WATER RESOURCES RESEARCH, VOL. 35, NO. 5, PAGES 1551–1557, MAY 1999

Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference

George Kuczera

Department of Civil, Surveying and Environmental Engineering, University of Newcastle
Callaghan, New South Wales, Australia

Abstract. In flood frequency applications where the design flood is required to have a specified exceedance probability, expected probability should be used. Its computation, however, presents formidable difficulties. This study presents a Monte Carlo Bayesian method for computing the expected probability distribution as well as quantile confidence limits for any flood frequency distribution using data on gauged flows, possibly corrupted by rating curve error, and on censored flows. This is achieved by a three-step process: (1) Formulate the likelihood function for the given data; (2) approximate the likelihood function using a multinormal distribution; and (3) integrate the expected probability integral using importance sampling. The FLIKE software for performing this is described, and an example is given.

1. Introduction

The primary objective of a flood frequency analysis, as the name suggests, is to provide an estimate of the exceedance probability for a given discharge. Curiously, much of the published research on flood frequency estimation methods has focused on providing accurate estimates of quantiles, namely, discharges corresponding to given exceedance probabilities.

In many, if not most, flood frequency applications, accurate estimates of exceedance probabilities rather than quantiles are required. *Stedinger* [1983, p. 520] succinctly summarized the situation as follows: "If a design event is to be a flood flow which will be exceeded with probability $1 - p$ (usually 1%), then one wants to determine the design event value which, given the available hydrologic information, will be exceeded with probability $1 - p$." *Beard* [1960] introduced the expected probability distribution which provides optimal (in a mean-squared-error sense) estimates of exceedance probabilities. He stressed the importance of expected probability estimation in flood damage assessment, noting that "average annual damages are linearly related to frequency" [*Beard*, 1987, p. 204]. Moreover, Australian Rainfall Runoff (ARR) [*Pilgrim*, 1987] suggests that in most situations expected probability be used in preference to quantile-based procedures.

Despite the relevance of the expected probability distribution, only limited progress has been made in developing estimation procedures for the probability models used in flood frequency analysis. The only distribution for which a rigorous expected probability result is available is the two-parameter lognormal. The expected probability method for the log-Pearson III distribution recommended in ARR [*Pilgrim*, 1987] is inadequate for two reasons: (1) It assumes the true log skew is known, and (2) it is only applicable to gauged data. A similar situation exists with the computation of confidence limits. Many approximations exist, but no general solution has yet been developed.

Although motivated by the need to improve expected probability estimation, this study in fact presents a comprehensive solution to the at-site parametric flood frequency problem. Building on previous developments in Bayesian flood frequency analysis and recent advances in Bayesian computational methods, it shows how expected probability distributions and quantile confidence limits can be derived using any probability distribution with virtually any kind of flood data, including gauged data affected by systematic rating error.

The main contribution of this study arises from the use of importance sampling, a Monte Carlo technique for sampling from a probability distribution. *Gelman et al.* [1997] provide a good overview of recent advances in Bayesian Monte Carlo techniques, including importance sampling, and demonstrate how the full power of the Bayesian paradigm can be applied to a wide range of difficult problems. The use of importance sampling enables virtually full application of the Bayesian paradigm to the at-site parametric flood frequency problem.

This paper is organized as follows: Following a brief overview of Bayesian at-site flood estimation, the expected probability distribution is formally defined, after which Monte Carlo importance sampling is introduced as a general method for computing the expected probability distribution and quantile confidence limits. An example and description of the FLIKE software conclude this study.

2. Bayesian Flood Estimation

The application of Bayesian methods to flood frequency estimation goes back more than two decades, with works such as those by *Wood and Rodriguez-Iturbe* [1975a, b] and *Vicens et al.* [1975] defining the principal elements of Bayesian flood frequency analysis. These and later works have shown that the Bayesian approach provides a general inference procedure for at-site flood frequency analysis using any probability model. Despite this, the Bayesian approach has seen only limited application in practice partly because it is poorly understood and, more significantly, because its implementation is typically numerically intensive when compared with the simpler classical methods such as described by the *Interagency Advisory Committee on Water Data* [1982] and in ARR [*Pilgrim*, 1987]. How-

Paper number 1999WR900012.
0043-1397/99/1999WR900012$09.00

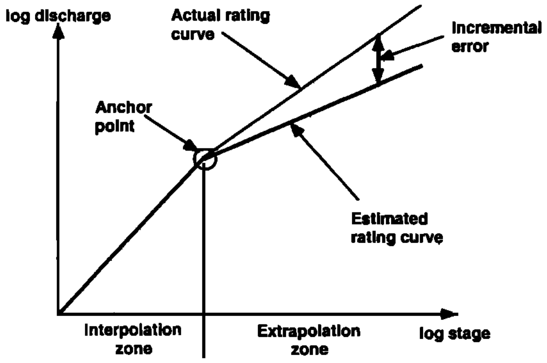

Figure 1. Rating curve extension error.

ever, with the widespread use of computers in hydrology, this obstacle is no longer relevant.

