Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Hydrology 278 (2003) 172–196
Journal of Hydrology
www.elsevier.com/locate/jhydrol

Data-based comparisons of moments estimators using historical and paleoflood data

John F. England Jr.[a,*], Robert D. Jarrett[b], José D. Salas[c]

[a]*US Bureau of Reclamation, P.O. Box 25007, D-8530, DFC, Denver, CO 80225, USA*
[b]*US Geological Survey, P.O. Box 25046, MS 412, DFC, Denver, CO 80225, USA*
[c]*Department of Civil Engineering, Colorado State University, Fort Collins, CO 80523, USA*

Received 28 November 2002; accepted 28 March 2003

Abstract

This paper presents the first systematic comparison, using historical and paleoflood data, of moments-based flood frequency methods. Peak flow estimates were compiled from streamflow-gaging stations with historical and/or paleoflood data at 36 sites located in the United States, Argentina, United Kingdom and China, covering a diverse range of hydrologic conditions. The Expected Moments Algorithm (EMA) and the Bulletin 17B historical weighting procedure (B17H) were compared in terms of goodness of fit using 25 of the data sets. Results from this comparison indicate that EMA is a viable alternative to current B17H procedures from an operational perspective, and performed equal to or better than B17H for the data analyzed. We demonstrate satisfactory EMA performance for the remaining 11 sites with multiple thresholds and binomial censoring, which B17H cannot accommodate. It is shown that the EMA estimator readily incorporates these types of information and the LP-III distribution provided an adequate fit to the data in most cases. The results shown here are consistent with Monte Carlo simulation studies, and demonstrate that EMA is preferred overall to B17H. The Bulletin 17B document could be revised to include an option for EMA as an alternative to the existing historical weighting approach. These results are of practical relevance to hydrologists and water resources managers for applications in floodplain management, design of hydraulic structures, and risk analysis for dams.

Keywords: Flood frequency analysis; Estimation; Historical information; Paleoflood data; Expected Moments Algorithm

1. Introduction

Historical and paleoflood data have been used to supplement peak flow estimates from existing stream gage records, and to extend those records in time. House et al. (2002) present recent developments and applications in paleoflood hydrology. There has been increased interest in obtaining and using historical and paleoflood data in flood frequency analysis (Jarrett and Tomlinson, 2000; O'Connell et al., 2002). Recently, a new parameter estimation procedure, called the Expected Moments Algorithm (EMA) (Cohn et al., 1997), was proposed as an improved alternative to the Bulletin 17B historical weighting procedure (B17H) (IACWD, 1982), and recommended for use by NRC (1999). These two

* Corresponding author. Tel.: +1-303-445-2541; fax +1-303-445-6351.
E-mail address: jengland@do.usbr.gov (J.F. England Jr.).

0022-1694/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0022-1694(03)00141-0

estimators are compared in this paper using at-site peak discharge, historical and paleoflood data.

This paper presents: (1) a data base of peak flows from streamflow-gaging station sites with historical and/or paleoflood data; (2) a data summary of peak flow characteristics; (3) a comparison of B17H and EMA at-site estimators in fitting selected data base samples; and (4) a demonstration of EMA performance by fitting sites with binomial and multiple censored data. The focus of the paper is on practical problems and issues that a practitioner may face. A data set consisting of historical, paleoflood, and systematic streamflow records was assembled to compare the two estimators. The focus was to use sites with available historical and paleoflood information adjacent to streamflow-gaging stations. Flood frequency analyses were conducted for each of the 36 data sets. Three calculations were made: (1) empirical frequency estimates of the observed floods; (2) a frequency curve utilizing the historical weighting procedure (B17H) from Bulletin 17B; and (3) a frequency curve based on the expected moments (EMA) method. Flood frequency estimation procedures and methods for comparison are discussed in Section 2. A summary of the systematic, historical and paleoflood data base used in the analysis is presented in Section 3. Limitations of the historical and paleoflood data base are discussed. Results and discussion are presented in Section 4. The results obtained as part of this paper are contrasted with those from previous Monte Carlo simulation experiments (Cohn et al., 1997; England, 1998).

The addition of historical and paleoflood data to frequency analysis is an essential element to obtaining realistic estimates of extreme flood quantiles (i.e. greater than a 500-year flood), rather than relying exclusively on model extrapolations based solely on gage data. Continuing efforts in paleoflood data collection (e.g. Baker, 1987; Jarrett, 1991; Enzel et al., 1993; Jarrett and Tomlinson, 2000; Levish et al., 2000) aid hydrologists and engineers to better understand the magnitude, occurrence, and distribution of extreme floods. These data can also help test flood frequency analysis assumptions, such as homogeneity and stationarity. In addition, one can investigate adequacy of distributions for fitting extreme quantiles, and the possible use of tail modeling procedures.

Several issues surrounding the quality of historical and paleoflood data in frequency analysis are discussed by others. Baker et al. (2002) briefly discuss two major issues: inaccuracies in flood age estimates and flood discharge reconstruction inaccuracies. Inaccuracies in age estimates have been reduced by using detailed soil sampling and radiocarbon techniques (e.g. Levish, 2002). Improved hydraulic techniques (e.g. Webb and Jarrett, 2002; Denlinger et al., 2002) and recent flood data interpretation (Jarrett and England, 2002; Yanosky and Jarrett, 2002) have helped to reduce these uncertainties. For the data presented in this paper, strict quality checking for erroneous individual historical/paleoflood discharge estimates was not conducted outside cursory reviews and discussions with other investigators for particular data sets. It was assumed that the records are correct and complete for all floods that exceed a threshold. Peak discharge estimates and historical and paleoflood age estimates and information were not independently verified, aside from cursory reviews. Historical flood discharge estimates have been shown to be in error and sometimes unreliable (e.g. Fuller, 1914, p. 569; Jarrett, 1987, 1994). Evidence of large floods over long time periods, however, such as historical peaks and/or paleofloods is invaluable and cannot be ignored as it can significantly extend the record length for frequency analysis. In certain cases, historical information consisted of a flood noted as the highest since some time in the past. No attempt was made to obtain unpublished historical data for individual sites, or to compare estimates with other sites in a region. The primary source for the systematic, annual peak discharge data was records from gaging stations published by the US Geological Survey, UNESCO (1976), and Rodier and Roche (1984). Historical and paleoflood information was obtained from journal articles, research reports, other technical reports, and communication with individuals performing paleoflood and historical flood research.

2. Estimation and comparison methods

The appropriate way to include historical and paleoflood data in flood frequency analysis is to consider that the data arise from a censored sample.

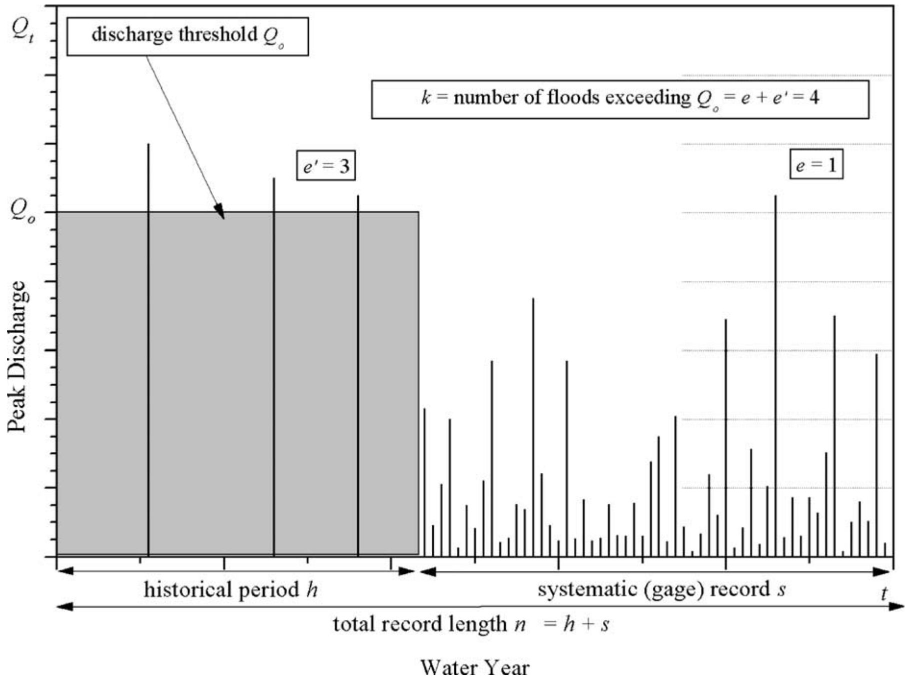

Fig. 1. Example peak discharge time series with historical period and discharge threshold Q_0. The shaded area represents floods of unknown magnitude less than Q_0.

The goal is to include positive evidence of large floods, or limits on flood magnitude, over longer time periods to extend peak discharge records from gaging stations. This concept is illustrated in Fig. 1. Following notation used by Hirsch and Stedinger (1987) and Guo and Cunnane (1991), we define a systematic (gage) record length (s), a historical/paleoflood record length (h), where $n = s + h$, and peak discharge threshold Q_0 that represents a censoring level. One knows the number of floods (e') (possibly their magnitude as well) in h, and the number of floods (e) in s that exceed Q_0. The number of floods that exceed the threshold is k, where $k = e + e'$, and the total number of floods is g, where $g = s + k - e$. Estimates of s, h, Q_0, e and e' were made for each data set based on the available information. Historical and paleoflood data can also be represented by binomial censoring, interval censoring and multiple censoring cases. Binomial-censored data (Stedinger and Cohn, 1986) are defined as the exact magnitude of a flood is unknown except that it exceeded a lower threshold. Interval censoring is utilized when the exact magnitude of a flood is unknown, but known to be within some upper and lower amount (Stedinger et al. 1988a; Cohn et al., 1997). Multiple censoring refers to cases where more than one peak discharge threshold (Q_0) is used to represent the historical and paleoflood data (e.g. Levish et al., 1994).

There is currently a lack of consensus in the hydrology community regarding what is the appropriate censored data model, Type I or Type II, for incorporating historical/paleoflood data in frequency analysis. In Type I censoring, the threshold Q_0 is fixed and the number of floods exceeding the threshold is a random variable. For Type II censoring, the number of floods exceeding Q_0 is fixed and Q_0 is a random variable. Previous investigators (e.g. Stedinger and Cohn, 1986; Hosking and Wallis, 1986a,b; Frances et al., 1994) made different assumptions about the type of censored data for simulation experiments, without presenting data sets to document their assumptions. Stedinger and Cohn (1986) assumed Type I censoring and found historical information to be valuable in virtually all cases considered. Hosking and Wallis (1986a,b) assumed Type II censored data and that only the largest flood was observed; their conclusions were nearly opposite to Stedinger and Cohn (1986). Stedinger and Baker (1987) explained that Hosking and Wallis' results were primarily due to the fact that they only used the systematic record to estimate the at-site scale parameter. The greatest impact of historical/paleoflood information should be in its ability to improve estimates of the at-site scale parameter (Stedinger and Baker, 1987). Guo and Cunnane (1991) indicated that Hosking and Wallis' results were due primarily to their Type II censoring assumption, and suggested that a Type II model may not be the best choice to represent the data. Frances et al. (1994) showed that historical/paleoflood data are valuable regardless if one assumes Type I or Type II censoring.

The log-Pearson Type III distribution (LP-III) was selected as the base flood frequency distribution to compare B17H and EMA. Logarithms $(X_1, \ldots, X_g)$ of the peak discharges $(Q_1, \ldots, Q_g)$ were fit to a Pearson Type III (P-III) distribution (IACWD, 1982). The P-III density function is defined as:

$$f(x|\tau, \alpha, \beta) = \begin{cases} \dfrac{\left(\dfrac{x-\tau}{\beta}\right)^{(\alpha-1)} \exp\left(-\dfrac{x-\tau}{\beta}\right)}{|\beta|\Gamma(\alpha)} & \left(\dfrac{x-\tau}{\beta}\right) \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where x is the peak discharge logarithm (random variable), τ, α, β are the location, shape, and scale parameters of the P-III distribution, and $\Gamma(\alpha)$ is the complete gamma function. The P-III distribution parameters (τ, α, β), expressed in terms of the first three population moments, are:

$$\tau = \mu - \alpha\beta \tag{2}$$

$$\alpha = \frac{4}{\gamma^2} \tag{3}$$

$$\beta = \mathrm{sign}(\gamma)\left(\frac{\sigma^2}{\alpha}\right)^{1/2} \tag{4}$$

where (μ, σ^2, γ) are the mean, variance, and coefficient of skew, respectively. The observed flood sample is used to estimate $(\hat{\mu}, \hat{\sigma}^2, \hat{\gamma})$ where the circumflexes (^) indicate that the quantities are estimates.

