# Bayes Estimates of Flood Quantiles using the Generalised Gamma Distribution*


Jan M. van Noortwijk**


July 20, 2004


**Abstract**

In this paper, a Bayesian approach is proposed to estimate flood quantiles while taking statistical uncertainties into account. Predictive exceedance probabilities of annual maximum discharges are obtained using the three- and four-parameter generalised gamma distribution (without and with location parameter respectively). The parameters of this distribution are assumed to be random quantities rather than deterministic quantities and to have a prior joint probability distribution. On the basis of observations, this prior joint distribution is then updated to the posterior joint distribution by using Bayes' theorem. An advantage is that the generalised gamma distribution fits well with the stage-discharge rating curve being an approximate power law between water level and discharge. Furthermore, since the generalised gamma distribution has three or four parameters, it is flexible in fitting data. Many well-known probability distributions which are commonly used to estimate quantiles of hydrological random quantities are special cases of the generalised gamma distribution. As an example, a Bayesian analysis of annual maximum discharges of the Rhine River at Lobith is performed to determine flood quantiles including their uncertainty intervals. The generalised gamma distribution can also be applied in lifetime and reliability analysis.



**Keywords**

Bayesian analysis; generalised gamma distribution; flood quantiles; Jeffreys prior; location parameter; river discharge; stage-discharge rating curve.


---


*Published 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, 2001.



**HKV Consultants, P.O. Box 2120, NL-8203 AC Lelystad, The Netherlands; Delft University of Technology, Department of Electrical Engineering, Mathematics and Computer Science, P.O. Box 5031, NL-2600 GA Delft, The Netherlands; E-mail: j.m.van.noortwijk@hkv.nl.

# 1 Introduction

The Dutch river dikes have to withstand water levels and discharges with an average return period up to 1250 years, where a downstream water level can be determined on the basis of the upstream discharge by using a river flow simulation model (see WL & EAC/RAND [34] and Walker et al. [33]). Therefore, a design discharge is defined as the annual maximum river discharge for which the probability of exceedance is 1/1250 per year. A common problem in obtaining design discharges is that there is a limited amount of observations available (for example, with respect to the Dutch Rhine River there are only 98 annual maximum discharges available).

The design discharge is usually estimated by extrapolating a probability distribution which is fitted to the observed discharges. The two-parameter gamma distribution (Pearson type III distribution) is among the distributions commonly used. When more shape flexibility is needed to fit the data, Ashkar & Ouarda [2] suggest the three-parameter generalised gamma distribution. In order to take the uncertainties into account when there is a small amount of data, they developed approximate confidence intervals for quantiles of the generalised gamma distribution. These confidence intervals have been obtained by applying techniques from classical statistics.

In this paper, an alternative way to take the statistical uncertainties into account is proposed: regarding the statistical parameters as being random quantities rather than deterministic quantities. On the basis of the observed annual maximum discharges, the prior density of these random quantities can be updated to the posterior density by using Bayes' theorem. In order to describe the a priori 'lack of knowledge', we use the non-informative Jeffreys prior for the scale and shape parameters of the three-parameter generalised gamma distribution. This Bayesian analysis is extended to estimating the location parameter of the four-parameter generalised gamma distribution as well.

To fit a generalised gamma distribution to discharge data, we refer to the relation between this distribution and the stage-discharge rating curve (an approximate power law in which the discharge is expressed in terms of the water level). Although it was not possible to derive the generalised gamma distribution solely on the basis of this physical law, the reason for using this distribution is in the spirit of the so-called 'engineering probability' (see Barlow [3] and Mendel [24]). According to Mendel [24] "engineering probability concerns the derivation of probabilistic models from the physical laws, geometric constraints, and other engineering knowledge concerning an engineering system". The observations that have been analysed are the annual maximum discharges of the Dutch Rhine River at Lobith (near the Dutch-German border).

This paper is set out as follows. In Section 2, an approximate rating curve is derived for the Rhine River at Lobith. The mathematical procedure to apply a Bayesian analysis to the observed annual maximum discharges can be found in Section 3. The Jeffreys prior of the three-parameter generalised gamma distribution is derived in Section 4. The posterior density of the scale and shape parameters of the three-parameter generalised gamma distribution is obtained in Section 5. Section 6 is devoted to taking account of the statistical uncertainty in the location parameter of the four-parameter generalised gamma distribution. Results for the Rhine River are presented in Section 7 and conclusions are in Section 8.

## 2 Stage-discharge rating curve

Although a flood wave is a gradually varied unsteady non-uniform flow, the uniform-flow condition is frequently assumed in the computation of flow in rivers. The results obtained from this assumption are recognised to be approximate and general, but they offer a relatively simple and satisfactory solution to many practical problems.

Under the condition of uniform open-channel flow, the average discharge can be determined using Manning's equation:

$$q = v\tilde{a} = \frac{\tilde{r}^{2/3}\tilde{a}\sqrt{s}}{n_m} = \frac{\tilde{a}^{5/3}\sqrt{s}}{\tilde{p}^{2/3}\,n_m},$$

where $q$ = discharge [m$^3$/s], $v$ = mean velocity [m/s], $\tilde{a}$ = cross-sectional flow area [m$^2$], $\tilde{p}$ = wetted perimeter [m], $\tilde{r} = \tilde{a}/\tilde{p}$ = hydraulic radius or hydraulic mean depth [m], $s$ = channel slope [-], and $n_m$ = Manning roughness coefficient [s/m$^{1/3}$] (see e.g. Chow [9, Chapters 5-6] or Chow, Maidment & Mays [10, Chapter 2]).

In the event of extreme discharges, a river breaks its banks and inundates the flood plain. Since the total width of the flood plain of the Rhine at Lobith is about 1150 m (including the main channel), the inundated flood plain can approximately be regarded as a wide rectangular cross-section with a width $w$ [m] much larger than its water depth $\tilde{d}$ [m]. In this situation, the hydraulic radius can be approximated by

$$\tilde{r} = \frac{\tilde{a}}{\tilde{p}} = \frac{w\tilde{d}}{w + 2\tilde{d}} \approx \tilde{d}$$

and, accordingly,

$$q \approx \frac{w\tilde{d}^{5/3}\sqrt{s}}{n_m}. \tag{1}$$

The stage-discharge rating curve (1) suggests that the rating curve can be approximated by the following power law between water level and discharge (see Shaw [29, Chapter 6] and Chow, Maidment & Mays [10, Chapter 9]):

$$q - q_0 \approx \alpha\,[h - h_0]^{\beta}, \qquad q \geq q_0,\ h \geq h_0,\ \alpha, \beta > 0, \tag{2}$$

where $q$ = discharge [m$^3$/s], $q_0$ = threshold value discharge [m$^3$/s], $h$ = water level [m +NAP], $h_0$ = threshold value water level [m +NAP], $\tilde{d} = h - h_0$ [m], and 1 m +NAP means 1 m above 'normal Amsterdam level'. By applying rating curve (2), regularly observed or continuously recorded water levels can be converted to corresponding discharge estimates. Hydrological statistical analyses are mostly performed on the basis of discharge data. As a result of Eq. (2), and by assuming the location parameter $q_0$ to be known, we fit a probability distribution to those discharges which are higher than the threshold $q_0 = 1750$ m$^3$/s (or, in other words, water levels which are higher than $h_0 = 9.0$ m +NAP). The threshold $q_0$ has been determined by maximising the marginal density of the observations in Eq. (5).

With the aid of a least-squares method, $\alpha$ and $\beta$ in Eq. (2) have been fitted to measurements and extrapolations of the Dutch Ministry of Transport, Public Works, and Water Management [26]. As it can be seen in Figure 1, the estimated values $\alpha = 207.9$ and $\beta = 1.917$ result in a good approximation of the actual rating curve of the Rhine at Lobith, especially for extreme discharges.