The Bayesian framework uses a probability distribution to describe what is known about the quantity of interest. In our case, interest focuses on the parameters of the flood probability model. The probability distribution for these parameters describes how precisely or imprecisely we know the true value of the parameters. The shape of this probability distribution is determined by the data relevant to the parameters. Following *Gelman et al.* [1997, p. 8], the centerpiece of Bayesian inference is the "basic property of conditional probability known as Bayes' rule," which expresses the posterior probability density of the parameters $\boldsymbol{\beta}$ given the at-site data $\mathbf{D}$ as

$$\xi(\boldsymbol{\beta}|\mathbf{D}) \propto f(\mathbf{D}|\boldsymbol{\beta})\xi(\boldsymbol{\beta}) \quad (1)$$

where $\xi(\boldsymbol{\beta})$, the prior probability density of the parameters, describes what is known about the flood model parameters prior to analysis of the data $\mathbf{D}$ (it may represent a noninformative situation or may represent prior knowledge arising from a regional analysis of flood data) and $f(\mathbf{D}|\boldsymbol{\beta})$ is the sampling distribution of the data given a chosen probability model and parameters $\boldsymbol{\beta}$. In application of Bayes' rule the data $\mathbf{D}$ are fixed, meaning they have been observed. In this context the function $f(\mathbf{D}|\boldsymbol{\beta})$ is regarded as a function of $\boldsymbol{\beta}$ and, to emphasize this point, is called the likelihood function of the parameters $\boldsymbol{\beta}$ given the data $\mathbf{D}$.

2.1. Flood Data

The annual maximum flood is defined as the largest discharge to occur during the year. The following data on floods at a particular site may be available:

1. Gauged data consist of a time series of flood discharge estimates. Usually, such estimates are based on observed peak stages (or water levels). A rating curve is used to transform stage observations to discharge estimates. When extrapolated, the rating curve can introduce large systematic error into discharge estimates.

2. Censored data consist of the number of floods above or below a threshold discharge. These arise in a number of ways. For example, prior to gauging, water level records may have been kept only for large floods above some perception threshold. Therefore all that may be known is that x floods were above this threshold. Sometimes, small floods below some threshold may be deliberately excluded because the fit to the right tail of flood distribution is being unduly influenced by the small floods.

We consider two formulations of the likelihood function. The first assumes there is no error in the flood data. The focus is on the contribution to the likelihood function made by gauged and censored data. The second generalizes the likelihood function to allow for error in discharge estimates.

2.2. Likelihood Function

Suppose the flood frequency model is described by the probability density function (pdf) $f(q|\boldsymbol{\beta})$, where q is an annual maximum discharge and $\boldsymbol{\beta}$ is a vector of parameters about which an inference is sought. Each flood discharge q is assumed statistically independent. Now suppose the following data are available: (1) a gauged record of n true flood peaks $\{q_1, \cdots, q_n\}$ and (2) m censored records in which u_i annual flood peaks in v_i years exceeded a threshold with true discharge Q_i, $i = 1, \cdots, m$. Denote the flood data as $\mathbf{D} = \{\mathbf{G}, \mathbf{C}\}$, where $\mathbf{G} = \{q_i, i = 1, \cdots, n\}$ and $\mathbf{C} = \{(u_i, v_i, Q_i), i = 1, \cdots, m\}$.

The sampling distribution of the gauged data $\mathbf{G}$ is the joint pdf of the n gauged floods. Given the statistical independence of annual maximum discharges, it follows that

$$f(\mathbf{G}|\boldsymbol{\beta}) = f(q_1, q_2, \cdots, q_n|\boldsymbol{\beta}) = \prod_{i=1}^{n} f(q_i|\boldsymbol{\beta}) \quad (2)$$

To derive the sampling distribution of the censored data, we note that the probability of observing u exceedances in v years given p is the probability of an exceedance is defined by the binomial probability mass function

$$f(u|v, p) = \binom{v}{u}(1-p)^{v-u}p^u, \qquad u = 0, 1, \cdots, v \quad (3)$$

Hence the sampling distribution of the censored data $\mathbf{C}$ becomes, upon invoking statistical independence of annual maxima [*Stedinger and Cohn*, 1986],

$$f(\mathbf{C}|\boldsymbol{\beta}) = \prod_{i=1}^{m} \binom{v_i}{u_i}[1 - F(Q_i|\boldsymbol{\beta})]^{u_i}F(Q_i|\boldsymbol{\beta})^{v_i-u_i} \quad (4)$$

where $1 - F(Q_i|\boldsymbol{\beta})$ is the probability of exceeding the threshold Q_i in any year, with $F(Q_i|\boldsymbol{\beta})$ defined as $P(q \leq Q_i|\boldsymbol{\beta})$.

The likelihood function of the parameter vector $\boldsymbol{\beta}$ given the gauged and censored data is

$$f(\mathbf{G}, \mathbf{C}|\boldsymbol{\beta}) = f(\mathbf{G}|\boldsymbol{\beta}, \mathbf{C})f(\mathbf{C}|\boldsymbol{\beta}) = f(\mathbf{G}|\boldsymbol{\beta})f(\mathbf{C}|\boldsymbol{\beta}) \quad (5)$$

This likelihood function assumes the discharges q and the threshold discharges Q are measured without error. This usually is far from the truth, especially for the biggest floods on record. *Potter and Walker* [1981, 1985] observe that flood discharge is inferred from a rating curve which is subject to discontinuous measurement error.

Consider Figure 1, which depicts a rating curve with two regions having different error characteristics. The interpolation region consists of that part of the rating curve well defined by discharge-stage measurements; typically, the error coefficient of variation (CV) would be small, say, up to 5%. In the extension region the rating curve is extended by methods such as slope-conveyance, log-log extrapolation, or fitting to indirect discharge estimates. Typically, such extensions are smooth and therefore can induce systematic underestimation or overestimation of the true discharge over a range of stage. In Figure 1 the extension systematically underestimates the true discharge.

The extension error CV is not well known, but *Potter and Walker* [1981, 1985] suggest it may be as high as 30%.

