Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Hydrology 274 (2003) 95–108

Journal of Hydrology
www.elsevier.com/locate/jhydrol

Bayesian POT modeling for historical data

Eric Parent*, Jacques Bernier

GRESE (Laboratory for Risk Analysis in Water Sciences), ENGREF, 19, Avenue du Maine, F-75732 Paris Cedex 15, France

Received 4 December 2001; revised 4 November 2002; accepted 15 November 2002

Abstract

When designing hydraulic structures, civil engineers have to evaluate design floods, i.e. events generally much rarer that the ones that have already been systematically recorded. To extrapolate towards extreme value events, taking advantage of further information such as historical data, has been an early concern among hydrologists. Most methods described in the hydrological literature are designed from a frequentist interpretation of probabilities, although such probabilities are commonly interpreted as subjective decisional bets by the end user. This paper adopts a Bayesian setting to deal with the classical Poisson–Pareto peak over treshold (POT) model when a sample of historical data is available. Direct probalistic statements can be made about the unknown parameters, thus improving communication with decision makers. On the Garonne case study, we point out that twelve historical events, however imprecise they might be, greatly reduce uncertainty. The 90% credible interval for the 1000 year flood becomes 40% smaller when taking into account historical data. Any kind of uncertainty (model uncertainty, imprecise range for historical events, missing data) can be incorporated into the decision analysis. Tractable and versatile data augmentation algorithms are implemented by Monte Carlo Markov Chain tools. Advantage is taken from a semi-conjugate prior, flexible enough to elicit expert knowledge about extreme behavior of the river flows. The data augmentation algorithm allows to deal with imprecise historical data in the POT model. A direct hydrological meaning is given to the latent variables, which are the Bayesian keytool to model unobserved past floods in the historical series.

Keywords: Bayesian models; Markov Chain Monte Carlo methods; Gibbs sampling; Data augmentation; Extreme value theory; Flood design; Historical information; Semi-conjugate prior

1. Introduction

In a recent survey, Berger (1999) emphasized the very large development of Bayes ideas and applications in the technical literature of many applied domains during the last 10 years. Berger argued that methodological developments are made possible through the use of Monte Carlo Markov Chains simulation techniques and that we are likely to see growth in application of Bayesian ideas for this reason (if no other). Strangely enough, hydrometeorological studies are poorly represented in this survey. For risk analyses of extreme environmental events for instance, with some exceptions such as Kuczera (1999), significant contributions such as Coles and Powell (1996) or Coles and Tawn (1996), have mostly been published in statistical reviews, and did not hold the attention of hydrologists. The Bayesian paradigm allows to revisit many old hydrological problems

* Corresponding author.
E-mail addresses: parent@engref.fr (E. Parent), jacques.bernier2@wanadoo.fr (J. Bernier).

0022-1694/03/$ - see front matter
doi:10.1016/S0022-1694(02)00396-7

within an open-minded and decision-oriented perspective. Among such classical problems are the assessment of risks induced by river floods. In this widely explored field, further difficulties arise when the available 'objective' information consists in systematic records completed with historical data.

The Bayesian point of view benefits from the long history of modeling floods in the hydrological literature (Bernier, 1967). In this paper we are not concerned with the general problem of modeling distributions of hydrological extremes by complete generalized Pareto distributions (NERC, 1975) or not (USWRC, 1992), validation of assumptions about thresholds (Davison and Smith, 1990), etc. There is also an important literature about incorporation of historical data following the seminal paper of Stedinger and Cohn (1986) until the last developments in Martins and Stedinger (2001a,b).

Specifically, the coherence of a complete Bayesian approach of risks analysis is herein exemplified for a case study with the particular and realistic Poisson–Pareto peak over threshold (POT) model. Our arguments are far from new but the availability of new estimation techniques: 'data augmentation' algorithms (Tanner, 1996), MCMC computation methods (Kuczera and Parent, 1998) together with a better understanding of Bayesian concepts in action, now allow new applications which, in turn, should give to such ideas more convincing support among the hydrologists community.

The paper is organized as follows. Sections 2 and 3 present the case study and the classical and Bayesian paradigms. Section 4 recalls the theoretical materials for a Bayesian to deal with Poisson–Pareto POT models. In Section 5, the historical data are modelled according to the simplest of two censoring approaches. In Section 6, advantage is taken from the intrinsic conditional structure of the model to implement Gibbs sampling combined with the 'data augmentation Algorithm (DAA)' of Tanner (1996). Numerical results from the Bayesian analysis are exposed in Section 7. It is found that the Bayesian 90% credible interval for the typical (1000 years return) design value is reduced by 60% when taking into account historical evidence. In Section 8, perspectives and limits of the use of data augmentation for incorporating historical data in extreme value analysis are discussed.

