Bayesian estimation of design loads

J.M. van Noortwijk[1,2], H.J. Kalk[1] and E.H. Chbab[3]

[1] HKV Consultants, Lelystad, The Netherlands.

[2] Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands.

[3] Institute for Inland Water Management and Waste Water Treatment, Ministry of Transport, Public Works, and Water Management, Lelystad, The Netherlands.

Probabilistic design of structures is usually based on estimates of design loads with a large 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. Nine predictive probability distributions are considered for determining extreme quantiles of loads: the exponential, Rayleigh, normal, lognormal, gamma, Weibull, Gumbel, generalised gamma and generalised extreme-value. The Bayesian method has been successfully applied to estimate the discharge of the rivers Rhine and Meuse with an average return period of 1,250 years while taking account of the statistical uncertainties involved. In order that the observations 'speak for themselves', the non-informative Jeffreys priors were chosen as priors. 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.

Key words: Bayesian analysis, non-informative Jeffreys prior, Bayes weights, river discharges, maximum likelihood.

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 observation period. For the rivers Rhine and Meuse, observed flood discharges are available for a period of 98 and 88 years only, respectively. 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 (Van Gelder, 1999; Van Gelder et al., 1999).

In this paper, a Bayesian method for estimating design loads is presented. Using Bayes' theorem, the prior distribution representing information about the uncertainty of the statistical parameters is updated to the posterior distribution as soon as data becomes available. Section 2 considers Bayesian estimation of quantiles associated with large average return periods. The Bayes estimates are compared to the classical maximum-likelihood estimates. In order that the observations 'speak for themselves', the non-informative Jeffreys priors are chosen as priors. Bayes factors are used to determine weights corresponding to how well a probability distribution fits the observed data. Section 3 and 4 are devoted to determining non-informative Jeffreys priors and Bayes weights, respectively. Section 5 presents a well-known Laplace expansion for the purpose of approximating the Bayes weights. The annual maximum discharges of the rivers Rhine and Meuse will be studied in Section 6. 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.

A 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 $\boldsymbol{\theta} = (\theta_1, ..., \theta_d)'$ 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 $\mathbf{x} = (x_1, ..., x_n)'$ when the parametric vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_d)'$ is given,
$\pi(\boldsymbol{\theta})$	=	the prior density of $\boldsymbol{\theta} = (\theta_1, ..., \theta_d)'$ before observing data $\mathbf{x} = (x_1, ..., x_n)'$,
$\pi(\boldsymbol{\theta}\|\mathbf{x})$	=	the posterior density of $\boldsymbol{\theta} = (\theta_1, ..., \theta_d)'$ after observing data $\mathbf{x} = (x_1, ..., x_n)'$, and
$\pi(\mathbf{x})$	=	the marginal density of the observations $\mathbf{x} = (x_1, ..., x_n)'$.

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 $\ell(x_1, ..., x_n|\boldsymbol{\theta})$ of the independent observations $\mathbf{x} = (x_1, ..., x_n)'$ is given by

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

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 $\ell(\mathbf{x}|\boldsymbol{\theta})$ 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

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

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. 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, the 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.

Unfortunately, Jeffreys priors don't always exist for all parameter values. Jeffreys (1961, pages 182-183) pointed out that his multi-parameter rule must be applied with caution, especially where scale and location parameters occur simultaneously. To counter this problem Jeffreys suggested: "We can then deal with location parameters, on the hypothesis that the scale and numerical parameters are irrelevant to them, by simply taking their prior probability uniform". In deriving Jeffreys priors, we have the experience that problems mainly occur in situations where a location parameter is bounded from below or above (e.g., is greater than zero or is less than the smallest possible observation). For probability distributions with a location parameter being bounded from above (such as the exponential, Rayleigh, gamma, Weibull and generalised gamma), we therefore follow Jeffreys' recommendation and assume the location parameter to be a priori independent of scale and shape parameters and take the uniform prior as a non-informative prior for the location parameter and the Jeffreys prior as a non-informative prior for the combination of scale and shape parameters.