2.1. Bulletin 17B historical (B17H) weighting method

The Bulletin 17B historical weighting procedure is presented in IACWD (1982). The B17H sample mean $(\hat{\mu})$, sample variance $(\hat{\sigma}^2)$ and coefficient of skew $(\hat{\gamma})$ estimates, neglecting low outliers, are:

$$\hat{\mu} = \frac{W\sum_{i=1}^{s-e} X_i + \sum_{j=s-e+1}^{g} X_j}{n} \tag{5}$$

$$\hat{\sigma}^2 = \frac{W\sum_{i=1}^{s-e} (X_i - \hat{\mu})^2 + \sum_{j=s-e+1}^{g} (X_j - \hat{\mu})^2}{n-1} \tag{6}$$

$$\hat{\gamma} = \frac{W\sum_{i=1}^{s-e} (X_i - \hat{\mu})^3 + \sum_{j=s-e+1}^{g} (X_j - \hat{\mu})^3}{(n-1)(n-2)\hat{\sigma}^3} \tag{7}$$

where the weighting factor W is defined as:

$$W = \frac{n-k}{s-e} \tag{8}$$

The weighting factor W is used to represent the unknown, below-threshold values and that they follow the same distribution as the below-threshold systematic observations (IACWD, 1982). Essentially the B17H adjustment 'fills in' the historical/paleoflood period with an appropriate number of replications of below-threshold (Q_0) portion of the gage record (Kirby, 1981; Thomas, 1985). Two shortcomings of this adjustment are: (1) the assumption that the gage record is representative of the entire historical period less the historical data; and (2) very little weight is given to the historical data (Lane, 1987). This second assumption is inappropriate for long historical/paleoflood periods in relation to the systematic record. For example, based on the Elkhead Creek data (site no. 18), $s = 41$, $h = 4959$, $e = 0$ and $e' = 1$; the gage record is weighted 122 times $(W = 121.9)$ to fill in the 4958 unobserved observations. Deficiencies of this estimator are discussed by Stedinger and Cohn (1986), Lane (1987), and England (1998). One needs to have at least one flood with historical information, and the flood estimate must be explicitly known for B17H to use the data. B17H cannot readily use binomial or interval censored floods, or data with multiple thresholds (England, 1998).

2.2. Expected Moments Algorithm method

The EMA (Cohn et al., 1997, 2001) is a new moments-based parameter estimation procedure that was designed to use the different types of gage, historical, and paleoflood data in flood frequency analysis. EMA was designed to handle historical and paleoflood data that the Bulletin 17B historical weighting procedure was never designed to use, such as binomial and interval censored data, and multiple thresholds. EMA is described in detail by Cohn et al. (1997, 2001) and England (1998).

The EMA sample mean, variance, and coefficient of skewness, including historical/paleoflood information about unknown flood magnitudes X_h less than X_0, are:

$$\hat{\mu}_m = \frac{\sum_{i=1}^{s-e} X_i + \sum_{j=s-e+1}^{g} X_j + (h-e')E[X_h]}{n} \tag{9}$$

$$\hat{\sigma}_m^2 = \frac{\left[\sum_{i=1}^{s-e}(X_i-\hat{\mu}_m)^2 + \sum_{j=s-e+1}^{g}(X_j-\hat{\mu}_m)^2\right] + (h-e')E[(X_h-\hat{\mu}_m)]^2}{n} \tag{10}$$

$$\hat{\gamma}_m = \frac{\left[\sum_{i=1}^{s-e}(X_i-\hat{\mu}_m)^3 + \sum_{j=s-e+1}^{g}(X_j-\hat{\mu}_m)^3\right] + (h-e')E[(X_h-\hat{\mu}_m)^3]}{n\hat{\sigma}_m^3} \tag{11}$$

The expression $E[\ ]$ is the expected value of a flood of unknown magnitude (X_h) during the historical period (h) given that it is below the threshold (X_0). The term is then weighted by the number of observations ($h-e'$) below X_0. The equation for the first expectation of a P-III distribution is (Cohn et al., 1997):

$$E[X_h|X_h \le x_0; \hat{\tau}, \hat{\alpha}, \hat{\beta}] = \hat{\tau} + \hat{\beta}\frac{\Gamma\left(\frac{x_0-\hat{\tau}}{\hat{\beta}}, \hat{\alpha}+1\right)}{\Gamma\left(\frac{x_0-\hat{\tau}}{\hat{\beta}}, \hat{\alpha}\right)} \tag{12}$$

The expectation for higher order moments is:

$$E[(X_h-\hat{\mu})^p|X_h \le x_0; \hat{\tau}, \hat{\alpha}, \hat{\beta}] = \sum_{j=0}^{p}\binom{p}{j}\hat{\beta}^j(\hat{\tau}-\hat{\mu})^{p-j}\left[\frac{\Gamma\left(\frac{x_0-\hat{\tau}}{\hat{\beta}}, \hat{\alpha}+j\right)}{\Gamma\left(\frac{x_0-\hat{\tau}}{\hat{\beta}}, \hat{\alpha}\right)}\right] \tag{13}$$

For binomial/interval censored data and multiple thresholds, Eq. (13) is generalized for a range $[x_0 \le X_h \le x_u]$ (Cohn et al., 1997):

$$E[(X_h-\hat{\mu})^p|x_0 \le X_h \le x_u; \hat{\tau}, \hat{\alpha}, \hat{\beta}] = \sum_{j=0}^{p}\binom{p}{j}\hat{\beta}^j(\hat{\tau}-\hat{\mu})^{p-j} \times \left[\frac{\Gamma\left(\frac{x_u-\hat{\tau}}{\hat{\beta}}, \hat{\alpha}+j\right) - \Gamma\left(\frac{x_0-\hat{\tau}}{\hat{\beta}}, \hat{\alpha}+j\right)}{\Gamma\left(\frac{x_u-\hat{\tau}}{\hat{\beta}}, \hat{\alpha}\right) - \Gamma\left(\frac{x_0-\hat{\tau}}{\hat{\beta}}, \hat{\alpha}\right)}\right] \tag{14}$$

2.3. Data comparison metrics

Three calculations were made for the single threshold comparison sites: (1) empirical frequency (plotting position) estimates of the observed floods; (2) a frequency curve for $(s+h)$ using the B17H historical weighting estimation procedure; and (3) a frequency curve for $(s+h)$ based on the EMA parameter estimation method. In all cases, the log-Pearson Type III distribution was assumed to provide an adequate fit to the data. This distribution assumption can influence the results. For the binomial and multiple censoring sites, the same calculations were made except that the B17H estimation was omitted.

Four data assumptions were made as part of this study: (1) peak discharge estimates are explicitly known (no measurement errors are incorporated), unless data are binomial or interval-censored; (2) the historical period h is known perfectly so there are no errors in estimating the historical/ paleoflood record length; (3) the flood records are approximately stationary and homogeneous; and (4) all data samples are from an LP-III parent distribution. While these four assumptions were made in this study for simplicity, all may possibly be violated in practice. Several investigators have questioned these operational assumptions and suggest that historical and paleoflood data can be caused by different mechanisms (Webb and Betancourt, 1992) and are non-stationary for these long time periods (Webb and Baker, 1987; Ely, 1992, 1997). O'Connell et al. (2002) present a Bayesian framework for incorporating data and model uncertainties in flood frequency analysis, but not climate or land use changes. NRC (1999) recommended using a dynamic flood frequency approach to handle non-stationary data. Currently, flood frequency methods that incorporate historical/paleoflood data and include trend or shifts in climate are under development.

Each estimation method was used to compute a flood discharge $\hat{Q}_i(T)$ for each observed flood (i) in each flood series, where the return period $T = (1/p(i))$. Comparisons were made between computed flood discharges $\hat{Q}_i(T)$ and data (assumed true) values $Q_i(T)$ for the g observed floods in each data set. This simple technique has frequently been used as a basis for comparing two estimation methods with empirical data sets (e.g. Benson, 1968; Bobée and Robitaille, 1977; IACWD, 1982; Jain and Singh, 1987).

Two metrics were used to compute the relative goodness of fit between computed discharges $\hat{Q}_i(T)$ and data (true) values $Q_i(T)$: a mean absolute relative deviation (ARD) and a mean squared relative deviation (MSD):

$$\mathrm{ARD} = \frac{1}{g}\sum_{i=1}^{g} |q_i(T)| \tag{15}$$

$$\mathrm{MSD} = \frac{1}{g}\sum_{i=1}^{g} [q_i(T)]^2 \tag{16}$$

and

$$q_i(T) = \frac{\hat{Q}_i(T) - Q_i(T)}{Q_i(T)} \tag{17}$$

The statistics ARD and MSD are objective indices of the goodness of fit of each method to sample data, throughout the recurrence intervals of interest for flood analysis (Bobée and Robitaille, 1977; Jain and Singh, 1987). Probability plots of the data and the fitted distributions were also used in a qualitative sense to assess the relative goodness of fit of each estimation procedure to the data.

Exceedance probability estimates $p(i)$ for the data values $Q(i)$ were computed via one of two plotting

$$p(i) = \begin{cases} \dfrac{i-\alpha}{k+1-2\alpha}(p_e) & i = 1, \ldots, k \\ p_e + (1-p_e)\dfrac{i-k-\alpha}{n+1-2\alpha} & i = k+1, \ldots, g \end{cases} \qquad \hat{p}_e = \frac{k}{n} \tag{18}$$

$$p(i) = \begin{cases} \dfrac{i-\alpha}{k+1-2\alpha}(p_e) & i = 1, \ldots, k \\ p_k + (1-p_k)\dfrac{i-k-\alpha}{n+1-2\alpha} & i = k+1, \ldots, g \end{cases} \qquad p_k = \frac{(k-\alpha)}{(k+1-2\alpha)}(p_e) \tag{19}$$

positions: (1) a Type I exceedance-based plotting position Eq. (18) for the Type I and binomial-censored samples (Hirsch and Stedinger, 1987); and (2) a Type II plotting position Eq. (19) for the Type II samples (Salas et al., 2002). Eqs. (18) and (19) give identical plotting position estimates $p(i)$ for the k largest floods. The expanded Type I formula (Hirsch and Stedinger, 1987, Appendix C) was used for the sites with multiple thresholds. To assess the effects of the plotting position distribution coefficient alpha (α) on the ARD and MSD results, calculations were made for three values that span the possible coefficient range ($\alpha = 0.00$, 0.4, and 0.5), for the Weibull, Cunnane, and Hazen formulas (Cunnane, 1978).

Because one is interested in the performance of the estimation procedure to fit the extreme events on record as well, three additional comparisons were made: (1) $q_i(T)$ for the largest flood was used as a basis for comparison; (2) an approximate non-exceedance probability test (Stedinger et al., 1988a) was used to assess the fit of the distribution to the largest observed flood (Q_{max}); and (3) ARD and MSD were estimated for only the k largest floods, as opposed to all g observations. Note that the first and third comparisons are identical when $k = 1$. The non-exceedance probability test is presented in Stedinger et al. (1988a) and was used by Stedinger et al. (1988b) to assess whether the fitted distribution is consistent with the observed magnitude of the largest flood for the period of record (n) (see also Conover, 1999, p. 145). Stedinger et al. (1988a) define the test statistic p^* as:

$$p^* = \mathrm{Prob}[X_i < X_{\max},\ i = 1, \dots, n] = F[X_{\max}|\hat{\alpha}, \hat{\beta}, \hat{\tau}]^n \tag{20}$$

In this case, the parameters are for a P-III distribution and $X = \log_e(Q)$. The null hypothesis for this test is that the fitted LP-III distribution quantile is consistent with the largest flood observation ($Q_{\max}$) in n years. The alternative hypothesis is the fitted distribution is inconsistent with the largest flood. The estimated value of p^* from the fitted distribution should fall in the interval (0.05, 0.95) with an approximate 90% probability.

3. Data utilized for comparison

A data base consisting of 36 at-site gaging station peak discharge estimate records where historical and/or paleoflood data were available was assembled. The stations represent different hydrologic conditions and climatic regions in the United States, Argentina, the United Kingdom, and China. The sites are located in the following states or countries, with number of sites at each location in parentheses: Alabama (1), Arizona (5), California (4), Colorado (10), Georgia (1), Iowa (1), Maryland (1), Pennsylvania (1), Tennessee (1), Texas (1), Utah (2), Virginia (2), Washington (1), West Virginia (1), Argentina (1) China (1), and United Kingdom (2). The majority of the sites are located in the United States. The gaging stations are located in diverse hydrologic regimes that range from humid to semi-arid, with drainage areas ranging from 65 to 1,950,000 km^2. Records include floods caused by rainfall from major synoptic-scale disturbances, tropical storms, and high intensity, local convective storms. Each station is numbered and listed by location in Table 1; systematic and historic record lengths and the sources of information used for each site are also included. Time series plots of the data at each site are shown in England (1998).

