Effect of Regional Heterogeneity on Flood Frequency Estimation

DENNIS P. LETTENMAIER,1 JAMES R. WALLIS,2 AND ERIC F. WOOD3

Recent work on regional flood frequency estimation has shown that accurate flood quantile estimates are possible when the underlying flood frequency distributions are identical at all sites in the region except for a scaling factor; particularly when the underlying distribution has a two-parameter form. The class of regional probability-weighted moment (PWM) estimators is investigated for robustness to misspecification of the assumed distributional form and to regional heterogeneity in moments of order higher than one. Whereas two-parameter distributions belonging to the extreme value family perform quite well when the form of the underlying distribution is close to that of the fitted distribution, large biases can result when the distribution is misspecified. The three-parameter generalized extreme value distribution (GEV), when fitted using the regional PWM method, has been shown to be relatively insensitive to violations of the distributional assumption, and to have low variability and bias. In this paper it is shown that regional estimation methods using the three-parameter GEV distribution are relatively insensitive to modest regional heterogeneity in the coefficient of variation and quite insensitive to regional variation in the skew coefficient. The key determinant of the performance of the regional estimators is shown to be the regional mean coefficient of variation. For high values of the mean coefficient of variation, such as might be encountered in arid regions, an alternate PWM estimation method based on the GEV distribution that accommodates the regional heterogeneity in the higher order moments is preferred. The trade-off between this alternate method and the approach that assumes regional homogeneity in moments higher than order one is sensitive to the record lengths.

INTRODUCTION

The estimation of flood risk is a problem that has motivated a multitude of journal articles, reports and conference proceedings which reflect a wide variety of approaches ranging from esoteric mathematical formulations to highly structured institutional guidelines. Despite the volume of literature that has resulted from both theoretical and applied investigations, there remains no consensus as to how best to proceed. Much of the theoretical work cannot be applied in practice, either because data limitations are ignored, or because the mathematical solutions require unrealistic assumptions [Greis, 1983]. On the other hand, the assumptions invoked by the two best known institutional approaches, U.S. Water Resources Council Bulletin 17B [Interagency Advisory Committee on Water Data, 1982] and the British Flood Studies Report National Environmental Research Council (NERC) [NERC, 1975], have been shown to be inappropriate [Hoskinson et al., 1985b; Wallis and Wood, 1985].

Although the estimation of flood risk at first glance would seem a straightforward exercise in statistical inference, it is complicated by two key factors: (1) the lack of a physical basis for determining the form of the underlying flood frequency distribution and (2) the necessity of evaluating flood risk for return periods that exceed the length of the observed record. Flood risk estimates are highly dependent on the form of that portion of the underlying flood frequency distribution (the right tail) which is most difficult to estimate from observed data. Further, there is no strong theory on which to base an appeal regarding the distributional form of the right hand tail. It is essential therefore that any proposed method be robust for a wide range of floodlike distributions. Much of the past theoretical work, as well as current uniform flood frequency estimation procedures, is deficient in this respect. In this paper we compare the robustness of a generalized procedure for flood frequency estimation and make some observations about the factors influencing the selection of efficient, consistent, and accurate flood frequency estimators.

PROBLEM STATEMENT

Consider a region with m steam gages, each with n i years of record, i = 1, ..., m. At each site, it is desired to estimate the T year flood, that is, the value x rT, for which F r(x rT ) = Pr(x r ≤ x rT ). While x rT is usually taken to be the peak runoff (either instantaneous or averaged over a fixed period, such as one day), it could also represent the runoff volume associated with a given event. The probability distributions F r are usually taken to represent the probability that the largest event in a given year will exceed x (annual maxima), and the discussion here makes this implicit assumption.

It is widely recognized that a short record of data from a floodlike distribution, when plotted on a probability scale, can display a behavior which is quite different from the x rT - T relationship of the underlying distribution. For this reason, at-site estimates of the T year event can be quite unreliable, and the recent thrust in flood frequency research has tended toward the development and evaluation of regionally derived flood frequency estimates which seek to estimate the x rT simultaneously. Regional estimation techniques introduce several new problems, however. Ideally, if the m sites simply represented alternate realizations of the same underlying population, then a straightforward pooling approach would be appropriate. The so-called index flood method, variations of which have been suggested by Dalrymple [1960], NERC [1975], Greis and Wood [1981; 1983], and others, make such an assumption. In the index flood methods, the data at each site are normalized, usually by the at-site mean, then the parameters of a regional (normalized) flood frequency distribution are estimated. The estimation procedures vary, depending on the assumed form of the underlying distribution. However, once the parameters of the regional distribution have been estimated, the flood frequency distribution at each of the sites is assumed to be identical, except for a scale factor.
Therefore the index flood methods contain an implicit assumption that the region from which the $m$ sites are drawn is homogeneous. That is, all moments of order higher than one are assumed to be identical (when corrected for scale). Hydrologically, this assumption is almost certain not to hold, particularly when the size of the catchments in the region varies. Lettenmaier and Potter [1985] have explored the robustness of some regional estimators, including the index flood method of Greis and Wood [1981; 1983] to heterogeneity of the region in terms of the second moment (coefficient of variation $C_v$). They found that the performance of the Greis-Wood estimator was degraded as either the regional mean $C_v$, or the site-to-site variation in the $C_v$ increased. However, Lettenmaier and Potter did not consider the performance of more sophisticated index flood estimators, such as those proposed by Hosking et al. [1985].