2. Garonne case study

The Garonne river is the most important river in the south-western part of France. Fig. 1 shows the sample of systematic data spanning over the period 1913–1977. They were recorded at a gauging station located near the city of Agen.

The 12 main historical events for the period 1770–1912 are very well documented, thanks to the long life work of Pardé (1935) devoted to the history and the hydrology of the Garonne river and its tributaries. The characteristics of the historical floods are summed up in Table 1. Many of the estimated flows are rather imprecise.

3. Classical versus Bayesian interpretations of probabilities

Basic theoretical framework of extreme value models and inferential techniques can be found in Coles (2001). Frequentist and Bayesian settings share common probabilistic tools for statistical modeling. However, the practical interpretation of these tools is different. For example, a Bayesian hydrologist would understand a 1000 year return flood Q_{1000} as the event which returns in the mean every 1000 year period. At the opposite, the bayesian interprets the return period as 'a quantitative degree of subjective belief' used as a guess in his decisional behavior and does not refer to a hypothetical. future 1000 year period.

Basic theoretical framework of extreme value models and inferential techniques can be found in Coles (2001). Frequentist and Bayesian statisticians share common probabilistic tools for statistical modeling. However, the practical interpretation of these tools is different because Bayesians and frequentists do differ when ascertaining values for unknown quantities. Consider the following statement: 'the 100 year return flood Q_{100} of the Garonne river has a confidence interval: 6070–7550 m^3/s with a probability greater than 90%'. Such a result can be obtained for instance from the software package Hyfran from INRS-EAU (Hyfran, 2000) and relies on maximum likelihood (ML) estimation techniques (Hosking, 1985; Martins and Clarke, 1993). The correct interpretation of such a frequentist confidence interval is that it reflects the asymptotic long-term

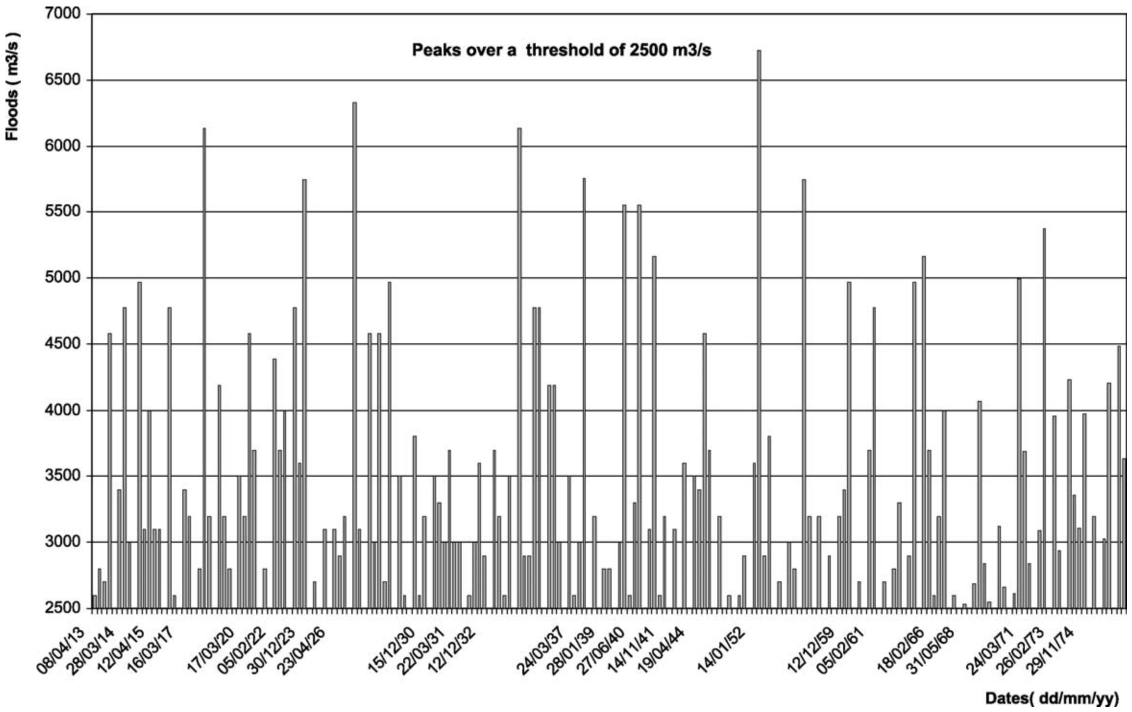

Fig. 1. 151 peaks over a 2500 m^3/s threshold for the Garonne river at Mass d'Agenais spanning over the period 1913–1977.

performance of the estimation procedure for the unknown Q_{100} over a large number of data sets. The problem is that most, if not all, decision makers would understand the previous statement as a direct probabilistic belief about the unknown parameter Q_{100}: 'there are 90% of chances that the random value Q_{100} lies between 6070 and 7550 m^3/s, given the sample data'. These misunderstandings distort the communication between classical hydrologists and decision makers. Indeed the frequentist interpretation of judgments becomes rather uneasy in the case of samples completed with partial, imprecise and scattered historical data. In addition, such situations of mixed information often depart from the mathematical setting of independently identically distributed samples of random variables for which the asymptotic properties of ML estimates have been established.

