Bayesian computation of design discharges

J.M. van Noortwijk
HKV Consultants, Lelystad, Netherlands

Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft, Netherlands

H.J. Kalk
HKV Consultants, Lelystad, Netherlands

E.H. Chbab
Ministry of Transport, Public Works, and Water Management, Institute for Inland Water Management and Waste Water Treatment, Lelystad, Netherlands

ABSTRACT: Probabilistic design of river dikes is usually based on estimates of a design discharge. Dutch design discharges are currently estimated using classical statistical methods. A shortcoming of this approach is that statistical uncertainties are not taken into account and that probability distributions are given equal weight. In the paper, a method based on Bayesian statistics is presented. Seven probability distributions for annual maxima are investigated for determining extreme quantiles of discharges: the exponential, Rayleigh, normal, lognormal, gamma, Weibull, and Gumbel. Bayes factors are used to determine weights corresponding to how well a probability distribution fits the observed data. Predictive exceedance probabilities are obtained using two different Bayesian computation methods: numerical integration and Markov Chain Monte Carlo (MCMC). MCMC methods can be used to draw samples from the posterior density. The pros and cons of numerical integration and MCMC are given and illustrated by estimating the discharge of the river Rhine with an average return period of 1,250 years.

1 INTRODUCTION

Probabilistic design of river dikes is usually based on estimates of the design discharge. In The Netherlands, the design discharge is defined as the discharge with an average return period of 1,250 years. Extreme quantiles, such as the design discharge are usually determined by fitting various probability distributions to the available observations. [See for example DH & EAC-RAND (1993), Castillo (1988), and Van Gelder (1999)]. Probability plots and goodness-of-fit tests (such as chi-square and Kolmogorov-Smirnov) are commonly used to select an appropriate distribution.

A major practical difficulty in fitting probability distributions is that there is often a limited amount of data for determining extreme quantiles. The associated return period is large compared with the length of the period of observation. In The Netherlands, observed flood discharges are available for a period of 98 years only. There is a large statistical uncertainty involved in estimating extreme quantiles when using these observations. The maximum-likelihood method has been recognised as one of the best parameter estimation methods (Galambos et al., 1994) and it is especially suitable when there is a large number of observations. A drawback of the maximum-likelihood method is that statistical uncertainties cannot be taken into account.

Another consequence of sparse data is that more than one probability distribution seems to fit the observations and only a few can be rejected. Different distributions usually lead to different extrapolated values and the goodness-of-fit tests for selecting the appropriate distribution are often inconclusive. The tests are more concentrated on the central part of the distribution than the tail. As an alternative, the Bayesian method can be used to determine weights for quantifying how well a probability distribution fits the observed data while taking account of the statistical uncertainties involved.

Predictive exceedance probabilities are obtained using two different Bayesian computation methods: numerical integration and Markov Chain Monte Carlo (MCMC). MCMC methods can be used to draw samples from the posterior density of the unknown statistical parameters; it is essentially Monte Carlo integration using cleverly constructed Markov chains. The chains are constructed so that the dependent samples are distributed according to the posterior distribution. The most widely used Markov chain simulation technique is the Metropolis algorithm.

The pros and cons of the numerical integration method and the MCMC Metropolis algorithm are given. Both Bayesian computational methods have been successfully applied to estimate the discharge of the river Rhine with an average return period of

1,250 years, while taking account of the statistical uncertainties involved.

In this paper, a Bayesian method for estimating design discharges is presented. Section 2 considers Bayesian estimation of both parameters and quantiles associated with large average return periods. Section 3 and 4 are devoted to determining non-informative prior distributions and Bayes weights, respectively. Section 5 presents two well-known computational methods for calculating posterior distributions and predictive exceedance probabilities: numerical integration and Markov Chain Monte Carlo. In Section 6, the two Bayesian computational methods are compared by performing a Bayesian statistical analyis of the annual maximum discharges of the river Rhine. Section 7 ends with conclusions.

2 BAYESIAN ESTIMATION

According to (amongst others) Slijkhuis et al. (1999) and Siu & Kelly (1998), uncertainties in risk analysis can primarily be divided into two categories: inherent uncertainties and epistemic uncertainties. Inherent uncertainties represent randomness or variability in nature. For example, even in the event of sufficient data, one cannot predict the maximum discharge that will occur next year. In this paper, we study inherent uncertainty in time (e.g., fluctuation of the discharge in time). Epistemic uncertainties represent the lack of knowledge about a physical system. In this paper, we study statistical uncertainty (due to lack of sufficient data); it includes parameter uncertainty (when the parameters of the distribution are unknown) and distribution-type uncertainty (when the type of distribution is unknown). Statistical uncertainty can be reduced as more data becomes available.