While an assumption of regional homogeneity may not be justified, the trade-off between the increased information provided by regional pooling of flood data, and site-to-site variations, is not clear. Further, the performance of regional estimation approaches based on recently developed robust estimators [e.g., Hosking et al., 1985b] has not been fully explored. Therefore in this paper, it is our objective to (1) explore the robustness of selected regional flood methods with respect to the assumed form of the regional flood distribution; (2) explore the robustness of selected regional flood methods with respect to regional heterogeneity in the at-site flood distributions; (3) explore the sensitivity of selected regional flood methods to record length; and (4) explore the performance of regional flood methods that provide for site-to-site variations in moments higher that the first order.

**EXPERIMENTAL DESIGN**

Since the true form of the underlying distribution of floods is unknown, the only viable approach for assessing alternate estimation procedures is Monte Carlo sampling. Although split sample tests of observed flood sequences have been used in attempts to compare the performance of alternate flood frequency estimation procedures [e.g., Beard, 1974], split sample tests can only assess the consistency and not the precision (e.g., variance) or accuracy (e.g., bias) of alternate estimators, since the true flood risk is unknown. Monte Carlo simulation procedures, on the other hand, can estimate both accuracy and precision for given assumptions about the underlying probability distributions.

The two key elements of a Monte Carlo experiment are the data generation model and the test (fitting) methods. To provide useful results it is important that the data generation model produce simulated series that are plausible representations of a real process. Specifically, the simulated data must have floodlike properties. While the underlying flood distribution is unknown, there are certain properties that can be attributed to the data generating mechanism. Landwehr et al. [1978] analyzed a large number of annual flood series from 13 hydrologic regions of the United States. They found that the estimated regional mean $C_v$ ranged from 0.39 to 0.85 for samples of length 10, and from 0.42 to 0.99 for samples of length 30, the regional mean skewness estimates ranged from 0.53 to 1.12 for samples of length 10, and from 0.93 to 1.97 for samples of length 30. Since the sample (moment) estimate of $C_v$ has relatively small bias, the regional averages probably give a reasonable idea of the range of population $C_v$'s. For the skew coefficient $\gamma$, the situation is more difficult. Wallis et al. [1978] showed that there is substantial downward bias in the sample skew coefficient and that the magnitude of the bias is strongly dependent on the form of the underlying probability distribution, as well as the record length and the magnitude of the true skew and kurtosis $\lambda$. Therefore given that the underlying probability distribution is unknown, little can be said about the population skew coefficients either, except that since the bias is downward, the underlying regional mean skewness coefficients probably exceed the average sample estimates.

Although sample estimates of moments of orders higher than two are both too variable and too biased, even when averaged regionally, to allow much to be said on the basis of recorded flood series, it is possible to make some statements about the general forms that underlying distributions might assume on the basis of their theoretical skew-kurtosis relationship. The family of extreme value distributions is the asymptotic result of the sampling of the $n$ largest events from a large pool of random variates which are independently and identically distributed and whose distributional form meets certain restrictions. While flood events cannot be expected to meet the requirements of extreme value theory strictly (for instance, because the underlying random process of the daily runoff series is neither independent, nor identically distributed, and...
because each extreme event may not be drawn from a large enough sample of independent events for the asymptotic results to apply) the skew-kurtosis relationship of the family of extreme value distributions of maxima (also termed the generalized extreme value distribution (GEV) by Jenkinson [1955]) provides a useful point of departure. Perhaps the best known member of the GEV family is the extreme value I distribution, EV1, which has a fixed skew coefficient of 1.1396, kurtosis of 5.4 (marked with a bold cross on Figure 1), and probability distribution function \( F(x) = \exp \left(-\frac{(x-u)}{a}\right) \). The extreme value II distribution EV2

\[ F(x) = \exp \left(-\frac{(x-u)}{a}\right)^{1/\theta} \quad \theta > 0 \quad a > 0 \]

has a fixed upper bound \((u + a/\theta)\), with skewness-kurtosis combinations less than that of the EV1, and is of interest primarily as a limiting case, although it has been found to be applicable for catchments with large amounts of upstream flood storage [Acreman and Sinclair, 1986]. The extreme value III distribution EV3

\[ F(x) = \exp \left(-\frac{(1 - g(x-u)/a)}{a}\right) \quad g > 0 \quad a > 0 \]

is perhaps of greater applicability to flood hydrology. It assumes skewness-kurtosis values greater than the EV1.