Kuczera [1996] generalizes the likelihood function (5) to explicitly incorporate information about systematic rating curve error. He shows that rating curve error has the potential to drastically devalue the information content of flood data.

2.3. Prior Distribution

The choice of prior distribution $\xi(\boldsymbol{\beta})$ is an essential step in any Bayesian analysis. In this study the emphasis is on at-site flood frequency analysis. Accordingly, it is assumed that the analyst is not prepared to use any informative prior knowledge about $\boldsymbol{\beta}$. All inference will be based on the at-site data $\mathbf{D}$. Therefore a locally uniform prior on $\boldsymbol{\beta}$ is used throughout.

Although a noninformative prior is assumed in this study, it needs to be stressed that an informative $\xi(\boldsymbol{\beta})$ can and should be used in practice. Regionalized flood frequency information can define an informative $\xi(\boldsymbol{\beta})$. If, as in this study, (1) is evaluated numerically, there is no practical restriction on the actual form of $\xi(\boldsymbol{\beta})$.

3. Expected Probability

The expected probability distribution is formally defined in terms of an integral equation. This lays the foundation for application of Monte Carlo importance sampling to evaluate the expected probability integral.

3.1. Design Flood Distribution

If the true value of $\boldsymbol{\beta}$ were known, then the pdf $f(q|\boldsymbol{\beta})$ could be used to estimate the design flood q_T, which has exceedance probability $1/T$, with T defined as the return period or average recurrence interval for the flood q_T. However, in practice this is not the case. The true value of $\boldsymbol{\beta}$ is unknown. All that is known about $\boldsymbol{\beta}$, given the data $\mathbf{D}$, is summarized by the posterior pdf $\xi(\boldsymbol{\beta}|\mathbf{D})$. This uncertainty must somehow be accounted for when making predictions, particularly when extrapolating to floods beyond the observed record.

The Bayesian predictive distribution explicitly accounts for this uncertainty. The Bayesian distribution of an annual maximum flood q is obtained by application of the total probability theorem, yielding the pdf

$$g(q|\mathbf{D}) = \int_{\boldsymbol{\beta}} f(q|\boldsymbol{\beta})\xi(\boldsymbol{\beta}|\mathbf{D})\, d\boldsymbol{\beta} \tag{6}$$

The integral (6) yields the expected pdf of q given the data $\mathbf{D}$. *Stedinger* [1983] refers to this distribution as the design flood distribution.

For the lognormal probability model fitted to n gauged annual maximum discharges and assuming a noninformative prior, (6) can be analytically evaluated, yielding

$$\frac{\log_e q - m}{s\sqrt{1 + \dfrac{1}{n}}} \sim t_{n-1} \tag{7}$$

where m and s are the mean and standard deviation of the log discharge, respectively, and t_{n-1} is a t random variable with $n - 1$ degrees of freedom. This result is somewhat counterintuitive. Even though the true flood model is lognormal, the t probability model should be used in estimating the design flood. When n and the exceedance probability are small, (7) can yield a substantially more conservative estimate than classical quantile estimators [*Stedinger*, 1983]. However, as n increases, m and s converge to their true values and t_{n-1} converges to the normal distribution. In other words, when n is large, we can use the classical estimator which assumes m and s are the true values.

Following *Krzysztofowicz and Yakowitz* [1980], the design flood q_T with exceedance probability $1/T$ is defined by

$$P(q > q_T|\mathbf{D}) = \int_{q_T}^{\infty} g(q|\mathbf{D})\, dq = \frac{1}{T} \tag{8}$$

Substituting (6) into (8) and changing the order of integration yields

$$P(q > q_T|\mathbf{D}) = \int_{\boldsymbol{\beta}} \left(\int_{q_T}^{\infty} f(q|\boldsymbol{\beta})\, dq \right) \xi(\boldsymbol{\beta}|\mathbf{D})\, d\boldsymbol{\beta}$$

$$= \int_{\boldsymbol{\beta}} P(q > q_T|\boldsymbol{\beta})\xi(\boldsymbol{\beta}|\mathbf{D})\, d\boldsymbol{\beta} \tag{9}$$

where $P(q > q_T|\boldsymbol{\beta})$ is the probability of q exceeding q_T given that $\boldsymbol{\beta}$ is the true parameter vector. Equation (9) makes clear that $P(q > q_T|\mathbf{D})$ is the expected exceedance probability for discharge q_T with the expectation taken over all feasible values of $\boldsymbol{\beta}$.

3.2. Optimal Estimation of Exceedance Probability

We have seen that the expected probability $P(q > q_T|\mathbf{D})$ is a natural consequence of allowing for uncertainty in $\boldsymbol{\beta}$. The expected probability has another important attribute, namely, that it is the optimal estimator of the exceedance probability of q_T in the mean-squared-error sense. The following discussion based on *DeGroot* [1970] expands upon this point.

Define the loss function $L(\hat{x}(\mathbf{D}), x(\boldsymbol{\beta}))$ as the loss incurred when the true value $x(\boldsymbol{\beta})$, which depends on $\boldsymbol{\beta}$, is estimated by the decision $\hat{x}(\mathbf{D})$, which depends on the data $\mathbf{D}$. Because the data do not exactly specify $\boldsymbol{\beta}$, the loss itself is not exactly known. This complicates finding the best decision.