On the other hand, the Bayesian paradigm gives firm and coherent theoretical justification to the decisional interpretation of probability as a subjective concept (Berger, 1985; Bernardo and Smith, 1994; Gelman et al., 1995). In our experience, decision

Table 1
Historical floods for the Garonne river at Mas d' Agenais (Miquel, 1984)

Date	Height (m)	Estimated flow (m^3/s)
April 1770	10.34	From 7000 to 7400
September 1772		6300
March 1783		From 7000 to 7200
May 1827		6500
May 1835		6400
Jan. 1843		6500
June 1855	9.96	7000
May 1856	9.62	6200
June 1856	9.88	6600
June 1875	10.56	From 7000 to 7500 (maybe 8000)
Jan. 1879	9.62	6300
Feb. 1879	10.02	7000
May 1918	9.51	6000
1913		*Systematic records of the river began*
March 1927	9.97	6700
March 1930	10.72	From 7000 to 7500 (maybe 8000)
March 1935	9.95	6700
Feb. 1952	10.26	From 6000 to 7000
Jan. 1955	9.32	from 5200 to 5700

makers have no difficulty to understand that preventive action can be triggered by a 'quantitative degree of subjective belief' concerning the occurrence of such events. Bayesian probabilistic judgments are relative to the object of interest (assumption or parameter interval) and conditional to the available information including historical data, even qualitative one such as the expert knowledge (Bernier and Parent, 2002).

Finally, much freedom in model design has been gained owing to the recent methodological advances in Bayesian computation with Monte Carlo Markov Chain (MCMC) methodology, which is by now the standard in the statistical profession. Asymptotic approximations are no longer necessary to avoid burdensome computations, inference can nowadays be achieved within large parametric families of models, and models become more and more realistic with the introduction of latent variables, such as missing data. Taking into account historical data in POT models can be formulated as a latent variable problem, to be solved by a Bayesian data augmentation algorithm. The basic idea behind the data augmentation algorithm is to complement the observed data (systematically measured or historically recorded) by the missing past data (unobserved or forgotten flood events) via appropriate Gibbs sampling simulations.

4. Bayesian modeling

4.1. POT likelihood model (Poisson—generalized Pareto)

The POT model (Rasmussen et al., 1994) is described by the generalized Pareto distribution with Poisson arrival rate, as studied by Wang (1991) following Pickands (1975). A regular time-spaced sequence of independently and identically distributed random variables $X_1, X_2, \ldots, X_i, \ldots, X_j, \ldots$ (river flows) defines a POT series by considering only the (flood) peaks $\ldots X_1, \ldots, X_j, \ldots$ above a specified threshold level u. The threshold u is assumed sufficiently high so that Pickhands' asymptotic theorem applies. We use a specific parametrization $\theta = (\mu, \rho, \beta)$ for this Poisson–Pareto model:

$$\Pr(X_j \leq x | X_j \geq u) = G(x|\rho, \beta, u)$$
$$= \begin{cases} (1 - (1 - \beta(x-u)^{\rho/\beta} & \text{for } \beta \neq 0 \\ 1 - \exp(-\rho(x-u) & \text{for } \beta = 0 \end{cases} \quad (1)$$

$$\Pr(N\{X_j \geq u\} = n | \text{in } T \text{ years})$$
$$= \frac{(\mu T)^n \exp(-\mu T)}{n!} \quad (2)$$

where $N\{X_j \geq u\}$ is the number of peaks over or equal to u in T years.

Hydrological presentations of this model usually refer to parameter $\xi = -\beta/\rho$ rather than β. Hydrologically, ξ is the natural dimensionless parameter determining the heaviness of the Pareto tail distribution. The main interest of the parametrization by ξ is its invariance with the threshold u. As we consider here a fixed level u, we nevertheless keep on working with the parameter $\theta = (\mu, \rho, \beta)$ which is far more advantageous for semi-conjugate Bayesian inference purposes, as shown later.