Distributions with heavier tails than the GEV family will all lie above the GEV curve in Figure 1, as will the presymptotic distribution of the largest from a finite number of random variates, each drawn from a heavy-tailed or mixed distribution. The presymptotic distribution of the largest of a finite number of random variates drawn from light-tailed distributions, on the other hand, lie below the GEV curve. Skewness-kurtosis curves defining many of the distributions that have been suggested for use in flood hydrology lie below the GEV curve [Wallis et al., 1974]. A regional analysis of California hourly and daily point rainfall (Figure 1) showed that all points plot in the vicinity of the theoretical GEV curve. Further, had bias corrections been attempted it would be expected that the points on Figure 1 would have plotted above and slightly to the right of their present positions and would, in fact, be either on or slightly above the GEV curve [Wallis, 1982].

The Wakeby distribution [Houghton, 1978a, b] is sufficiently flexible (five parameters) to assume values over much, but not all, of the permissible skewness-kurtosis domain (Figure 2). The Wakeby distribution might have been used as the underlying flood generator in this study. However, the logistical complications introduced by such a generalization were not felt to be warranted, and the GEV skewness-kurtosis curve was considered to represent a sufficient diversity of possible hydrologies. If desired, however, the experiments could be generalized over the feasible (Wakeby) skewness-kurtosis domain. A partial investigation of this extended domain has been made [Hosking et al., 1985b] with results that are consistent with those presented in Figure 2.

The second key element of a Monte Carlo study such as this concerns the choice of fitting methods. Although it is infeasible to explore all at-site and regional methods that have been proposed for flood frequency estimation, it is possible to dismiss a number of the methods based on the results of previous work. For instance, both the United States Interagency Advisory Committee on Water Data [Bulletin 17B, 1982] and British [NERC, 1975] institutional methods, variations of which are used in much of the world, have been shown to perform poorly in comparison with index flood methods based on probability-weighted moment (PWM) estimators of the GEV and Wakeby distributions [Wallis and Wood, 1985; Hosking et al., 1985b]. Further, Lettenmaier and Potter [1985] have shown that the empirical Bayes regionalization approach suggested by Kuczera [1982], which allows for site-to-site variations in the parameters of the two-parameter lognormal distribution, is quite sensitive to violations of the distributional assumption. Likewise, methods based on normalizing transformations of the raw data have rarely been found to give satisfactory estimates of flood quantiles and such methods are not explored here.

The flood risk estimation procedures to be tested are all based on PWM parameter estimators, as introduced by Greenwood et al. [1979]. PWM's are defined generally as

\[
M_{\text{PWM}} = \frac{\mathbb{E}[X^p F(X)]}{1 - \mathbb{E}[F(X)]^p}
\]

where \( F(X) \) is the probability distribution function of \( X \). The special cases \( M_{\text{PWM}}^1 = M_{\text{PWM}} \), and \( M_{\text{PWM}}^2 = M_{\text{PWM}}^3 \), are of particular interest for parameter estimation. For those distributions for which \( F(x) \) can be inverted explicitly so that \( x = x(F) \), parameter estimates can be derived in terms of either \( M_{\text{PWM}}^1 \) or \( M_{\text{PWM}}^2 \) by substituting a sample estimate of \( M_{\text{PWM}}^1 \) or \( M_{\text{PWM}}^2 \) for its popu-
lation value in the theoretical relationship between the PWM's and the parameters of the selected distribution. Various sample estimates of the PWM's might be used. Based on the results of Landwehr et al. [1979b] the estimator used here is $M_{j*} = 1/n \sum_{i=1}^{n} (P_i/M_i)$ and $M_{k*} = 1/n \sum_{i=1}^{n} (1 - P_i)/x_i$, where $P_i$ is the sample percentage point of the $i$th largest event $X_i$ in a sample of size $n$ and is defined by $P_i = (i - 0.35)/n$. Subsequent estimators are expressed in terms of $M_{j*}$ and $M_{k*}$, and the superscript is dropped for convenience.