One approach to finding the best decision is to seek the decision $\hat{x}(\mathbf{D})_{\text{opt}}$, which minimizes the expected loss defined by

$$\hat{x}(\mathbf{D})_{\text{opt}} \leftarrow \min \int_{\boldsymbol{\beta}} L(\hat{x}(\mathbf{D}), x(\boldsymbol{\beta}))\xi(\boldsymbol{\beta}|\mathbf{D})\, d\boldsymbol{\beta} \tag{10}$$

If the loss function is quadratic, namely,

$$L(\hat{x}(\mathbf{D}), x(\boldsymbol{\beta})) = \alpha(\hat{x}(\mathbf{D}) - x(\boldsymbol{\beta}))^2 \tag{11}$$

where α is a constant, *DeGroot* [1970] shows that the optimal decision is the expected value of $x(\boldsymbol{\beta})$. It follows that the optimal estimate (in a mean-squared-error sense) of $P(q > q_T|\boldsymbol{\beta})$ is its expected value $P(q > q_T|\mathbf{D})$.

The use of a quadratic loss function to justify the use of expected probability can be challenged as being unrealistic. In particular, the assumption of symmetry is questionable given that the consequence of underestimation typically differs from that of overestimation [*Slack et al.*, 1975]. However, if the user can specify a realistic loss function, the search for $\hat{x}(\mathbf{D})_{\text{opt}}$ should be relatively straightforward given the availability of random samples drawn from $\xi(\boldsymbol{\beta}|\mathbf{D})$ (see section 5).

4. Importance Sampling

Apart from (7), there are no known analytical expressions for (6) for the other probability models commonly used in flood frequency analysis. As a result, numerical techniques must be used.

Importance sampling is a Monte Carlo technique [*Geweke*, 1989; *Tanner*, 1992] for random sampling from a probability distribution. It enables integration of equations like (6) for most probability distributions for which the integrals exist. It possesses the disadvantages and advantages of most Monte Carlo schemes: Although it will only yield results in practice accurate to two, possibly three, significant figures, it is simple and robust even as the dimension of the integral increases. To integrate (9) using importance sampling, reexpress (9) as

$$P(q > q_T|\mathbf{D}) = \int_{\boldsymbol{\beta}} P(q > q_T|\boldsymbol{\beta}) \frac{\xi(\boldsymbol{\beta}|\mathbf{D})}{I(\boldsymbol{\beta})} I(\boldsymbol{\beta})\, d\boldsymbol{\beta} \tag{12}$$

where $I(\boldsymbol{\beta})$ is the importance pdf, typically a convenient multivariate distribution of $\boldsymbol{\beta}$ which provides a reasonable approximation to $\xi(\boldsymbol{\beta}|\mathbf{D})$ and spans the domain of $\xi(\boldsymbol{\beta}|\mathbf{D})$.

In this study the multivariate normal was chosen as the importance distribution. Graphical study of the posterior probability distribution $\xi(\boldsymbol{\beta}|\mathbf{D})$, illustrated in the case study, can be used to assess how well the multinormal distribution approximates $\xi(\boldsymbol{\beta}|\mathbf{D})$ and, in particular, spans the domain of $\xi(\boldsymbol{\beta}|\mathbf{D})$. Indeed, importance sampling was selected in preference to the more general Metropolis technique (for example, see *Gelman et al.* [1997] for a general discussion or *Kuczera and Parent* [1998] for hydrologic applications) because it is expected to be more efficient. On the basis of the author's graphical study of $\xi(\boldsymbol{\beta}|\mathbf{D})$ for flood frequency data sets, the multinormal approximation to $\xi(\boldsymbol{\beta}|\mathbf{D})$ can be reasonably good as the number of gauged data approach 50.

Tanner [1992] presents the following Monte Carlo scheme to evaluate (12):

1. Draw N independent and identically distributed samples from $I(\boldsymbol{\beta})$, yielding $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_N\}$.
2. Evaluate

$$\hat{P}(q > q_T|\mathbf{D}) = \left(\sum_{i=1}^{N} P(q > q_T|\boldsymbol{\beta}_i) w_i \right) \Big/ \sum_{i=1}^{N} w_i \tag{13}$$

where the weight w_i is $\xi(\boldsymbol{\beta}_i|\mathbf{D})/I(\boldsymbol{\beta}_i)$.

3. Evaluate the asymptotic standard error of $\hat{P}(q > q_T|\mathbf{D})$ using

$$\frac{1}{N^{1/2}} \left(\left[\left(\sum_{i=1}^{N} w_i P(q > q_T|\boldsymbol{\beta}_i)^2 \right) \Big/ \sum_{i=1}^{N} w_i \right] - \hat{P}(q > q_T|\mathbf{D})^2 \right)^{1/2}$$

Note that this standard error differs from that quoted by *Tanner* [1992].

The importance pdf $I(\boldsymbol{\beta})$ can be derived from the second-order approximation to the posterior pdf $\xi(\boldsymbol{\beta}|\mathbf{D})$. Following *Gelman et al.* [1997], the second-order Taylor series expansion about $\hat{\boldsymbol{\beta}}$, the most probable value of $\boldsymbol{\beta}$, yields

$$\log_e \xi(\boldsymbol{\beta}|\mathbf{D}) \to \log_e \xi(\hat{\boldsymbol{\beta}}|\mathbf{D}) + \tfrac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \left. \frac{\partial^2 \log_e \xi(\boldsymbol{\beta}|\mathbf{D})}{\partial \boldsymbol{\beta}^2} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \tag{14}$$

noting that the vector $\partial \log_e \xi(\boldsymbol{\beta}|\mathbf{D})/\partial \boldsymbol{\beta}|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$ equals zero at $\hat{\boldsymbol{\beta}}$. It follows that the second-order approximation is the multivariate normal distribution

$$\boldsymbol{\beta}|\mathbf{D} \to N[\hat{\boldsymbol{\beta}}, \Sigma_D] \tag{15}$$

where the covariance matrix Σ_D is defined by the observed information matrix

$$\Sigma_D = \left(- \left. \frac{\partial^2 \log_e \xi(\boldsymbol{\beta}|\mathbf{D})}{\partial \boldsymbol{\beta}^2} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right)^{-1} \tag{16}$$