The likelihood for $\theta = (\mu, \rho, \beta)$ based on an observed series x of n data $x_1, x_2, \ldots, x_n$ over the threshold u during T years is given by the probability density function $f(x|\theta)$ of x given θ. This function can be seen as the product of two parts:

- the probability of N, let: $P(n|\mu, u) = \Pr(N\{X_j \geq u\} = n | \text{in } T \text{ years})$ and given by 2,
- the probability density $g(x|\theta, n)$ of the sample $x_1, x_2, \ldots, x_n$ given n, such as:

$$g(\mathbf{x}|\theta, n) = \prod_{i=1}^{i=n} \frac{\mathrm{d}G(x_i|\rho, \beta, u)}{\mathrm{d}x_i}$$
$$= \rho^n \exp(\rho - \beta) S_n(\mathbf{x}, \beta)$$

owing to the mutual independence of members of the sample where

$$S_n(\mathbf{x}, \beta) = \frac{1}{\beta} \sum_{i=1}^{n} \log(1 - \beta(x_i - u))$$

and

$$S_n(\mathbf{x}, 0) = -\sum_{i=1}^{n}(x_i - u)$$

Then the likelihood is:

$$f(\mathbf{x}|\theta) = P(n|\mu, u) \times g(\mathbf{x}|\theta, n)$$
$$= \left[\frac{(\mu T)^n \exp(-\mu T)}{n!}\right][\rho^n \exp(\rho - \beta) S_n(\mathbf{x}, \beta)] \tag{3}$$

A useful property of the POT model concerns the change of threshold from u to $u^* \geq u$. The new model of exceedance remains a process of the same type, but with parameters $\theta^* = (\mu^*, \rho^*, \beta^*)$ such that

$$\mu^* = \mu(1 - G(u^*|\rho, \beta, u)) = \mu(1 - \beta(u^* - u))^{\rho/\beta} \tag{4}$$

$$\rho^* = \frac{\rho}{1 - \beta(u^* - u)} \tag{5}$$

$$\beta^* = \frac{\beta}{1 - \beta(u^* - u)} \tag{6}$$

The usual annual quantile $q(p, \theta)$ corresponds to the design value with annual failure probability $1 - p$. Evaluating the inverse cumulative function for the maximum of random POT values with $T = 1$ (1 year) leads to the analytical expression of the flood with return period $1/p$

$$q(p, \theta) = u + \frac{1}{\beta}\left[1 - \left(-\frac{\log(p)}{\mu}\right)^{\beta/\rho}\right] \tag{7}$$

4.2. Semi-conjugate prior structure

According to the Bayesian line of reasoning, the prior distribution is part of model specification.

We assume that prior beliefs about $\theta = (\mu, \rho, \beta)$ can be represented by the following probability density function:

$$\pi(\theta) = \frac{\psi^\nu}{\Gamma(\nu)}\mu^{\nu-1}\exp(-\psi\mu)\frac{\varphi^{\gamma(\beta)}}{\Gamma(\gamma(\beta))}\rho^{\gamma(\beta)-1}$$
$$\times \exp(-\varphi(\beta)\rho)\pi_0(\beta) \tag{8}$$

In other words, the marginal prior for β is $\pi_0(\beta)$. $\pi_0(\beta)$ is arbitrary and its functional shape is left to the expert to encode his prior knowledge. Conditionally to these prior beliefs for β, the pdf for ρ is approximated by a gamma distribution with hyperparameters $(\gamma(\beta), \varphi(\beta))$ that are functions of β. If not, β and ρ are a priori independent. It is also assumed that the priors for μ and (β, ρ) are independent, which is realistic as many experts often proceed separately when dealing with the yearly expected number of floods and when estimating their intensity once they occur. Prior belief about μ is taken in the conjugate Poisson family, that is a gamma distribution with hyperparameters (ν, ψ).

Arguments for choosing such a functional form as Eq. (8) are:

- The prior model (8) is quite flexible and the quantities $(\pi_0(\beta), \nu, \psi, \gamma(\beta), \varphi(\beta))$ need to be elicitated from the experts' knowledge about extreme behavior of the river flows or chosen according to a non-informative structure. They can also be estimated through a regional analysis of extreme events in the vicinity of the river basin or themselves appear as a first layer of at-site effects in a hierarchical model.
- Advantage is taken from partial conjugate structure; as the likelihood (3) belongs to a partly exponential family (conditioned upon β), prior and posterior pdfs for θ will exhibit conjugate (given β) properties for ρ and μ.