The only statistical theory which combines modelling inherent uncertainty and statistical uncertainty is Bayesian statistics. The theorem of Bayes (1763) provides a solution to the problem of how to learn from data. In the framework of estimating the parameters $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_d)'$ of a probability distribution $\ell(x|\boldsymbol{\theta})$, Bayes' theorem can be written as

$$\pi(\boldsymbol{\theta}|\mathbf{x}) = \frac{\ell(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} \ell(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} = \frac{\ell(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{x})} \tag{1}$$

with

$\ell(\mathbf{x}|\boldsymbol{\theta})$ = the likelihood function of the observations $\mathbf{x} = (x_1,\ldots,x_n)'$ when the parametric vector $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_d)'$ is given,

$\pi(\boldsymbol{\theta})$ = the prior density of $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_d)'$ before observing data $\mathbf{x} = (x_1,\ldots,x_n)'$,

$\pi(\boldsymbol{\theta}|\mathbf{x})$ = the posterior density of $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_d)'$ after observing data $\mathbf{x} = (x_1,\ldots,x_n)'$, and

$\pi(\mathbf{x})$ = the marginal density of the observations $\mathbf{x} = (x_1,\ldots,x_n)'$.

The likelihood function $\ell(x|\boldsymbol{\theta})$ represents the inherent uncertainty of a random variable X when $\boldsymbol{\theta}$ is given, whereas the prior density $\pi(\boldsymbol{\theta})$ and the posterior density $\pi(\boldsymbol{\theta}|\mathbf{x})$ represent the statistical uncertainty in $\boldsymbol{\theta}$. This statistical uncertainty in $\boldsymbol{\theta}$ is parameter uncertainty. Using Bayes' theorem, we can update the prior distribution to the posterior distribution as soon as new observations become available. The more observations that are available, the smaller the parameter uncertainty. If a random variable X has a probability density function $\ell(x|\boldsymbol{\theta})$ depending on the parametric vector $\boldsymbol{\theta}$, then the likelihood function $\ell(x_1,\ldots,x_n|\boldsymbol{\theta})$ of the independent observations $\mathbf{x} = (x_1,\ldots,x_n)'$ is given by

$$\ell(\mathbf{x}|\boldsymbol{\theta}) = \ell(x_1,\ldots,x_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} \ell(x_i|\boldsymbol{\theta}). \tag{2}$$

The marginal density $\pi(\mathbf{x})$ is obtained by integrating the likelihood $\ell(\mathbf{x}|\boldsymbol{\theta})$ over $\boldsymbol{\theta}$. Note that the maximum-likelihood estimate of the parametric vector $\boldsymbol{\theta}$ is defined as the estimate $\hat{\boldsymbol{\theta}}$, which maximises the likelihood function $\ell(x|\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$.

The cumulative distribution function and the survival function of the random variable X are denoted by $F(x|\boldsymbol{\theta})$ and $\overline{F}(x|\boldsymbol{\theta})$, respectively. The posterior predictive probability of exceeding x_0 is

$$\begin{aligned}\Pr\{X > x_0|\mathbf{x}\} &= \int_{\boldsymbol{\theta}} \Pr\{X > x_0|\boldsymbol{\theta}\}\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta} = \\ &= \int_{\boldsymbol{\theta}} \overline{F}(x_0|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}.\end{aligned} \tag{3}$$

Besides representing parameter uncertainty on the basis of Bayesian statistics, distribution-type uncertainty can also be taken into account using so-called Bayes factors or Bayes weights (see Section 4).

3 NON-INFORMATIVE PRIORS

For the purpose of flood prevention, we would like the observations to 'speak for themselves', especially in comparison to the prior information. This means that the prior distribution should describe a certain 'lack of knowledge' or, in other words, should be as 'vague' as possible. For this purpose, so-called non-informative priors have been developed.

The pioneer in using non-informative priors was Bayes (1763) who considered a uniform prior. However, the use of uniform priors is criticised because of a lack of invariance under one-to-one transformations. The physicist Sir Jeffreys (1961, Chapters 3-4) was the first to produce an alternative to solely using uniform non-informative priors. His main motivation for deriving non-informative priors (currently

known as Jeffreys priors) were invariance requirements for one-to-one transformations. For decades, there has been a discussion going on whether the multivariate Jeffreys rule is appropriate. We believe that the following statement made by Dawid (1999) is right: "we do not consider it as generally appropriate to use other improper priors than the Jeffreys measure for purposes of 'fully objective' formal model comparison". The main advantage of the Jeffreys prior is that it is always both invariant under transformations and dimensionless. For examples of Jeffreys priors, see the Appendix.

In explaining the derivation of non-informative Jeffreys priors, we refer to Box & Tiao (1973, Section 1.3). Let $\mathbf{x} = (x_1, \ldots, x_n)'$ be a random sample from a multi-parameter probability distribution with likelihood function $\ell(x|\boldsymbol{\theta})$. When the probability distribution obeys certain regularity conditions, then for sufficiently large n, the posterior density function of parametric vector $\boldsymbol{\theta}$ is approximately normal, and remains approximately normal under mild one-to-one transformations of $\mathbf{x} = (x_1, \ldots, x_n)'$. As a consequence, the prior distribution for $\boldsymbol{\theta}$ is approximately non-informative if it is taken proportional to the square root of Fisher's information for a single observation. The elements of Fisher's information matrix are

$$I_{ij}(\boldsymbol{\theta}) = E\left(-\frac{\partial^2 \log \ell(X|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right), \quad i, j = 1, \ldots, d,$$

and the corresponding non-informative Jeffreys prior is defined by

$$J(\boldsymbol{\theta}) = \sqrt{|I(\boldsymbol{\theta})|} = \sqrt{\det I_{ij}(\boldsymbol{\theta})}, \quad i, j = 1, \ldots, d.$$

4 BAYES FACTORS AND BAYES WEIGHTS

The Bayesian approach to hypothesis testing originates from the work of Jeffreys (1961). He developed a methodology for quantifying the evidence in favour of a scientific theory using the so-called Bayes factor. This factor is the posterior odds of the null hypothesis when the prior probability on the null is one-half. A recent overview on Bayes factors can be found in Kass & Raftery (1995).

