The use of Bayes factors for model selection in structural reliability

J.M. van Noortwijk, H.J. Kalk, M.T. Duits
HKV Consultants, Lelystad, Netherlands

E.H. Chbab
Ministry of Transport, Public Works, and Water Management, Lelystad, Netherlands

ABSTRACT: Probabilistic design of structures is usually based on estimates of a design load with a high average return period. Design loads are often estimated using classical statistical methods. A shortcoming of this approach is that statistical uncertainties are not taken into account. In this paper, a method based on Bayesian statistics is presented. Using Bayes' theorem, the prior distribution representing information about the uncertainty of the statistical parameters can be updated to the posterior distribution as soon as data becomes available. Seven predictive probability distributions are considered for determining extreme quantiles of loads: the exponential, Rayleigh, normal, lognormal, gamma, Weibull and Gumbel. The Bayesian method has been successfully applied to estimate the design discharge of the river Rhine while taking account of the statistical uncertainties involved. As a prior the non-informative Jeffreys prior was chosen. The Bayes estimates are compared to the classical maximum-likelihood estimates. Furthermore, so-called Bayes factors are used to determine weights corresponding to how well a probability distribution fits the observed data; that is, the better the fit, the higher the weighting.

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 an extreme 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. These 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.

In this paper, a Bayesian method for estimating the design discharges is presented. Statistical uncertainties will be the subject of Section 2. Section 3 considers Bayesian estimation of both parameters and quantiles associated with large average return periods. Section 4 and 5 are devoted to determining non-informative prior distributions and Bayes weights, respectively. Section 6 presents a well-known Laplace expansion for the purpose of approximating the Bayes weights. Computational aspects that are important to cope with non-informative priors are treated in Section 7. The annual maximum discharges of the river Rhine will be studied in Section 8. Section 9 ends with conclusions.

2 STATISTICAL UNCERTAINTIES

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. The two main types of inherent uncertainty are inherent uncertainty in time (e.g., fluctuation of the discharge in time) and inherent uncertainty in space (e.g., fluctuation of a dike height in space). It is not possible to reduce inherent uncertainty in time. Epistemic uncertainties represent the lack of knowledge about a (physical) system. The two main types of epistemic uncertainty are statistical uncertainty (due to lack of sufficient data) and model uncertainty (due to lack of understanding the physics). Statistical uncertainty can be parameter uncertainty (when the parameters of the distribution are unknown) and distribution type uncertainty (when the type of distribution is unknown). In principle, epistemic uncertainties can be reduced as knowledge increases and more data becomes available.

3 BAYESIAN ESTIMATION

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 $\bar{F}(x|\boldsymbol{\theta})$, respectively. The posterior predictive probability of exceeding x_0 is

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

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.

4 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. A disadvantage of most non-informative priors is that these priors can be improper; that is, they often do not integrate to one. This disadvantage can be resolved by focussing on the posterior distributions rather than the prior distributions. As a matter of fact, formally carrying out the calculations of Bayes' theorem by combining an improper prior with observations often results in a proper posterior.

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. As an example, let us consider an unknown parameter θ and suppose the problem has been parameterised in terms of $\phi = \exp\{\theta\}$. This is a one-to-one transformation, which should have no bearing on the ultimate result. The Jacobian of this transformation is given by $d\theta/d\phi = d\log\phi/d\phi = 1/\phi$. Hence,

if the non-informative prior for θ is chosen to be uniform (constant), then the non-informative prior for ϕ should be proportional to $1/\phi$ to maintain consistency. Unfortunately, we cannot maintain consistency and choose both the non-informative priors for θ and ϕ to be constant.

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. In a multi-parameter setting, Jeffreys prior takes account of dependence between the parameters. 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.

As an example, the multivariate Jeffreys prior for the normal model with unknown mean μ and unknown standard deviation σ is

$$J(\mu,\sigma)\,d\mu d\sigma = \frac{\sqrt{2}}{\sigma^2}\,d\mu d\sigma .$$

It can be easily seen that the above prior is dimensionless: i.e., $d\mu$, $d\sigma$, and σ have the same dimension. For other examples, see the Appendix. Because the non-dimensionality argument is rather sound (from a physics point of view), we propose to use the multivariate Jeffreys measure for the purpose of model comparison.