One main criterion used to select sites was the length of the historical or paleoflood period h. The record lengths for the 36 data sets span from 100 to 10,000 years (Fig. 2); note that a variable class width is used to represent the distribution. There is roughly equal representation of sites with record lengths less than and greater than 500 years. Many additional sites exist in the United States where historical flood data in the 100 to 200-year range are available (e.g. Thomson et al., 1964). Although 18 out of 36 sites that were obtained have record lengths greater than 500 years, we were unable to rapidly obtain published data at other locations to supplement these sites.

One limiting factor in this study was the number of sites with paleoflood information. Systematic efforts, methods and protocols on a national (US) or international scale do not exist for the collection, archival and retrieval of historical and paleoflood peak discharge data. Historical and paleoflood data generally are not readily available or easily accessible in a convenient form for the practitioner to use, as compared to a large, systematic streamflow record

Table 1
Streamflow gaging station sites with historical and/or paleoflood data

Site no.	Location and historical/paleoflood data references	Drainage area (km^2)	Systematic record		Historical/paleoflood record	
			Years	s	Years AD (BP as noted)	h
1	James River at Richmond, Virginia USGS Gage No. 02037500; Williams and Guy (1973)	17,503	1935–1993	59	1607–1935	328
2	Passage Creek near Buckton, Virginia USGS Gage No. 01635500; Hupp (1987, 1988)	227	1933–1993	61	1720–1933	213
3	Western Run at Western Run, Maryland USGS Gage No. 01583500; Costa (1974, 1978a)	155	1945–1993	49	2100 BP–1945	2051
4	Shenandoah River at Milleville, West Virginia USGS Gage No. 01636500; Fuertsch (1992), Fanok and Wohl (1997)	7874	1896–1993	77	1740–1896	177
5	Susquehanna River at Harrisburg, Pennsylvania USGS Gage No. 01570500; Murphy (1905), Benson (1950)	62,419	1891–1993	103	1780–1891	111
6	Savannah River at Augusta, Georgia USGS Gage No. 02197000; Sanders et al. (1990)	19,446	1876–1951	76	1720–1876	156
7	Big Sandy River at Bruceton, Tennessee USGS Gage No. 03606500; IACWD (1982)	531	1930–1987	58	1860–1930	70
8	Alabama River near Montgomery, Alabama USGS Gage No. 02420000; Pearman et al. (1991)	39,075	1928–1993	66	1800–1928	128
9	Floyd River at James, Iowa USGS Gage No. 06600500; IACWD (1982), USGS (1959)	2295	1935–1993	59	1850–1935	85
10	Pecos River near Langtry, Texas USGS Gage Nos. 08447400, 08446500; Kochel et al. (1982), Baker et al. (1983), Lane (1987)	91,069	1900–1982	83	4450 BP–1900	4367
11	North Fork Cache La Poudre River near Livermore, Colorado USGS Gage No. 06751500; England (1998)	1469	1929–1931, 1947–1965, 1976	23	1850–1947	109
12	Cache La Poudre River at mouth of canyon near Fort Collins, Colorado USGS Gage No. 06752000; England (1998)	2735	1883–1993	110	1850–1883	34
13	Big Thompson River at mouth of canyon near Drake, Colorado USGS Gage No. 06378000; Costa (1978a,b), Jarrett and Costa (1988)	790	1888–1993	70	5000 BP–1888	4930
14	Clear Creek near Golden, Colorado USGS Gage No. 06719500; Baker (1974)	1036	1911–1974	64	6000 BP–1911	5936
15	First Creek below Buckley Road near Rocky Mountain Arsenal, Colorado USGS Gage No. 06720460; Capesius (1996)	65	1982–1993	11	1893–1982	89
16	Bear Creek at Morrison, Colorado USGS Gage No. 06710500; Grimm (1993), Grimm et al. (1995)	425	1896–1993	83	8500 BP–1896	8417
17	Muddy Creek at Kremmling, Colorado USGS Gage No. 09041500; Jarrett (1996)	751	1982–1994	12	5000 BP–1982	4988
18	Elkhead Creek near Elkhead, Colorado USGS Gage No. 09245000; Jarrett and Tomlinson (2000)	166	1953–1993	41	5000 BP–1953	4959
19	Animas River at Howardsville, Colorado USGS Gage No. 09357500; Pruess (1996)	145	1936–1982	47	1000 BP–1936	953
20	Junction Creek near Durango, Colorado USGS Gage No. 09361400; Pruess (1996)	68	1960–1965, 1980–1993	19	1000 BP–1960	981
21	S. Fork Ogden River near Huntsville, Utah USGS Gage No. 10137500; Ostenaa and Levish (1995), Ostenaa et al. (1997)	355	1921–1994	74	2500 BP–1921	2425
22	Escalante River near Escalante, Utah USGS Gage No. 09337500; Enzel et al. (1993), Webb and Baker (1987), Webb et al. (1988)	829	1943–1955, 1972–1993	42	2100 BP–1910	2058
23	Virgin River at Littlefield, Arizona USGS Gage No. 09508500; Enzel et al. (1994)	13,183	1930–1993	63	1250 BP–1930	1437

(*continued on next page*)

Table 1 (*continued*)

Site no.	Location and historical/paleoflood data references	Drainage area (km^2)	Systematic record		Historical/paleoflood record	
			Years	*s*	Years AD (BP as noted)	*h*
24	Colorado River at Lee's Ferry, Arizona USGS Gage No. 09380000; Enzel et al. (1993), Smith et al. (1995)	289,561	1884–1963	42	4000 BP–1884	3958
25	Salt River near Roosevelt, Arizona USGS Gage No. 09482500; Partridge and Baker (1987), Baker et al. (1987), Stedinger et al. (1988b)	11,152	1924–1993	69	2000 BP–1924	1931
26	Verde River below Tangle Creek above Horseshoe Dam, Arizona USGS Gage No. 09508500; Baker et al. (1987), Stedinger et al. (1988b), House et al. (1995)	15,172	1925–1993	69	2000 BP–1925	1931
27	Santa Cruz River at Tucson, Arizona USGS Gage No. 09482500; Webb and Betancourt (1992), Smith et al. (1995)	5755	1915–1993	78	1850–1915	66
28	Santa Ana River near Riverside Narrows, California USGS Gage No. 11066500; Beattie and Beattie (1939), Sidler (1968)	2214	1928–1973	45	1860–1928	69
29	Santa Ynez River at Bradbury Dam Site Santa Barbara County, California USGS Gage No. 11126000; Levish et al. (1994), Ostenaa et al. (1996)	1093	1907–1993	85	2920 BP–1907	2835
30	Trinity River above Coffee Creek near Trinity Center, California USGS Gage No. 11523200; Helley and LaMarche (1973)	386	1956–1994	37	1500–1956	463
31	Coffee Creek near Trinity Center, California USGS Gage No. 11523700; Stewart and LaMarche (1967)	308	1958–1966	9	1200–1958	791
32	Skagit River near Concrete, Washington USGS Gage No. 12194000; Stewart and Bodhaine (1961)	7089	1925–1994	69	1800–1925	125
33	Parana River at Corrientes, Argentina; Salas et al. (2002)	1,949,743	1904–1989	86	1800–1904	104
34	Trent River at Trent Bridge, United Kingdom; UNESCO (1976)	7511	1915–1969	52	1700–1915	218
35	Thames River at Teddington, United Kingdom; UNESCO (1976)	9661	1884–1971	88	1700–1884	184
36	Changjiang (Yangtze River) at Yichang, Hubei, China; Rodier and Roche (1984)	1,010,095	1877–1976	91	1100–1877	786

data base (e.g. Slack and Landwehr, 1993). So far as the current authors are aware, there are no programs in place to systematically document collection methods and data quality assurance/control. Current activities in the United States should improve this situation. A global paleoflood databank is currently being developed at the Laboratory of Tree Ring Research (LTRR), Arizona in cooperation with the Bureau of Reclamation, and will eventually be archived at the National Geophysical Data Center's (NGDC) World Data Center for broad scientific use (Hirschboeck, 1999). Ongoing paleoflood data collection efforts by Reclamation (e.g. Ostenaa et al. 1996, 1997; Levish et al., 2000) will enhance the databank.

Data compiled as part of this study show that both Type I and Type II censoring assumptions may apply. A physical criterion (e.g. bridge or building markers, terraces, etc.) was used to select the Type I model and differentiate from Type II cases. Type II sites were selected based on the largest flood criterion, where the information indicates that the largest flood since some time has occurred, without relation to some fixed threshold. For the data sets obtained, there are 13 Type I sites and 12 Type II sites. In addition, multiple censoring (more than one threshold) and binomial-censored data (Stedinger and Cohn, 1986) are observed at nine and two sites, respectively. Both Type I and II censoring assumptions may be needed when multiple thresholds

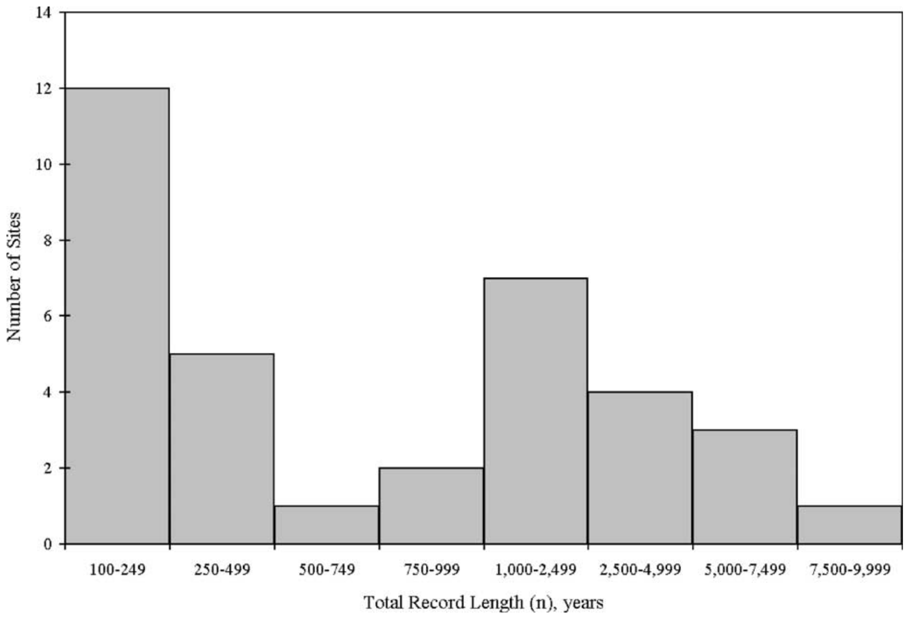

Fig. 2. Distribution of the total record length (n) for the 36 sites.

are present, as documented by Stedinger et al. (1988b) for the Salt and Verde Rivers. In practice, the censoring assumption, discharge threshold level, and historical period are estimated based on the available information. The issues in practice are really to identify the threshold level(s) Q_0, number of flood exceedances k and if the magnitude of Q_0 is nearly equal to or substantially less than the kth largest flood, rather than identifying the censoring type.

In addition to censoring type, there has been controversy regarding the number of historical floods or paleofloods that equal or exceed a discharge threshold. Typically, for a Type I censored data model the number of threshold exceedance floods (k) is a function of the return period of the threshold and the length of record (n), whereas k is fixed in advance for a Type II model. Previous researchers (e.g. Hosking and Wallis' (1986a,b) model), who have assumed only the largest flood was recorded in a historical period, have not adequately represented the potential information available. For the 27 sites with a single threshold compiled as part of this investigation (Table 2), two or more historical and/or paleofloods had exceeded the discharge threshold at 16 sites. For the 12 locations with Type II censoring, more than one exceedance flood ($k > 1$) was observed at two of these sites. As noted above, both censored data models, cases with multiple thresholds (nine sites) and binomial censored data (two sites) need to be considered. In general, however, it is difficult to infer the censoring type and discharge threshold when one knows only that a single large flood has occurred. In many practical situations Q_0 is assumed equal to the largest flood, without any additional information to indicate otherwise.