Geweke (1989) observes that it is important for convergence of the Monte Carlo scheme that the tails of $I(\boldsymbol{\beta})$ not decay faster than the tails of $\xi(\boldsymbol{\beta}|\mathbf{D})$. To this end, it is suggested that the importance distribution be scaled to give

$$I(\boldsymbol{\beta}) \sim N[\hat{\boldsymbol{\beta}}, \gamma^2 \Sigma_D] \tag{17}$$

where γ is a scaling factor typically in the range 1.5–2.5. The purpose of γ is to scale the covariance Σ_D to ensure samples are drawn well outside the usual region drawn from $N[\hat{\boldsymbol{\beta}}, \Sigma_D]$. This ensures that $I(\boldsymbol{\beta})$ representatively draws samples from the whole posterior pdf $\xi(\boldsymbol{\beta}|\mathbf{D})$.

The Monte Carlo scheme is particularly robust in the presence of constraints on $\boldsymbol{\beta}$. Whenever an infeasible $\boldsymbol{\beta}_i$ is drawn, the associated posterior pdf $\xi(\boldsymbol{\beta}_i|\mathbf{D})$ is set to zero.

5. Confidence Limits and Optimal Quantiles

Importance sampling generates a sequence of parameters and their normalized weights $\{\boldsymbol{\beta}_i, p_i, i = 1, \cdots, N\}$, where the normalized weight $p_i = w_i/\sum_{j=1}^{n} w_j$. Sampling from the N values of $\boldsymbol{\beta}_i$ according to the probability p_i yields samples drawn from $\xi(\boldsymbol{\beta}|\mathbf{D})$. Accordingly, posterior distribution of functions dependent on $\boldsymbol{\beta}$ can be readily approximated.

For example, $100(1 - \alpha)\%$ quantile confidence limits (strictly speaking, they should be called credible limits in recognition of the different interpretation of confidence limits in classical statistics) can be obtained as follows: Define the T-year quantile $q_T(\boldsymbol{\beta})$ such that $P[q > q_T(\boldsymbol{\beta})] = 1/T$. Rank the N quantiles $\{q_T(\boldsymbol{\beta}_i), i = 1, \cdots, N\}$. The upper and lower confidence limits are the quantiles whose exceedance probabilities are nearest to $\alpha/2$ and $1 - (\alpha/2)$, respectively.

It follows from the discussion in section 3.2 that the optimal (in a mean-squared-error sense) quantile estimate is the expected value of the quantile defined as

$$E[q_T(\boldsymbol{\beta})|\mathbf{D}] = \int_{\boldsymbol{\beta}} q_T(\boldsymbol{\beta}) \xi(\boldsymbol{\beta}|\mathbf{D})\, d\boldsymbol{\beta} \tag{18}$$

provided the integral is finite. This can be readily estimated by importance sampling using

$$E[q_T(\boldsymbol{\beta})|\mathbf{D}] = \sum_{i=1}^{n} p_i q_T(\boldsymbol{\beta}_i) \tag{19}$$

However, the practical value of (19) is limited by the fact that the integral (18) may not necessarily be bounded for some of the probability distributions used in flood frequency analysis [*Stedinger*, 1983]. Such situations may be detected by computing the asymptotic coefficient of variation of the estimate (19). When the integral (18) is unbounded, the coefficient of variation does not converge to zero for large n.

Table 1. Comparison of 90% Quantile Confidence Limits for Lognormal Distribution Fitted to 30 Years of Gauged Data

Return Period, years	Maximum Likelihood Quantile, m³/s	Exact 90% Quantile Confidence Limits, m³/s		Percent Error in Quantile Confidence Limits Using Importance Sampling With $\gamma = 1.5$			
				n = 10,000		n = 50,000	
50	3.34×10^5	1.90×10^5	8.69×10^5	0.02	−0.71	−0.09	0.12
100	4.80×10^5	2.61×10^5	1.38×10^6	−0.21	−0.97	−0.29	0.10
500	1.00×10^6	4.95×10^5	3.53×10^6	−0.83	−1.23	−0.65	0.23

6. Practical Implementation

All the Bayesian procedures described in this paper have been incorporated in a program called FLIKE, which is an event-driven Windows 95/NT application that fully supports a graphical user interface. At present, FLIKE can perform a full flood frequency analysis on five distributions: two-parameter lognormal, log-Pearson III, Gumbel, GEV, and generalized Pareto with support for uniform and multinormal priors. FLIKE can be obtained free from the author upon request.

Two significant difficulties were encountered during the development of FLIKE. These are discussed along with the adopted solutions in the following:

1. Developing a robust search method to find the most probable parameters proved to be a challenge. The use of unconstrained gradient-based local search methods occasionally failed to converge for distributions such as GEV and log-Pearson III, which have upper or lower bounds depending on the shape parameter. The problem was attributed to the use of initial parameter estimates based on the method of moments or probability-weighted moments. These estimates sometimes located the starting point of the search algorithm close to the edge of the feasible region, causing it to terminate prematurely. This problem was overcome by assigning initial parameters well removed from the infeasible parameter region. In the case of the GEV the shape parameter is initially assigned 0, whereas for the log-Pearson III the log skew is assigned 0. This difficulty may be the reason why previous investigators found maximum likelihood methods somewhat unreliable. An alternative solution, also implemented in FLIKE, is to use a probabilistic search method such as the shuffled complex evolution (SCE) algorithm of *Duan et al.* [1992]; though less efficient than gradient-based searches, it locates the environs of the global optimum with high reliability.