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 $\ell(x|\theta)$. When the probability distribution obeys certain regularity conditions, then for sufficiently large n, the posterior density function of parametric vector θ is approximately normal, and remains approximately normal under mild one-to-one transformations of θ. As a consequence, the prior distribution for θ is approximately non-informative if it is taken proportional to the square root of Fisher's information for one observation. The elements of Fisher's information matrix are

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

and the corresponding non-informative Jeffreys prior is defined by

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

5 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 $\sum_{j=1}^m p(H_j) = 1$ and $\sum_{j=1}^m p(H_j|\mathbf{x}) = 1$. These posterior probabilities can be obtained using Bayes' theorem as follows:

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

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 H_k, $\pi(\mathbf{x}|H_k)$, can be obtained by integrating with respect to the probability distribution of the uncertain parametric vector $\theta_k = (\theta_{1k},\ldots,\theta_{dk})$ with number of parameters d:

$$\pi(\mathbf{x}|H_k) = \int \ell(\mathbf{x}|\theta_k, H_k)\,\pi(\theta_k|H_k)\,d\theta_k, \qquad (5)$$

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

A difficulty in using non-informative improper priors for calculating Bayes factors is that the prior odds, and thus the Bayes factor, may be undefined. The reason for this is that strictly speaking, the prior probability $p(H_k)$ is defined as

$$p(H_k) = w(H_k)\int J(\theta_k|H_k)\,d\theta_k,$$

where the integral over the non-informative Jeffreys prior $J(\theta_k|H_k)$ is often infinite and $w(H_k)$ is the prior weight. However, according to Dawid (1999), this problem can be resolved by redefining the posterior odds as

$$\frac{p(H_j|\mathbf{x})}{p(H_k|\mathbf{x})} = \frac{\pi(\mathbf{x}|H_j)}{\pi(\mathbf{x}|H_k)} \times \frac{w(H_j)}{w(H_k)}, \quad j,k = 1,\ldots,m. \qquad (6)$$

This posterior odds is well-defined so long as both integrals in it converge, which will typically be the case so long as the sample size n is large enough. For the seven probability distributions con-

sidered in this paper (see the Appendix), the marginal densities of the data do indeed converge. Using Eqs. (4) and (6), the posterior probability of model H_k being correct can now rewritten as

$$p(H_k|\mathbf{x}) = \frac{\pi(\mathbf{x}|H_k)w(H_k)}{\sum_{j=1}^{m}\pi(\mathbf{x}|H_j)w(H_j)}, \quad k=1,\ldots,m. \tag{7}$$

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

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

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

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

Both the marginal densities and the predictive exceedance probabilities have been obtained by numerical integration.

6 APPROXIMATE BAYES WEIGHTS

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

$$\log(\pi(\mathbf{x}|H)) \approx \frac{d}{2}\log(2\pi) - \frac{d}{2}\log(n) + \log(\ell(\mathbf{x}|\hat{\theta},H)) \tag{9}$$

for $n\to\infty$, where $\hat{\theta}$ 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

$$\pi(\mathbf{x}|H) \approx \left(\frac{n}{2\pi}\right)^{-\frac{d}{2}} \ell(\mathbf{x}|\hat{\theta},H) \tag{10}$$

for $n\to\infty$. The second and third terms on the right-hand side of Eq. (9) 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. For an example, see Section 8.

An advantage of the above Laplace expansion is the possibility to use output of classical statistics software (maximum-likelihood estimators). Another advantage of the Laplace approximation is the independence of the prior distribution (which, of course, can also be seen as a disadvantage). 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.

7 COMPUTATION OF BAYES WEIGHTS

The method of Bayes weights to select a probability model has been tested with Monte Carlo simulations. Samples were drawn from a known probability distribution and the corresponding Bayes weights were determined using numerical integration.

Two computational methods were considered. In both methods the numerical integration is performed with respect to suitably chosen lower and upper bounds of the posterior density $\pi(\theta|\mathbf{x})$. The first method uses the normalised Jeffreys prior; that is, given the lower and upper bounds of the statistical parameters, the Jeffreys prior is normalised to a proper probability density. The second method calculates these weights without normalising the Jeffreys prior; that is, the marginal density of the data in Eq. (5) is obtained by integrating with respect to the non-normalised Jeffreys prior.