# 3 Bayesian analysis of discharges

In this section, it is shown that the generalised gamma distribution fits well with the stage-discharge rating curve in terms of the power law between water level and discharge. As a matter of fact, if the discharge is assumed to have a generalised

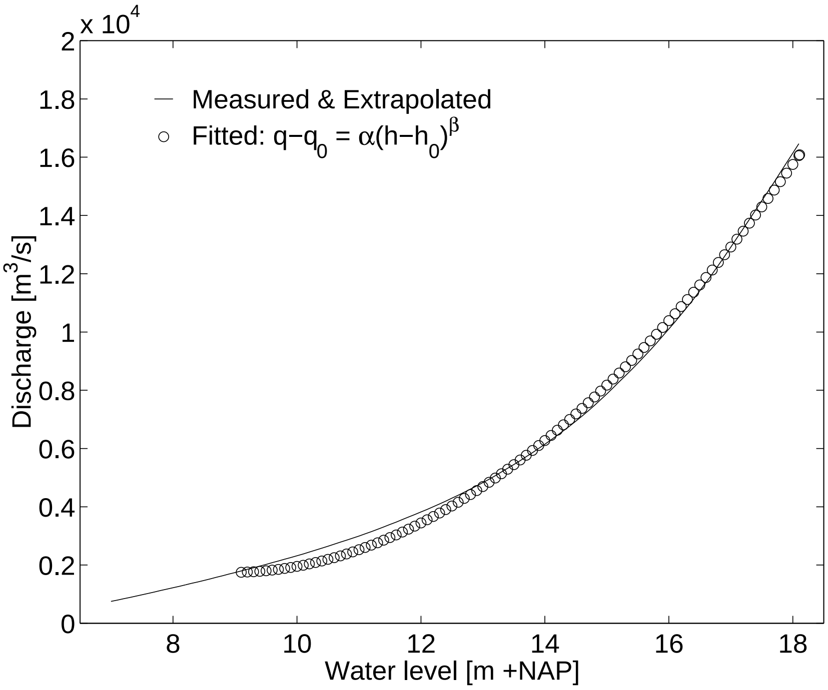


Figure 1: Actual (measured and extrapolated) and fitted rating curve of the Rhine River at Lobith.

gamma distribution and if the power law holds, then the water level has a generalised gamma distribution as well (and this also applies in reverse). Furthermore, we regard the scale and shape parameters of the three-parameter generalised gamma distribution to be unknown, having a joint probability distribution.

Because the rating curve (2) approximately holds for the thresholds $q_0$ and $h_0$, the following question must be answered. What is the probability distribution of the annual maximum discharge $Q$ for $Q \geq q_0$ while taking the statistical uncertainties into account? (or, in other words, what is the probability distribution of the annual maximum water level $H$ for $H \geq h_0$?) This question can be answered using Bayesian statistics (see e.g. Bernardo & Smith [6] and Carlin & Louis [8]).

The probability distribution of the annual maximum discharge at Lobith is chosen so that the type of distribution remains the same when the transformation (2) is applied from discharge to water level. If the probability density functions of the discharge $Q$ and the water level $H$ are given by $p(q - q_0)$ and $\tilde{p}(h - h_0)$, respectively, then

$$\tilde{p}\,(h - h_0) = \alpha\beta\,[h - h_0]^{\beta - 1}\,p\left(\alpha\,[h - h_0]^{\beta}\right)$$

for $h \geq h_0$ or, in terms of the discharge,

$$p\,(q - q_0) = \tfrac{1}{\beta}\alpha^{-\frac{1}{\beta}}\,[q - q_0]^{\frac{1}{\beta} - 1}\,\tilde{p}\left(\alpha^{-\frac{1}{\beta}}\,[q - q_0]^{\frac{1}{\beta}}\right)$$

for $q \geq q_0$. A probability density function of $Q$ for which $p(\cdot)$ and $\tilde{p}(\cdot)$ belong to the same type of distribution is the three-parameter generalised gamma distribution (see

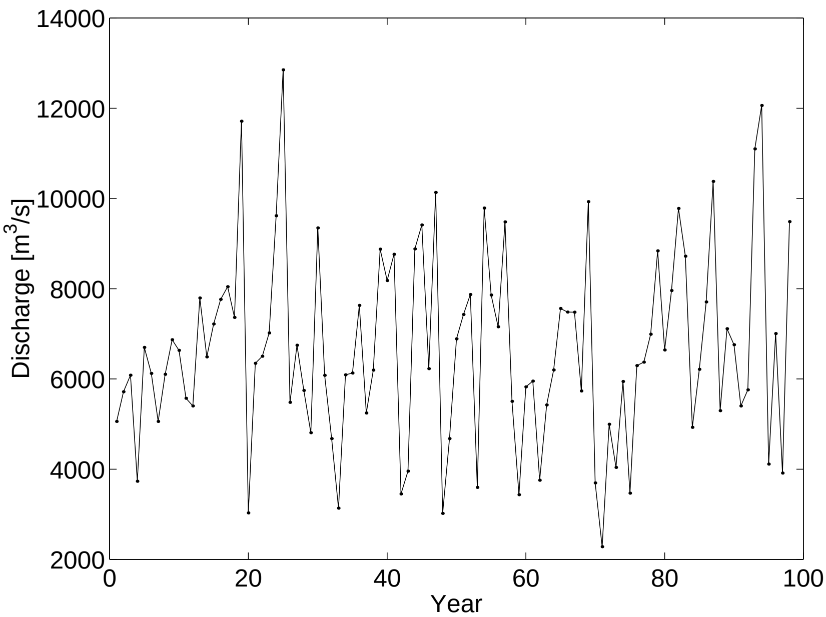


Figure 2: Observed annual maximum discharges of the Rhine River at Lobith during 1901-1998.

Stacy [31], Parr & Webster [27], Lienhard & Meyer [21], Hager & Bain [13] and Johnson, Kotz & Balakrishnan [17, Chapter 17]):

$$\ell\,(q - q_0|\, a, b, c) = \mathrm{Gga}\,(q - q_0|\, a, b, c) =$$
$$= \frac{c}{b\Gamma(a)} \left[\frac{q - q_0}{b}\right]^{ca-1} \exp\left\{- \left[\frac{q - q_0}{b}\right]^{c}\right\} I_{(0,\infty)}(q - q_0) \tag{3}$$

with parameters $a, b, c > 0$. Indeed, the probability density function of $h$ is also a generalised gamma distribution:

$$\ell\left(h - h_0 \,\middle|\, a, [b/\alpha]^{1/\beta}, \beta c\right) = \mathrm{Gga}\left(h - h_0 \,\middle|\, a, [b/\alpha]^{1/\beta}, \beta c\right).$$

The three-parameter generalised gamma distribution has been successfully fitted to annual maximum discharges by Ashkar & Ouarda [2]. Since it has three parameters, it includes many well-known probability distributions like the exponential distribution ($a = 1$ and $c = 1$), the gamma distribution ($c = 1$), the chi-square distribution ($a = t/2$, $b = 2$, and $c = 1$), the Weibull distribution ($a = 1$), the $l_c$-isotropic distribution ($a = 1/c$), the Rayleigh distribution ($a = 1$ and $c = 2$), and the Maxwell distribution ($a = 3/2$ and $c = 2$). In addition, the lognormal distribution is a limiting special case when $a$ tends to infinity. The proof follows by noting that the logarithm of a

Table 1: Cross-sectional parameters, as well as Bayes estimates of $a$, $b$, and $c$ for the three-parameter generalised gamma distribution.