In explaining the derivation of non-informative Jeffreys priors, we refer to Box & Tiao (1973, Section 1.3). Let $\mathbf{x} = (x_1,...,x_n)'$ be a random sample from a multi-parameter probability distribution with likelihood function $\ell\left(x|\boldsymbol{\theta}\right)$. When the probability distribution obeys certain regularity conditions, then for sufficiently large n, the posterior density function of the parametric vector $\boldsymbol{\theta}$ is approximately normal, and remains approximately normal under mild one-to-one transformations of $\mathbf{x} = (x_1,...,x_n)'$. 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

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

and the corresponding non-informative Jeffreys prior is defined by

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

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 $\mathbf{x} = (x_1,...,x_n)'$ 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:

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

where

$$B_{jk} = \frac{\pi(\mathbf{x}|H_j)}{\pi(\mathbf{x}|H_k)}, \quad j, k = 1,\ldots,m, \tag{5}$$

is denoted by the Bayes factor. The posterior probability $p(H_k|\mathbf{x})$ is also called the Bayes weight attached to model k. 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 $\boldsymbol{\theta}_k = (\theta_{1k},...,\theta_{dk})'$ with number of parameters d:

$$\pi(\mathbf{x}|H_k) = \int \ell(\mathbf{x}|\boldsymbol{\theta}_k, H_k)\pi(\boldsymbol{\theta}_k|H_k)d\boldsymbol{\theta}_k, \tag{6}$$

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.

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

$$p(H_k) = w(H_k)\int J(\boldsymbol{\theta}_k|H_k)d\boldsymbol{\theta}_k,$$

where the integral over the non-informative Jeffreys prior $J(\boldsymbol{\theta}_k|H_k)$ 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

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

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. Using Eqs. (4) and (7), the posterior probability of model H_k being correct can now be rewritten as

$$p(H_k|\mathbf{x}) = \frac{\pi(\mathbf{x}|H_k)w(H_k)}{\sum_{j=1}^{m}\pi(\mathbf{x}|H_j)w(H_j)}, \quad k = 1,\ldots,m. \tag{8}$$

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

$$\Pr\{X > x_0|\mathbf{x}\} = \sum_{k=1}^{m} p(H_k|\mathbf{x}) \Pr\{X > x_0|H_k, \mathbf{x}\}, \tag{9}$$

where $\Pr\{X > x_0|H_k, \mathbf{x}\}$ is the predictive probability of exceeding x_0 under likelihood model H_k, $k = 1,\ldots,m$. The predictive exceedance probabilities have been obtained by numerical integration.

5 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. (6) may 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

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

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

$$\pi(\mathbf{x}|H) = \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \ell(\mathbf{x}|\hat{\boldsymbol{\theta}}, H) \tag{11}$$

for $n \to \infty$. The second and third terms on the right-hand side of Eq. (10) 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). Another advantage of the Laplace approximation is the independence of the prior distribution (which, of course, can also be seen as a disadvantage). This approximation appears to work well in practice (see Van Noortwijk et al., 2001). In this paper, the Laplace expansion has been used to approximate the Bayes weights.

6 Discharges of the rivers Rhine and Meuse

In this section, the discharges of the rivers Rhine and Meuse with an average return period of 1,250 years are determined. 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 and of the Meuse at Borgharen at 3,800 m^3/s (Van De Langemheen & Berger, 2002).

6.1 ***River Rhine at Lobith***

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. (8) 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 were attached to the seven individual marginal densities.
This paper extends the Bayesian analysis of Van Noortwijk et al. (2001) in the sense that the generalised gamma and generalised extreme-value distribution have been added and that the statistical uncertainty in the location parameter has been taken into account in a Bayesian manner. In order to improve the Bayes estimates of the exceedance frequencies and to link up with the maximum-likelihood method, the location parameters are assumed to be unbounded from below. As a matter of fact, both the maximum-likelihood and Bayes estimates of the location parameters can be negative. Given equal prior weights, the Bayes posterior weights of the nine 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.

Table 4: Prior and posterior Bayes weights as well as the maximum-likelihood and Bayes estimates of the 1/1,250 quantile for the annual maximum river Rhine discharge [m³/s]

Probability distribution	**Bayes weight**		**1/1,250 quantile estimate**	
	Prior	**Posterior**	**maximum-likelihood**	**Bayes**
Exponential	0.1111	0.0000	>22,000	>22,000
Rayleigh	0.1111	0.2177	15,450	15,868
Normal	0.1111	0.0720	13,363	13,593
Lognormal	0.1111	0.1270	15,243	15,290
Gamma	0.1111	0.1361	15,113	15,153
Weibull	0.1111	0.1769	14,163	14,436
Gumbel	0.1111	0.0858	18,999	19,340
Generalised gamma	0.1111	0.0454	14,287	14,808
Generalised extreme-value	0.1111	0.1391	14,330	15,785
Bayes combination	1.0000	1.0000	15,508	15,941