For the GEV distribution

$$F(x) = \exp \left[ -\left( 1 - \frac{g(x - u)}{a} \right)^{1/g} \right] \quad g \neq 0 \quad (2)$$

$$F(x) = \exp \left[ -\exp \left( -\frac{x - u}{a} \right) \right] \quad g = 0$$

with $x$ bounded by $u + a/g$ from above if $g > 0$ and from below if $g < 0$. Here $u$ and $a$ are location and scale parameters, respectively, and $g$ is the shape parameter. Closed-form solutions for the parameters of the GEV distribution can most conveniently be derived in terms of $M_j$, $j = 0, 1, 2$ as described by Hosking et al. [1985a]. For the special case $g = 0$, the GEV distribution becomes the EV1. Greenwood et al. [1979] derived the relationship between the parameters $a$, $u$, and $M_0$, and $M_1$ for an EV1 distribution. For small to moderate sample sizes, Greis and Wood [1981, 1983] have shown that the root-mean-squared error of estimated flood quantiles for return periods of typical hydrologic interest is slightly less when parameters are estimated using PWM's as compared with maximum likelihood.

Two classes of estimators are considered here: at-site and regional. In the experiments reported later, the at-site estimators were applied independently to each site, while the regional estimators, all of which are based on an index flood approach, made use of the data from all the sites simultaneously as described below. The estimators considered were as follows.

1. EV1/AS, an at-site estimate of the EV1 distribution using PWM estimators as described above, resulting in $a = (M_{0*} - 2M_{1*})/\ln(2); u = M_{0*} - 0.57721a$.
2. GEV/AS, an at-site estimate of the GEV distribution also using PWM estimators. As shown by Hosking et al. [1985b] at-site estimates of the GEV distribution can be obtained using

$$c = (2M_{1*} - M_{0*})/(3M_{2*} - M_{0*}) - \ln(2)/\ln(3) \quad (3a)$$

$$g = 7.8590c - 2.9554c^2 \quad (3b)$$

$$a = (2M_{1*} - M_{0*})g/(\Gamma(1 + g)(1 - 2^{-g})) \quad (3c)$$

$$u = M_{0*} - a(\Gamma(1 + g) - 1)/g \quad (3d)$$

3. EV1/R, a regional EV1 estimator based on the PWM's. This estimator is computed as follows. First, the at-site means $M_{0*}$ are estimated at all $m$ sites, and the data at each site are normalized by their mean. Then, $M_{1*}$ is computed at each site, and a weighted regional average of the $M_{1*}$ for the normalized sequences is computed as

$$M_{1r*} = \frac{\sum \sum n_i}{\sum n_i} M_{1r*} \quad (4)$$

From $M_{0r*} = 1.0$ and $M_{1r*}$, the parameters $a$, and $u$, of a regional flood frequency curve are computed, using the same relationships as for the at-site parameter estimates described in estimator 1 above. Then, the flood frequency distribution at each site is estimated as $F_i(x) = F_i(x/M_{0r*})$, where $F_i$ is the regional normalized flood frequency distribution based on $a$, and $u$. This procedure is essentially the same as that used by Greis and Wood [1981, 1983], except that they assume $M_{0r*} = 1.0, \forall i$; therefore their results do not reflect the uncertainties in flood quantile estimates that result from scaling the regional flood frequency estimates by the at-site means $M_{0*}$.

As our results will show, this is a critical source of uncertainty, particularly for regions where the mean $Cv$ is large.

4. EV2/R, a regional EV2 estimator with the skewness fixed at 3.0. This estimator is most easily expressed by considering the EV2 as a special case of the GEV distribution. The regional estimation procedure is identical to method 3, except that the relationship between the parameters $a$, and $u$, and the normalized regional PWM's $M_{0r*} = 1.0$, and $M_{1r*}$ is based on the properties of the GEV distribution (see method 6) rather than the EV1 distribution. The third regional parameter $g$, is determined by the unique relationship between the assumed regional skew of 3.0, and $g$ and results in $g = -0.1769$. Subsequently, the flood frequency distributions at the individual sites are estimated as $F_i(x) = F_i(x/M_{0r*})$.

Method 4 can be considered to be a near equivalent of the regional exponential distribution methodology of Damazio and Kelman [1982].

5. EV3/R, a regional PWM estimator for the EV3 distribution, with the skewness fixed at 0.9. This estimator is identical to method 4, except that $g = 0.0435$, is used at all sites within the region rather than being calculated from the sample data as in method 6. This approach might be applicable to very large catchments or catchments with substantial upstream flood storage.

6. GEV-1, a regional PWM estimator based on the GEV distribution. This method is similar to methods 4 and 5, except that $M_{j*}$, $j = 0, 1, 2$ are computed, rather than just
Then, the regional parameters \(a, u, \) and \(g\) are computed using the three regional PWM's. Flood frequency distributions at the individual sites are estimated by scaling by the at-site means, as in methods 3–5.