The pertinent peak discharge data characteristics are summarized in Tables 2 and 3 for single threshold sites (including binomial censoring) and multiple threshold sites, respectively. Two simple ratios, shown in Tables 2 and 3, were estimated to provide a rough empirical measure of the historical/paleoflood information content. The threshold ratio is defined as the ratio of the largest observed flood (Q_{max}) to the discharge threshold (Q_0). A similar concept was first expressed by Potter and Walker (1985) to describe the uncertainty in measured versus observed systematic discharge values. Matthai (1990) noted that the magnitudes of the largest Holocene flood peak discharge estimates on two rivers in Arizona were not much greater than two times the largest systematic flood peaks. The historical ratio is defined herein as the ratio of the historical record length (h) to the systematic

Table 2
Streamflow data summary, single threshold sites

Site no.	Site name	Years		Ratio h/s	Discharge threshold		No. floods $k > Q_0$		Max. observed discharge (m^3/s)		Ratio Q_{max}/Q_0
		s	h		Q_0 (m^3/s)	Ret. Per. T	e	e'	Q_h	Q_s	
Type I censoring (13 sites)											
5	Susquehanna River	103	111	1.1	10,870	9	17	6	18,500	28,883	2.66
7	Big Sandy River	58	70	1.4	510	46	0	3	708	481	1.39
8	Alabama River	66	128	1.9	7930	65	1	2	9120	8014	1.15
11	N. Fork Poudre River	23	109	5.1	184	100	2	2	241	268	1.46
12	Cache La Poudre River	110	34	0.3	283	100	3	1	340	340	1.20
15	First Creek	11	89	8.1	28.32	33	0	3	99.1	8.69	3.50
16	Bear Creek	83	8417	101.4	105	2100	3	1	146	146	1.39
28	Santa Ana River	45	69	1.5	9030	114	0	1	9060	2832	1.00
30	Trinity River	37	463	12.5	589	250	2	0	–	750	1.27
32	Skagit River	69	125	1.8	5660	32	0	6	14,160	4361	2.50
33	Parana River	86	104	1.2	50,000	38	2	3	60,200	60,215	1.20
34	Trent River	52	218	4.2	1100	67	1	3	1420	1110	1.29
36	Yangtze River	91	786	8.6	80,000	110	0	8	110,000	71,100	1.38
Type II censoring (12 sites)											
1	James River	59	328	5.6	6170	129	3	0	–	6286	1.02
3	Western Run	49	2051	41.9	1076	2100	1	0	–	1076	1.00
6	Savannah River	76	156	2.0	4530	14	11	5	7930	9910	2.19
9	Floyd River	59	85	1.4	2030	144	1	0	–	2025	1.00
13	Big Thompson R.	70	4930	70.4	883	5000	1	0	–	883	1.00
14	Clear Creek	64	5936	92.8	1420	6000	0	1	1420	167	1.00
17	Muddy Creek	12	4988	415.7	142	5000	0	1	142	47.3	1.00
18	Elkhead Creek	41	4959	121.0	142	5000	0	1	142	80.7	1.00
19	Animas River	47	953	20.3	69.9	1000	0	1	69.9	56.1	1.00
20	Junction Creek	19	981	51.6	86.9	1000	0	1	86.9	17.0	1.00
27	Santa Cruz River	78	66	0.9	1490	144	1	0	–	1492	1.00
31	Coffee Creek	9	791	87.9	501	800	1	0	–	501	1.00
Binomial censoring (two sites)											
2	Passage Creek	61	213	3.5	155.7	15	6	12	651.3	594.7	4.18
35	Thames River	88	184	2.1	1048	54	1	4	1060	1060	1.01

record (s). Frances et al. (1994) suggested that this ratio (their r) was one factor that described the statistical gain of Type I censored samples. As the historical ratio increases, the historical and paleoflood data provide more information about extreme floods and as a basis for extrapolation.

Flood frequency analyses were conducted for each of the 36 data sets as a practical way to compare B17H and EMA estimation procedures. The data were placed in three groups: (1) single threshold sites, either Type I or Type II censoring, where the discharge estimates were approximately known (Table 2, 25 sites); (2) binomial-censored data sets (Table 2, two sites); and (3) multiple-censored data locations (Table 3, nine sites). The first group was used as the major basis for the comparison. The B17H procedure can incorporate only a single threshold, and is unable to utilize binomial or multiple censored data. Data from the second two groups were utilized to demonstrate EMA performance in these practical situations.

4. Results and discussion

The single threshold EMA-B17H comparison results (Type I and Type II data) are presented and

Table 3
Streamflow data summary, multiple threshold locations (nine sites)

Site no.	Site name (number of censoring thresholds)	Years		Ratio h_j/s	Discharge threshold		No. floods $k_j > Q_0$		Max. observed discharge (m^3/s)		Ratio Q_{max}/Q_{oj}
		s	h_j		Q_{oj} (m^3/s)	Ret. Per. T	e	e'	Q_h	Q_s	
4	Shenandoah River (2 thresholds)	77	129	1.7	2350	12	12	9	4900	6500	2.78
			48	0.6	1150	4	17	3	–	2150	1.87
10	Pecos River (6 thresholds)	83	1890	22.7	11,330	1100	2	2	11,300	26,800	2.37
			260	3.1	10,760	776	0	1	10,800	16,300	1.52
			990	11.9	8500	289	0	5	8500	–	1.00
			280	3.4	6800	201	0	2	6800	–	1.00
			650	7.8	6090	113	0	4	6100	–	1.00
			300	3.6	5100	52	0	4	5100	–	1.00
21	S. Fork Ogden River (3 thresholds)	74	2025	27.0	115	>2020	0	0	–	–	0.00
			400	5.3	70	>400	0	0	–	–	0.00
			27	0.4	40	11	7	0	–	53	1.33
22	Escalante River (2 thresholds)	42	1980	47.2	699	2100	0	1	725	600	1.04
			76	1.8	501	29	4	0	–	600	1.20
23	Virgin River (4 thresholds)	63	1070	16.9	1600	750	0	2	1750	–	1.09
			100	1.6	1250	276	0	1	1400	–	1.12
			200	3.2	1100	80	0	3	1100	–	1.00
			72	1.1	600	13	3	6	850	997	1.42
24	Colorado River (3 thresholds)	42	3890	92.6	13,610	4000	0	1	13,900	–	1.02
			50	1.2	7930	109	0	1	8500	–	1.07
			20	0.5	6230	40	1	0	–	6230	1.00
25	Salt River (3 thresholds)	69	1400	20.3	4100	2000	0	1	4100	–	1.00
			477	6.9	3000	95	4	2	3500	4050	1.35
			54	0.8	1980	12	6	3	–	2950	1.49
26	Verde River (3 thresholds)	69	1000	14.5	5000	2000	0	1	5200	–	1.04
			894	13.0	3650	400	2	0	4100	4100	1.12
			37	0.5	1950	11	5	4	2720	2830	1.45
29	Santa Ynez River (3 thresholds)	85	2090	24.6	2550	>2900	0	0	–	–	0.00
			700	8.2	1980	830	1	0	–	2300	1.16
			45	0.5	1270	31	4	0	–	1560	1.22

discussed in Section 4.1. Binomial and multiple censoring results for EMA are discussed in Section 4.2. Probability plots of the data and LP-III distributions using each estimation procedure for the 13 Type I and 12 Type II censored data sets are shown in England (1998). Plots for selected data sets are shown in Section 4.1. In each plot, a filled square symbol is used to plot the k (threshold exceedance) flood observations at each site; an open triangle is used to plot the $(g-k)$ below threshold observations. LogNormal probability paper was chosen to graphically display the observed data and fitted distributions for three reasons: (1) this paper is the traditional choice to graphically display the empirical cumulative distribution of flood data; (2) the paper uses distorted axes to present a curve that may appear approximately linear; and (3) one may observe whether the data can readily be fit by some standard probability distributions. There is a considerable amount of distortion in this type of graph paper and the hydrologist should resist the false security that it may convey (Cudworth, 1989, p. 194). The paper gives more weight to the median (0.5) probability values and severely distorts the upper and lower tails of the data and fitted distributions.

4.1. Single threshold

The results of the EMA-B17H average relative deviation (ARD) comparison for the 25 sites are presented in Table 4 for both the Type I and Type II sites. In this case, the plotting position coefficient α was assumed to be equal to 0.4. The number of observed floods (g) and number of threshold exceedance floods (k) are listed for each location. The estimation method that corresponds to the lower value for ARD is considered best.

For the 13 Type I censored data sites considered here, the results in Table 4 indicate that EMA and B17H perform about equally well to fit the g observed floods. Strictly speaking, B17H fit 7 of 13 sites best using the ARD(g) criterion, and EMA fit six of 13 sites best. However, the absolute difference between the ARD results for the two estimators for these sites is practically negligible. Overall, EMA provided a slightly better fit than B17H as shown by the ARD sums for all Type I sites (Table 4). Minimal differences between the EMA and B17H frequency curves for the g flood observations were seen for most sites (England, 1998).

Table 4
Average relative deviation (ARD) results for the 25 comparison sites (plotting position $\alpha = 0.4$)

Site name	No. observed floods (g)	Floods $\geq Q_0(k)$	ARD (g)		ARD (k floods)		ARD (Q_{max})	
			EMA	B17H	EMA	B17H	EMA	B17H
Type I censoring (13 sites)								
Susquehanna River	109	23	0.0321	0.0299	0.0406	0.0459	0.1855	0.2173
Big Sandy River	61	3	0.0360	0.0366	0.0368	0.0515	0.0764	0.0960
Alabama River	68	3	0.0410	0.0399	0.0421	0.0474	0.0577	0.0695
N. Fork Poudre R.	25	4	0.1650	0.1537	0.1922	0.1890	0.4766	0.2610
Cache La Poudre R.	111	4	0.0277	0.0279	0.1271	0.1314	0.1708	0.1500
First Creek	14	3	0.1775	0.1833	0.1423	0.2410	0.1474	0.2881
Bear Creek	84	4	0.1067	0.2011	0.0984	2.879	0.1394	4.079
Santa Ana River	46	1	0.1480	0.1468	0.0081	0.1769	0.0081	0.1769
Trinity River	37	2	0.0928	0.0944	0.0664	0.0660	0.0862	0.1261
Skagit River	75	6	0.0347	0.0331	0.0640	0.1012	0.1601	0.2185
Parana River	89	5	0.0142	0.0127	0.0408	0.0524	0.0517	0.0206
Trent River	55	4	0.0233	0.0243	0.0224	0.0268	0.0225	0.0436
Yangtze River	99	8	0.0322	0.0270	0.0527	0.1345	0.0947	0.2158
ARD Sum for 13 Type I sites			0.9312	1.011	0.9339	4.143	1.677	5.962
Type II censoring (12 sites)								
James River	59	3	0.0556	0.0564	0.0908	0.1086	0.0972	0.0413
Western Run	49	1	0.0703	0.0692	0.0528	0.4098	0.0528	0.4098
Savannah River	81	16	0.0693	0.0467	0.0895	0.0898	0.1379	0.0433
Floyd River	59	1	0.0796	0.0751	0.1804	0.3500	0.1804	0.3500
Big Thompson R.	70	1	0.0644	0.0590	0.1998	0.4200	0.1998	0.4200
Clear Creek	65	1	0.0751	0.0650	0.0735	0.6270	0.0735	0.6270
Muddy Creek	13	1	0.0669	0.0762	0.1173	0.3326	0.1173	0.3326
Elkhead Creek	42	1	0.0580	0.0485	0.0383	0.2576	0.0383	0.2576
Animas River	48	1	0.0296	0.0302	0.0581	0.1103	0.0581	0.1103
Junction Creek	20	1	0.1194	0.0836	0.0575	0.6461	0.0575	0.6461
Santa Cruz River	78	1	0.0674	0.0664	0.0507	0.1123	0.0507	0.1123
Coffee Creek	9	1	0.0941	0.0940	0.1928	0.5126	0.1928	0.5126
ARD Sum for 12 Type II sites			0.8497	0.7703	1.2015	3.9767	1.2563	3.8629

Note: sites where $k = 1$ ARD (Q_{max}) equals ARD (k floods).

When one examines the Type I above-threshold floods (k and Q_{max}), EMA does a better job at predicting the largest floods, 11 out of 13 sites using the k observations and 10 out of 13 sites using the Q_{max} observation (Table 4). B17H appeared to perform marginally better than EMA for the North Fork Poudre and Parana River sites when one uses the ARD (g, k, and Q_{max}) numerical criteria. However, probability plots for these two sites (not shown) indicate that the performance of the estimators is nearly indistinguishable and most likely the differences are well within data sampling variability. In contrast, B17H performs poorly as compared to EMA for several sites where the historical/paleoflood period is long, especially for Bear Creek (Fig. 3), and substantially overestimates the largest floods. This finding is similar to the Monte Carlo simulation results reported by England (1998) for Type I censored samples and long record lengths.

The EMA-B17H comparison for Type II censored samples indicated somewhat similar results as for Type I samples. It appears that B17H does a marginally better job at fitting the g observed floods (nine out of 12 sites) as indicated in Table 4; but again the differences between the methods are practically negligible. B17H showed a slight advantage when the ARD(g) results were summed for the Type II sites, but performed worse than EMA for the large floods (Table 4). Based on the results shown in Table 4, the largest difference between the two estimators, where B17H performed better than EMA, was observed for the Savannah River site. However, the EMA estimator over-predicted the largest flood by about 13 percent, which is typically within measurement error for the largest floods. The two estimated curves are very similar in shape and both provide a reasonable fit to the data (Fig. 4).