Assume the data $\mathbf{x} = (x_1, \ldots, x_n)'$ to have arisen under one of m models H_k, $k = 1, \ldots, m$. These hypotheses represent m marginal probability densities $\pi(\mathbf{x}|H_k)$, $k = 1, \ldots, m$. Given the prior probabilities $p(H_k)$, $k = 1, \ldots, m$, the data produce the posterior probabilities $p(H_k|\mathbf{x})$, $k = 1, \ldots, m$, where

$$\sum_{j=1}^{m} p(H_j) = 1 \text{ and } \sum_{j=1}^{m} p(H_j|\mathbf{x}) = 1.$$

These posterior probabilities can be obtained using Bayes' theorem as follows:

$$p(H_k|\mathbf{x}) = \frac{\pi(\mathbf{x}|H_k)p(H_k)}{\sum_{j=1}^{m} \pi(\mathbf{x}|H_j)p(H_j)}, \quad k = 1, \ldots, m, \tag{4}$$

where

$$B_{jk} = \frac{\pi(\mathbf{x}|H_j)}{\pi(\mathbf{x}|H_k)}, \quad j, k = 1, \ldots, m,$$

is denoted by the Bayes factor. The marginal densities of the data under model H_k, $\pi(\mathbf{x}|H_k)$, can be obtained by integrating with respect to the probability distribution of the uncertain parametric vector $\boldsymbol{\theta}_k = (\theta_{1k}, \ldots, \theta_{dk})'$ with number of parameters d:

$$\pi(\mathbf{x}|H_k) = \int \ell(\mathbf{x}|\boldsymbol{\theta}_k, H_k)\pi(\boldsymbol{\theta}_k|H_k)\, d\boldsymbol{\theta}_k, \tag{5}$$

where $\pi(\boldsymbol{\theta}_k|H_k)$ is the prior density of $\boldsymbol{\theta}_k$ under model H_k and $\ell(\mathbf{x}|\boldsymbol{\theta}_k, H_k)$ is the likelihood function of the data $\mathbf{x}$ given $\boldsymbol{\theta}_k$ and H_k.

If the prior distribution is the non-informative, improper, Jeffreys prior then the marginal density of the data $\mathbf{x} = (x_1, \ldots, x_n)'$ given in Eq. (5) can be difficult to compute. A possible solution is to approximate the logarithm of the marginal density using the Laplace expansion (De Bruijn, 1981, Chapter 4). The logarithm of the marginal density of the data can then be approximated by

$$\log(\pi(\mathbf{x}|H)) \approx \frac{d}{2}\log(2\pi) - \frac{d}{2}\log(n) + \log(\ell(\mathbf{x}|\hat{\boldsymbol{\theta}}, H)) \tag{6}$$

for $n \to \infty$, where $\hat{\boldsymbol{\theta}}$ is the maximum-likelihood estimator for the probability model H, d is the number of parameters of the probability model H, and n is the number of observations [see Tierney & Kadane (1986), Draper (1995), and Dawid (1999)]. Accordingly, the marginal density can be approximated by

$$\pi(\mathbf{x}|H) \approx \left(\frac{n}{2\pi} \right)^{-\frac{d}{2}} \ell(\mathbf{x}|\hat{\boldsymbol{\theta}}, H) \tag{7}$$

for $n \to \infty$. The second and third terms on the right-hand side of Eq. (6) form the Bayesian information criterion for model selection (Schwarz, 1978). The first term on the right-hand side, $(d/2)\log(2\pi)$, has been mostly omitted. However, we confirm the statement of Draper (1995) that its inclusion improves the accuracy of approximations to the marginal density.

An advantage of the above Laplace expansion is the possibility to use output of classical statistics software (maximum-likelihood estimators). Despite the fact that the relative error in the Bayes factor using the Laplace expansion has, in general, an accuracy of $O(1)$, the approximation appears to work rather well in practice (see Van Noortwijk et al., 2001).

It remains to choose the prior weights $p(H_k)$. For formal model comparison, we propose to use equal prior weights: i.e., $p(H_k) = 1/m,\ k = 1,\ldots,m$.

The posterior predictive probabilities of exceeding x_0 are calculated using the non-informative Jeffreys prior. Using the Bayes weights $p(H_k|\mathbf{x})$, $k = 1,\ldots,m$, the weighted predictive probability of exceeding x_0 is then defined by

$$\Pr\{X > x_0|\mathbf{x}\} = \sum_{k=1}^{m} p(H_k|\mathbf{x}) \Pr\{X > x_0|H_k, \mathbf{x}\}, \qquad (8)$$

where $\Pr\{X > x_0|H_k, \mathbf{x}\}$ is the predictive probability of exceeding x_0 under likelihood model H_k, $k = 1,\ldots,m$.