7. WAK/R, a regional PWM estimator for the Wakeby distribution [Wallis, 1980, 1981]. This method is similar to method 6, except that since the Wakeby has five parameters, PWM’s up through order four must be estimated. Further, it is more convenient to use \(M_{0j}^*\) rather than the \(M_j^*\), as was shown by Greenwood et al. [1979]. The explicit relationships between the \(M_j^*\) and the parameters of the distribution are quite complicated, and the reader is referred to Greenwood et al. [1979a, b] for their explicit form. Had a flood generator other than GEV been used, it is to be expected that WAK/R would outperform GEV-1. However, because the generated data were always GEV the results for this estimator were very similar to those for the GEV-1 estimator, a finding previously reported [Hosking et al., 1985b]. Therefore although all the numerical experiments included this estimator, the results are not reported here, since they were entirely consistent with the earlier work.

8. GEV-2, a regional PWM estimator based on the GEV distribution, where only the third parameter \(g\) is based on the regional estimate \(g_j\), which, in turn, is based on \(M_{0j}^*\), \(j = 0, 1, 2\). The parameters \(a\) and \(u\) are based on the at-site PWM’s \(M_{0j}^*\) and \(M_{1j}^*\). Since the third parameter \(g\) is invariant with respect to scaling, there is no necessity to scale by the regional mean, and the estimated flood frequency distributions at each site are based on \(a_j, u_j,\) and \(g_j\) that is, the at-site estimates of \(a\) and \(u\), and the regional estimate of \(g\). Therefore there is no implicit assumption in this method that the flood series at the sites, when scaled by the mean, are homogeneous, but rather only that the third parameter (hence the skew coefficient at each site) can be represented by a regional average.

\[ \text{Estimated bias (dashed curve) and upper and lower confidence limits (dotted curves) at site 11 of 21 site region for two-parameter regional PWM fitting methods EV1/R, EV2/R, and EV3/R applied to EV1, EV2, and EV3 parents. EV2/R and EV3/R fitting methods assume fixed skew coefficients of } 3.0 \text{ and } 0.9, \text{ respectively. Sample sizes and parent coefficients of variation and skew coefficients (EV2) as given in Figure 3; EV3 parent skew coefficients varied from 1.0 (site 1) to 0.75 (site 21).} \]

**DATA GENERATION**

Several flood populations were considered. A region with 21 sites was hypothesized. Three underlying flood distributions were considered: EV1, EV2, and EV3. Population skewnesses, record lengths, and CV’s varied by site. Record lengths ranged uniformly from 10 at site 1 to 30 at site 21. The population skew coefficient for the EV1 worlds was constrained by the distributional form to 1.1396; for the EV2 and EV3 distributions, the variation of skew coefficient by site is shown in Figure 3 for the case of a regional median, \(M(CV) = 0.5\), and a regional range in \(Cv, R(CV) = 0.15\). For convenience, the regional range can be normalized as \(R^*(CV) = R(CV)/M(CV)\) and throughout this paper the variation in \(CV\) within a region will be defined by \(M(CV)\) and \(R^*(CV)\). For the EV2 worlds, the shortest record lengths are associated with the highest skew coefficients. In most real cases, high skewnesses are associated with small catchments (since a greater fraction of the area is likely to be affected by extreme precipitation events). Small catchments, in turn, tend to have shorter record lengths. For the EV3 worlds, however, the situation is reversed; the sites with the shortest records have the smallest skew coefficients. While this situation is unusual, it has been observed in some regions with large numbers of lakes (see, for example, Acreman and Sinclair [1986]). The generated floods were independent in time and space. The temporal dependence of annual flood series is usually a minor concern in flood frequency estimation. Although an argument can be made in some cases for positive correlation of annual flood series based on year-to-year carryover of basin water storage, this dependence is usually so weak as to have negligible effect on the quantile estimates compared to the other factors investigated. Spatial correlation is potentially a greater concern. Because large floods may be derived from the same storm event, particularly in regions where the annual flood maxima are the result of frontal storms, significant spatial correlations may be present in annual flood series. Various formulas have been proposed to estimate the effect of spatial correlation on regional flood quantile estimation procedures [e.g., Stedinger, 1983]. Hosking and Wallis [1985] have shown that the relationship between intersite correlation and regionally derived quantile estimates is more complex than these formulas imply. Further, while these formulas may represent the approximate loss of information for regional flood frequency distributions \(F_n\), they are likely to result in severe over estimates of the loss of information for individual site estimates \(F_n(x) = F_n(x/M_{0n}^*)\). For this reason, the investigations reported here are based upon spatially as well as temporally independent series.