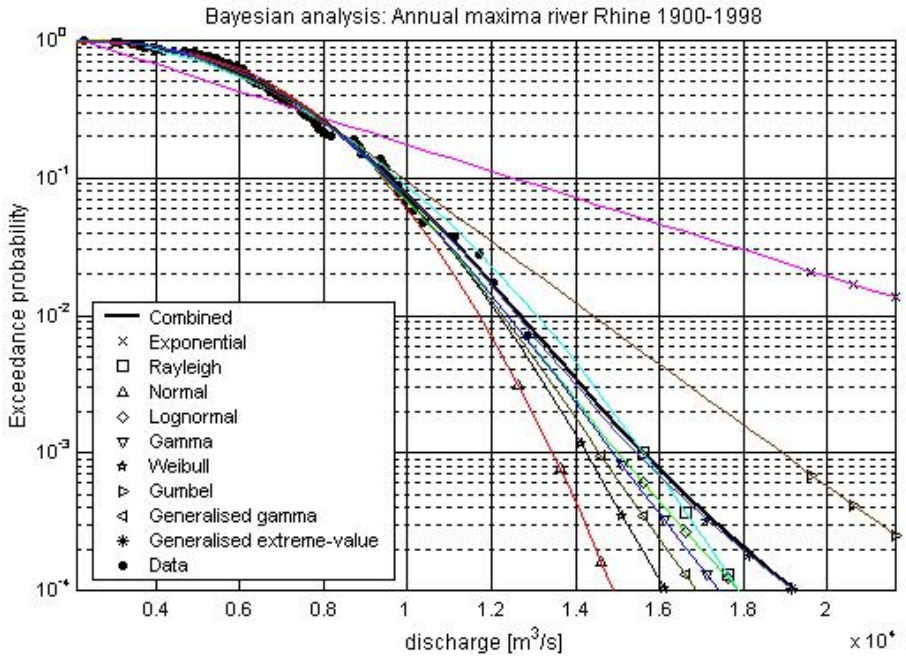

Figure 5: Predictive exceedance probability of annual maximum river Rhine discharge.

Using the Laplace expansion for the Bayes weights, the Rayleigh distribution appears to fit best with a Bayes weight of 22%. The second, third, fourth, and fifth best Bayes fits are the Weibull, generalised extreme-value, gamma, and lognormal distribution.

The Rayleigh distribution is a special case of the Weibull distribution, which also belongs to the family of generalised gamma distributions. Because the Rayleigh distribution fits best with the data, the question now arises how to deal with a probability distribution which is a special case of another probability distribution. The Bayes weights show a tendency to be dependent on the number of parameters of the probability distributions that are considered. That is, if a two-parameter probability distribution (such as the Rayleigh) appears to fit the observations well, then it gets a higher weight than the three-parameter probability distribution (such as the Weibull) of which it is a member. Mathematically this can be illustrated by considering the Laplace expansions of the logarithm of the marginal densities $\ell\left(\mathbf{x}|\hat{\boldsymbol{\theta}}_1, H_1\right)$ and $\ell\left(\mathbf{x}|\hat{\boldsymbol{\theta}}_2, H_2\right)$, where d_1 is the dimension of $\boldsymbol{\theta}_1$ and d_2 the dimension of $\boldsymbol{\theta}_2$. Let model H_1 be a member of model H_2, where dimension d_1 is less than d_2. When the data $\mathbf{x}$ originates from model H_1, the likelihood functions of both models are identical. Hence, the Laplace expansion of the Bayes factor of the two models (5) can be rewritten as

$$\log B_{12} = \log\left(\frac{\pi\left(\mathbf{x}\middle|H_1\right)}{\pi\left(\mathbf{x}\middle|H_2\right)}\right) \approx \log\left[\frac{\ell\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}_1, H_1\right)}{\ell\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}_2, H_2\right)}\right] + \frac{d_2 - d_1}{2}\log\left(\frac{n}{2\pi}\right) = \frac{d_2 - d_1}{2}\log\left(\frac{n}{2\pi}\right).$$

As $n \to \infty$, this Bayes factor signifies that the lower-dimensional model H_1 is relatively more probable than the higher-dimensional model H_2.

Although the generalised extreme-value distribution is especially useful for fitting annual maxima of loads, it receives less weight than would have been expected from the theoretical point of view. This is probably because the conditions for which the generalised extreme-value distribution can be applied don't hold in this situation. In mathematical terms, the generalised extreme-value distribution is obtained as a limiting distribution of the maximum values in a random sample of increasing size. In order to apply the generalised extreme-value distribution, the annual maxima must approximately be composed of an 'infinite' number of loads which can be treated as statistically independent. However, the smaller the units of time for which river discharges are considered, the more dependent they are.