| parameter | description | value | dimension |
|---|---|---|---|
| $w$ | width main channel and flood plain | 1147 | m |
| $\alpha$ | parameter rating curve | 207.9 | - |
| $\beta$ | parameter rating curve | 1.917 | - |
| $q_0$ | threshold value discharge | 1750 | m$^3$/s |
| $h_0$ | threshold value water level | 9.015 | m +NAP |
| $n$ | number of observations | 98 | - |
| $a_L$ | lower bound for $a$ | 0.01 | - |
| $a_U$ | upper bound for $a$ | 6 | - |
| $k$ | number of subdivisions for $a$ | 100 | - |
| $c_L$ | lower bound for $c$ | 0.01 | - |
| $c_U$ | upper bound for $c$ | 6 | - |
| $m$ | number of subdivisions for $c$ | 100 | - |
| $E(A\|\, \mathbf{q} - q_0)$ | posterior expectation of $a$ | 1.380 | - |
| $E(B\|\, \mathbf{q} - q_0)$ | posterior expectation of $b$ | 4936 | m$^3$/s |
| $E(C\|\, \mathbf{q} - q_0)$ | posterior expectation of $c$ | 2.310 | - |

random quantity with a generalised gamma distribution has a generalised Gompertz distribution. Ahuja & Nash [1] proved that the so-obtained generalised Gompertz distribution is asymptotically normal as $a \to \infty$. An interesting physical characterisation of generalised gamma distributions on the basis of statistical mechanics can be found in Lienhard & Meyer [21]. As a special case, Lienhard [20] derived a generalised gamma distribution (with $c = 2$) to describe rainfall run-off from a watershed.

Subsequently, we determine the moments and the exceedance probability of the generalised gamma distribution. For this purpose, the incomplete gamma function is defined as

$$\Gamma(x,y) = \int_{t=y}^{\infty} t^{x-1} e^{-t}\, dt, \qquad x > 0,\ y \geq 0.$$

By means of the gamma function, $\Gamma(x) = \Gamma(x,0)$, the $n$th moment of $Q - q_0$ can be written as

$$E\left([Q - q_0]^n \,|\, a,b,c\right) = b^n\, \Gamma\left(\tfrac{n}{c} + a\right) \Big/ \Gamma\left(a\right)$$

for $q \geq q_0$ and $n \geq 0$. Similarly, the conditional exceedance probability follows:

$$\Pr\left\{Q > q |\, a,b,c\right\} = 1 - \Gamma\left(a, \left[\frac{q - q_0}{b}\right]^c\right) \Big/ \Gamma\left(a\right) \tag{4}$$

for $q \geq q_0$. Note that the generalised gamma distribution takes its name from the fact that $Y = [(Q - q_0)/b]^c$ has a gamma distribution with shape parameter $a$ and scale parameter 1. A random quantity $X$ has a gamma distribution with shape parameter $\nu > 0$ and scale parameter $\mu > 0$ if its probability density function is given by:

$$\mathrm{Ga}(x|\, \nu,\mu) = [\mu^\nu / \Gamma(\nu)]\, x^{\nu-1} \exp\{-\mu x\}\, I_{(0,\infty)}(x).$$

Recall that $n$ annual maximum discharges have been observed. Since these discharges are conditionally independent when the values of the parameters $a$, $b$ and $c$ are given, the likelihood function of the observations $q_j$, $j = 1,\ldots,n$, can be written as

$$\ell\left(\mathbf{q} - q_0 |\, a,b,c\right) = \prod_{j=1}^n \ell\left(q_j - q_0 |\, a,b,c\right),$$

where $\mathbf{q} = (q_1,\ldots,q_n)'$. In order to quantify the uncertainty in the parameters $a$, $b$, and $c$, we assume that they have a prior probability distribution. The marginal density $\pi(\mathbf{q} - q_0)$ of the observations $\mathbf{q}$ is obtained by integrating the likelihood over $(a,b,c)$:

$$\pi(\mathbf{q} - q_0) = \int_{a=0}^{\infty} \int_{b=0}^{\infty} \int_{c=0}^{\infty} \ell(\mathbf{q} - q_0 |\, a,b,c) \pi(a,b,c)\, da db dc. \tag{5}$$

# 4 Non-informative Jeffreys prior

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 unity. 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 [4] who considered a uniform prior. However, the use of uniform priors is criticised because of a lack of invariance under one-to-one transformations. For example, let us consider an unknown parameter $\theta$ 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 $\theta$ is chosen to be uniform (constant), then the non-informative prior for $\phi$ should be proportional to $1/\phi$ to maintain consistency. Unfortunately, we cannot maintain consistency and choose both the non-informative priors for $\theta$ and $\phi$ to be constant.

An illustration of the danger of using uniform distributions as non-informative priors for the parameters $(a, b, c)$ of the generalised gamma distribution is given below. Let us make a change of variables from $(a, b, c)$ to $(\theta, \lambda, \phi) = (a^{-1}, ac, b^{-c})$ and assume the non-informative prior of $(\theta, \lambda, \phi)$ to be $\pi(\theta, \lambda, \phi) \propto \phi^{-1}$. Hence, the marginal priors of $\theta$, $\lambda$, and $\log\phi$ are uniform. Accounting for the Jacobian of this transformation, the corresponding non-informative prior of $(a, b, c)$ has the form $\pi(a, b, c) \propto (c/a) \cdot b^{-1}$. The marginal prior of $c$ is all but non-informative!

The physicist Sir Jeffreys [16, Chapters 3-4] was the first to produce an alternative to solely using uniform non-informative priors. His main reason for deriving non-informative priors (now 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 discussion going on as to whether the multivariate Jeffreys rule is appropriate (see e.g. Kass & Wasserman [19]). We concur with the following statement made by Dawid [12]: "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". When a location parameter, which is bounded from below or above, is involved, it may be recommended to modify the Jeffreys prior (see Section 6). The main advantage of the Jeffreys prior is that it is always both dimensionless and invariant under transformations.

As an example, the multivariate Jeffreys prior for the normal model with unknown mean $\mu$ and unknown standard deviation $\sigma$ 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 $\sigma$ have the same units. For another example, see the Jeffreys prior of the three-parameter generalised gamma distribution in Eq. (7). Because the non-dimensionality argument is sound (from a physics point of view), the multivariate Jeffreys prior is used as a non-informative prior for the three-parameter generalised gamma distribution.

To explain the derivation of non-informative Jeffreys priors, see Box & Tiao [7, Section 1.3]. Let $\mathbf{x} = (x_1, \ldots, x_n)'$ be a random sample from a multi-parameter probability distribution with likelihood function $\ell(x|\boldsymbol{\theta})$ with parametric vector $\boldsymbol{\theta}$. When the probability distribution obeys certain regularity conditions, then for sufficiently large $n$, the posterior density function of $\boldsymbol{\theta}$ is approximately normal and remains approximately normal under mild one-to-one transformations of $\boldsymbol{\theta}$. 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 measure for a single observation. In mathematical terms, the elements of this matrix are

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

where $r$ is the number of parameters. The corresponding non-informative Jeffreys prior is defined by the square root of the determinant of Fisher's information matrix for a single observation:

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

According to Hager & Bain [13], the Fisher information matrix of the three-parameter generalised gamma distribution is found to be

$$I(a,b,c) = \begin{pmatrix} \psi'(a) & \dfrac{c}{b} & -\dfrac{\psi(a)}{c} \\ \dfrac{c}{b} & \dfrac{c^2 a}{b^2} & -\dfrac{1+a\psi(a)}{b} \\ -\dfrac{\psi(a)}{c} & -\dfrac{1+a\psi(a)}{b} & \dfrac{1+2\psi(a)+a\psi'(a)+a[\psi(a)]^2}{c^2} \end{pmatrix}. \tag{6}$$

After straightforward algebra, the determinant of Fisher's information matrix can be written as

$$|I(a,b,c)| = \frac{[a\psi'(a)]^2 - \psi'(a) - 1}{b^2}.$$

Hence, the Jeffreys prior of the three-parameter generalised gamma distribution is

$$J(a,b,c) = \sqrt{|I(a,b,c)|} = \frac{\sqrt{[a\psi'(a)]^2 - \psi'(a) - 1}}{b} = \frac{J(a)}{b}. \tag{7}$$

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 [5] and Schneider [28], respectively.