For a likelihood function with one parameter θ, the difference between these two methods becomes clear when comparing the *normalised* marginal density

$$\tilde{\pi}(\mathbf{x}) \approx \sum_{j=1}^{r} \ell(\mathbf{x}|\theta_j)p(\theta_j) = \frac{\sum_{j=1}^{r}\ell(\mathbf{x}|\theta_j)J(\theta_j)}{\sum_{j=1}^{r}J(\theta_j)} \tag{11}$$

with the *non-normalised* marginal density

$$\pi(\mathbf{x}) \approx \sum_{j=1}^{r} \ell(\mathbf{x}|\theta_j)J(\theta_j)\Delta\theta, \tag{12}$$

where r is the number of subdivisions and $\Delta\theta$ is the corresponding step size. The logarithm of the ratio of the normalised and non-normalised marginal density gives us

$$\log(\pi(\mathbf{x})) \approx \log(\tilde{\pi}(\mathbf{x})) + \log\left(\sum_{j=1}^{r} J(\theta_j)\Delta\theta\right).$$

For convenience, the approximation method has been explained for a likelihood function having only one unknown parameter θ. For multi-parameter distributions, the above formulas can be easily extended.

Table 1. Bayes weights for samples from exponential distribution with mean 100 (left: normalised Jeffreys; right: non-normalised Jeffreys).

	n=10	n=20	n=50	n=100
Exponential	0.38 0.19	0.46 0.29	0.55 0.45	**0.58 0.55**
Rayleigh	0.11 0.04	0.03 0.01	0.00 0.00	0.00 0.00
Normal	0.03 0.04	0.01 0.01	0.00 0.00	0.00 0.00
Lognormal	0.19 0.19	0.16 0.15	0.07 0.06	0.02 0.01
Gamma	0.13 0.23	0.17 0.25	0.20 0.24	0.21 0.22
Weibull	0.11 0.23	0.15 0.25	0.18 0.25	0.19 0.22
Gumbel	0.05 0.07	0.03 0.03	0.00 0.00	0.00 0.00

Table 2. Bayes weights for samples from Rayleigh distribution with mean 100 (left: normalised Jeffreys; right: non-normalised Jeffreys).

	n=10	n=20	n=50	n=100
Exponential	0.04 0.02	0.01 0.01	0.00 0.00	0.00 0.00
Rayleigh	0.39 0.15	0.47 0.24	0.58 0.42	**0.65 0.56**
Normal	0.15 0.15	0.11 0.11	0.05 0.05	0.02 0.01
Lognormal	0.15 0.15	0.12 0.11	0.05 0.04	0.01 0.01
Gamma	0.06 0.17	0.06 0.17	0.06 0.14	0.05 0.10
Weibull	0.10 0.21	0.11 0.22	0.15 0.24	0.18 0.24
Gumbel	0.12 0.16	0.12 0.14	0.11 0.11	0.09 0.07

Table 3. Bayes weights for samples from normal distribution with mean 100 and standard deviation 20 (left: normalised Jeffreys; right: non-normalised Jeffreys).

	n=10	n=20	n=50	n=100
Exponential	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00
Rayleigh	0.01 0.00	0.00 0.00	0.00 0.00	0.00 0.00
Normal	0.31 0.21	0.34 0.23	0.44 0.31	**0.57 0.42**
Lognormal	0.27 0.18	0.24 0.17	0.20 0.14	0.15 0.09
Gamma	0.04 0.19	0.04 0.20	0.05 0.21	0.05 0.21
Weibull	0.18 0.24	0.22 0.27	0.25 0.29	0.22 0.27
Gumbel	0.18 0.17	0.15 0.13	0.06 0.05	0.01 0.01

Table 4. Bayes weights for samples from lognormal distribution with mean 100 and standard deviation 20 (left: normalised Jeffreys; right: non-normalised Jeffreys).

	n=10	n=20	n=50	n=100
Exponential	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00
Rayleigh	0.01 0.00	0.00 0.00	0.00 0.00	0.00 0.00
Normal	0.27 0.18	0.25 0.17	0.21 0.14	0.14 0.09
Lognormal	0.30 0.20	0.32 0.23	0.40 0.29	**0.52 0.37**
Gamma	0.04 0.20	0.04 0.22	0.05 0.26	0.07 0.30
Weibull	0.13 0.18	0.11 0.14	0.04 0.05	0.01 0.01
Gumbel	0.25 0.24	0.28 0.25	0.29 0.25	0.26 0.22

Table 5. Bayes weights for samples from gamma distribution with mean 100 and standard deviation 20 (left: normalised Jeffreys; right: non-normalised Jeffreys).