The distribution of \(CV\) over the sites reflects the degree of heterogeneity of the sites within the region (recall that for fitting methods 3–7, the sites are assumed homogeneous, and these methods can be expected to perform at their best when this assumption is not violated). For the base runs, two values of \(M(CV), 0.5 \text{ and } 1.0, \text{ and two values of } R^*(CV), 0.3 \text{ and } 0.5, \text{ were considered. For each set } (M(CV), R^*(CV)), \text{ the population } CV \text{ for site } 11 \text{ was set to } M(CV), \text{ and the corresponding parameters } a_{11} \text{ and } u_{11} \text{ were determined. Next, the site } 1 \text{ CV was fixed at } M(CV)(1 + R^*(CV))/2, \text{ and } a_1 \text{ and } u_1 \text{ were determined. Likewise, the site } 21 \text{ CV was fixed at } M(CV)(1 - R^*(CV))/2, \text{ and}
Fig. 5. Root-mean-squared error (rmse) and bias as function of $R^*(Cv)$ and $M(Cv)$ for site 11 of 21 site region. Population is EV2; sample size and skew variation by site as given in Figure 3.

$a_{21}$ and $u_{21}$ were determined. The parameters $a$ and $u$ for the remaining sites were determined by linearly interpolating between the parameters for sites 1, 11, and 21.

In addition to the resulting 12 base runs ($3$ distributions $\times 2M(Cv)\times 2R^*(Cv)$) additional runs were made to determine the sensitivity of the results to $M(Cv)$. For these runs, the EV2 generator was used, with $R^*(Cv)=0.3$ and $0.5$ and $M(Cv)=0.5, 0.75, 1.0, 1.5,$ and $2.0$. Three other runs were made for $M(Cv)=0.5, R^*(Cv)=0.3$ with $u$ constant for all sites and varying over the three runs as $(20, 50, 100)$.

In all cases, biases and normalized root-mean-squared errors were estimated at each site and for each estimation method. Biases were estimated as

$$\frac{1}{N} \sum_{p=1}^{N} \frac{x_{tpq} - x_{rq}}{x_{rq}}$$  \hspace{1cm} (5)

where $p$ is the Monte Carlo simulation index, $q$ is the site number ($1 \leq q \leq 21$), $r$ is the estimation method ($1 \leq r \leq 8$), and $x_{rq}$ is the true flood quantile at site $q$. Normalized root-mean-squared errors were estimated as

$$\left[ \frac{1}{N} \sum_{p=1}^{N} (x_{tpq} - x_{rq})^2 \right]^{1/2}$$  \hspace{1cm} (6)

In most cases, the number of Monte Carlo sequences $N$ was 50,000. The exception was that in cases where any of the at-site estimated means $M_{ij}$ were less than 0.1 of the population mean, such cases were eliminated from the simulations. These cases only occurred at high values of the regional mean $Cv$ (e.g., $>1.5$) and resulted from a high percentage of negative "floods" being generated at one or more of the sites. This procedure avoided the instability that otherwise occurred in the Monte Carlo estimates of bias and variability. It should be noted that this is primarily a computational issue that occurs...
RESULTS

The relative performance of the estimation methods depends on (1) the distribution of sample size among the sites (2) $M(C_v)$, (3) $R(C_v)$, (4) the return period $T$, and (5) the parent distribution. Clearly, the combinatorial problem would become overwhelming if an exhaustive analysis were attempted, and instead, it is necessary to examine cross-sections of the sample space, where all but one or two factors are held constant. We have elected to show five sets of results (indexed in Table 1), as follows.

1. The estimated expected value and two standard deviation confidence bounds of the quartile estimates for fixed $M(C_v)$ and $R(C_v)$ (0.5 and 0.3, respectively) for the 11th site in a 21-site region, for nine parent-fitting method combinations (EV1, EV2, and EV3 parent distributions each with the regional two-parameter EV1/R, EV2/R, and EV3/R estimators). This set of results is intended to show the trade-off between bias and root-mean-squared error (rmse) that occurs when relatively inflexible two-parameter estimation methods are used and the shape of the fitted distribution deviates from the parent.

2. The estimated root-mean-squared error and bias as a function of return period for the 11th site in a 21-site region for two-parameter EV1 and three-parameter GEV at-site estimators, EV1/AS and GEV/AS, and a three-parameter regional estimator, GEV-1, when the parent is EV2 and $M(C_v)$ varies from 0.5 to 1.5 with $R(C_v)$ of 0.3 and 0.5. This set of results is intended to show the effect of heterogeneity in the basin mix upon the relative performance for selected regional and at-site methods.