5 BAYESIAN COMPUTATION

In this paper, predictive exceedance probabilities are obtained using two different Bayesian computation methods: numerical integration and Markov Chain Monte Carlo (MCMC).

5.1 *Numerical integration*

In situations where the posterior joint probability density function of a parametric vector $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_d)'$ cannot be expressed in explicit form, we have to resort to approximations. For low-dimensional parametric vectors, we may define discrete distributions in the same way as they were applied to quantify the uncertainty in the shape parameter of a Weibull distribution in Soland (1969) and Mazzuchi & Soyer (1996a, 1996b). For convenience, the approximation method will be explained for a likelihood function $\ell(x|\theta)$ having only one unknown parameter θ. For multi-parameter distributions, the formulas can be easily extended. Let us assume the prior density of θ to be the non-informative Jeffreys prior $J(\theta)$ and the posterior density of θ to be based on the conditionally independent observations $\mathbf{x} = (x_1,\ldots x_n)'$.

By replacing the integral in Bayes' theorem (1) with a summation, the continuous posterior density function $\pi(\theta|\mathbf{x})$ can be discretised as

$$\Pr\{\Theta = \theta_j|\mathbf{x}\} = p(\theta_j|\mathbf{x}) =$$

$$= \frac{\ell(\mathbf{x}|\theta_j)p(\theta_j)}{\sum_{j=1}^{k} \ell(\mathbf{x}|\theta_j)p(\theta_j)} = \frac{\ell(\mathbf{x}|\theta_j)J(\theta_j)}{\sum_{j=1}^{k} \ell(\mathbf{x}|\theta_j)J(\theta_j)},$$

where

$$\theta_j = \theta_L + \frac{2j-1}{2}\frac{\theta_U - \theta_L}{k} = \theta_L + \left[j - \frac{1}{2}\right]\Delta\theta,$$

$j = 1,\ldots,k$, with θ_L and θ_U being the lower and upper bound for θ, respectively. In order to obtain a close approximation to the posterior density function, the interval $(\theta_L; \theta_U)$ should be chosen as wide as necessary and numerically possible. Suitable bounds can be determined on the basis of an approximate posterior density such as a (transformed) normal distribution with mean equal to the posterior mode and covariance matrix equal to the inverse of the expected Fisher information matrix. Convergence to normality of the posterior can often be improved by transformation (e.g., by taking the logarithm of a non-negative parameter). Also the number of subdivisions k should be suitably large. The non-informative prior probability function of θ is a normalised Jeffreys prior; that is,

$$\Pr\{\Theta = \theta_j\} = p(\theta_j) = \frac{J(\theta_j)}{\sum_{j=1}^{k} J(\theta_j)}.$$

The predictive mean of a function $f(\theta)$ can now be easily approximated by

$$E(f(\Theta)|\mathbf{x}) \approx \sum_{j=1}^{k} f(\theta_j)p(\theta_j|\mathbf{x}).$$

5.2 *Markov Chain Monte Carlo*

Markov Chain Monte Carlo (MCMC) methods can be used to draw samples from the posterior density. MCMC is essentially Monte Carlo integration using cleverly constructed Markov chains. The chains are constructed so that the dependent samples are distributed according to the posterior distribution of the unknown statistical parameters. The most widely used Markov chain simulation technique is the Metropolis algorithm (Metropolis et al., 1953). Clearly written overviews on MCMC methods can be found in Gelman et al. (1995, Chapter 11) and Carlin & Louis (2000, Chapter 5).

MCMC is based on sampling from a Markov chain with the property that after a sufficiently long burn-in period, the samples will be approximately from the posterior distribution of the parameters. Let us generate a sequence of parametric vectors $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots\}$, such that at each time $t \geq 0$, the next parametric vector $\boldsymbol{\theta}^{(t+1)}$ is drawn from the transition distribution $p(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)})$ depending only on the current state of the chain $\boldsymbol{\theta}^{(t)}$. This means that given $\boldsymbol{\theta}^{(t)}$, the next state $\boldsymbol{\theta}^{(t+1)}$ does not depend on the further history of the chain $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(t-1)}\}$. This sequence is called a Markov chain with transition distribution or kernel $q(\cdot|\cdot)$. Due to the Markovian structure, the samples are dependent.