The EMA estimator performed particularly well in fitting the largest observed floods (Q_{max}) at all 11 sites where $k = 1$, as compared to the B17H estimator. This finding is consistent with the simulation results presented by England (1998). The practical differences between B17H and EMA for Type II samples are illustrated by an example. For the Western Run

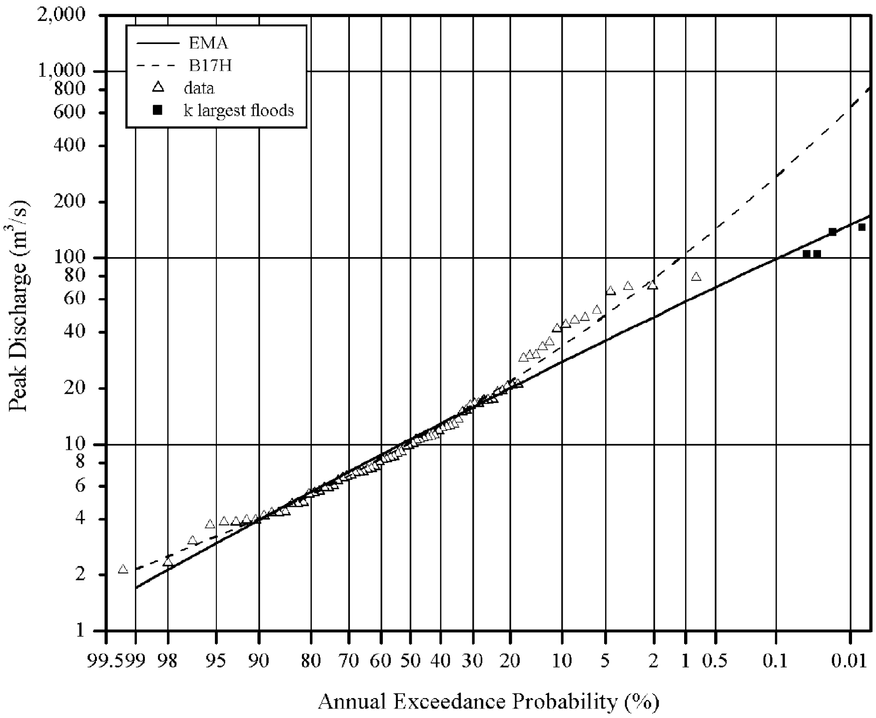

Fig. 3. Bear Creek flood frequency plot (Type I censoring, site 16).

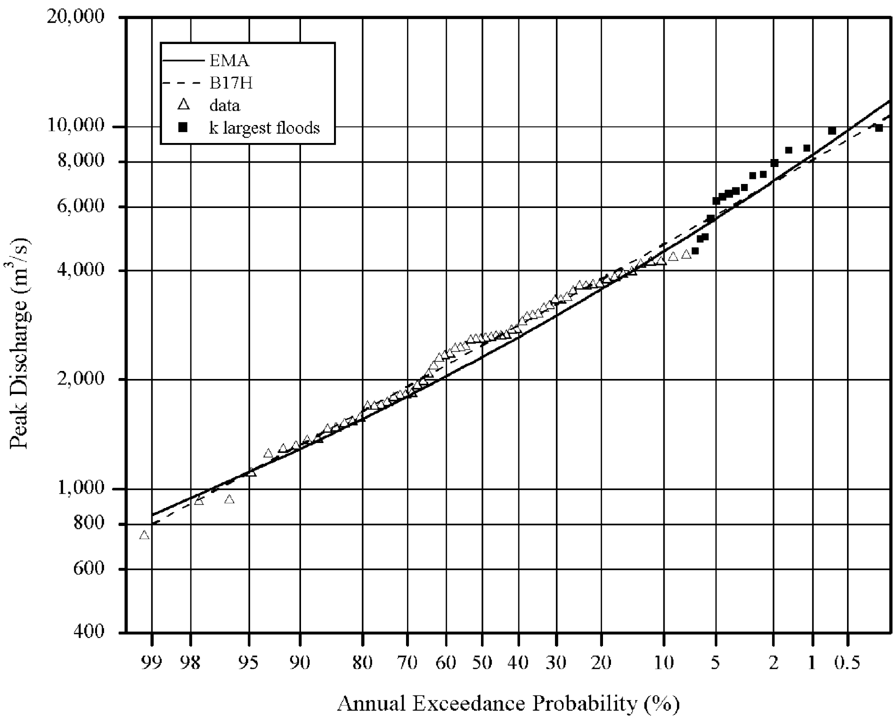

Fig. 4. Savannah River flood frequency plot (Type II censoring, site 6).

data, the estimation procedure results were nearly identical for the *g* observations, with B17H performing slightly better than EMA; frequency curves for this site are shown in Fig. 5. However, B17H underestimated the maximum peak at Western Run by over 40 percent (Table 4), which is the opposite situation from Bear Creek (Fig. 3).

Three Type II data sets have particularly short peak discharge gage records (less than or equal to 20 years): Muddy Creek, Junction Creek, and Coffee Creek. For these sites, estimation of flood exceedance probabilities, for return periods greater than 40 years (about twice the record length), is highly problematic. The addition of the largest flood known in some long time period can substantially improve extreme flood probability estimates. The B17H estimator for these three data sets underutilizes the largest flood information because it represents a time period longer than 500 years, and underestimates the largest flood magnitude by 33–64 percent (Table 4). These practical results agree with Type II Monte Carlo simulation results (England, 1998).

One interesting result is that EMA does not appear to provide a good fit to the entire range of flood observations, and over-predicts flood magnitudes in the 10–5 percent exceedance range, for the three short records mentioned above and for Clear Creek (site 14). However, B17H ignores the largest floods. EMA pays more attention to the largest flood, fits it well, and appears to provide more accurate peak discharge estimates for rare exceedance probabilities, for these sites. One plausible explanation for this supposed lack of fit for certain ranges of the probability distribution is data from mixed populations (e.g. Jarrett and Tomlinson, 2000). The probability plots show that the largest flood is inconsistent or does not follow the general trend of the data and may indicate floods from different mechanisms. One alternate possibility is that uncertainties in paleoflood discharge and date estimation have not been taken into account, or that the estimates are of poor quality. Other potential factors include short streamflow records, inaccurate maximum discharge estimates and/or flood dating errors, different flood parent distributions (i.e. the data are not LP-III),

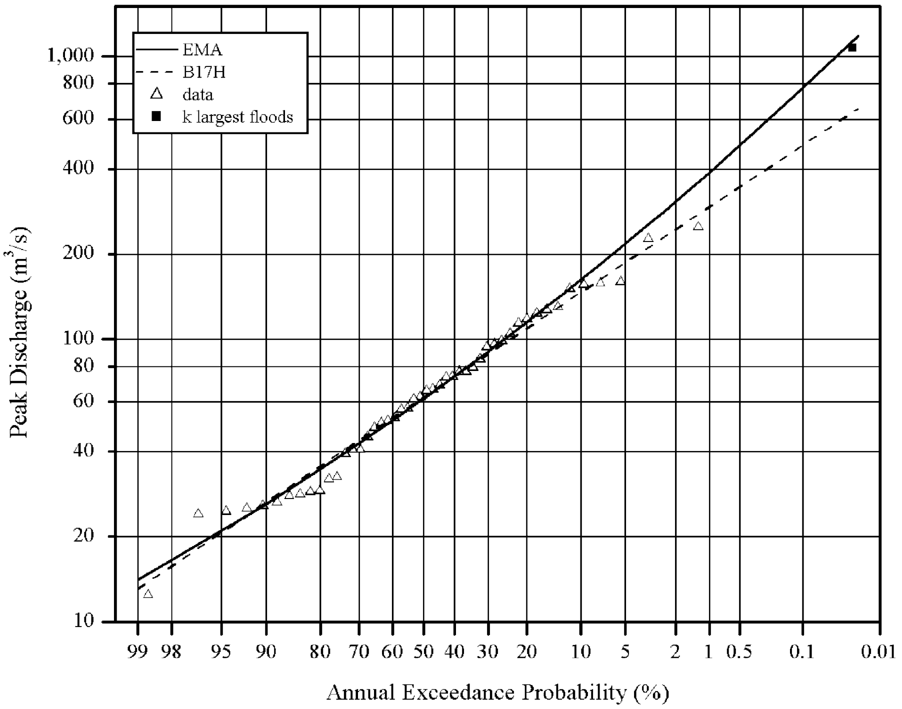

Fig. 5. Western Run flood frequency plot (Type II censoring, site 3).

non-stationary data, and low outliers. Further work is needed to assess the possible impact of these factors on making extreme flood probability statements.

The second numerical criterion used to compare the fit of the distributions was the mean squared relative deviation (MSD). Similar results to the ARD comparison were obtained for the MSD metric (Table 5) for the 25 data sets, but the differences were accentuated. EMA performed slightly better overall than B17H for the 13 Type I sites using the MSD(g), MSD(k floods) and MSD(Q_{max}) criteria. The differences between the estimators for the Type I sites were small except for the Bear Creek site, where EMA performed superior than B17H. For the Type I MSD(g) criterion at 13 sites, B17H fit five sites best, EMA fit four sites best, and the estimators were equal for the remaining four sites. The MSD(k floods) criterion indicated EMA fit best at 12 out of 13 sites. As in the ARD comparison, the estimators performed about the same for the 12 Type II sites using the MSD(g) criterion. EMA performed better at all 12 Type II sites for the MSD(k floods) criterion. Overall, EMA performs better than B17H for fitting the largest floods based on the MSD results.

The Weibull and Hazen plotting position parameter values ($\alpha = 0.0$ and 0.5, respectively) were used to estimate different $p(i)$ values for each data set to determine the effects of a plotting position on the ARD and MSD results presented in Tables 4 and 5. Although not shown, it was found that the choice of α had a minimal effect on the results when the 13 Type I and 12 Type II data sets were each considered as a group. Recall that the Type I or Type II exceedance-based plotting position estimate of the largest flood (Q_{max}), for samples where $k = 1$, is independent of α and the sample censoring type. For the 11 data sets where $k = 1$ (Table 4), the α change had some effect on the ARD(g) and MSD(g) estimates, but not on ARD(Q_{max}) and MSD(Q_{max}).

The descriptive ability of the method to match the extreme flood of record (Q_{max}) and the above-threshold floods is of interest, in addition to matching the range of quantiles as discussed above. Benson (1968, p. 904) suggested that future work be conducted to handle

Table 5
Mean squared relative deviation (MSD) results for the 25 comparison sites (plotting position $\alpha = 0.4$)

Site name	No. observed floods (g)	Floods $\geq Q_0(k)$	MSD (g)		MSD (k floods)		MSD (Q_{max})	
			EMA	B17H	EMA	B17H	EMA	B17H
Type I censoring (13 sites)								
Susquehanna River	109	23	0.0018	0.0017	0.0031	0.0040	0.0344	0.0472
Big Sandy River	61	3	0.0021	0.0021	0.0021	0.0036	0.0058	0.0092
Alabama River	68	3	0.0032	0.0032	0.0025	0.0033	0.0033	0.0048
N. Fork Poudre R.	25	4	0.0439	0.0334	0.0643	0.0420	0.2272	0.0681
Cache La Poudre R.	111	4	0.0017	0.0016	0.0182	0.0187	0.0292	0.0225
First Creek	14	3	0.0401	0.0458	0.0212	0.0710	0.0217	0.0830
Bear Creek	84	4	0.0229	0.4287	0.0116	8.8060	0.0194	16.6340
Santa Ana River	46	1	0.0388	0.0335	0.0001	0.0313	0.0001	0.0313
Trinity River	37	2	0.0173	0.0181	0.0048	0.0080	0.0074	0.0159
Skagit River	75	6	0.0029	0.0028	0.0067	0.0140	0.0256	0.0477
Parana River	89	5	0.0004	0.0004	0.0021	0.0034	0.0027	0.0004
Trent River	55	4	0.0009	0.0009	0.0006	0.0010	0.0005	0.0019
Yangtze River	99	8	0.0016	0.0021	0.0038	0.0206	0.0090	0.0466
MSD Sum for 13 Type I sites			0.1776	0.5743	0.1411	9.0269	0.3863	17.0126
Type II censoring (12 sites)								
James River	59	3	0.0046	0.0050	0.0088	0.0148	0.0094	0.0017
Western Run	49	1	0.0123	0.0105	0.0028	0.1679	0.0028	0.1679
Savannah River	81	16	0.0063	0.0037	0.0088	0.0093	0.0190	0.0019
Floyd River	59	1	0.0193	0.0131	0.0326	0.1225	0.0326	0.1225
Big Thompson R.	70	1	0.0074	0.0086	0.0399	0.1764	0.0399	0.1764
Clear Creek	65	1	0.0173	0.0114	0.0054	0.3932	0.0054	0.3932
Muddy Creek	13	1	0.0085	0.0142	0.0138	0.1106	0.0138	0.1106
Elkhead Creek	42	1	0.0052	0.0048	0.0015	0.0664	0.0015	0.0664
Animas River	48	1	0.0013	0.0014	0.0034	0.0122	0.0034	0.0122
Junction Creek	20	1	0.0376	0.0284	0.0033	0.4174	0.0033	0.4174
Santa Cruz River	78	1	0.0080	0.0072	0.0026	0.0126	0.0026	0.0126
Coffee Creek	9	1	0.0135	0.0334	0.0372	0.2628	0.0372	0.2628
MSD Sum for 12 Type II sites			0.1413	0.1417	0.1601	1.7661	0.1709	1.7456

Note: sites where $k = 1$ MSD (Q_{max}) equals MSD (k floods).