3. The estimated root-mean-squared error and bias at site 11 of a 21-site region as a function of $M(C_v)$ for $R(C_v) = 0.3$ and 0.5; and as a function of $R(C_v)$ for $M(C_v) = 0.5$ and 1.0, for the $T = 100$ year flood with and EV2 parent distribution, and GEV/AS, and two GEV regional fitting methods, GEV-1 and GEV-2. These results allow exploration of the relative accuracy of fitting methods as a function of the average site variability and regional heterogeneity.

4. The estimated root-mean-squared error and bias for each site in a 21-site region for $M(C_v) = 0.5, 1.0,$ and 1.5; for $R(C_v)$ of 0.3 and 0.5, for the 100-year flood with at-site and two regional GEV fitting methods. This set of results is intended to show the variation in the statistical performance of the same three estimators used in sets II and III across the sites in regions of varying heterogeneity and variability.

5. The estimated root-mean-squared error and bias at site 11 of a 21-site region as a function of record lengths for the $T = 100$ year flood and the same fitting methods used in II, III, and IV, with an EV2 parent, for the four combinations $M(C_v) = 0.5$ and 1.0 and $R(C_v) = 0.3$ and 0.5. This set of results is intended to show the effect of sample size on the performance of the at-site and regional GEV fitting methods for a similar range in variability and regional heterogeneity investigated in sets II, III, and IV.

Plots of the results are given in Figures 4–8 corresponding to sets I–V, respectively. From Figure 4, which shows all combinations of EV1, 2, and 3 worlds with two-parameter regional fitting methods based on the EV1, EV2, and EV3 distribution,
Fig. 7. Root-mean-squared error (rmse) as function of site $M(C_v)$ and $R'(C_v)$ for $T = 100$ and EV2 population. Sample size and skew coefficient as given in Figure 3 (left column); results in right column are for same population with all sample sizes increased by 30.
the principal conclusion to be drawn is that two-parameter fitting methods perform well only when the tails of the parent distribution are similar to those of the distribution assumed by the fitting method; that is, they have poor robustness. In those cases where the assumptions of the fitting method are met, the two-parameter methods have relatively low rmse and low bias, and the scaled rmse for these methods tends to stay constant or decrease as the return period increases (the scaled rmse for three-parameter methods usually increases with return period; see, for example, Hosking et al. [1985b]). However, all of the two-parameter methods investigated suffer from large biases when the parent distribution differs from the fitted distribution, and the bias tends to increase in absolute value for return periods greater than the average record length.

Figure 5 (result set II) shows the relative rmse's and biases of the two parameter, EV1/AS, and a three-parameter at-site, GEV/AS, and the three-parameter regional estimation method GEV-1. The latter two methods therefore assumed the correct distributional form as the generator was EV2, a member of the GEV family of distributions. It is clear from the results that the rmse of the three-parameter at-site fitting method is much larger than its competitors when $M(Cv)$ and $R*(Cv)$ are modest. The EV1 at-site method performs quite well in terms of rmse, especially for the high $M(Cv)$ cases, but, for the reasons discussed in the assessment of the set I results, tends to have large downward biases. The regional GEV fitting method is clearly preferred for all but extremely variable and heterogeneous regions. This result is consistent with Lettenmaier and Foster [1985], who found, for a more limited set of fitting methods, that the performance of an index flood method (equivalent to method 3 here) was degraded for large $M(Cv)$. This result can be attributed to the domination of the rmse by the errors associated with the at-site mean estimate when $M(Cv)$ is large.

Figure 6 further investigates the effect of the mean variability and regional heterogeneity, focusing on three estimation methods that were found to offer the best performances across the $M(Cv) - R*(Cv)$ range investigated. All three methods use three-parameter PWM estimators based on the GEV distribution. They include the index flood regional estimator (GEV-1, method 6); a modified PWM method that estimates only the shape parameter of the GEV distribution from the regional PWM's, with zeroth and first order at-site PWM's used to estimate the remaining parameters (GEV-2, method 8); and the at-site estimator (GEV/AS, method 2). The results show that the index flood estimator performs best over most of the $M(Cv)$ and $R*(Cv)$ combinations considered. The modified index flood estimator is motivated by the premise that heterogeneity in the high-order moments (e.g., skewness) is relatively unimportant if heterogeneity in the low-order moments ($Cv$) is accommodated. This estimator is, as expected, preferred for high $M(Cv)$, although it should be noted that $Cv$'s in the range for which this method is preferred are hydrologically rare and occur primarily in arid climates. The at-site estimator performs poorly except for the most extreme $M(Cv)$ and $R*(Cv)$ values. This is also to be expected, given the set I results.