	n=10	n=20	n=50	n=100
Exponential	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00
Rayleigh	0.01 0.00	0.00 0.00	0.00 0.00	0.00 0.00
Normal	0.29 0.19	0.28 0.20	0.30 0.20	0.29 0.20
Lognormal	0.29 0.20	0.30 0.21	0.36 0.26	**0.46** 0.30
Gamma	0.04 0.20	0.04 0.22	0.06 0.27	0.08 **0.35**
Weibull	0.15 0.20	0.14 0.17	0.09 0.10	0.04 0.05
Gumbel	0.22 0.21	0.23 0.20	0.19 0.16	0.13 0.10

Table 6. Bayes weights for samples from Weibull distribution with mean 100 and standard deviation 20 (left: normalised Jeffreys; right: non-normalised Jeffreys).

	n=10	n=20	n=50	n=100
Exponential	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00
Rayleigh	0.01 0.00	0.00 0.00	0.00 0.00	0.00 0.00
Normal	0.34 0.22	0.37 0.25	0.40 0.28	0.35 0.26
Lognormal	0.24 0.16	0.18 0.12	0.08 0.06	0.01 0.01
Gamma	0.04 0.18	0.03 0.16	0.02 0.10	0.01 0.04
Weibull	0.23 0.30	0.33 0.39	0.48 0.55	**0.62 0.69**
Gumbel	0.15 0.13	0.09 0.07	0.02 0.01	0.00 0.00

Table 7. Bayes weights for samples from Gumbel distribution with mean 100 and standard deviation 20 (left: normalised Jeffreys; right: non-normalised Jeffreys).

	n=10	n=20	n=50	n=100
Exponential	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00
Rayleigh	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00
Normal	0.24 0.16	0.17 0.12	0.07 0.05	0.01 0.01
Lognormal	0.32 0.22	0.32 0.24	0.30 0.24	0.25 0.21
Gamma	0.03 0.20	0.04 0.19	0.03 0.15	0.02 0.09
Weibull	0.10 0.14	0.06 0.07	0.01 0.01	0.00 0.00
Gumbel	0.31 0.29	0.41 0.37	0.59 0.55	**0.72 0.70**

In order to exclude statistical variability, the determination of the Bayes weights has been performed 1,000 times and the mean values are presented in Tables 1-7. In each cell the value on the left-hand side represents the Bayes weight according to the normalised Jeffreys prior; the value on the right-hand side according to the non-normalised Jeffreys prior. The formulas of the corresponding normalised and non-normalised marginal density can be found in Eqs. (11) and (12), respectively. For seven probability distributions, Bayes weights have been computed for four sample sizes (i.e., n = 10, 20, 50, 100). For sample sizes of one hundred, the largest Bayes weights are shown in bold format.

From the Monte Carlo simulations, two conclusions can be drawn. Firstly, the results of the two computation methods differ significantly. Secondly, the method based on non-normalised Jeffreys priors performs better. For the gamma distribution, the normalised Jeffreys prior even wrongly points to the lognormal distribution. In Section 8, however, there appears to be an excellent agreement between the Bayes weights on the basis of the non-normalised Jeffreys prior and the approximate Bayes weights based on the Laplace expansion in Eq. (9).

Summarising, we recommend to use the non-normalised Jeffreys prior for formal model selection.

8 DESIGN DISCHARGE OF THE RIVER RHINE

A statistical analysis based on a modified maximum-likelihood method applied to annual maximum discharges of the river Rhine at Lobith resulted in three distributions which could not be rejected. These three distributions are the lognormal, gamma and Gumbel. Because none of these distributions could be clearly identified as being the best, DH & EAC-RAND (1993) gave these three distributions equal weight. The design discharge is then defined as the average of the three corresponding design discharges. This method has been used to determine the current Dutch design discharges.

Recently, Van Gelder (1999) proposed to fit various probability distributions to the data and to attach different weights to these distributions according to how good the fits are. This can be done either by using Bayes weights (Van Gelder, 1999) or linear regression weights (Tang, 1970). Van Gelder et al. (1999) showed that Bayes weights perform better than linear regression weights.