2. Evaluation of Σ_D requires numerical computation of the Hessian of $\log_e \xi(\boldsymbol{\beta}|\mathbf{D})$. When using a finite difference scheme to evaluate the Hessian, care must be exercised in selecting the perturbation on $\boldsymbol{\beta}$. This is because $\hat{\boldsymbol{\beta}}$ can sometimes lie at the edge of the feasible region for distributions with lower or upper bounds. In such cases a one-sided finite difference scheme must be employed, with the perturbation taken in a direction toward the interior of the feasible region.

7. Example

A case study for the Hunter River at Singleton, Australia, is presented to illustrate various aspects of the importance sampling scheme. The data consist of 30 years of gauged flows and censored data based on the fact that one historic flood in 117 years exceeded the largest gauged flood, which occurred in 1955. A uniform prior was used.

Table 1 compares 90% quantile confidence limits obtained from fitting the lognormal distribution to the 30 years of gauged data. The exact confidence limits were obtained using *Interagency Advisory Committee on Water Data* [1982], whereas approximate limits were obtained using importance sampling with a scaling factor γ of 1.5 for 10,000 and 50,000 samples. From a practical viewpoint the approximate limits based on 10,000 samples adequately describe the uncertainty.

Figure 2 presents the log-Pearson III fit to 30 years of gauged data. It displays the classical (maximum likelihood) and expected probability results along with 90% quantile confidence limits. Figure 3 shows a plot of the posterior distribution

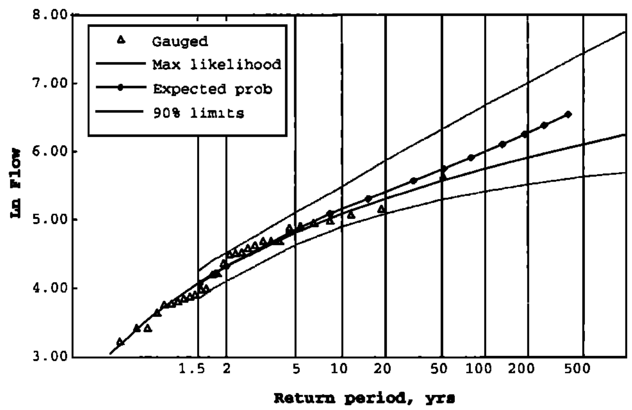

Figure 2. Flood frequency plot for Hunter River at Singleton, Australia, showing maximum likelihood and expected probability distributions along with 90% quantile confidence limits: log-Pearson III fitted to 30 years of gauged data on Gumbel probability paper.

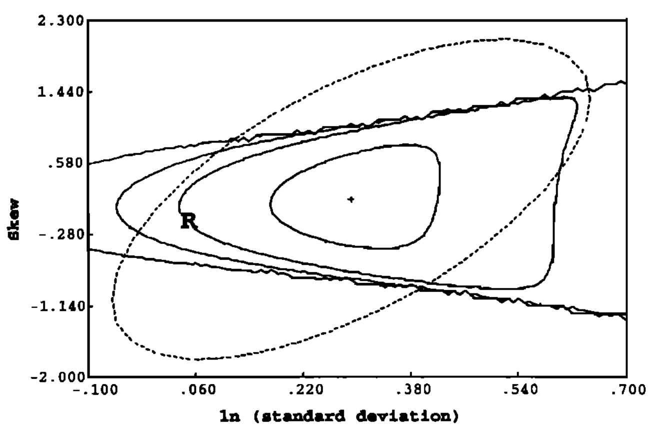

Figure 3. Plot of conditional posterior probability density function (pdf) for log-Pearson III standard deviation and skew parameters fitted to 30 years of gauged data along with multinormal 90% probability region.

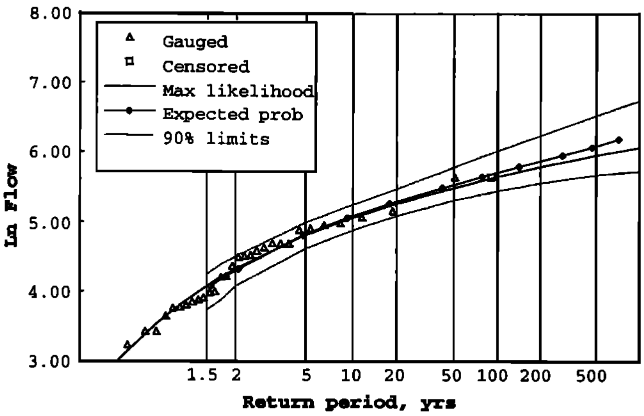

Figure 4. Flood frequency plot for Hunter River at Singleton, Australia, showing maximum likelihood and expected probability distributions along with 90% quantile confidence limits: log-Pearson III fitted to 30 years of gauged and censored data on Gumbel probability paper.

for the standard deviation and skew parameters with the mean held fixed at its most probable value. The dashed ellipse encloses the 90% conditional probability region based on the multinormal approximation (15), whereas the contour labeled by R should coincide with the dashed ellipse if the multinormal approximation were adequate. Figures 4 and 5 present similar results for the log-Pearson III fitted to both the gauged and censored data.