The basic idea of the Metropolis algorithm will be described next. In this algorithm the next state $\boldsymbol{\theta}^{(t+1)}$ is chosen by first sampling a candidate point $\boldsymbol{\theta}^{*}$ from a so-called *proposal density* $q(\boldsymbol{\theta}^{*}|\boldsymbol{\theta}^{(t)})$ that depends on the current state $\boldsymbol{\theta}^{(t)}$. The Metropolis algorithm assumes the proposal density to be symmetric; that is, $q(\boldsymbol{\theta}^{*}|\boldsymbol{\theta}^{(t)}) = q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{*})$ for all $\boldsymbol{\theta}^{*}$ and $\boldsymbol{\theta}^{(t)}$. The proposal density is also called the candidate or jumping density. An example of a proposal density

is a multivariate normal distribution with mean $\boldsymbol{\theta}^{(t)}$ and suitably chosen fixed covariance matrix. The candidate $\boldsymbol{\theta}^*$ is accepted with probability $\alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^*)$ where

$$\alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^*) = \min\left\{\frac{\pi(\boldsymbol{\theta}^*|\mathbf{x})}{\pi(\boldsymbol{\theta}^{(t)}|\mathbf{x})}, 1\right\}.$$

If the candidate is accepted, the next state becomes $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^*$. If the candidate is rejected, the chain remains at $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$. Both acceptance and rejection counts as an iteration of the algorithm. When a candidate is sampled for which $\pi(\mathbf{x}|\boldsymbol{\theta}^*)$ and thus $\pi(\boldsymbol{\theta}^*|\mathbf{x})$ is zero, we must continue sampling until we have a candidate for which $\pi(\mathbf{x}|\boldsymbol{\theta}^*) > 0$. Remarkably, the stationary distribution (for which the Markov chain gradually 'forgets' its initial state) is exactly the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{x})$. The stationary distribution from which we would like to have samples is also called the *target* distribution. The proof and regularity conditions that the so-obtained distribution $p^{(t)}(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(0)})$ converges to the stationary distribution can be found in Tierney (1994, 1996) and Roberts (1996).

Although the proposal density can be arbitrarily chosen, the convergence largely depends on the proposal density. On the one hand, a proposal density with large jumps to places far from the support of the posterior has low acceptance rate and causes the Markov chain to stand still most of the time. On the other hand, a proposal density with small jumps and high acceptance rate may cause the chain to move slowly and to get stuck in one state. A great advantage of the Metropolis algorithm is that it only depends on the posterior density through ratios of the form $\pi(\boldsymbol{\theta}^*|\mathbf{x})/\pi(\boldsymbol{\theta}^{(t)}|\mathbf{x})$. Hence, the posterior density $\pi(\boldsymbol{\theta}|\mathbf{x})$ only needs to be known up to a normalising constant!

Although general advice can not be given, many studies pointed out that taking a multivariate normal distribution as proposal density often leads to accurate results [see, e.g., Tierney (1994) and Gelman et al. (1996)]. Suppose a likelihood model has a d-dimensional parametric vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)'$ (possibly after transformation) defined on the d-dimensional Euclidean space IR^d. Useful transformations are taking logarithms of non-negative quantities and taking logits of quantities that lie between 0 and 1. When the number of observations n is sufficiently large and the prior is non-informative, all probability distributions that are used to estimate design discharges obey regularity conditions for which the posterior distribution of $\boldsymbol{\theta}$ is approximately normal and remain approximately normal under mild one-to-one transformations of $\boldsymbol{\theta}$ (see Section 3). Because the posterior distribution is approximately normal, the proposal density should ideally also be normal with the same covariance structure as the posterior. This covariance matrix $\boldsymbol{\Sigma}$ can be approximated by the inverse Fisher information matrix for a sample of n observations evaluated at the maximum-likelihood estimator $\hat{\boldsymbol{\theta}}$; that is, $\boldsymbol{\Sigma} = [nI(\hat{\boldsymbol{\theta}})]^{-1}$ [see, e.g., Carlin & Louis (2000, Section 5.2)]. Recall that the Fisher information matrix for a single observation is denoted by $I(\boldsymbol{\theta})$ (see Section 3).

It should be noted that Gelman et al. (1996) proposed a slightly different proposal density. For posterior distributions being a d-dimensional spherical normal distribution, they showed that the optimal symmetric proposal density has the following properties. Firstly, its covariance matrix is approximately $[2.4]^2/d$ times the covariance matrix of the posterior. Secondly, the acceptance rate of the corresponding Metropolis algorithm is approximately 44% for $d = 1$ and 35% for $d = 2$. Note that these acceptance rates apply to posterior densities having *independent* components. Summarising, Gelman et al. (1996) advice to "choose the scaling of the proposal density so that the average acceptance rate of the algorithm is roughly ¼". A suitable starting point of the Metropolis algorithm is the maximum-likelihood estimator; that is, $\boldsymbol{\theta}^{(0)} = \hat{\boldsymbol{\theta}}$.

Summarising, the proposal density can best be a d-dimensional normal distribution with mean $\boldsymbol{\theta}^{(t)}$ and fixed covariance matrix $\boldsymbol{\Sigma} = [nI(\boldsymbol{\theta})]^{-1}$ multiplied by $[2.4]^2/d$. How to draw random samples from a multivariate normal distribution is explained in Carlin & Louis (2000, Section A.2.2). For the estimation of design discharges, the proposal density presented above appears to perform well.

6 DISCHARGE OF THE RIVER RHINE

In Van Noortwijk et al. (2001), a Bayesian analysis using numerical integration has been applied to the annual maximum discharges of the river Rhine at Lobith during the period 1901-1998. The Bayes weights in Eq. (4) were determined for seven probability distributions: the exponential, Rayleigh, normal, lognormal, gamma, Weibull and Gumbel. On the basis of a statistical analysis, the location parameter was chosen to be 2,125 m^3/s. This location parameter followed by maximising the weighted marginal density of the observations, where Bayes weights have been attached to the seven individual marginal densities. A Bayesian method of taking account of the statistical uncertainty in the location parameter will be presented in a future paper.