# 5 Posterior density

On the basis of the observations, we obtain the posterior conditional probability density function of $b$ when the values of $a$ and $c$ are given, as well as the posterior joint probability density function of $a$ and $c$. On the basis of the Jeffreys prior for the generalised gamma distribution, the posterior distribution of $b$, given $(a, c, \mathbf{q} - q_0)$, can be expressed in explicit form:

$$\pi\left(b \middle| a, c, \mathbf{q} - q_0\right) = \tag{8}$$
$$= \frac{\left(\sum_{j=1}^{n}\left[q_j - q_0\right]^c\right)^{na}}{\Gamma\left(na\right)} c b^{-nca-1} \exp\left\{-\sum_{j=1}^{n}\left[\frac{q_j - q_0}{b}\right]^c\right\}.$$

In a similar manner, the marginal density of the observations, given $a$ and $c$, can be expressed in explicit form

$$\pi\left(\mathbf{q} - q_0 \middle| a, c\right) = \tag{9}$$
$$= \int_{b=0}^{\infty} \ell\left(\mathbf{q} - q_0 \middle| a, b, c\right) b^{-1}\, db = c^{n-1} \cdot \frac{\Gamma\left(na\right)}{\left[\Gamma\left(a\right)\right]^n} \cdot \frac{\prod_{j=1}^{n}\left[q_j - q_0\right]^{ca-1}}{\left(\sum_{j=1}^{n}\left[q_j - q_0\right]^c\right)^{na}}.$$

Since the posterior joint probability density function of $a$ and $c$ cannot be expressed in explicit form, we have to resort to approximations. For this, it is convenient to 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 [30] and Mazzuchi & Soyer [22, 23]. Using Eq. (9) and Bayes' theorem, it follows that

$$p\left(a_h, c_i \middle| \mathbf{q} - q_0\right) = \frac{\pi\left(\mathbf{q} - q_0 \middle| a_h, c_i\right) p\left(a_h\right) p\left(c_i\right)}{\sum_{h=1}^{k}\sum_{i=1}^{m} \pi\left(\mathbf{q} - q_0 \middle| a_h, c_i\right) p\left(a_h\right) p\left(c_i\right)}, \tag{10}$$

where

$$\begin{aligned} a_h &= a_L + \left[(2h-1)/2\right] \cdot \left[(a_U - a_L)/k\right], \quad h = 1, \ldots, k, \\ c_i &= c_L + \left[(2i-1)/2\right] \cdot \left[(c_U - c_L)/m\right], \quad i = 1, \ldots, m, \end{aligned}$$

with $a_L$ and $a_U$ being the lower and upper bounds for $a$, and $c_L$ and $c_U$ being the lower and upper bounds for $c$. Suitably wide integration bounds may be determined on the basis of an approximate posterior density such as a (transformed) normal density with a mean equal to the maximum-likelihood estimator $\hat{\boldsymbol{\theta}}$ and a covariance matrix equal to $[nI(\hat{\boldsymbol{\theta}})]^{-1}$, being the inverse Fisher information matrix for a sample of $n$ observations evaluated at $\hat{\boldsymbol{\theta}}$. Convergence to normality of the approximate posterior density can often be improved by transformation (e.g., by taking the logarithm of a non-negative parameter).

The non-informative prior probability functions of $a$ and $c$ are based on the Jeffreys prior (7); that is,

$$\begin{aligned} p(a_h) &= \Pr\{A = a_h\} = J(a_h)/[\textstyle\sum_{h=1}^{k} J(a_h)], \qquad h = 1, \ldots, k, \\ p(c_i) &= \Pr\{C = c_i\} = 1/m, \qquad i = 1, \ldots, m. \end{aligned}$$

Alternative forms of the prior distributions of $a$, $b$, and $c$ are also possible; they can be informative rather than non-informative as well.

Finally, the predictive expected exceedance probability of the annual maximum discharge can be determined by integrating Eq. (4) over the random quantities $a$, $b$, and $c$:

$$\Pr\{Q > q\} = \sum_{h=1}^{k} \sum_{i=1}^{m} \int_{b=0}^{\infty} \Pr\{Q > q |\, a_h, c_i, b\}\, p\left(a_h, c_i, b |\, \mathbf{q} - q_0\right)\, db, \tag{11}$$

where

$$p\left(a_h, c_i, b |\, \mathbf{q} - q_0\right) = \pi\left(b |\, a_h, c_i, \mathbf{q} - q_0\right) p\left(a_h, c_i |\, \mathbf{q} - q_0\right) \tag{12}$$

(the product of Eqs. (8) and (10), respectively) for $h = 1, \ldots, k$ and $i = 1, \ldots, m$. Note that the prior independence between the random quantities $a$, $b$, and $c$ converts to posterior dependence given the observations.

# 6 Location parameter

This section discusses how to extend the Bayesian analysis from the three-parameter generalised gamma distribution (without location parameter) to the four-parameter generalised gamma distribution (with location parameter). The likelihood function of the four-parameter generalised gamma distribution with scale parameter $b$, shape parameters $a$ and $c$, and location parameter $d$ is defined as (see Harter [14])

$$\begin{aligned} &\ell\left(q |\, a, b, c, d\right) = \mathrm{Gga}\left(q - d |\, a, b, c\right) = \\ &= \frac{c}{b\Gamma(a)} \left[\frac{q-d}{b}\right]^{ca-1} \exp\left\{- \left[\frac{q-d}{b}\right]^{c}\right\} I_{(d,\infty)}(q). \end{aligned} \tag{13}$$

For determining the Jeffreys prior in the presence of location parameter $d$, we have to extend Fisher's information matrix with seven elements (see the Appendix) of which three are equal to each other (due to the symmetry of Fisher's information matrix). Theoretically speaking, the Jeffreys prior of the four-parameter generalised gamma distribution with location parameter $d$ may be obtained by taking the square root of the determinant of the four-dimensional Fisher information matrix (whose elements are given in Eq. (6) and in the Appendix). Unfortunately, a serious disadvantage of the so-obtained Jeffreys prior is that it is only defined for $ac > 2$ (see the Appendix) thus possibly excluding valuable distributions such as the exponential and Rayleigh distribution. For this reason, maximum-likelihood estimation for the four-parameter generalised gamma distribution is called *regular* if and only if the product of the two shape parameters is greater than two. For the definition of regular estimation problems, see e.g. Cox & Hinkley [11, pages 106-113].

The problem of a non-existing Jeffreys prior notably arises when parameters of different kinds are considered simultaneously. Jeffreys [16, pages 182-183] pointed out that his multi-parameter rule must be applied with caution, especially where scale and location parameters co-exist. 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, serious problems may occur in situations where a location parameter is bounded from below or above (e.g., is non-negative).

For the four-parameter generalised gamma distribution with a location parameter being bounded from below, we follow Jeffreys' recommendation and assume the location parameter $d$ to be a priori independent of the scale and shape parameters $(a, b, c)$ and take the uniform prior as a non-informative prior for $d$ and the Jeffreys prior (7) as a non-informative prior for $(a, b, c)$; that is,

$$J(a, b, c, d) = J(a, b, c) = \frac{J(a)}{b}. \tag{14}$$

On the basis of the Jeffreys prior for the four-parameter generalised gamma distribution, the posterior distribution of $(a, b, c, d)$ given the observations $\mathbf{q}$ remains to be calculated. This posterior can be determined by extending the formulas in Section 5 with a summation over the discretised prior distribution of $d$, which is defined by

$$p(d_j) = \Pr\{D = d_j\} = 1/l, \quad j = 1, \ldots, l,$$

where

$$d_j = d_L + [(2j - 1)/2] \cdot [(d_U - d_L)/l], \quad j = 1, \ldots, l,$$

with $d_L$ and $d_U$ being the lower and upper bounds for $d$. Obviously, river flow physics suggests the lower bound of $d$ to be zero.