Bayesian analysis has been applied to the annual maximum discharges of the river Rhine at Lobith during the period 1901-1998. The Bayes weights (7) have been determined for seven probability distributions: the exponential, Rayleigh, normal, lognormal, gamma, Weibull and Gumbel. The Bayes weights largely depend on the location parameters of the probability distributions. For proper model selection, we therefore propose to use the same location parameter for all seven distributions. On the basis of a statistical analysis, the location parameter can best be chosen to be 2,125 m^3/s. This location parameter follows by maximising the weighted marginal density of the observations, where Bayes weights have been attached to the seven individual marginal densities. The method of determining the location parameter, as well as a sensitivity analysis, will be presented in a future paper.

For location parameter 2,125 m^3/s, the Bayes weights of the seven probability distributions can be found in Table 8. The Bayes weights have been computed on the basis of normalised Jeffreys priors, non-normalised Jeffreys priors, and the Laplace approximation. Because of the agreement between the Bayes weights based on non-normalised Jeffreys priors and the approximate Bayes weights, we recommend to use the non-normalised Jeffreys prior for model selection. Normalisation of the Jeffreys prior doesn't affect the predictive exceedance probability of a distribution; it only affects the Bayes weights. Recall that the Laplace approximation can be applied when the number of observations is large.

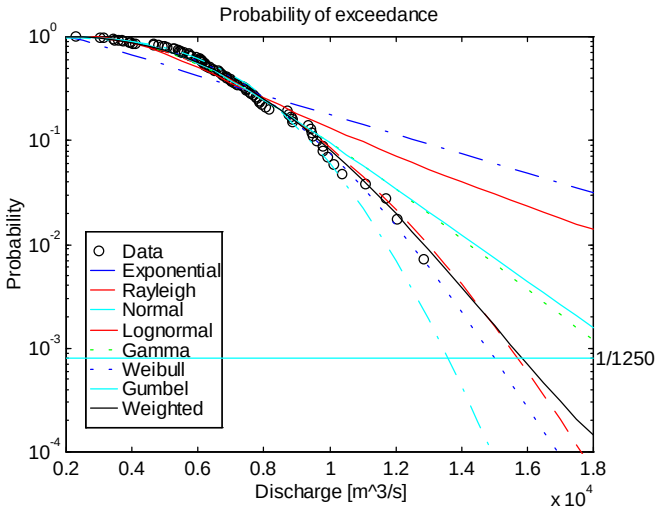

Figure 1. Predictive exceedance probability of annual maximum river Rhine discharge.

For a non-normalised Jeffreys prior, the Rayleigh and Weibull distribution appear to fit best with Bayes weights of 57% and 32%, respectively. The Bayes estimate of the design discharge with an average return period of 1,250 years is 15,845 m^3/s. Figure 1 shows both the empirical exceedance probability and the predictive exceedance probabilities. Using the maximum-likelihood method combined with the Bayes weights, the estimate of the design discharge decreases to 15,534 m^3/s. As expected, taking account of parameter uncertainty results in larger design discharges. The Bayesian approach seems to be promising for determining future design discharges.

Table 8. Prior and posterior Bayes weights for the annual maximum river Rhine discharge.

Bayes weights	Prior	Posterior		
		Normalised Jeffreys prior	Non-normalised Jeffreys prior	Laplace approximation
Exponential	0.1429	0.0000	0.0000	0.0000
Rayleigh	0.1429	0.6607	0.5718	0.5719
Normal	0.1429	0.0556	0.0481	0.0480
Lognormal	0.1429	0.0000	0.0000	0.0000
Gamma	0.1429	0.0033	0.0072	0.0072
Weibull	0.1429	0.2204	0.3173	0.3156
Gumbel	0.1429	0.0600	0.0555	0.0573

9 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. For formal model comparison, the use of the (non-normalised) Jeffreys prior is recommended. The design discharge increases when taking the statistical uncertainties properly into account.

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.

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.

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.

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.

Tang, W.H. (1980). Bayesian frequency analysis. *Journal of the Hydraulics Division*, 106(HY7):1203-1218.

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 Gelder, P.H.A.J.M., Van Noortwijk, J.M., and Duits, M.T. (1999). Selection of probability distributions with a case study on extreme Oder river discharges. 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 1475-1480. Rotterdam: 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. Special care has been given to deriving possible constants in the Jeffreys prior. The reason for this is that we agree with Dawid (1999), who stated that, "for the purposes of 'objective' model comparison, there is nothing to be gained by rescaling (...), and that the actual Jeffreys measure should be used".

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 μ 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

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

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

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