In this paper, the results of numerical integration and MCMC are compared. To assure a proper comparison, we use the same location parameters as in Van Noortwijk et al. (2001); that is, 2,125 m^3/s. Given equal prior weights, the Bayes posterior weights of the seven probability distributions can be found in Table 1. They have been computed on the basis of the Laplace approximation. Recall that the

Laplace approximation can be applied when the number of observations is large.

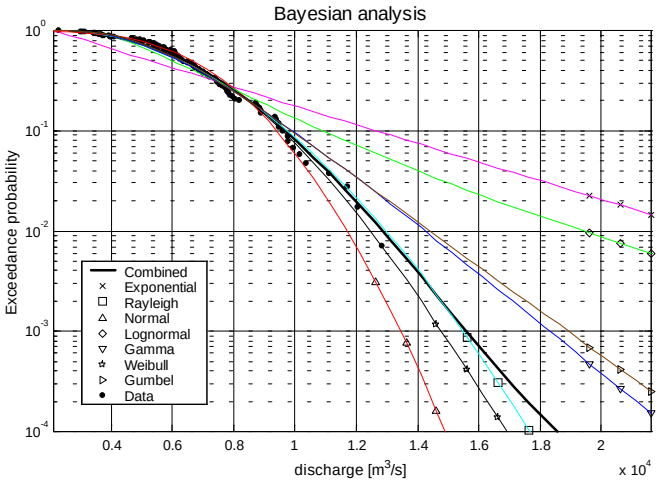

Figure 1. Predictive exceedance probability of annual maximum river Rhine discharge using numerical integration.

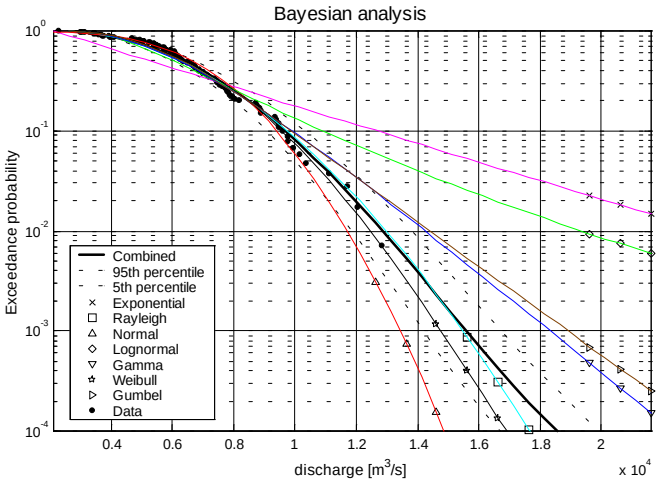

Figure 2. Predictive exceedance probability of annual maximum river Rhine discharge using Markov Chain Monte Carlo.

Using the Laplace expansion for the Bayes weights, the Rayleigh and Weibull distribution appear to fit best with Bayes weights of 57% and 32%, respectively. The Bayes estimate of the discharge with an average return period of 1,250 years is 15,856 m^3/s using numerical integration and 15,857 m^3/s using MCMC with 20,000 samples. Figures 1 and 2 show both the empirical exceedance probability and the predictive exceedance probabilities using numerical integration and MCMC, respectively. Using the maximum-likelihood method combined with the obtained Bayes weights, the estimate of the discharge with an average return period of 1,250 years decreases to 15,546 m^3/s. As expected, taking account of parameter uncertainty results in larger design discharges.

The results for numerical integration coincide with the results for MCMC. Both computational techniques utilise important results from the maximum-likelihood estimation. For example, the expected Fisher information matrix is used both for determining suitable numerical integration bounds and for specifying the MCMC proposal densities.

Disadvantages of numerical integration are that we must specify lower and upper bounds, as well as the method of discretisation, beforehand. These difficulties can be overcome by Markov chain simulation though in this situation the proposal density must be assessed. Because the number of statistical parameters is rather small (maximally two), numerical integration outperforms MCMC in computing time. Another advantage of numerical integration is that it is much more intuitive than MCMC. The advantage of MCMC, however, is that uncertainty bounds (for example, in terms of the 5th and 95th percentile; see Figure 2) can be easily obtained from the MCMC samples.

Table 1. Prior and posterior Bayes weights as well as the 1/1250 quantile for the annual maximum river Rhine discharge.

Distribution	Prior weight	Posterior weight (Laplace)	1/1250 quantile of discharge	
			Numerical integration	MCMC
Exponential	0.1429	0.0000	>22,000	>22,000
Rayleigh	0.1429	0.5719	15,715	15,717
Normal	0.1429	0.0480	13,593	13,583
Lognormal	0.1429	0.0000	>22,000	>22,000
Gamma	0.1429	0.0072	18,702	18,713
Weibull	0.1429	0.3156	14,996	14,984
Gumbel	0.1429	0.0573	19,340	19,349
Bayes combination	1.0000	1.0000	15,856	15,857

Finally, we remark that the estimated discharges in this paper are results of the proposed Bayesian method, and are therefore not statutory. The statutory Design Discharge of the Rhine at Lobith is currently set at 16,000 m^3/s (Van De Langemheen & Berger, 2002).