In the simple case of a systematic sample $\mathbf{x}$ (with no historical data), the joint posterior distribution $\pi(\theta|\mathbf{x})$ has the same conditional structure as the prior pdf:

- μ remains a posteriori independent from (β, ρ). It follows a gamma distribution with hyperparameters $(\nu + n, \psi + T)$.
- Conditionally upon β, ρ is also gamma distributed with updated hyperparameters $(\gamma(\beta) + n, \varphi(\beta) - S_n(\mathbf{x}, \beta)$.
- The marginal posterior density of β, $\pi(\beta|\mathbf{x})$ and the posterior pdf of β conditioned on ρ, $\pi(\beta|\rho, \mathbf{x})$ are explicitly known up to a constant of normalization.

The conjugate prior distribution is chosen for mathematical convenience (cf. Raiffa and Schlaifer, 1961), but this general form is flexible enough to represent a large family of quantified expert beliefs. We refer to Bernier and Parent (2002) for a presentation of proper priors based on expert assessments for the same data. In this paper, a non-informative prior (Box and Tiao, 1973) is derived by letting ($\pi_0 \to 1, \nu \to 0, \psi \to 0, \gamma(\beta) \to 0, \varphi(\beta) \to 0$) in order not to hide the effect of historical information by any prior observation. Note that historical evidence is generally a valuable source to help subjective introspection when attempting to formulate a prior in Bayesian contexts. In the present approach, we deliberately choose not to consider historical information for prior elicitation but we will directly incorporate it into the analysis as part of the objective data set.

5. Models of historical data

We do not deal here with paleo floods and we assume that the historical sample consists of r historical maxima $\mathbf{y} = (y_1, y_2, \ldots, y_r)$, although not necessarily annual maxima as in Smith (1986). They have been observed and assessed by various methods (Pardé, 1935) during a historical period H ($r \le H$). For the time being, we make the additional hypothesis that the historical data were assessed without errors.

The POT likelihood as formulated by Eq. (3) does not apply to the Garonne data containing both systematic records (x) and historical values (y). In the flood literature, two ways of incorporating historical evidence in extreme value analysis are currently proposed. They correspond to two different censoring procedures:

1. *Censoring by threshold*. A perception level u^* is considered and it is *assumed* that the historical floods were recorded because they were the only ones that overcome this perception level u^*. With this assumption, the component of the likelihood due to the historical data $\mathbf{y}$ is just a reformulation of Eq. (3) with $\theta^* = (\mu^*, \rho^*, \beta^*)$ given by Eq. (4) + Eq. (5) + Eq. (6):

$$f^{(h)}(\mathbf{y}; \mu^*, \rho^*, \beta^*) = \frac{(\mu^* H)^r \exp(-\mu^* H)}{r!} (\rho^*)^r \times \exp\left[\frac{(\rho^* - \beta^*)}{\beta^*} \sum_{j=1}^{r} \log(1 - \beta^*(y_j - u^*))\right] \quad (9)$$

This expression is rather simple but needs two additional parameters: H and u^*.

2. *Censoring by number*. Here the only assumption made is that the historical sample contains all the r highest events from the historical period. To evaluate the historical contribution to the likelihood, it is useful to condition upon the number of floods that exceed the threshold u during H years. The number K of past exceedances follows the Poisson distribution with parameter $H\mu$. Given $K = k$, the historical sample consists of the r highest values and all we know about the remaining $k - r$ values is that they stand below $y_{(1)}$ (with probability $G(y_{(1)}|\rho, \beta, u)$ since each of them is a POT variable but can only occur anywhere in the interval $[u, y_{(1)}]$).

$$f^{(h)}(\mathbf{y}; \mu, \rho, \beta) = \sum_{k=r}^{+\infty} \frac{(\mu H)^k \exp(-\mu H)}{k!} \frac{k!}{(k-r)!} \times G(y_{(1)}|\rho, \beta, u)^{k-r} \prod_{i=1}^{r} g(y_i|\rho, \beta, u) \quad (10)$$

The summation is straightforward:

$$f^{(h)}(\mathbf{y}; \mu, \rho, \beta) = (\mu\rho H)^r \exp[-\mu H(1 - G(y_{(1)}|\rho, \beta, u))] \times \exp((\rho - \beta) S_r(y, \beta)) \quad (11)$$

Although the generalized extreme value cdf $G(y_{(1)}|\rho, \beta, u)$ in Eq. (11) is known in closed form, this second model for including historical information is often believed to be less tractable that the one given by Eq. (9). But it is more parsimonious since only H is to be included in the study and it does not need to estimate the poorly defined parameter u^*. Furthermore, the two modeling approaches (11) and (9) merge when:

$$u^* = y_{(1)}$$

The way to incorporate historical information belongs to the modeling process. It requires assumptions and new external parameters such as H and the perception levels u^*. But both H and u^* are uncertain parameters and referring to the Bayesian paradigm, we could use priors to take their uncertainty into account. In many cases, little is known about them, it is preferable to deal with their uncertainties by a mere sensitivity analysis. In the Garonne case study there are hydrological and historical justifications (Pardé, 1935) to let the historical period begin in 1770.