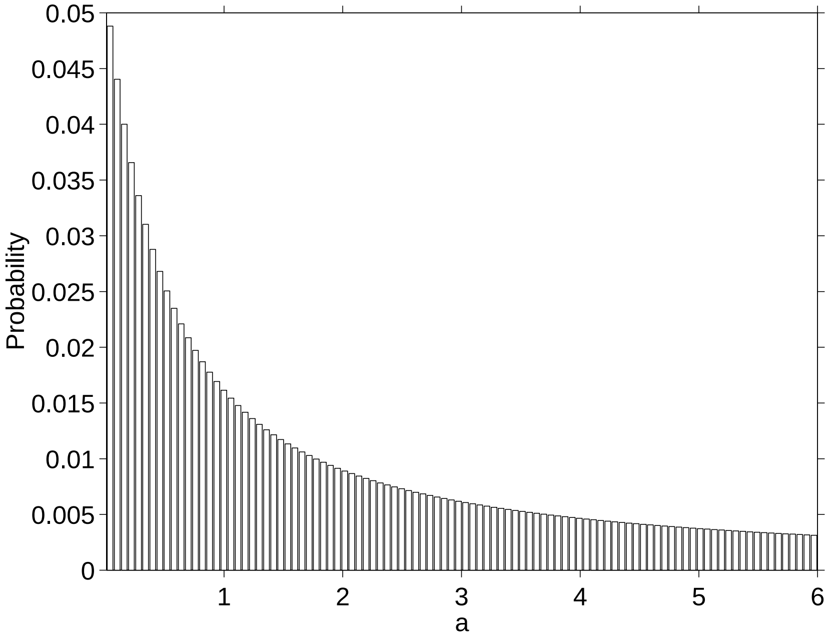


Figure 3: Prior probability function of $a$: $p(a_h)$, $h = 1, \ldots, k$.

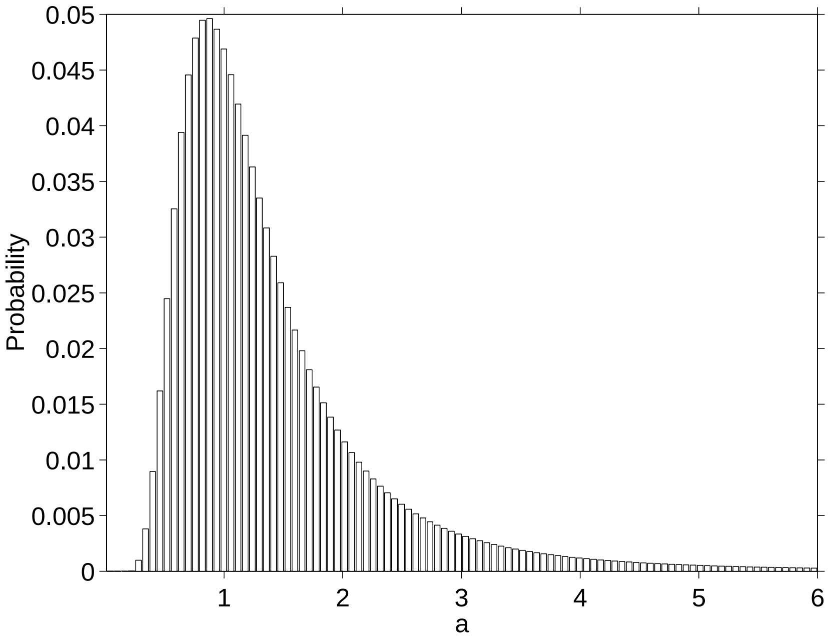


Figure 4: Posterior probability function of $a$: $p(a_h | \mathbf{q} - q_0)$, $h = 1, \ldots, k$.

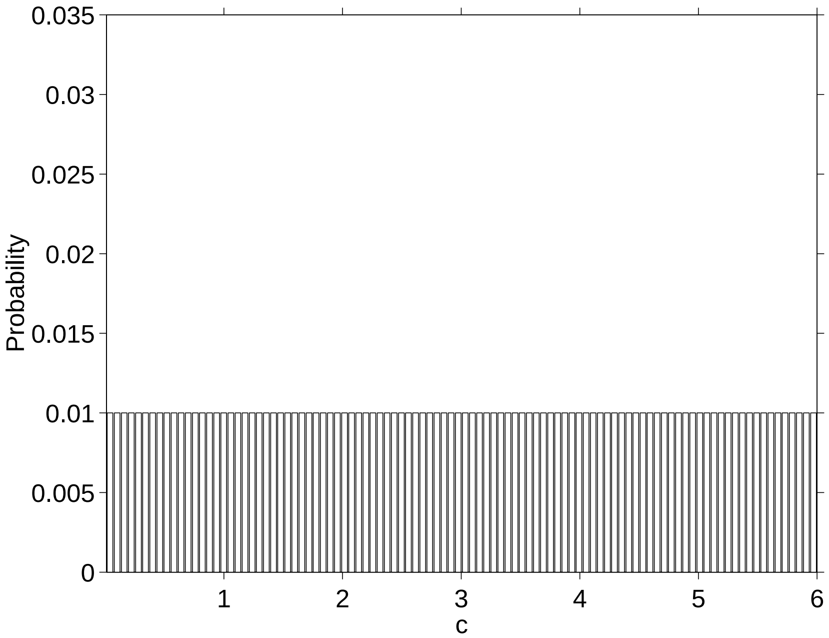


Figure 5: Prior probability function of $c$: $p(c_i)$, $i = 1, \ldots, m$.

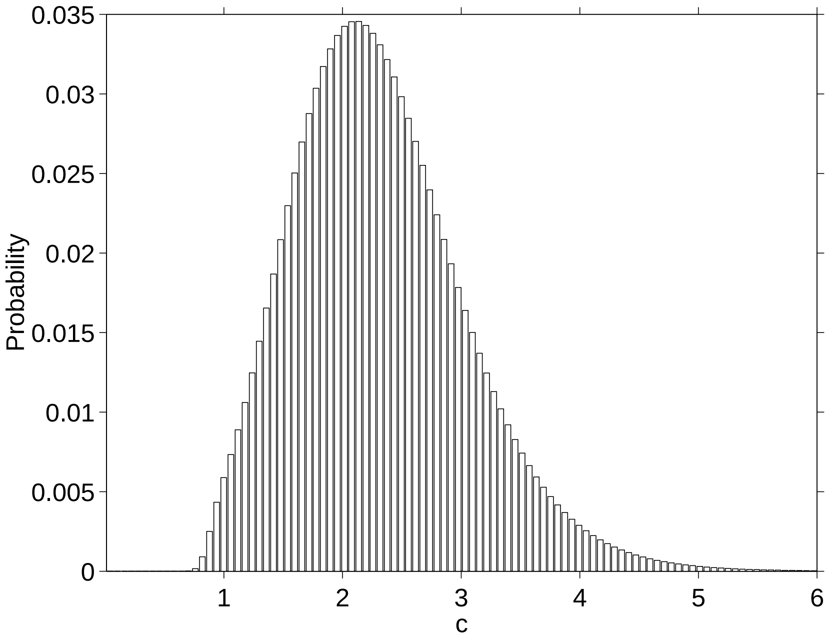


Figure 6: Posterior probability function of $c$: $p(c_i | \, \mathbf{q} - q_0)$, $i = 1, \ldots, m$.

# 7 Results: Design discharge of the Rhine River

The Bayesian updating approach has been applied to the annual maximum discharges of the Rhine River at Lobith during the period 1901-1998 (see Figure 2). The four largest discharges in descending order are 12,849 $m^3/s$ (1926), 12,060 $m^3/s$ (1995), 11,712 $m^3/s$ (1920), and 11,100 $m^3/s$ (1993). The Bayesian analysis of the annual maxima has been performed with respect to both the three-parameter generalised gamma distribution (without location parameter) and the four-parameter generalised gamma distribution (with location parameter).