The generalised gamma distribution – with the Rayleigh, Weibull, gamma and lognormal distribution as special cases – gets a very low Bayes weight. Despite the flexibility of the four-parameter generalised gamma distribution, it apparently doesn't fit well with the data and results in a 1/1,250 quantile being too low (14,808 m^3/s). The four-parameter generalised gamma distribution was also studied by Van Noortwijk (2001), who assumed the location parameter to be greater than zero. Other probability distributions with zero or very low Bayes weights are the exponential, normal and Gumbel.

The Bayes estimate of the river Rhine discharge at Lobith with an average return period of 1,250 years is 15,941 m^3/s. Figure 1 shows both the empirical exceedance probability and the predictive exceedance probabilities computed using numerical integration. Using the maximum-likelihood method combined with the Bayes weights, the estimate of the discharge with an average return period of 1,250 years decreases to 15,508 m^3/s. As expected, taking account of parameter uncertainty results in larger design discharges.

*6.2 **River Meuse at Borgharen***

A similar statistical analysis has been applied to the annual maximum discharges of the river Meuse at Borgharen during the period 1911-1998. Using the Laplace expansion for the Bayes weights, the Rayleigh distribution fits best with a Bayes weight of 34%. The Gumbel distribution has a Bayes weight of 13%, the lognormal, gamma and Weibull each a weight of 12%, and the generalised extreme-value a weight of 11%. The remaining probability distributions – the exponential, normal and generalised gamma – have zero or very low Bayes weights.

The Bayes estimate of the river Meuse discharge at Borgharen with an average return period of 1,250 years is 3,846 m^3/s. Figure 2 shows both the empirical exceedance probability and the predictive exceedance probabilities computed using numerical integration. Using the maximum-likelihood method combined with the Bayes weights, the estimate of the discharge with an average return period of 1,250 years decreases to 3,724 m^3/s.

Table 5: Prior and posterior Bayes weights as well as the maximum-likelihood and Bayes estimates of the 1/1,250 quantile for the annual maximum river Meuse discharge [m^3/s]

Probability distribution	**Bayes weight**		**1/1,250 quantile estimate**	
	Prior	**Posterior**	**maximum-likelihood**	**Bayes**
Exponential	0.1111	0.0000	>5,000	>5,000
Rayleigh	0.1111	0.3420	3,546	3,653
Normal	0.1111	0.0281	3,115	3,176
Lognormal	0.1111	0.1205	3,635	3,680
Gamma	0.1111	0.1196	3,577	3,611
Weibull	0.1111	0.1179	3,393	3,512
Gumbel	0.1111	0.1306	4,428	4,518
Generalised gamma	0.1111	0.0328	3,452	3,577
Generalised extreme-value	0.1111	0.1085	3,529	3,944
Bayes combination	1.0000	1.0000	3,724	3,846

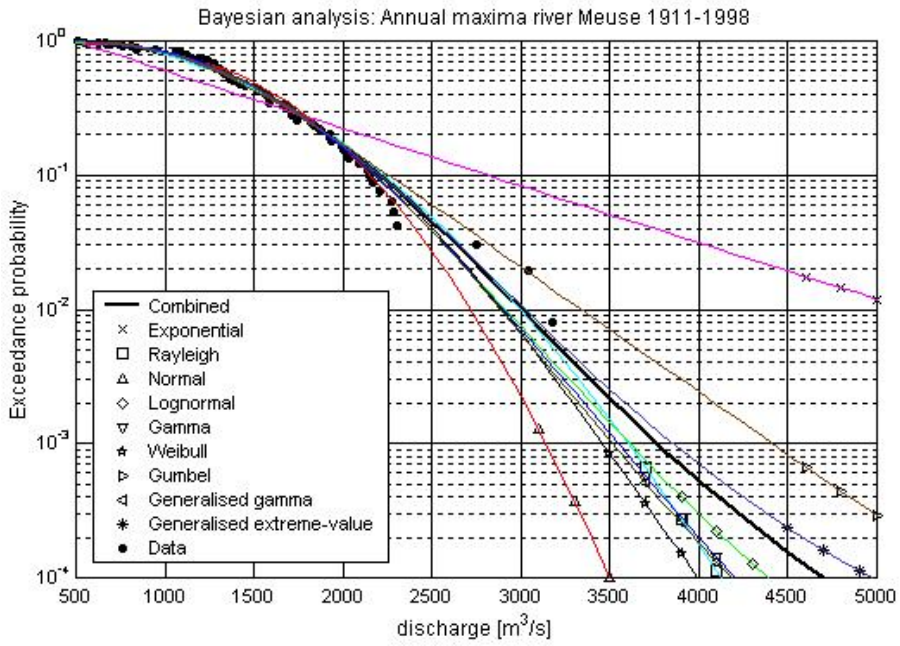

Figure 6: Predictive exceedance probability of annual maximum river Meuse discharge.