The following three observations are made:

1. The expected probability distribution lies above the classical distribution. This is most noticeable in the right-hand tail. It deviates farther from the classical distribution for the gauged data case because parameter uncertainty is greater than in the gauged and censored data case.

2. The upper 90% confidence limit starts to diverge for a return period T of about 10 years for the gauged data case. In contrast, the addition of censored data postpones divergence until the return period exceeds about 100 years.

3. The multinormal approximation provides, at best, a fair approximation to the actual posterior distribution which describes uncertainty about the model parameters $\boldsymbol{\beta}$. Nevertheless, the approximate 90% probability regions virtually enclose the R contour. This suggests that a scaling factor γ of 1 would ensure the multinormal distribution adequately spans the posterior distribution, thereby ensuring convergence of importance sampling algorithm. Note the dramatic reduction in parameter uncertainty when censored data are included.

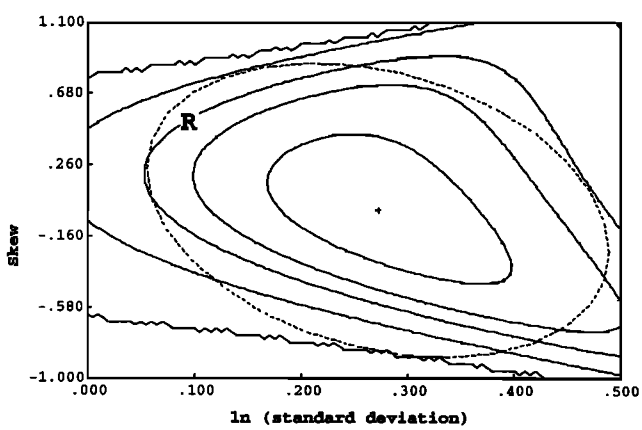

Figure 5. Plot of conditional posterior pdf for log-Pearson III standard deviation and skew parameters fitted to 30 years of gauged and censored data with multinormal 90% probability region.

The sensitivity of expected return periods, defined as the inverse of the expected exceedance probabilities, to choice of scaling factors γ is assessed in Table 2, which reports expected return periods and their 95% asymptotic probability limits for 50,000 Monte Carlo samples with γ equal to 1.0 and 2.0 for the gauged data case. Observe that the 95% probability limits for the expected return periods overlap, suggesting insensitivity to the choice of γ and confirming the visual interpretation of Figure 3 that the multinormal approximation adequately spans the posterior distribution.

A similar finding was found for the gauged and censored data case. However, because the data were more informative, the expected probability distribution was closer to the distribution based on the maximum likelihood estimate. For example, for the 1000-year maximum likelihood discharge, the expected return period jumped from 191 years for the gauged data to 450 years for the gauged and censored data case.

Execution of the importance sampling algorithm is quick. For example, it took just over 40 s on a 75-MHz Pentium with 16 Mb of RAM to perform 50,000 Monte Carlo samples for the entries in Table 2.

The efficacy of importance sampling deteriorates when the number of gauged data are reduced. Figure 6 shows a plot of the conditional posterior distribution for the standard deviation and skew parameters for the fit to 10 years of gauged data. The multinormal approximation is poor, with the shape of the R contour departing markedly from that of the dashed ellipse. One would expect many of the importance samples to correspond to infeasible or highly improbable parameters, resulting in considerable inefficiency. Not surprisingly, Table 3 reveals that the expected probabilities have become sensitive to the choice of scaling factors γ. For γ equal to 1.5 and 2, the expected probabilities, though not consistent, agree reasonably well. One could conclude that the scaling factor γ of 1.5 offers some protection against the multinormal distribution poorly approximating the posterior distribution. More important, this highlights the need to perform a visual inspection of the posterior distribution to identify situations where importance sampling may only yield results of fair accuracy. Of course, one could argue this example is contrived; significant extrapolation of flood frequency curves fitted to only 10 years of gauged data borders on being a pointless exercise.

Table 2. Log-Pearson III Expected Return Periods for Selected Maximum Likelihood Quantiles Based on 30 Years of Gauged Data and 50,000 Monte Carlo Samples

Maximum Likelihood Quantile, m^3/s	Return Period, years	Expected Return Period and Its 95% Probability Limits as Function of γ	
		$\gamma = 1.0$	$\gamma = 2.0$
3.70×10^5	50	31.2 ± 0.2	31.4 ± 0.3
5.53×10^5	100	51.1 ± 0.5	51.4 ± 0.5
1.27×10^6	500	134.1 ± 1.8	134.3 ± 1.9
1.76×10^6	1000	191.6 ± 3.0	190.9 ± 3.1

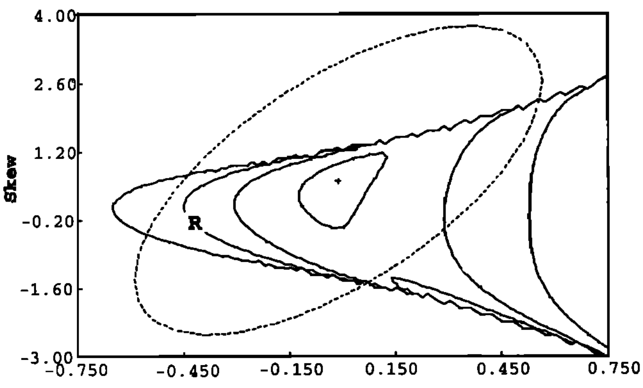

Figure 6. Plot of conditional posterior pdf for log-Pearson III standard deviation and skew parameters fitted to 10 years of gauged data with multinormal 90% probability region.