The prior and posterior probability functions of the shape parameters $a$ and $c$ for the three-parameter gamma distribution are shown in Figures 3 to 6. The parameters of the prior probability distributions of $a$ and $c$, as well as the posterior expectations of $a$, $b$, and $c$ can be found in Table 1. Figure 7 presents the empirical cumulative probability distribution, the predictive cumulative probability distribution, and the 90 per cent uncertainty intervals (Bayesian credible sets or regions). These 90 per cent uncertainty intervals (the 5th and 95th percentile) are determined using Monte Carlo simulation on the basis of 10,000 samples. The empirical and predictive probabilities of exceedance, according to Eq. (11), are displayed in Figure 8 including their 90 per cent

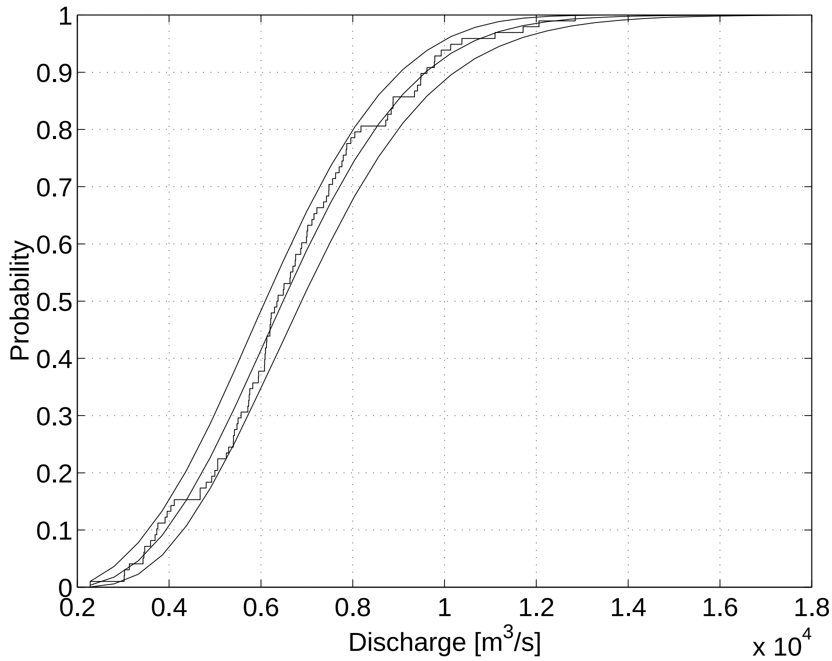


Figure 7: Empirical and predictive cumulative probability distribution of the annual maximum discharge of the Rhine River at Lobith, including their 90 per cent uncertainty interval, for the three-parameter generalised gamma distribution.

uncertainty intervals. Note that the empirical probabilities of exceedance are calculated with the aid of Chegodayev's formula: $\Pr\{Q > q_{z-u+1:z}\} = [u - 0.3]/[z + 0.4]$, where $q_{z-u+1:z}$ is the $u$th largest observation among the sample of $z$ observations ordered by descending magnitude (see e.g. Chow, Maidment & Mays [10, Chapter 12]). For the three-parameter generalised gamma distribution, the Bayes estimate of the design discharge (i.e. the discharge with an exceedance probability of 1/1250) is 15,150 $m^3/s$ with a 90 per cent uncertainty interval of (12,950; 16,950).

In order to take account of the uncertainty in the location parameter, the four-parameter generalised gamma distribution has been studied as well. The posterior probability functions of the shape parameters $a$ and $c$, and the location parameter $d$ are displayed in Figures 9 to 11, as well as the empirical and predictive exceedance probability and its 5th and 95th percentiles in Figure 12. The Bayes estimates of the four parameters are summarised in Table 2. For the four-parameter generalised gamma distribution, the Bayes estimate of the design discharge is 15,200 $m^3/s$ with a 90 per cent uncertainty interval of (13,000; 16,950). Surprisingly, the uncertainty in the location parameter does not seem to have a big influence. It should be noted however, that it is not yet known whether the generalised gamma distribution performs the best in fitting the observed annual maximum discharges. This is currently being investigated using Bayes factors (Kass & Raftery [18]).

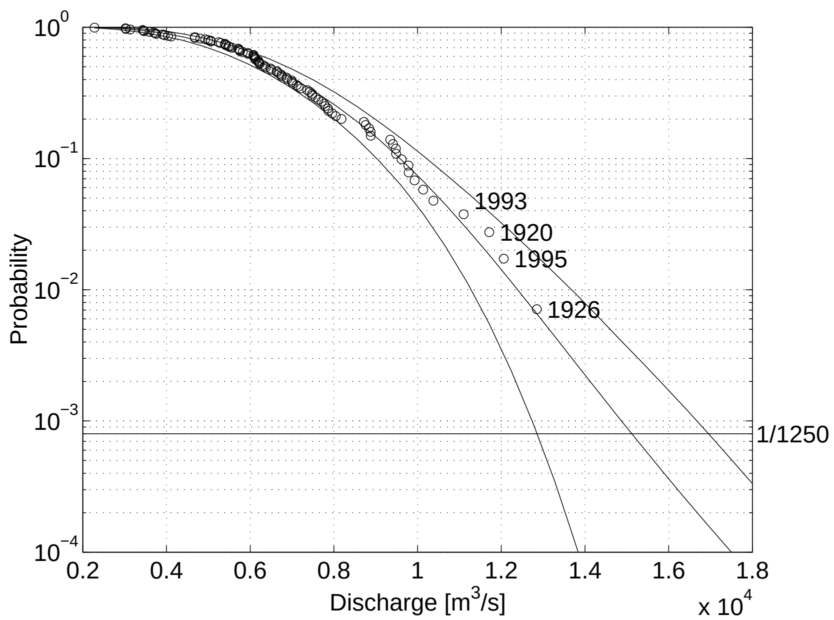


Figure 8: Empirical and predictive probability of exceedance of the annual maximum discharge of the Rhine River at Lobith, including their 90 per cent uncertainty interval, for the three-parameter generalised gamma distribution.

Using the maximum-likelihood method, the estimate of the design discharge decreases to 14,285 $m^3/s$ for the three-parameter generalised gamma distribution and 14,289 $m^3/s$ for the four-parameter generalised gamma distribution. As expected, taking account of parameter uncertainty results in larger design discharges.

Finally, I would like to mention that more or less the same results have been obtained with the Markov chain Monte Carlo method (see e.g. Carlin & Louis [8, Chapter 5] and Tierney [32]). The Metropolis algorithm—developed by Metropolis et al. [25] and generalised by Hastings [15]—was used to determine the predictive exceedance probabilities and their 5th and 95th percentiles. As a proposal or jumping density, the symmetric three- or four-dimensional normal density was chosen with mean equal to the parameter values at the current state and covariance matrix equal to the inverse Fisher information matrix for a sample of $n$ observations evaluated at the maximum-likelihood estimator. Probably due to the possible non-regularity of the four-parameter generalised gamma distribution, numerical integration in combination with Monte Carlo performed better than Markov chain Monte Carlo.