'outliers' or rare floods, and cautioned that "in any case, any major modifications… would have to meet the test of conforming to the data satisfactorily." Similarly, the NRC recommended focusing on the extreme tails of the probability distribution as one of three ways to better estimate probabilities of extreme floods (NRC, 1988). The relative deviations for the return period of the largest flood $q_i(Q_{max})$ and for the k threshold exceedance floods for each of the 25 comparison data sets are shown in Tables 4 and 5. The results indicate that the EMA estimator is superior, as compared to B17H, for the most extreme (largest magnitude) observations. These practical findings match Monte Carlo simulation results (England, 1998).

In addition to the $q_i(Q_{max})$ comparison, a non-exceedance probability test was used to roughly indicate whether the LP-III fitted distribution is consistent with Q_{max} for the period n. The non-exceedance probability plot test results for the 25 comparison sites are shown in Table 6. EMA appears to provide a suitable fit to the largest flood for all 25 sites. On the other hand, the results indicate that an LP-III distribution with parameters estimated by the B17H method does not adequately describe the largest flood at eight sites (denoted in italics). One may confirm these results by examining the probability plots. In one case (Bear Creek), B17H severely over-estimates the largest flood exceedance probability; one infers that the flood has a smaller return period.

Table 6
Non-exceedance probability test results for the 25 comparison sites. Numbers in italics indicate the fitted distribution is inconsistent with the largest observation

Site name	Q_{max} (m^3/s)	EMA		B17H	
		Return period	p^*	Return period	p^*
Type I censoring (13 sites)					
Susquehanna River	28,880	1150	0.830	1700	0.879
Big Sandy River	708	165	0.458	154	0.433
Alabama River	9118	217	0.408	202	0.383
N. Fork Poudre River	268	90	0.230	129	0.357
Cache La Poudre River	339.8	126	0.318	135	0.344
First Creek	99.1	150	0.511	308	0.722
Bear Creek	146	7200	0.307	209	*0.000*
Santa Ana River	9060	226	0.603	293	0.678
Trinity River	750	2500	0.815	4700	0.899
Skagit River	14,160	620	0.731	870	0.800
Parana River	1705	221	0.423	279	0.505
Trent River	1420	590	0.631	730	0.692
Yangtze River	3110	6700	0.876	> 100,000	*1.000*
Type II censoring (12 sites)					
James River	8860	1600	0.785	940	0.663
Western Run	1080	3400	0.543	111,000	*0.982*
Savannah River	9910	213	0.336	305	0.467
Floyd River	2025	515	0.756	1300	0.896
Big Thompson River	883	5500	0.402	3500	0.242
Clear Creek	1420	9700	0.539	> 1,000,000	*0.998*
Muddy Creek	142	4400	0.320	> 1,000,000	*0.999*
Elkhead Creek	1410	6700	0.475	> 1,000,000	*1.000*
Animas River	69.9	1000	0.369	650	0.214
Junction Creek	86.9	1600	0.532	> 1,000,000	*1.000*
Santa Cruz River	1490	340	0.654	429	0.715
Coffee Creek	501	1000	0.448	125,000	*0.994*

For the other seven cases, B17H underestimates the flood hazard at rare exceedance probabilities.

EMA performed equally as well or slightly better overall than B17H for the 25 data sets, based on three numerical metrics (ARD, MSD and p^*) and one qualitative criterion (probability plots) considered. Based on the data and comparison presented in this study, EMA performs comparably to the existing B17H approach for the range of quantiles and distribution considered, and provides a better fit to the largest observed floods. Several factors that could have a small impact on the results presented here were not investigated. The effects of the distribution (LP-III) assumption were not considered here; however, based on probability plots (not shown), the distribution choice appears to be adequate (England, 1998). For ease of comparison and simplicity, regional coefficients of skew and low outlier adjustments were not considered here.

Thomas (1985) raised concerns about potential computational disadvantages with alternative approaches such as censoring theory, and that they must be weighed against any improvement in accuracy. The censored data EMA is an iterative solution. For example, the number of iterations for solution convergence was six and 36 for the Big Sandy River and Elkhead Creek data sets, respectively. On a modern desktop personal computer, individual data set run times were less than about 2 s on average for the 36 data sets analyzed. EMA computational disadvantages, as compared to B17H, are negligible for these data sets.

Table 7a
Results for the two binomial and nine multiple censoring sites using EMA (average relative deviation (ARD) and mean squared relative deviation (MSD) criteria (plotting position $\alpha = 0.4$))

Site name	No. observed floods (g)	Total floods $\geq Q_0(k)$	ARD			MSD		
			(g)	(k floods)	(Q_{max})	(g)	(k floods)	(Q_{max})
Binomial censoring (two sites)								
Passage Creek	73	18	0.1310	0.2127	0.0664	0.0236	0.0506	0.0044
Thames River	92	5	0.0628	0.1352	0.3122	0.0081	0.0276	0.0975
Multiple censoring (nine sites)								
Shenandoah River	89	41	0.0504	0.0499	0.1203	0.0040	0.0034	0.0145
Pecos River	101	20	0.1284	0.1221	0.3236	0.0254	0.0209	0.1047
S. Fk Ogden River	47	7	0.0519	0.0697	0.0419	0.0055	0.0059	0.0018
Escalante River	43	5	0.2076	0.4828	0.5287	0.0826	0.2458	0.2795
Virgin River	75	16	0.0848	0.1312	0.3512	0.0122	0.0233	0.1234
Colorado River	44	3	0.0793	0.0417	0.0480	0.0146	0.0026	0.0023
Salt River	75	16	0.1338	0.2129	0.5716	0.0301	0.0659	0.3267
Verde River	74	12	0.1175	0.1550	0.1342	0.0252	0.0286	0.0180
Santa Ynez River	85	4	0.1336	0.0777	0.0265	0.0245	0.0082	0.0007

4.2. Binomial censored cases and multiple threshold cases

Data from 11 sites are used to illustrate the application of the EMA estimation procedure with binomial, interval and multiple-censored data sets. Two simple criteria are used to assess the performance of EMA: the probability plot (qualitative), and the non-exceedance probability test (quantitative). ARD and MSD metrics were also estimated for the EMA curve as above; however, results cannot be used in a strict quantitative fashion due to the presence of binomial and/or interval censored data. EMA results for the two binomial censoring sites and nine multiple threshold sites are displayed via probability plots in England (1998). Results from selected sites are shown below. Binomial and interval censored data are shown in a range with vertical bars on the frequency plots. Because peak discharge estimates are not precisely known, one must use caution when comparing data plotting position estimates to the computed frequency curve.

ARD and MSD results are shown with p^* estimates in Tables 7a and b, respectively. The LP-III distribution appears to fit through the median portion of the data reasonably well for all sites. Computed ARD(g) values for the 11 sites are generally low (less than 0.3). The estimated distribution also provides a somewhat reasonable fit to the k above-threshold floods; ARD(k) results are relatively low except for the Escalante River data. As shown in Fig. 6, the k flood magnitudes for this site are nearly equal, and are disjoint as compared to the remainder of the data. Webb and Baker (1987) suggest that the flood record at this site is non-stationary. These large floods may also have been produced from some different storm mechanism than the other, lower magnitude flows

Table 7b
Results for the two binomial and nine multiple censoring sites using EMA (non-exceedance probability test criterion; numbers in italics indicate the fitted distribution is inconsistent with the largest observation)

Site name	Q_{max} (m^3/s)	Return period	p^*
Binomial censoring (two sites)			
Passage Creek	651	570	0.617
Thames River	1060	128	0.119
Multiple censoring (nine sites)			
Shenandoah River	6513	290	0.412
Pecos River	26,844	77,000	0.944
South Fork Ogden River	53.5	71	0.348
Escalante River	725	910	0.100
Virgin River	1800	780	0.146
Colorado River	13,875	12,000	0.715
Salt River	4103	390	*0.006*
Verde River	5207	1600	0.276
Santa Ynez River	2491	1300	0.537

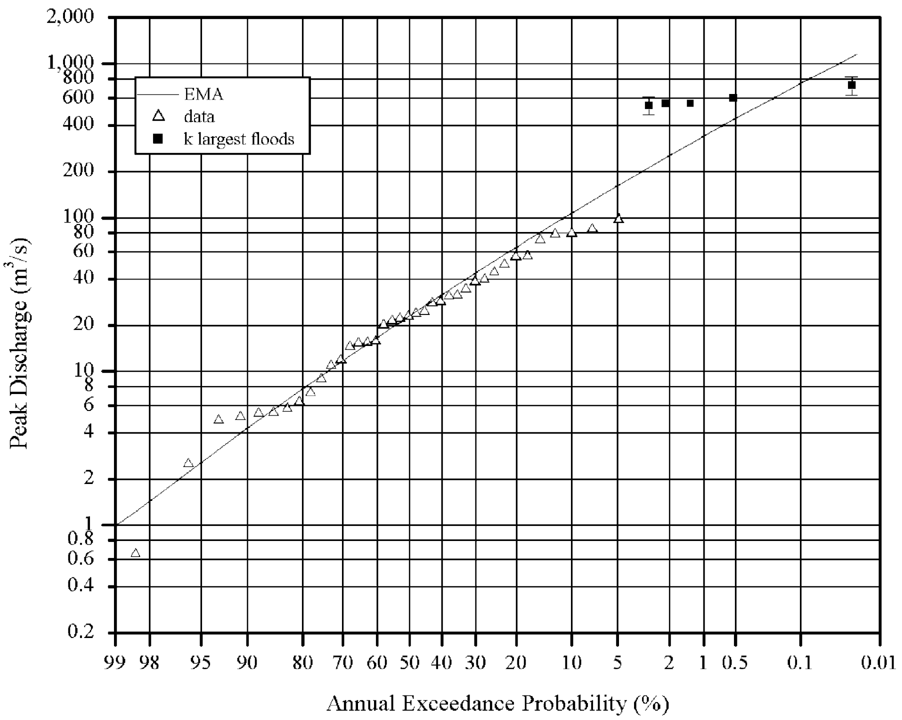

Fig. 6. Escalante River flood frequency plot (multiple censoring, site 22).

(mixed population). In any case, it is virtually impossible to fit a single, commonly used probability distribution to the entire Escalante River data set without investigating and separating mixed population data (if it exists), and/or correcting for stationarity. The data do not appear to obey the traditional distribution requirement that observations are 'monotonically increasing.'

The probability plots reveal discrepancies in the k exceedance flood observations that are not necessarily reflected in the ARD(k) results. In addition to the Escalante River data, at three sites (Passage Creek, Salt River and Verde River), it appears a single distribution cannot accurately fit the entire range of data, probably due to mixed populations. For example, EMA generally under-fits the Passage Creek binomial and interval censored data in the 5–0.5 percent exceedance range, but does fit Q_{max} well. At both the Salt River (Fig. 7) and Verde River sites the data appear to have a prominent S-shape; the k largest floods are not particularly well fit by the LP-III distribution. It is virtually impossible to fit the entire range of the Salt and Verde River data with one of the commonly used 2- or 3-parameter distributions. Stedinger et al. (1988b) censored about 70 percent of the lower flows to provide a better fit to the upper portion of the data.