Although model (9) is widely used in most extreme values studies from the North American literature, the perception parameter u^* is questionable, both in theory and in practice. Such a level of perception may have varied between time periods. It can also be very difficult to assess since the only relevant information concerning u^* is carried by $y_{(1)}$, the smallest record in the historical sample. Consequently, in what follows, we prefer censoring by number.

6. Bayesian estimation by Gibbs and 'data augmentation'

6.1. The Gibbs sampling algorithm complements the historical records

Bayesian analysis does not suffer from the usual difficulties brought by eventual non-regularity of the likelihood (3) pointed out by Smith (1985) or of the likelihood completed by any form of the two historical contributions (9) or (11). We even avoid to implement the computation of these likelihoods by relying on the data augmentation algorithm. The intuitive idea behind the data augmentation algorithm used in this case is to complement the observed data $(\mathbf{x},\mathbf{y}) = y_1, y_2, \ldots, y_r,\ x_1, x_2, \ldots, x_n$ (historically recorded or systematically measured) by the missing past data $Z_1, Z_2, \ldots, Z_{K-r}$ (unobserved or forgotten flood events)so as to regenerate a complete sample on the period $H+T$, i.e.K events during the historical period H and n events during the recent period T. As an illustrative example, such a data augmented sample could re-assemble the series as: $(Z_1, Z_2, y_1, y_2, Z_3, Z_4, y_3, \ldots, Z_5, y_r, Z_6, \ldots, Z_{K-r}, x_1, x_2, \ldots, x_n)$.

Gibbs sampling is an iterative algorithm allowing computation of a multi-dimensional k-joint distribution $\Pr(\zeta_1, \ldots, \zeta_i, \ldots, \zeta_k)$ from the knowledge of the k one-dimensional distributions $\Pr(\zeta_i | \{\zeta_{j\neq i}\}$, called *complete conditionals*. Its theoretical justification is that, under suitable regularity conditions, such a joint distribution can be considered as the ergodic limit of an homogeneous Markov chain having a transition probability density defined as the product of the complete conditional densities $\Pr(\zeta_i | \{\zeta_{j\neq i}\}$ (Gelfand and Smith, 1990). This is still true when using a block-component $\{\zeta_1, \ldots, \zeta_i, \ldots, \zeta_k\}$ decomposition of the argument of the multi-dimensional distribution.

A Gibbs sampling algorithm to evaluate the joint pdf of $(\theta, K, \mathbf{Z} | (\mathbf{X}=\mathbf{x}, \mathbf{Y}=\mathbf{y}))$ can be implemented as follows (with the previous notations: $k=3$, $\zeta_1 = \theta | \mathbf{x}, \mathbf{y}$, $\zeta_2 = K | \mathbf{x}, \mathbf{y}$, $\zeta_3 = \mathbf{Z} | \mathbf{x}, \mathbf{y}$):

Let $\theta^{(0)}, k^{(0)}, z^{(0)}$ denote arbitrary starting values and $\theta^{(s-1)}, k^{(s-1)}, z^{(s-1)}$ the values generated at step $s-1$. Step s of the Gibbs sampler consists of the following three phases:

(1) using the structural conditional decomposition of the posterior $\pi(\theta | \mathbf{x}, \mathbf{y}, z^{(s-1)})$, draw $\partial\varepsilon\theta^{(s)} = (\mu^{(s)}, \rho^{(s)}, \beta^{(s)})$
(2) based on the Poisson distribution with parameter $H\mu^{(s)}$, generate the number $K = k^{(s)}$ of data during period H with $K \geq r$; the number of missing values for the unobserved sample $\mathbf{Z}$ is $(k^{(s)} - r)$.
(3) sample the $(k^{(s)} - r)$ values $Z_1, Z_2, \ldots, Z_{(k^{(s)}-r)}$ independently from the truncated distributions
    (a) $g(z|\rho^{(s)}, \beta^{(s)}, u)/G(u^*|\rho^{(s)}, \beta^{(s)}, u)$ defined on the interval $[u, u^*]$ for the model (9) with historical data censored by threshold u^*,
    (b) $g(z|\rho^{(s)}, \beta^{(s)}, u)/G(y_{(1)}|\rho^{(s)}, \beta^{(s)}, u)$ defined on the interval $[u, y_{(1)}]$ if the model (11) with historical data censored by number is chosen.