# 8 Conclusions

In this paper, the discharge of the Rhine at Lobith with an average return period of 1250 years has been determined on the basis of a Bayesian analysis. Both the

Table 2: Bayes estimates of $a$, $b$, $c$, and $d$ for the four-parameter generalised gamma distribution.

| parameter | description | value | dimension |
|---|---|---|---|
| $a_L$ | lower bound for $a$ | 0.01 | - |
| $a_U$ | upper bound for $a$ | 11 | - |
| $k$ | number of subdivisions for $a$ | 100 | - |
| $c_L$ | lower bound for $c$ | 0.01 | - |
| $c_U$ | upper bound for $c$ | 7 | - |
| $m$ | number of subdivisions for $c$ | 100 | - |
| $d_L$ | lower bound for $d$ | 0 | $m^3/s$ |
| $d_U$ | upper bound for $d$ | 2280 | $m^3/s$ |
| $l$ | number of subdivisions for $d$ | 100 | - |
| $E(A\vert \mathbf{q})$ | posterior expectation of $a$ | 2.348 | - |
| $E(B\vert \mathbf{q})$ | posterior expectation of $b$ | 4363 | $m^3/s$ |
| $E(C\vert \mathbf{q})$ | posterior expectation of $c$ | 2.113 | - |
| $E(D\vert \mathbf{q})$ | posterior expectation of $d$ | 1155 | $m^3/s$ |
| $d^*$ | posterior mode of $d$ | 1653 | $m^3/s$ |

three-parameter generalised gamma distribution (without location parameter) and the four-parameter generalised gamma distribution (with location parameter) were used to obtain predictive exceedance probabilities of annual maximum discharges. In order to take account of the statistical uncertainties, the parameters of the generalised gamma distribution are assumed to be unknown and to have a prior joint probability distribution.

As prior density, the non-informative Jeffreys prior has been used. On the basis of the observed annual maximum discharges at Lobith, this prior density has been updated to the posterior density by using Bayes' theorem. An advantage is that the generalised gamma distribution fits well with the stage-discharge rating curve being an approximate power law between water level and discharge. Furthermore, since the generalised gamma distribution has three or four parameters, it is flexible in fitting data. Many well-known probability distributions which are commonly used to estimate quantiles of hydrological random quantities, are special cases of the generalised gamma distribution.

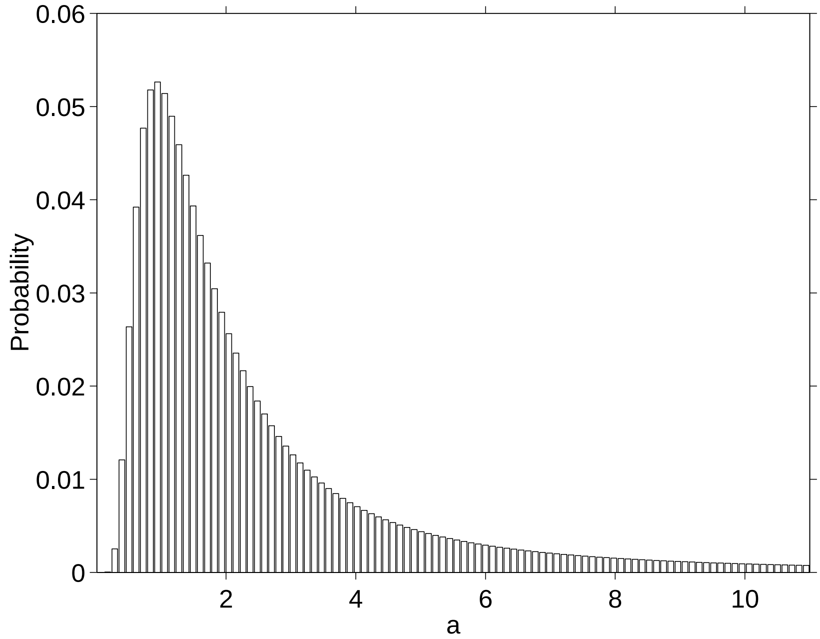


Figure 9: Posterior probability function of $a$: $p(a_h|\,\mathbf{q})$, $h = 1, \ldots, k$.

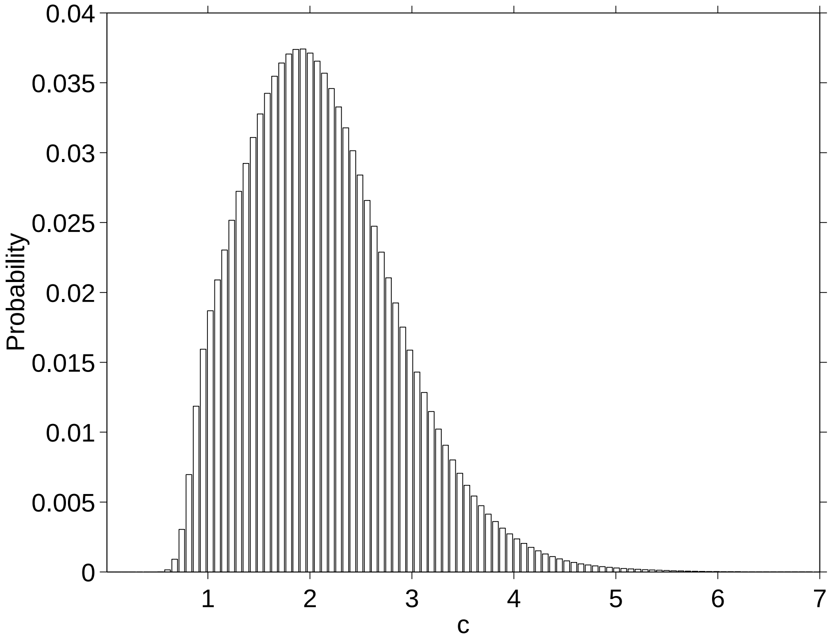


Figure 10: Posterior probability function of $c$: $p(c_i|\,\mathbf{q})$, $i = 1, \ldots, m$.

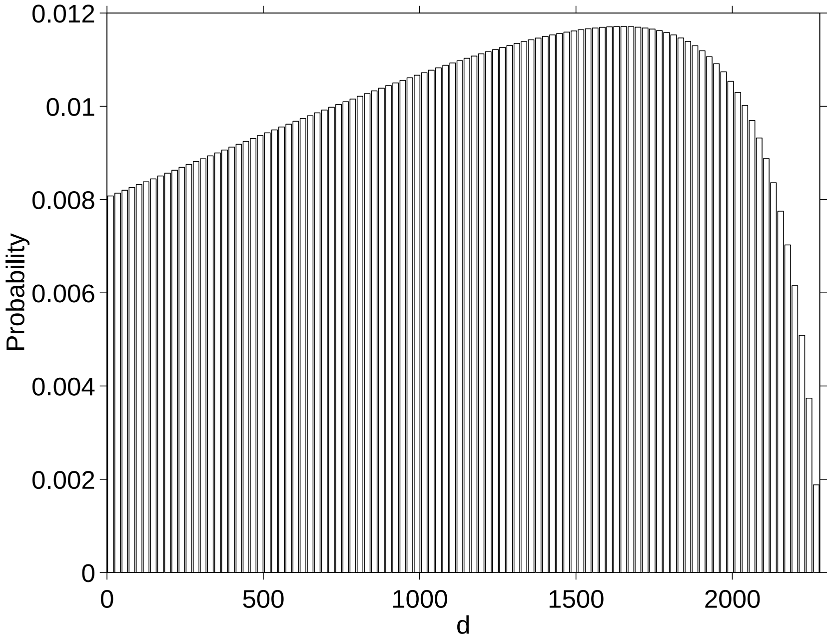


Figure 11: Posterior probability function of $d$: $p(d_j\,|\,\mathbf{q})$, $j = 1, \ldots, l$.