7 CONCLUSIONS

In this paper, the discharge of the Rhine at Lobith with an average return period of 1,250 years has been determined taking account of the statistical uncertainties involved. Statistical uncertainty occurs due to a lack of data. It can be subdivided into parameter uncertainty (when the parameters of a distribution are unknown) and distribution-type uncertainty (when the type of distribution is unknown). Bayes estimates and Bayes weights can be used to account for parameter uncertainty and distribution-type uncertainty respectively. Using Bayes weights, it is possible to discriminate between different probability models and to quantify how well a distribution fits the data.

Predictive exceedance probabilities can be obtained using either numerical integration or Markov

Chain Monte Carlo (MCMC). The results for both computational methods coincide. Because the number of statistical parameters is rather small, numerical integration requires less computing time than MCMC. Another advantage of numerical integration is that it is much more intuitive than MCMC. The advantage of MCMC is that uncertainty bounds can be easily obtained from the MCMC samples.

REFERENCES

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370-418.

Bernardo, J.M. (1976). Algorithm AS 103: Psi (digamma) function. *Applied Statistics*, 25:315-317.

Box, G.E.P., and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. New York: John Wiley & Sons.

Carlin, B.P. , and T.A. Louis (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman & Hall.

Castillo, E. (1988). Extreme value theory in engineering. New York: Academic Press.

Dawid, A.P. (1999). The trouble with Bayes factors. Research Report No. 202, Department of Statistical Science, University College London.

De Bruijn, N.G. (1981). *Asymptotic Methods in Analysis*. New York: Dover Publications Inc.

DH & EAC-RAND (1993). Toetsing uitgangspunten rivierdijkversterkingen; Deelrapport 2: Maatgevende belastingen [Investigating basic principles of river dike improvement; Supporting Volume 2: Design loads], Delft Hydraulics (DH) & European American Center (EAC) for Policy Analysis/RAND, Delft, The Netherlands.

Draper, D. (1995). Assessment and propagation of model uncertainty (with Discussion). *Journal of the Royal Statistical Society B*, 57(1):45-97.

Galambos, J., Lechner, J., and Simiu, E. (1994). Extreme Value Theory and Applications. Dordrecht: Kluwer Academic Publishers.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin (1995). *Bayesian Data Analysis*. London: Chapman & Hall.

Gelman, A., G.O. Roberts, and W.R. Gilks (1996). Efficient Metropolis jumping rules. In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 5*, pages 599-607. Oxford: Oxford University Press.

Jeffreys, H.J. (1961). *Theory of Probability; Third Edition*. Oxford: Clarendon Press.

Kass, R.E., and Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773-795.

Mazzuchi, T.A., and R. Soyer (1996a). Adaptive Bayesian replacement strategies. In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 5*, pages 667-674. Oxford: Oxford University Press.

Mazzuchi, T.A., and R. Soyer (1996b). A Bayesian perspective on some replacement strategies. *Reliability Engineering and System Safety*, 51:295-303.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087-1092.

Roberts, G.O. (1996). Markov chain concepts related to sampling algorithms. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pp. 45-57. London: Chapman & Hall.

Schneider, B.E. (1978). Algorithm AS 121: Trigamma function. *Applied Statistics*, 27:97-99.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461-464.

Siu, N.O., and Kelly, D.L. (1998). Bayesian parameter estimation in probabilistic risk assessment. *Reliability Engineering and System Safety*, 62:89-116.

Slijkhuis, K.A.H., Van Gelder, P.H.A.J.M., Vrijling, J.K., and Vrouwenvelder, A.C.W.M. (1999). On the lack of information in hydraulic engineering models. In G.I. Schuëller and P. Kafka, editors, *Safety and Reliability, Proceedings of ESREL '99 - The Tenth European Conference on Safety and Reliability, Munich-Garching, Germany, 1999*, pages 713-718. Rotterdam: Balkema.

Soland, R.M. (1969). Bayesian analysis of the Weibull process with unknown scale and shape parameters. *IEEE Transactions on Reliability*, 18(4):181-184.

Tierney, L. (1994) Markov chains for exploring posterior distributions (with Discussion). *The Annals of Statistics*, 22(4):1701-1762.

Tierney, L. (1996). Introduction to general state-space Markov chain theory. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pp. 59-74. London: Chapman & Hall.

Tierney, L., and J.B. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82-86.

Van De Langemheen, W., and H.E.J. Berger (2002). Hydraulische randvoorwaarden 2001: maatgevende afvoeren Rijn en Maas [Hydraulic boundary conditions: design discharges Rhine and Meuse]. Technical Report RIZA 2002.014, Ministry of Transport, Public Works, and Water Management, Institute for Inland Water Management and Waste Water Treatment (RIZA), Lelystad, the Netherlands.

