals with conditional frequencies and considers the real position of historical floods. Conditional probability analysis for extreme events and their risk is new in risk analysis and hazard management (Karlsson et al., 1988), because extreme events may be governed by some internal or external conditions (some causal chains may exist), for instance, dam failure, nuclear power plant accidents and other system accidents. One problem with this method is that, when the number of discrete segments is large, the mathematical formulae become very complicated, but they can be solved on computers by using recurrence formulae.

ACKNOWLEDGEMENT

Special thanks are due to Prof. Samuel J. Ratick in Center for Technology, Environment, and Development of Clark University, who commented on a large portion of the manuscript.

TABLE 1—Conditional Flood Frequency Estimation For Annual Maximum Daily Volume of Flood.

Year	flood order	number of discrete section	flood volume (10^8 m^3)	conditional flood frequency
1842	1	2	4.32	0.25
1903	1	1	2.62	1.60
1921	2	1	2.32	2.90
1938	3	1	2.00	4.20
1956	4	1	1.93	5.60
taken	(1)	—	—	—
1967	2	0	1.32	9.5
1975	3	0	1.22	13.5
1977	4	0	1.22	17.4
1961	5	0	1.08	21.4
1963	6	0	1.08	25.3
1973	7	0	1.07	29.2
1966	8	0	1.06	33.1
1974	9	0	0.88	37.1
1962	10	0	0.87	41.1
1958	11	0	0.86	45.0
1968	12	0	0.84	49.0
1955	13	0	0.72	52.8
1959	14	0	0.67	56.8
1957	15	0	0.64	60.8
1960	16	0	0.63	64.6
1954	17	0	0.62	68.7
1972	18	0	0.42	72.6
1964	19	0	0.37	76.5
1971	20	0	0.31	80.5
1969	21	0	0.26	84.4
1976	22	0	0.25	88.4
1970	23	0	0.21	92.4
1965	24	0	0.20	96.3

REFERENCES

Ashkar, F.; Bobee, B., 1988: Confidence intervals for flood events under a Pearson Type III or Log Pearson Type III distribution. *Water Resources Research, 24(3):* 639-650.

Baecher, B.; Pater, M. E.; Neufville, R. D. E., 1980: Risk of Dam failure in benefit-cost analysis. *Water Resources Research, 16(3).*

Bao, Yixing; Yeou-Koung Tung; Hasfurher, V. R., 1987: Evaluation of uncertainty in flood magnitude estimator on annual expected damage costs of hydraulic structures. *Water Resources Research, 23(11):* 2023-2029.

Bensom, M. A., 1950: Use of historical data in flood frequency analysis. *Eos Trans. AGR., 31:* 419-424.

Bodo, B.; Unny, T. E., 1976: Model uncertainty in flood frequency analysis and frequency-based design. *Water Resources Research, 12(6):* 1109-1117.

Byer, P., 1979: Screening objectives and uncertainties in water resources planning. *Water Resources Research, 15(4):* 768.

Costa, J. E., 1987: Holocene stratigraphy in flood frequency analysis. *Water Resources Research, 14:* 626-632.

Cover, K. A.; Unny, T. E., 1986: Application of computer intensive statistics to parameter uncertainty in streamflow synthesis. *Water Resources Bulletin, 22(3):* 495-507.

Crawford, N. H.; Linsley, R. K., 1986: Digital simulation in hydrology: Stanford watershed model iv. *Tech. Rep. 39, Dep. of Civ. Eng.*, Stanford University., Stanford Calif.

Dandy, G. C., 1985: An approximate method for the analysis of uncertainty in benefit-cost ratios. *Water Resources Research, 21(3):* 267-371.

Greis, N. P.; Wood, E. F., 1981: Regional flood frequency estimation and network design. *Water Resources Research, 17:* 1167-1177.

Hosking, J. R. M.; Wallis, J. R., 1986: Paleoflood hydrology and flood frequency analysis. *Water Resources Research, 22(4):* 543-550.

Karlsson, P. O.; Haimes, Y. Y., 1988: Risk-based analysis of extreme events. *Water Resources Research, 24(1):* 9-20.

Kitanidis, P. K., 1986: Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resources Research 22(4):* 499-507.

Kucera, G., 1982: Robust flood frequency models. *Water Resources Research, 18:* 315-324.

Leach, M. R.; Haimes, Y. Y., 1987: Multiobjective risk-impact analysis method. *Risk Analysis, vol. 7(2):* 225-241.

Lee, J. L.; Mays, L. W., 1983: Improved risk and reliability model for hydraulic structures. *Water Resources Research, 19(6):* 1415-1422.

Leese, M. N., 1973: Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series. *Water Resources Research, 9:* 1534-1542.

Loaiciga, H. A.; Marino, M. A., 1986: Risk analysis for reservoir operation. *Water Resources Research, 22(4):* 483-488.

Matalas, N. C.; Jacobs, B., 1964: A correlation procedure for augmenting hydrologic date. *U.S. Geol. Survey Professional Paper:* 434-E.

National Research Council, 1988: *Estimating probabilities of extreme floods — methods and recommended research*, National Academy Press, Washington D.C.

Natural Environment Research Council, 1975: *Flood Studies Report*, 1, London.

Ratick, S. J.; Wei Du, 1988: Guidelines for Evaluating Risk and Uncertainty in Local Flood Damage Protection Projects, Draft Technique *Report* to the U.S. Army Corps of Engineers.

Reid, S. G., 1987: Frequency-cost curves and derivative risk profile. *Risk Analysis, 7(2):* 261-267.

Stedinger, J. R., 1983: Confidence intervals for design events. *J. Hydraul. Eng., 109(1):* 13-27.

Stedinger, J. R., 1983: Design events with specified flood risk. *Water Resources Research 19(2):* 511–522.
Tung, Y. K.; Mays, L. W., 1980: Optimal risk-based design of water resource engineering projects. *Technique Report:. CRWR-171,* Center for Research in Water Resources, Univ. of Tex., Austin.
U.S. Water Resources Council, 1982: Guidelines for determining flood flow frequency, *Bulletin #17B,* Appendix 9, Hydrology Committee, Washington D.C.
U.S. Water Resources Council, 1982: Guidelines for determining flood flow frequency, *Bulletin #17B,* Hydrology Committee, Washington D.C.
Wallis, J. R., 1982: *Hydrologic problems associated with oilshale development. In:* Environmental Systems and Management, ed. by Srinaldi. pp. 85–102, North-Holland, Amsterdam.
Wicens, G. J.; Rodriguez-Iturbe, I., 1975: Bayesian framework for the use of regional information in hydrology. *Water Resources Research, 11(3):* 405–414.
Wood, E. F.; Rodriguez-Iturbe, I.: Bayesian inference and decision making for extreme hydrologic events. *Water Resources Research, 11(4):* 533–542.