The ARD(Q_{max}) and p^* results shown in Tables 7a and b indicate that the LP-III (EMA) distribution may not provide an adequate fit to the maximum observation at five sites (Thames, Pecos, Escalante, Virgin and Salt Rivers). The p^* results suggest that the fitted distribution is inconsistent with and over predicts Q_{max} for the Salt River. The ARD(Q_{max}) and p^* results also suggest the distribution over predicts the largest flood for the Thames, Escalante, and Virgin Rivers and significantly under predicts the largest flood on the Pecos River. However, the magnitude of the largest peak discharge on the Thames River is not known with certainty; the apparent lack of fit at this site may be an artifact of not knowing the magnitude or being able to constrain it to a narrower range. The lack of fit to the largest flood at the other three sites may be due to several factors, such as mixed-population data, and/or the fact that nature may not follow simple 3-parameter

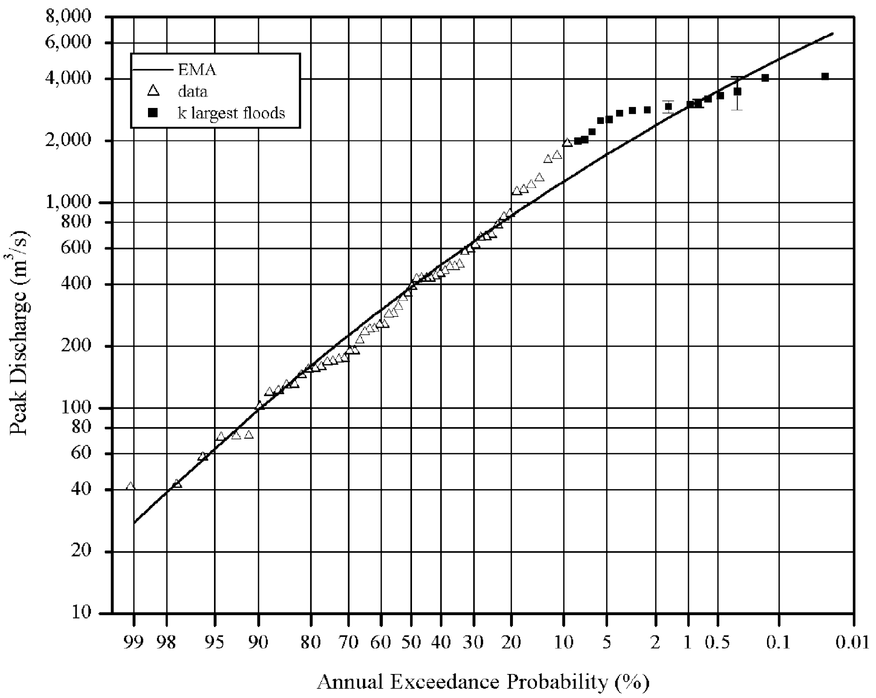

Fig. 7. Salt River flood frequency plot (multiple censoring, site 25).

probability distributions to describe peak discharge probability relations for data from these long time frames.

Moments-based estimation procedures, including EMA, may not work so well for some paleoflood non-exceedance bound cases on snowmelt-dominated streams, without severe censoring. For example, there are two paleoflood non-exceedance bounds in the South Fork Ogden River data set (Ostenaa et al., 1997). The main portion of the data indicated a negative $\log_e$-space coefficient of skew (-1.08) that reflects a snowmelt-dominated record (Fig. 8). The two paleohydrologic bounds appear to be inconsistent with the general trend of the gage data and potentially represent a different, larger flood mechanism (mixed population) than snowmelt that has caused the largest floods on record in the basin. The estimated upper bound of the LP-III (EMA) fitted distribution ($\exp(\tau) = 67\ m^3/s$) is inconsistent with the paleohydrologic bounds for the South Fork Ogden River, because it is smaller than the lower 400-year paleohydrologic bound discharge ($70\ m^3/s$). As noted by Cohn et al. (1997, p. 2091), when one uses the method of moments (or L-moments) to estimate three parameters of a probability distribution that is bounded such as LP-III, GEV, P-III, etc. one or more observations may lie outside the upper (lower) bound, and thus outside distribution support. The main problem with fitting these data with any three parameters function is the apparent non-homogeneity or lack of including floods that exceeded the paleohydrologic bounds (Fig. 8). Censoring lower flows using EMA may provide a better fit to the observed data. If there were paleofloods that exceeded one or more of the thresholds at this site, the fit to the paleoflood bounds would most likely improve. Further work is needed to understand the relations between paleohydrologic bounds, paleofloods and gaging station flows, especially for mixed-population data sets. Additional explorations in using EMA are warranted, such as fitting longer, multiple threshold historical and paleoflood data sets, and for lower flow censoring (e.g. NRC, 1999).

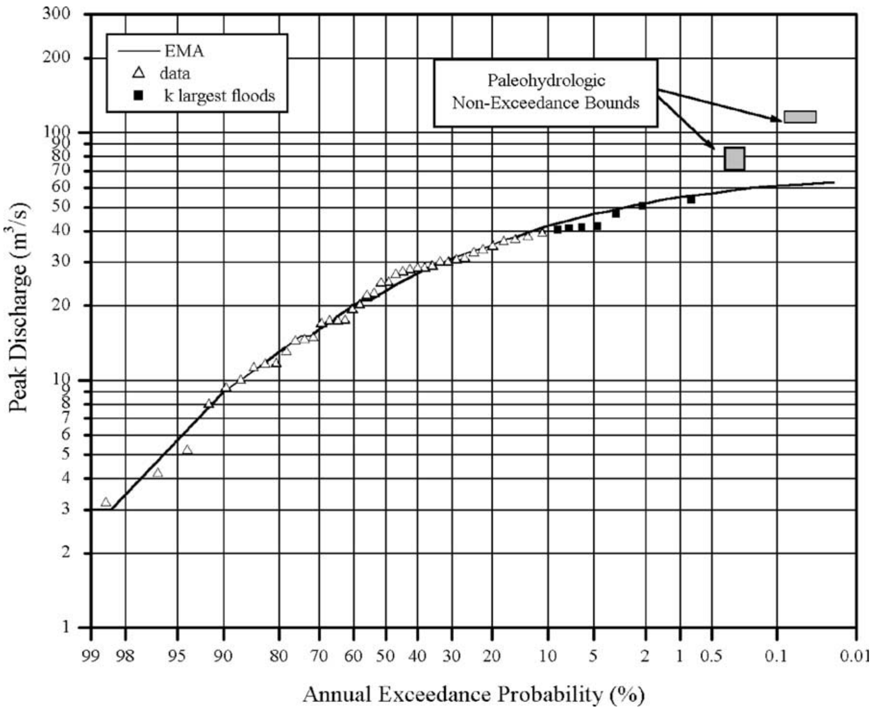

Fig. 8. South Fork Ogden River flood frequency plot (multiple censoring, site 21).

5. Conclusions

A data base consisting of historical, paleoflood and peak discharge information at 36 sites was compiled for flood frequency analysis. Data from 25 sites were utilized to compare B17H and EMA moments estimators. Information at 11 sites consisted of binomial and/or multiple censored data that B17H could not use, but can be used by EMA. In terms of practical application of these censored data models, compilation of this 36-site data base revealed several points. It was shown that both Type I and Type II censored data are prevalent; thus, both types need to be considered. Paleoflood information was available at 19 of the 36 sites, thus indicating there are some paleoflood data available, but the number of sites is severely limited compared to gaging station records. In addition, there were five or less floods with historical/paleoflood information at 26 out of 36 sites. On the basis of this data base, additional efforts should be invested in obtaining additional historical and paleofloods, determining discharge censoring thresholds and providing a physical understanding for the flood recording mechanisms.

Three quantitative metrics of comparison (ARD, MSD, and p^*) and one qualitative method (probability plots) were used to compare EMA and B17H. Goodness-of-fit results for 25 data sets revealed that EMA and B17H estimators perform equally well for fitting the entire range of floods observed at each site based on the two metrics, probability plots and distribution considered. EMA outperformed B17H for fitting the largest historical/paleofloods (k above-threshold floods) and the largest flood at 11 out of 13 Type I sites and 10 out of 12 Type II sites.

Results from this data comparison study indicate that EMA appears to be a viable alternative to current B17H procedures from an operational perspective, and performed equally or better than the existing approach for the 25 data sets analyzed. This data comparison study reinforces prior Monte Carlo simulation results. An additional 11 binomial and multiple censored data sets were utilized to demonstrate EMA performance for these cases. It was shown that the EMA estimator readily incorporates these types of information and the LP-III distribution provides an adequate fit to the data in most cases. The results shown here are encouraging. As EMA is moments-based, it is consistent with the Bulletin 17B guidelines. The Bulletin 17B document could be revised to include an option for EMA as an alternative or replacement to the existing historical weighting approach.

Acknowledgements

The financial support of the United States Society on Dams in the form of a graduate student scholarship for the first author is gratefully acknowledged. NSF Grant CMS-9625685 on 'Uncertainty and Risk Analysis Under Extreme Hydrologic Events' provided partial support of the third author. The advice and encouragement from William L. Lane and Timothy A. Cohn in support of this work are appreciated.

References

Baker, V.R., 1974. Paleohydraulic interpretation of Quaternary alluvium near Golden, Colorado. Quat. Res. 4, 94–112.

Baker, V.R., 1987. Paleoflood hydrology and extraordinary flood events. J. Hydrol. 96 (1–4), 79–99.

Baker, V.R., Kochel, R.C., Patton, P.C., Pickup, G., 1983. Paleohydrologic analysis of Holocene flood slack-water sediments. Spec. Publ. Int. Ass. Sediment. 6, 229–239.

Baker, V.R., Ely, L.L., O'Connor, J.E., Partridge, J.B., 1987. Paleoflood hydrology and design applications. In: Singh, V.P., (Ed.), Regional Flood Frequency Analysis, Reidel, Dordrecht, pp. 339–353.

Baker, V.R., Webb, R.H., House, P.K., 2002. The scientific and societal value of paleoflood hydrology. Am. Geophys. Union Water Sci. Appl. Ser. 5, 1–19.

Beattie, G.W., Beattie, H.P., 1939. Heritage of the Valley: San Bernadino's First Century, San Pasqual Press, Pasadena, CA.

Benson, M.A., 1950. Use of historical data in flood frequency analysis. Trans. Am. Geophys. Union 31 (3), 419–424.

Benson, M.A., 1968. Uniform flood-frequency estimating methods for Federal agencies. Water Resour. Res. 4 (5), 891–908.

Bobée, B.B., Robitaille, R., 1977. The use of the Pearson Type III and Log Pearson Type III distributions revisited. Water Resour. Res. 13 (2), 427–443.

Capesius, J., 1996. A regional analysis of eastern Colorado streams and a paleoflood analysis of First Creek, Rocky Mountain Arsenal, Commerce City, Colorado. MS Thesis, Department of Geology, University of Denver, Denver, CO, 178 p.

Cohn, T.A., Lane, W.L., Baier, W.G., 1997. An algorithm for computing moments-based flood quantile estimates when historical information is available. Water Resour. Res. 33 (9), 2089–2096.

Cohn, T.A., Lane, W.L., Stedinger, J.R., 2001. Confidence intervals for Expected Moments Algorithm flood quantile estimates. Water Resour. Res. 37 (6), 1695–1706.

Conover, W.J., 1999. Practical Nonparametric Statistics, third ed, Wiley, New York, 584 p.

Costa, J.E., 1974. Response and recovery of a Piedmont watershed from Tropical Storm Agnes, June 1972. Water Resour. Res. 10 (1), 106–112.

Costa, J.E., 1978a. Holocene stratigraphy in flood frequency analysis. Water Resour. Res. 14 (4), 626–632.

Costa, J.E., 1978b. Colorado Big Thompson flood: geologic evidence of a rare hydrologic event. Geology 6, 617–620.

Cudworth, A.G. Jr., 1989. Flood Hydrology Manual. A Water Resources Technical Publication, US Department of Interior, Bureau of Reclamation, Denver, CO, 243 p.

Cunnane, C., 1978. Unbiased plotting positions—a review. J. Hydrol. 37, 205–222.

Denlinger, R.P., O'Connell, D.R.H., House, P.K., 2002. Robust estimation of stage and discharge: an example from an extreme flood on the Verde River, Arizona. Am. Geophys. Union Water Sci. Appl. Ser. 5, 127–146.

Ely, L.L., 1992. Large floods in the southwestern United States in relation to late Holocene climate variations. PhD dissertation, Department of Geosciences, University of Arizona, Tucson, 326 p.

Ely, L.L., 1997. Response of extreme floods in the southwestern United States to climate variations in the late Holocene. Geomorphology 19, 175–201.

England, J.F., Jr., 1998. Assessment of historical and paleohydrologic information in flood frequency analysis. MS Thesis, Department of Civil Engineering, Colorado State University, Fort Collins, CO, 292 p.

Enzel, Y., Ely, L.L., House, P.K., Baker, V.R., Webb, R.H., 1993. Paleoflood evidence for natural upper bound to flood magnitudes in the Colorado River Basin. Water Resour. Res. 29 (7), 2287–2297.

Enzel, Y., Ely, L.L., Martinez-Goytre, J., Vivian, R.G., 1994. Paleofloods and a dam-failure flood on the Virgin River, Utah and Arizona. J. Hydrol. 153, 291–315.

Fanok, S.F., Wohl, E.E., 1997. Assessing the accuracy of paleohydrologic indicators, Harpers Ferry, West Virginia. J. Am. Water Resour. Assoc. 33 (5), 1091–1102.

Frances, F., Salas, J.D., Boes, D.C., 1994. Flood frequency analysis with systematic and historical or paleoflood data based on the two-parameter general extreme value models. Water Resour. Res. 30 (6), 1653–1664.

Fuertsch, S.J., 1992. A paleohydrologic investigation in the vicinity of Harpers Ferry, West Virginia. MS Thesis, Department of Earth Resources, Colorado State University, Fort Collins, CO, 113 p.