Figure 7 (result set IV) shows the variation of the bias and rmse by site for the same three fitting methods used in set III for a range of $M(Cv)$ and $R*(Cv)$. The results for bias are especially interesting. Except in extreme cases, the rmse tends to be dominated by the variance of the estimates, so the biasing at the extreme sites (that is, those sites with underlying distributions that are much different than the median site) tends not to be evidenced in either the rmse's and biases for

\begin{align*}
M(Cv) &= 1.0, \quad R'(Cv) = 0.3 \\
M(Cv) &= 0.5, \quad R'(Cv) = 0.3 \\
M(Cv) &= 1.0, \quad R'(Cv) = 0.5 \\
M(Cv) &= 0.5, \quad R'(Cv) = 0.5
\end{align*}
the median site or for averages over all sites, on which result sets I, II, and III are based. What is clear from Figure 7 is that for highly heterogeneous regions, severe biases can occur. Clearly, one solution is to select regions that are sufficiently homogeneous that such problems can be minimized. However, in an actual situation the problem may be complicated somewhat by the necessity to include enough sites in a region so that the advantages of regionalization can be enjoyed. While it is easy to develop screening methods based on sample statistics for classifying regions, and in many cases such approaches will suffice, more work remains to be done to develop schemes for site classification when the data base is highly heterogeneous.

Figure 8 (result set 5) shows the relative performance of the three fitting methods investigated in result sets II-IV for varying sample size and a representative range of \( M(Cv) \) and \( R*(Cv) \). As the sample size increases, it becomes more advantageous to use the GEV-2 estimator (method 8), which employs at-site PWM's of order zero and one and regionalizes only the shape parameter of the GEV. This occurs because the accuracy of the at-site PWM's improves with sample size, and the heterogeneity that is ignored by the GEV-1 (which implicitly assumes a homogeneous region) begins to dominate the rmse of this method. This implies that for moderate to large sample sizes, it is possible to make use of the modified method which accommodates some of the regional heterogeneity; hence the importance of choosing relatively homogeneous regions is somewhat reduced as the sample size increases. This result is perhaps of less importance in evaluating the methods than for illustrating the relative importance of the variables controlling the rmse's and biases. In the vast majority of practical applications, average regional sample sizes will be in the range 20–40, and result sets I-IV should provide a reasonable basis for assessing the relative performance of the methods.

CONCLUSIONS

The experiments described herein were designed to explore the robustness of selected regional and at-site flood frequency estimation procedures with respect to (1) the underlying flood distribution, (2) regional heterogeneity in the moments of the underlying flood distribution, (3) variations in record length over the region, and (4) to explore the performance of regional flood frequency estimation methods that provide for site-to-site variations in moments higher than the first order. The results of the experiments have shown that two-parameter regional flood frequency estimation methods can perform quite well if the assumed distributional form is similar to the population, but that large biases may result if the distributional form is misspecified. The three-parameter generalized extreme value distribution tends to give excessively variable flood quantile estimates for at-site applications. However, when incorporated in a regional estimation scheme, it performs nearly as well as the two parameter estimators in terms of root-mean-error, and much better in terms of bias. In addition, the regional index flood PWM estimation methods using the GEV distribution are relatively insensitive to modest regional heterogeneity in the coefficient of variation, when performance is measured in terms of the regional average root-mean-square error. The relative insensitivity of the aggregate or regional average root mean square error does obscure, to some extent, biases that occur at sites that are most dissimilar from the regional average. A more important determinant of the relative performance of the estimators, though, is the regional mean coefficient of variation \( M(Cv) \). For high values of \( M(Cv) \) the advantage of the methods that assume regional homogeneity in moments above the first order is reduced, and at very high \( M(Cv) \)'s, such as might be encountered in arid regions, alternate estimators that accommodate the regional heterogeneity in the higher moments are preferred. Finally, the trade-off between the index flood PWM estimators, which assume regional homogeneity, and the alternate methods that allow for regional variation in moments above the first order moves toward lower values of \( M(Cv) \) as the regional average record length is increased. However, within the range of record lengths that are commonly available, record length dependence is likely to be less important than geographic differences in \( M(Cv) \).

REFERENCES


Beard, L. R., Flood Flow Frequency Techniques, Center for Research in Water Resources, The University of Texas at Austin, 1974.


D. P. Lettenmaier, Department of Civil Engineering, FX-10, University of Washington, Seattle, Washington 98195.

J. R. Wallis, Mathematical Sciences Department, IBM Research, Yorktown Heights, NY 10598.

E. F. Wood, Department of Civil Engineering, Princeton University, Princeton, NJ 08544.

(Received February 10, 1986; revised September 26, 1986; accepted October 1, 1986.)