No difficulty is encountered when running the third phase of the Gibbs sampler since GEV distributions are available in closed form and random generation is directly based on inverse cumulative density function like Eq. (7). In practice, as u^* is often elicited close to $y_{(1)}$ the two possible procedures of the third phase do not differ much.

Iterating these three phases guarantees that the Gibbs sampler chain converges to the correct equilibrium distribution of $\Pr(\theta, K, \mathbf{Z}|(\mathbf{x}, \mathbf{y}))$. The posterior sample from $\Pr(\theta|(\mathbf{x}, \mathbf{y}))$ is straightforwardly derived by subsequent direct marginalisation. Demonstrations of MCMC ergodic properties and practical implementation of such techniques are now extensively published and so will not be recalled here. Details can be found for instance in Robert and Casella (1998) or in Tanner (1996).

6.2. Benefits from Gibbs sampling and a data augmentation algorithm

6.2.1. A missing value perspective for hydrological interpretation

Missing values Z can be interpreted as latent (i.e. unobserved) variables which can be put in the model as conditionally linked between observed variables and parameters. This contributes to a simpler interpretation of the model and suggests the coherent Gibbs computation method of marginal posterior distributions via complete conditional ones.

6.2.2. Imprecision can be dealt with an additional loop in the Gibbs sampling scheme

Now a look at Table 1 reveals that historical floods are not as precise as everyday data since they have not been evaluated with the same measuring devices. Most of them are just past experts assessments with a rule of thumb or present evaluation of their magnitude on the basis of past damages reported in archives. In addition, at times of extremely high water levels, gauging stations, if any, did not work well or did not work at all (as it is generally still the case nowadays!). It is fairly easy to represent the poor quality of the historical data y_j in our algorithms by assuming that all we know about it is that it has occurred in the range $[y_i^{\min}, y_j^{\max}]$. We therefore consider that historical values stem from a random mechanism on the interval $[y_i^{\min}, y_j^{\max}]$, for instance a uniform draw or a truncated distribution centered on a reference value $\bar{y}_i$. An additional stage in the Gibbs sampler is necessary to introduce this uncertain historical evidence: at each loop a value for y_j is drawn accordingly. The independence of draws is assumed here for illustration purposes only. The data augmentation algorithm would work with any other model of imprecision. However, assessing a model for the imprecision of the historical data is a rather uneasy task because strong modeling assumptions are to made. Using this very parsimonious model simply (but quantitatively) illustrates the decreasing value of historical information with the decreasing precision.

6.2.3. Decision analysis as a side-product of the Gibbs sampler

Finally the same algorithms are suitable for a complete decisional analysis of the risk problem. With regards to the balance between the cost of protection and the possible damages, the failure level $1 - p$ is generally set to probabilities typically as small as 1/100 or 1/1000 in flood risk analysis. This means extrapolating the model far away from the range of values for which it was adjusted (very few gauged stations possess as much as 100 years of data and 1000 years is always out of reach). Consequently, one should not forget that the values given by Eq. (7) are blurred by a lot of uncertainties. Of course, the decision maker wishes to adopt only one single design value d which mitigates the possible damaging consequences. Without entering here into a complete analysis of these consequences, a *conventional* 'cost' function like Eq. (12) can roughly sketch the consequences of errors between d and $q(p, \theta)$.

$$u_p(d, \theta) = (1 - c)[\max(0, d - q(p, \theta))]^\lambda + c[\max(0, q(p, \theta) - d)]^\lambda \tag{12}$$

For instance, were an overestimation of $q(p, \theta)$ assumed to be less damaging than an underestimation, then the weighting coefficient c should be chosen such that $(1 - c) > c$. The index λ is used to tune the relative compensations between high and small deviations. The traditional least squares criterion corresponds to $c = 1/2$ and $\lambda = 2$. Bayesian decision theory (Tribus, 1972; Berger, 1985) deals with uncertainty about θ by picking the decision d that minimizes the expected predictive cost:

$$E^{(d)}(u|x,y) = \int u_p(d, \theta\pi(\theta|x,y)\mathrm{d}\theta \qquad (13)$$

The computation of $E^{(d)}(u|x,y)$ is a fairly easy by-product of the MCMC algorithm that provides samples from $\pi(\theta|x,y)$.

7. Results

7.1. Inference with systematic data only

7.1.1. Implementation

When dealing with data systematically recorded, the posterior distribution $\pi(\theta|\mathbf{x})$ is known up to a constant. However, as a practical consequence of the semi-conjugate structure when conditioning upon β, a very simple random sampling scheme can be achieved: an easy univariate random generation of β by a numerical inverse probability method , followed by a gamma generation for ρ (conditioned on β) and then μ is drawn independently from its own posterior gamma distribution.