Fuller, W.E., 1914. Flood flows. Trans. Am. Soc. Civ. Eng., Paper No. 1293 77, 564–617.

Grimm, M.M., 1993. Paleoflood history and geomorphology of Bear Creek basin, Colorado. MS Thesis, Department of Earth Resources, Colorado State University, Fort Collins, CO, 126 p.

Grimm, M.M., Wohl, E.E., Jarrett, R.D., 1995. Coarse-sediment distribution as evidence of an elevation limit for flash flooding, Bear Creek, Colorado. Geomorphology 14, 199–210.

Guo, S.L., Cunnane, C., 1991. Evaluation of the usefulness of historical and palaeological floods in quantile estimation. J. Hydrol. 129, 245–262.

Helley, E.J., LaMarche, V.C., 1973. Historic flood information for northern California streams from geological and botanical evidence. US Geological Survey Professional Paper 485-E, 16 p.

Hirsch, R.M., Stedinger, J.R., 1987. Plotting positions for historical floods and their precision. Water Resour. Res. 30 (6), 1653–1664.

Hirschboeck, K.K., 1999. The Global Paleoflood Databank: A Hands-on Overview and Input Workshop, Program and Abstracts, Second International Paleoflood Conference, Prescott, AZ, September 26–29, 1999 19 p.

Hosking, J.R.M., Wallis, J.R., 1986a. Paleoflood hydrology and flood frequency analysis. Water Resour. Res. 22 (4), 543–550.

Hosking, J.R.M., Wallis, J.R., 1986b. The value of historical data in flood frequency analysis. Water Resour. Res. 22 (11), 1606–1612.

House, P.K., Pearthree, P.A., Fuller, J.E., 1995. Hydrologic and paleohydrologic assessment of the 1993 floods on the Verde River, Arizona. Arizona Geological Survey Open-File Report 95–20, 23 p.

House, P.K., Webb, R.H., Baker, V.R., Levish, D.R., 2002. Ancient floods, modern hazards: principles and applications of paleoflood hydrology. Am. Geophys. Union Water Sci. Appl. Ser. 5, 385.

Hupp, C.R., 1987. Botanical evidence of floods and paleoflood history. In: Singh, V.P., (Ed.), Regional Flood Frequency Analysis, Reidel, Dordrecht, pp. 355–369.

Hupp, C.R., 1988. Plant ecological aspects of flood geomorphology and paleoflood history. In: Baker, V.R., Kochel, R.C., Patton, P.C. (Eds.), Flood Geomorphology, Wiley, New York, pp. 335–356.

Interagency Committee on Water Data (IACWD), 1982. Guidelines for determining flood flow frequency: Bulletin 17-B. Hydrology Subcommittee, March 1982 (revised and corrected), 28 p. and appendices.

Jain, D., Singh, V.P., 1987. Comparison of some flood frequency distributions using empirical data. In: Singh, V.P., (Ed.), Hydrologic Frequency Modeling, Reidel, Dordrecht, pp. 467–485.

Jarrett, R.D., 1987. Flood hydrology of foothill and mountain streams in Colorado. PhD Dissertation, Department of Civil Engineering, Colorado State University, Fort Collins, CO, 239 p.

Jarrett, R.D., 1991. Paleohydrology and its value in analyzing floods and droughts. US Geological Survey Water-Supply Paper 2375, pp. 105–116.

Jarrett, R.D., 1994. Historic-flood evaluation and research needs in mountainous areas. In: Cotroneo, G.V., Rumer, R.R. (Eds.), Hydraulic Engineering'94: National Conference on Hydraulic Engineering, August 1–5, 1994, Buffalo, NY, vol. 1. ASCE, pp. 875–879.

Jarrett, R.D., 1996. Interdisciplinary paleoflood investigation of the Muddy Creek Basin for Ritschard Dam near Kremmling, Colorado. Report submitted to the Colorado River Water Conservation District, dated April 24, 1996. US Geological Survey, Denver, Colorado, 21 p.

Jarrett, R.D., Costa, J.E., 1988. Evaluation of the flood hydrology in the Colorado Front Range using precipitation, streamflow, and paleoflood data for the Big Thompson River Basin. US Geological Survey Water-Resources Investigations Report 87-4117, 37 p.

Jarrett, R.D., England, J.F. Jr., 2002. Reliability of paleostage indicators for paleoflood studies. Am. Geophys. Union Water Sci. Appl. Ser. 5, 91–109.

Jarrett, R.D., Tomlinson, E.M., 2000. Regional interdisciplinary paleoflood approach to assess extreme flood potential. Water Resour. Res. 36 (10), 2957–2984.

Kirby, W.H., 1981. Annual flood frequency analysis using US Water Resources Council Guidelines (Program J407). Section C in Lepkin (and others), WATSTORE User's Guide vol. 4, Chapter 1, US Geological Survey Open-File Report 79-1336-I, pp. C-1 to C-56.

Kochel, R.C., Baker, V.R., Patton, P.C., 1982. Paleohydrology of southwestern Texas. Water Resour. Res. 18 (4), 1165–1183.

Lane, W.L., 1987. Paleohydrologic data and flood frequency estimation. In: Singh, V.P., (Ed.), Regional Flood Frequency Analysis, Reidel, Dordrecht, pp. 287–298.

Levish, D.R., 2002. Paleohydrologic bounds-nonexceedance information for flood hazard assessment. Am. Geophys. Union Water Sci. Appl. Ser. 5, 175–190.

Levish, D.R., Ostenaa, D.A., O'Connell, D.R.H., 1994. A non-inundation approach to paleoflood hydrology for the event-based assessment of extreme flood hazards. Proceedings of the Association of State Dam Safety Officials 1994 Annual Conference, Boston, Massachusetts, 69–82.

Levish, D.R., England, J.F., Jr., Klawon, J.E., O'Connell, D.R.H., 2000. Probabilistic flood hazard analysis for Seminoe and Glendo Dams, Kendrick Project and Pick-Sloan Missouri Basin Program, Wyoming. Draft Report, Bureau of Reclamation, Technical Service Center, Denver, various paging.

Matthai, H.F., 1990. Floods. Geol. Soc. Am., Geol. N. Am. 0-1, 97–120.

Murphy, E.C., et al., 1905. Destructive floods in the United States in 1904. US Geological Survey Water-Supply Paper 147, 206 p.

National Research Council (NRC), 1988. Estimating Probabilities of Extreme Floods: Methods and Recommended Research, National Academy Press, Washington, DC, 141 p.

National Research Council (NRC), 1999. Improving American River Flood Frequency Analysis, National Academy Press, Washington, DC, 120 p.

O'Connell, D.R.H., Ostenaa, D.A., Levish, D.R., Klinger, R.E., 2002. Bayesian flood frequency analysis with paleohydrologic bound data. Water Resour. Res. 38 (5), doi: 10.1029/2000WR000028.

Ostenaa, D.A., Levish, D.R., 1995. The application of a non-inundation approach of paleoflood hydrology for the assessment of extreme flood hazards on the Ogden River, Utah. Utah Geol. Surv. Publ. 24, 495–508.

Ostenaa, D.A., Levish, D.R., O'Connell, D.R.H., 1996. Paleoflood study for Bradbury Dam, Cachuma Project, California.

Bureau of Reclamation Seismotectonics Report 96-3, Denver, CO, 86 p.
Ostenaa, D.A., Levish, D.R., O'Connell, D.R.H., Cohen, E.A., 1997. Paleoflood study for Causey and Pineview Dams, Weber Basin and Ogden River Projects, Utah. Bureau of Reclamation Seismotectonics Report 96-6, Denver, CO, 69 p.
Partridge, J., Baker, V.R., 1987. Palaeoflood hydrology of the Salt River, Arizona. Earth Surf. Process. Landforms 12, 109–125.
Pearman, J.L., Sedberry, F.C., Stricklin, V.E., Cole, P.W., 1991. Water Resources Data, Alabama, Water Year 1990. US Geological Survey Water Data-Report AL-90-1, 407 p.
Potter, K.W., Walker, J.F., 1985. An empirical study of flood measurement error. Water Resour. Res. 21 (3), 403–406.
Pruess, J., 1996. Paleoflood reconstructions within the Animas River basin upstream from Durango, Colorado. MS Thesis, Department of Earth Resources, Colorado State University, Fort Collins, CO, 192 p.
Rodier, J.A., Roche, M., 1984. World catalogue of maximum observed floods. IAHS-AISH Publ. No. 143, 354.
Salas, J.D., Smith, R.A., Tabios, G.Q., Heo, J.H., 2002. Statistical Computer Techniques in Water Resources and Environmental Engineering, Draft of Book, Department of Civil Engineering, Colorado State University, Fort Collins, CO.
Sanders, C.L., Kubik, H.E., Hoke, J.T., Kirby, W.H., 1990. Flood frequency of the Savannah River at Augusta, Georgia. US Geological Survey Water-Resources Investigations Report 90-4024, 87 p.
Sidler, W.A., 1968. Agua Mansa and the Flood of January 22, 1862, Santa Ana River. San Bernadino County Flood Control District, 10 p. and 2 figures.
Slack, J.R., Landwehr, J.M., 1993. Hydro-climatic data network (HCDN): a US Geological Survey streamflow data set for the United States for the study of climatic variations. US Geological Survey Water-Resources Investigations Report 93-4076 (CD-ROM).
Smith, C.F., Anning, D.W., Duet, N.R., Fisk, G.G., McCormack, H.F., Pope, G.L., Rigas, P.D., Wallace, B.L., 1995. Water Resources Data, Arizona, Water Year 1994. US Geological Survey Water Data-Report AZ-94-1, 320 p.
Stedinger, J.R., Baker, V.R., 1987. Surface water hydrology: historical and paleoflood information. Rev. Geophys. 25 (2), 119–124.
Stedinger, J.R., Cohn, T.A., 1986. Flood frequency analysis with historical and paleoflood information. Water Resour. Res. 22 (5), 273–286.
Stedinger, J.R., Surani, R., Therivel, R., 1988a. Max user's guide: a program for flood frequency analysis using systematic-record, historical, botanical, physical paleohydrologic and regional hydrologic information using maximum-likelihood techniques. Department of Environmental Engineering, Cornell University, Ithaca, NY, 51 p.
Stedinger, J.R., Therivel, R., Baker, V.R., 1988b. Flood frequency analysis with historical and paleoflood information, Salt and Verde Rivers, Arizona. Proceedings of the Eighth Annual Meeting of the US Committee on Large Dams, Salt River Project, Phoenix, Arizona, pp. 3.1–3.35.
Stewart, J.E., Bodhaine, G.L., 1961. Floods in the Skagit River Basin, Washington. US Geological Survey Water-Supply Paper 1527, 66 p.
Stewart, J.H., LaMarche, V.C., 1967. Erosion and deposition produced by the flood of December 1964 on Coffee Creek, Trinity County, California. US Geological Survey Professional Paper 422-K, 22 p.
Thomas, W.O. Jr., 1985. A uniform technique for flood frequency analysis. J. Water Resour. Plann. Mgmt., ASCE 111 (3), 321–337.
Thomson, M.T., Gannon, W.B., Thomas, M.P., Hayes, G.S., 1964. Historical floods in New England. US Geological Survey Water-Supply Paper 1779-M, 105.
UNESCO, 1976. World Catalogue of Very Large Floods, The UNESCO Press, 424 p.
US Geological Survey (USGS), 1959. Floods of June 1953 in northwestern Iowa. US Geological Survey Water-Supply Paper 1320-A, pp. 1–68.
Webb, R.H., Baker, V.R., 1987. Changes in hydrologic conditions related to large floods on the Escalante River, south-central Utah. In: Singh, V.P., (Ed.), Regional Flood Frequency Analysis, Reidel, Dordrecht, pp. 309–323.
Webb, R.H., Betancourt, J.L., 1992. Climatic variability and flood frequency of the Santa Cruz River, Pima County, Arizona. US Geological Survey Water-Supply Paper 2379, 40 p.
Webb, R.H., Jarrett, R.D., 2002. One-dimensional estimation techniques for discharges of paleofloods and historical floods. Am. Geophys. Union Water Sci. Appl. Ser. 5, 111–126.
Webb, R.H., O'Connor, J.E., Baker, V.R., 1988. Paleohydrologic reconstruction of flood frequency on the Escalante River, south-central Utah. In: Baker, V.R., Kochel, R.C., Patton, P.C. (Eds.), Flood Geomorphology, Wiley, New York, pp. 403–418.
Williams, G.P., Guy, H.P., 1973. Erosional and depositional aspects of Hurricane Camille in Virginia, 1969. US Geological Survey Professional Paper 804, 80 p.
Yanosky, T.M., Jarrett, R.D., 2002. Dendrochronologic evidence for the frequency and magnitude of paleofloods. Am. Geophys. Union Water Sci. Appl. Ser. 5, 77–89.