Van Gelder, P.H.A.J.M. (1999). Statistical Methods for the Risk-Based Design of Civil Structures. PhD Thesis, Delft University of Technology, Delft, The Netherlands.

Van Noortwijk, J.M., H.J. Kalk, M.T. Duits, and E.H. Chbab (2001). The use of Bayes factors for model selection in structural reliability. In R.B. Corotis, G.I. Schuëller, and M. Shinozuka, editors, *Structural Safety and Reliability; Proceedings of the Eighth International Conference on Structural Safety and Reliability (ICOSSAR), Newport Beach, California, U.S.A., 17-22 June 2001*. Lisse: Balkema.

APPENDIX: PROBABILITY DISTRIBUTIONS AND THEIR JEFFREYS PRIORS

This Appendix contains the probability distributions which are considered in the statistical analysis of the annual maximum discharges, as well as their non-informative Jeffreys priors.

Exponential distribution

A random variable X has an exponential distribution with scale parameter $\theta > 0$ if the probability density function of X is given by

$$\mathrm{Ex}(x|\theta) = \frac{1}{\theta}\exp\left\{-\frac{x}{\theta}\right\} I_{(0,\infty)}(x),$$

where $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ if $x \notin A$ for every set A. The Jeffreys prior is

$$J(\theta) = \frac{1}{\theta}.$$

Rayleigh distribution
A random variable X has a Rayleigh distribution with quasi-scale parameter $\theta > 0$ if the probability density function of X is given by

$$\mathrm{Ra}(x|\theta) = \frac{2x}{\theta} \exp\left\{-\frac{x^2}{\theta}\right\} I_{(0,\infty)}(x).$$

The Jeffreys prior for the Rayleigh distribution is

$$J(\theta) = \frac{1}{\theta}.$$

Normal distribution
A random variable X has a normal distribution with mean m and precision $r > 0$ if the probability density function of X is given by

$$\mathrm{N}(x|m,r) = \left(\frac{r}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{r}{2}(x-m)^2\right\}.$$

The joint Jeffreys prior of the mean m and precision r of a normal distribution is

$$J(m,r) = \frac{1}{\sqrt{2r}}.$$

Lognormal distribution
A random variable X has a lognormal distribution with parameters m and $r > 0$ if the probability density function of X is given by

$$\mathrm{LN}(x|m,r) = \left(\frac{r}{2\pi}\right)^{\frac{1}{2}} \frac{1}{x} \exp\left\{-\frac{r}{2}(\log(x)-m)^2\right\} I_{(0,\infty)}(x).$$

Hence, if $\log(X)$ has a normal distribution, then X has a lognormal distribution. The joint Jeffreys prior of the parameters m and r of a lognormal distribution is

$$J(m,r) = \frac{1}{\sqrt{2r}}.$$

Gamma distribution
A random variable X has a gamma distribution with shape parameter $a > 0$ and scale parameter $b > 0$ if the probability density function of X is given by

$$\mathrm{Ga}(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp\{-bx\} I_{(0,\infty)}(x),$$

where

$$\Gamma(a) = \int_{t=0}^{\infty} t^{a-1} e^{-t} dt$$

is the gamma function for $a > 0$. The Jeffreys prior for the gamma distribution is

$$J(a,b) = \frac{\sqrt{a\psi'(a) - 1}}{b}.$$

The function $\psi'(a)$ is the first derivative of the digamma function:

$$\psi'(a) = \frac{\partial \psi(a)}{\partial a} = \frac{\partial^2 \log \Gamma(a)}{\partial a^2}$$

for $a > 0$. It is called the trigamma function. The digamma function and the trigamma function can be accurately computed using algorithms developed by Bernardo (1976) and Schneider (1978), respectively.

Weibull distribution
A random variable X has a Weibull distribution with shape parameter $a > 0$ and scale parameter $b > 0$ if the probability density function of X is given by

$$\mathrm{We}(x|a,b) = \frac{a}{b}\left[\frac{x}{b}\right]^{a-1} \exp\left\{-\left[\frac{x}{b}\right]^a\right\} I_{(0,\infty)}(x).$$

The Jeffreys prior for the Weibull distribution is

$$J(a,b) = \frac{1}{b}\frac{\pi}{\sqrt{6}}.$$

Remark: An alternative parameterisation of the Weibull distribution is using the quasi-scale parameter

$$\tilde{b} = b^a$$

instead of the scale parameter b. For calculating posterior distributions and predictive exceedance probabilities using numerical integration, the parameterisation in terms of the scale parameter is to be recommended. This is because the dependency between scale parameter and shape parameter is smaller than the dependency between quasi-scale parameter and shape parameter.

Gumbel distribution
A random variable X has a Gumbel distribution with location parameter a and scale parameter $b > 0$ if the probability density function of X is given by

$$\mathrm{Gu}(x|a,b) = \frac{1}{b} \exp\left\{-\frac{x-a}{b}\right\} \exp\left\{-\exp\left\{-\frac{x-a}{b}\right\}\right\}.$$

The Jeffreys prior for the Gumbel distribution is

$$J(a,b) = \frac{1}{b^2}\frac{\pi}{\sqrt{6}}.$$