8. Conclusions

In flood frequency applications where the design flood is required to have a specified exceedance probability, it can be argued that the expected probability distribution should be used. This study presents, for the first time, a general method for computing the expected probability distribution as well as quantile confidence limits for any flood probability model using virtually any kind of flood data.

The generality of the approach comes from the use of Bayesian inference, which allows the use of gauged and censored flood data possibly corrupted by systematic rating error. In addition, prior information derived from regional analysis can be easily incorporated into the Bayesian analysis; although not considered in this study, the use of informative priors is supported by the FLIKE program. The ability to compute the expected probability distribution and quantile confidence limits for any flood probability model is made possible by use of importance sampling, which is a general Monte Carlo algorithm for random sampling from the Bayesian posterior distribution.

The success of importance sampling depends on how well the posterior probability distribution can be approximated by the so-called importance distribution. In this study the multivariate normal distribution was adopted as the importance distribution. It is suggested that this distribution will be adequate in most flood frequency applications. It is recommended that a covariance scaling factor γ of 1.5 be used to ensure that the posterior distribution is adequately spanned by the multinormal importance distribution. Furthermore, it is prudent to examine graphically the posterior distribution to check that the posterior domain is adequately spanned and to identify important features such as multiple optima and infeasible parameter regions.

Table 3. Log-Pearson III Expected Return Periods for Selected Maximum Likelihood Quantiles Based on 10 Years of Gauged Data and 50,000 Monte Carlo Samples

Maximum Likelihood Quantile, m^3/s	Return Period, years	Expected Return Period and Its 95% Probability Limits as Function of γ		
		$\gamma = 1.0$	$\gamma = 1.5$	$\gamma = 2.0$
1.28×10^5	50	32.0 ± 0.5	29.6 ± 0.4	30.6 ± 0.5
1.90×10^5	100	48.2 ± 0.8	42.8 ± 0.8	45.1 ± 0.8
4.41×10^5	500	104.9 ± 2.1	83.7 ± 2.0	92.0 ± 2.1

Acknowledgments. Constructive and insightful comments by reviewers Jacques Bernier, Vincent Fortin, and Roman Krzysztofowicz are gratefully acknowledged.

References

Beard, L. R., Probability estimates based on small normal-distribution samples, *J. Geophys. Res.*, *65*, 2143–2148, 1960.

Beard, L. R., Potential and application of flood frequency determinations, in *Application of Frequency and Risk in Water Resources*, edited by V. P. Singh, pp. 199–205, D. Reidel, Norwell, Mass., 1987.

DeGroot, M. H., *Optimal Statistical Decisions*, McGraw-Hill, New York, 1970.

Duan, Q., S. Sorooshian, and V. Gupta, Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031, 1992.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, p. 526, Chapman and Hall, New York, 1997.

Geweke, J., Bayesian inference in econometric models using Monte Carlo integrations, *Econometrica*, *24*, 1317–1339, 1989.

Interagency Advisory Committee on Water Data, Guidelines for determining flood flow frequency, *Bull. 17B*, Hydrol. Subcomm., Off. of Water Data Coord., Geol. Surv., U.S. Dep. of Interior, Washington, D. C., 1982.

Krzysztofowicz, R., and S. Yakowitz, Large-sample methods for decision analysis of gamma variates, *Water Resour. Res.*, *16*(3), 491–500, 1980.

Kuczera, G., Correlated rating curve error in flood frequency inference, *Water Resour. Res.*, *32*(7), 2119–2127, 1996.

Kuczera, G., and E. Parent, Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.*, *211*, 69–85, 1998.

Pilgrim, D. H., *Australian Rainfall and Runoff: A Guide to Flood Estimation*, Inst. of Eng., Canberra, A. C. T., Australia, 1987.

Potter, K. W., and J. F. Walker, A model of discontinuous measurement error and its effects on the probability distribution of flood discharge measurements, *Water Resour. Res.*, *17*(5), 1505–1509, 1981.

Potter, K. W., and J. F. Walker, An empirical study of flood measurement error, *Water Resour. Res.*, *21*(3), 403–406, 1985.

Slack, J. R., J. R. Wallis, and N. C. Matalas, On the value of information to flood frequency analysis, *Water Resour. Res.*, *11*(5), 629–647, 1975.

Stedinger, J. R., Design events with specified flood risk, *Water Resour. Res.*, *19*(2), 511–522, 1983.

Stedinger, J. R., and T. A. Cohn, Flood frequency analysis with historical and paleoflood information, *Water Resour. Res.*, *22*(5), 785–793, 1986.

Tanner, M. H., *Tools for Statistical Inference: Observed Data and Data Augmentation Methods*, Lect. Notes in Stat., vol. 67, Springer-Verlag, New York, 1992.

Vicens, G. J., I. Rodriguez-Iturbe, and J. C. Schaake Jr., A Bayesian framework for the use of regional information in hydrology, *Water Resour. Res.*, *11*(3), 405–414, 1975.

Wood, E. F., and I. Rodriguez-Iturbe, A Bayesian approach to analyzing uncertainty among flood frequency models, *Water Resour. Res.*, *11*(6), 839–843, 1975a.

Wood, E. F., and I. Rodriguez-Iturbe, Bayesian inference and decision making for extreme hydrologic events, *Water Resour. Res.*, *11*(4), 533–542, 1975b.

G. Kuczera, Department of Civil, Surveying and Environmental Engineering, University of Newcastle, Callaghan, NSW 2308, Australia. (cegak@cc.newcastle.edu.au)

(Received May 21, 1998; revised January 5, 1999;
accepted January 6, 1999.)