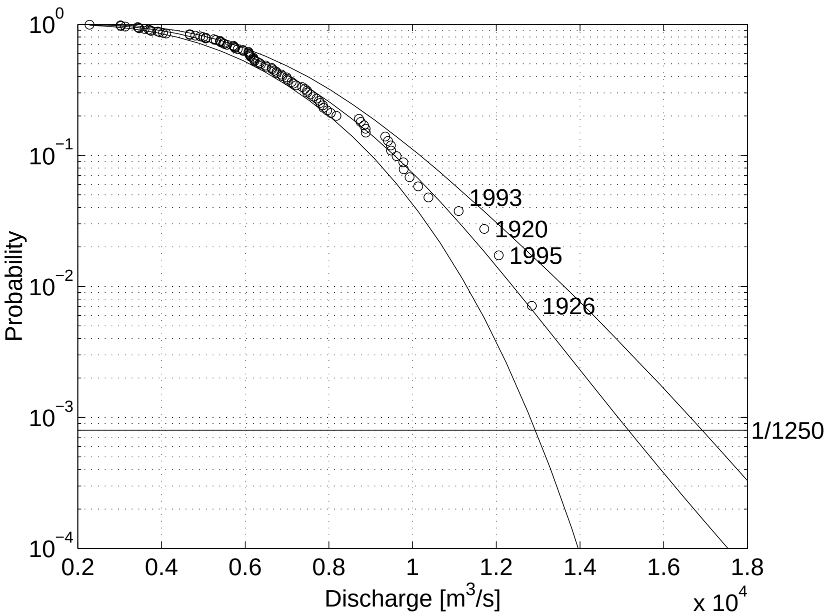


Figure 12: Empirical and predictive probability of exceedance of the annual maximum discharge of the Rhine River at Lobith, including their 90 per cent uncertainty interval, for the four-parameter generalised gamma distribution.

# Appendix

In this Appendix, the four extra elements of the symmetric Fisher information matrix are derived when the three-parameter generalised gamma distribution (3) is extended to the four-parameter generalised gamma distribution (13) with scale parameter $b$, shape parameters $a$ and $c$, and location parameter $d$; that is, when $\boldsymbol{\theta} = (a, b, c, d)'$. By applying the transformation $T = [(Q-d)/b]^c$ and using

$$\textstyle\int_{t=0}^{\infty} t^a e^{-t} \log t \, dt = \Gamma'(a+1) = \Gamma(a) + a\Gamma'(a), \qquad a > 0,$$

the four elements of Fisher's information matrix for location parameter $d$ are

$$\begin{aligned}
E\left(-\frac{\partial^2 \log \ell(Q|\boldsymbol{\theta})}{\partial a \partial d}\right) &= \frac{c}{b} E\left(T^{-\frac{1}{c}}\right) = \frac{c\Gamma(a-c^{-1})}{b\Gamma(a)}, \\
E\left(-\frac{\partial^2 \log \ell(Q|\boldsymbol{\theta})}{\partial b \partial d}\right) &= \frac{c^2}{b^2} E\left(T^{1-\frac{1}{c}}\right) = \frac{c(ca-1)\Gamma(a-c^{-1})}{b^2\Gamma(a)}, \\
E\left(-\frac{\partial^2 \log \ell(Q|\boldsymbol{\theta})}{\partial c \partial d}\right) &= \frac{1}{b} E\left(aT^{-\frac{1}{c}} - T^{1-\frac{1}{c}} - T^{1-\frac{1}{c}} \log T\right) = \\
&= \frac{(1-c)\Gamma(a-c^{-1}) - (ca-1)\Gamma'(a-c^{-1})}{bc\Gamma(a)}, \\
E\left(-\frac{\partial^2 \log \ell(Q|\boldsymbol{\theta})}{\partial d^2}\right) &= E\left(\frac{ca-1}{b^2} T^{-\frac{2}{c}} + \frac{c(c-1)}{b^2} T^{1-\frac{2}{c}}\right) = \\
&= \frac{(c^2a - 2c + 1)\Gamma(a - 2c^{-1})}{b^2\Gamma(a)}. \qquad (15)
\end{aligned}$$

The other elements of Fisher's information matrix can be found in Eq. (6). Because $\Gamma(a - 2c^{-1})$ in Eq. (15) only exists for $ac > 2$, the determinant of Fisher's information matrix for a single observation is only finite for $ac > 2$.

# References


[1] J.C. Ahuja and S.W. Nash. The generalized Gompertz-Verhulst family of distributions. *Sankya: The Indian Journal of Statistics, Series A*, 29:141–156, 1967.

[2] F. Ashkar and T.B.M.J. Ouarda. Approximate confidence intervals for quantiles of gamma and generalized gamma distributions. *Journal of Hydrologic Engineering*, 3(1):43–51, 1998.

[3] R.E. Barlow. *Engineering Reliability*. Philadelphia: American Statistical Association (ASA) & Society for Industrial and Applied Mathematics (SIAM), 1998.

[4] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.

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

[6] J.M. Bernardo and A.F.M. Smith. *Bayesian Theory.* Chichester: John Wiley & Sons, 1994.

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

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

[9] V.T. Chow. *Open-Channel Hydraulics.* Singapore: McGraw-Hill, 1959.

[10] V.T. Chow, D.R. Maidment, and L.W. Mays. *Applied Hydrology.* Singapore: McGraw-Hill, 1988.

[11] D.R. Cox and D.V. Hinkley. *Theoretical Statistics.* London: Chapman & Hall, 1974.

[12] A.P. Dawid. The trouble with Bayes factors. Technical Report No. 202, Department of Statistical Science, University College London, 1999.

[13] H.W. Hager and L.J. Bain. Inferential procedures for the generalized gamma distribution. *Journal of the American Statistical Association*, 65(332):1601–1609, 1970.

[14] H.L. Harter. Maximum-likelihood estimation of the parameters of a four-parameter generalized gamma population from complete and censored samples. *Technometrics*, 9(1):159–165, 1967.

[15] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[16] H.J. Jeffreys. *Theory of Probability; Third Edition.* Oxford: Clarendon Press, 1961.

[17] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, Volume 1;* Second Edition. New York: John Wiley & Sons, 1994.

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

[19] R.E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1370, 1996.

[20] J.H. Lienhard. A statistical mechanical prediction of the dimensionless unit hydrograph. *Journal of Geophysical Research*, 69(24):5231, 1964.

[21] J.H. Lienhard and P.L. Meyer. A physical basis for the generalized gamma distribution. *Quarterly of Applied Mathematics*, 25(3):330–334, 1967.

[22] T.A. Mazzuchi and R. Soyer. 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, 1996.

[23] T.A. Mazzuchi and R. Soyer. A Bayesian perspective on some replacement strategies. *Reliability Engineering and System Safety*, 51:295–303, 1996.

[24] M. Mendel. The case for engineering probability. In R. Cooke, M. Mendel, and H. Vrijling, editors, *Engineering Probabilistic Design and Maintenance for Flood Protection*, pages 1–22. Dordrecht: Kluwer Academic Publishers, 1997.

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

[26] Ministry of Transport, Public Works, and Water Management. *Year-book Monitoring Rivers and Canals 1993 [Jaarboek Monitoring Rijkswateren 1993 (in Dutch)].* The Hague, The Netherlands, 1994.

[27] V.B. Parr and J.T. Webster. A method for discriminating between failure density functions used in reliability predictions. *Technometrics*, 7(1):1–10, 1965.

[28] B.E. Schneider. Algorithm AS 121: Trigamma function. *Applied Statistics*, 27:97–99, 1978.

[29] E.M. Shaw. *Hydrology in Practice;* Second Edition. London: Chapman & Hall, 1988.

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

[31] E.W. Stacy. A generalization of the gamma distribution. *Annals of Mathematical Statistics*, 33:1187–1192, 1962.

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

[33] W.E. Walker, A. Abrahamse, J. Bolten, J.P. Kahan, O. van de Riet, M. Kok, and M. den Braber. A policy analysis of Dutch river dike improvements: trading off safety, cost, and environmental impacts. *Operations Research*, 42(5):823–836, 1994.

[34] Waterloopkundig Laboratorium (WL) and European American Center (EAC) for Policy Analysis/RAND. Investigating basic principles of river dike improvement; Supporting Volume 2: Design loads [Toetsing uitgangspunten rivierdijkversterkingen; Deelrapport 2: Maatgevende belastingen (in Dutch)]. Delft, The Netherlands, 1993.

# Index