7 Conclusions

In this paper, the river discharges of the Rhine at Lobith and the Meuse at Borgharen with an average return period of 1,250 years have been determined taking account of the statistical uncertainties in location, scale and shape parameters. 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-informative Jeffreys prior is recommended. The Bayes weights have been approximated by the Laplace expansion and the predictive exceedance probabilities have been computed using numerical integration. The design discharge increases when taking the statistical uncertainties properly into account. For both rivers, the Rayleigh distribution appears to fit best with the annual maximum discharges. Because the Rayleigh distribution fits well, the method of Bayes weights attaches a higher weight to this lower-dimensional probability distribution than to the higher-dimensional distributions of which it is a member (such as the Weibull and the generalised gamma).

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.

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 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.

Van Noortwijk, J.M. (2001). Bayes estimates of flood quantiles using the generalised gamma distribution. In Y. Hayakawa, T. Irony, and M. Xie, editors, *System and Bayesian Reliability; Essays in Honor of Professor Richard E. Barlow on His 70th Birthday*, pages 351-374. Singapore: World Scientific Publishing.

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. 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".

For probability distributions with a location parameter being bounded from above by the smallest possible observation, we follow the recommendation of Jeffreys (1961, pages 182-183) and assume the location parameter to be a priori independent of scale and shape parameters. The approximate Jeffreys prior then follows by taking the uniform prior as a non-informative prior for the location parameter and the Jeffreys prior as a non-informative prior for the combination of scale and shape parameters.

Exponential distribution

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

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

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

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

Rayleigh distribution

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

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

The approximate Jeffreys prior for the Rayleigh distribution is

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

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 Jeffreys prior for the normal distribution is

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

Lognormal distribution

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

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

Hence, if $\log(X - a)$ has a normal distribution, then $X - a$ has a lognormal distribution. The approximate Jeffreys prior for the lognormal distribution is

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

Gamma distribution

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

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

where

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

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

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

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 $a > 0$, scale parameter $b > 0$, and location parameter c if the probability density function of X is given by

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

The approximate Jeffreys prior for the Weibull distribution is

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

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

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

Generalised gamma distribution

A random variable X has a generalised gamma distribution with scale parameter $b > 0$, shape parameters $a > 0$ and $c > 0$, and location parameter d if the probability density function of X is given by

$$\mathrm{Gga}\left(x\middle|a,b,c,d\right)=\frac{c}{b\Gamma(a)}\left[\frac{x-d}{b}\right]^{ca-1}\exp\left\{-\left[\frac{x-d}{b}\right]^{c}\right\}I_{(d,\infty)}(x)\,.$$

The approximate Jeffreys prior for the generalised gamma distribution is

$$J(a,b,c,d)\approx J(a,b,c)=\frac{\sqrt{[a\psi'(a)]^2-\psi'(a)-1}}{b}\,.$$

Generalised extreme-value distribution

A random variable *X* has a generalised extreme-value distribution with location parameter *a*, scale parameter $b>0$ and shape parameter *c* if the probability density function of *X* is given by

$$\mathrm{Gev}\left(x\middle|a,b,c\right)=\begin{cases}\dfrac{1}{b}\left[1-\dfrac{c(x-a}{b}\right]_{+}^{\frac{1}{c}-1}\ exp\left\{-\left[1-\dfrac{c(x-a)}{b}\right]_{+}^{\frac{1}{c}}\right\}, & c\neq 0,\quad b>0,\\ \dfrac{1}{b}\exp\left\{-\dfrac{x-a}{b}\right\}\exp\left\{-\exp\left\{-\dfrac{x-a}{b}\right\}\right\}, & c=0,\quad b>0,\end{cases}$$

where $[y]_{+}=\max\{0,y\}$. The variable *x* is bounded by *a* + *b/c* from below for *c* < *0* and from above for *c* > 0; that is, $a+b/c<x<\infty$ for $c<0$ and $-\infty<x<a+b/c$ for $c>0$. The case *c* = 0, which is the Gumbel distribution, is the limiting distribution as $c\to 0$ where $-\infty<x<\infty$. For $c<1/2$, the Jeffreys prior of the generalised extreme-value distribution is

$$J(a,b,c)=\frac{|1-c|}{(bc)^2}\sqrt{\left\{\Gamma(1-2c)-\Gamma^2(1-c)\right\}\frac{\pi^2}{6}-\Gamma^2(1-c)\left\{\psi(1-c)+\gamma\right\}^2}$$

with Euler's constant $\gamma=0.5772$.