Table 2
Posterior characteristics for model parameters based on systematic data

	Mean	Median	Sup 95%	Inf 95%
μ	2.33	2.32	2.65	2.02
1000β	0.11	0.12	0.20	−0.01
1000ρ	0.83	0.82	0.99	0.68
q_{10}	5550	5520	6010	5195
q_{100}	7165	7005	8520	6355
q_{1000}	8395	8000	11,090	6930

7.1.2. Parameter and percentile inference

Table 2 shows posterior pdf for parameters and usual design percentiles. One cannot exclude that parameter β may be negative, leading to a Weibull bounded type distribution for floods. Fig. 2 shows a sample of the percentiles drawn from their posterior distribution. They are not bell-shaped but highly skewed.

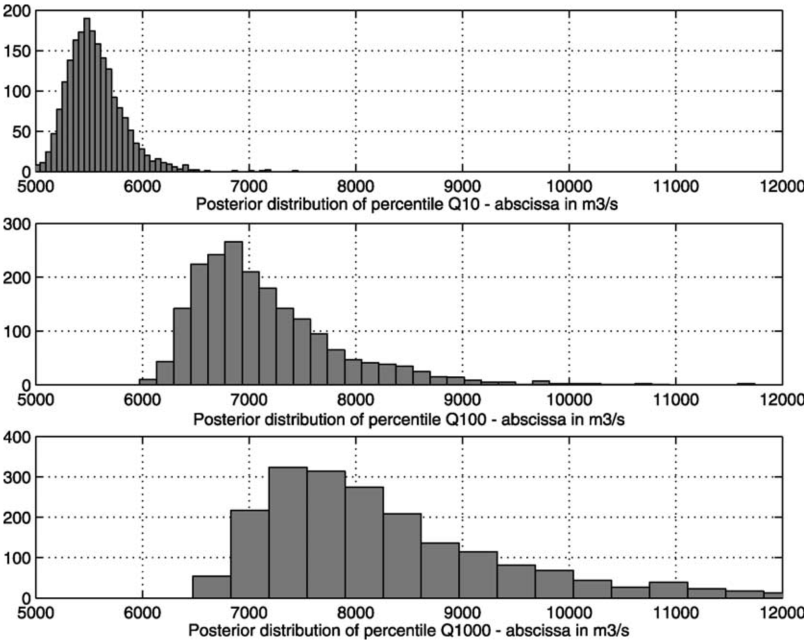

Fig. 2. Posterior samples of quantiles based on (systematic) data spanning on the period 1913–1977.

The range of the 90% credible interval for the 1000 year return flood is close to 4000 m^3/s, to be compared with the expected value $E_{\theta|\mathbf{x}}(q(10^{-3}, \theta)) = 8400$ m^3/s.

7.2. Inference with historical data

7.2.1. Parameter and quantile inference

Table 3 shows 90% credible intervals for parameters and usual design percentiles when dealing with the historical + systematic information based on the posterior pdf $\pi(\theta|\mathbf{x}, \mathbf{y})$. It is now very likely that parameter β is positive, thus excluding finite end point distributions for maximum floods. Fig. 3 compares the parameter posterior marginal distributions with and without historical data. The pdfs for parameters β and ρ are less diffuse, as expected, when both types of information are take into account. μ posterior marginal pdf remains unaffected except in terms of a little shift in the mean. This pdf does not depend on the very values of the historic sample but on the number of historical observations and the length of the historical period H.

The range of the 90% credible interval for the 1000 year return flood is now close to 2600 m^3/s, i.e. 65% of the corresponding range when only systematic data are used. The mean value $E_{\theta|\mathbf{x}}(q(10^{-3}, \theta)) = 8600$ m^3/s is a larger than in the previous case. Fig. 4 shows the increase of precision for various return periods on a log scale.

Table 3
Posterior characteristics for model parameters based on systematic + historical data

	Mean	Median	Sup 95%	Inf 95%
μ	2.33	2.32	2.65	2.02
1000β	0.11	0.12	0.17	0.03
1000ρ	0.77	0.77	0.93	0.64
q_{10}	5740	5740	6020	5480
q_{100}	7420	7360	8200	6890
q_{1000}	8640	7970	10,300	7640

7.2.2. Decision making and predictive analysis

Fig. 5 plots the expected costs (Eq. (13)) with $c = 2/3$, $\lambda = 1.5$ and $c = 2/3$, $\lambda = 3$ for designing a 100 and a 1000 year return flood. In uncertain situations, depending how they weight under and overestimations, decision makers should generally pick a design value larger than the previous posterior mean estimates.

The aim of current engineering practice is to protect against a maximum probable flood on a given period of time. Fig. 6 plots maximum flood predictive pdfs for various spanning periods, with and without

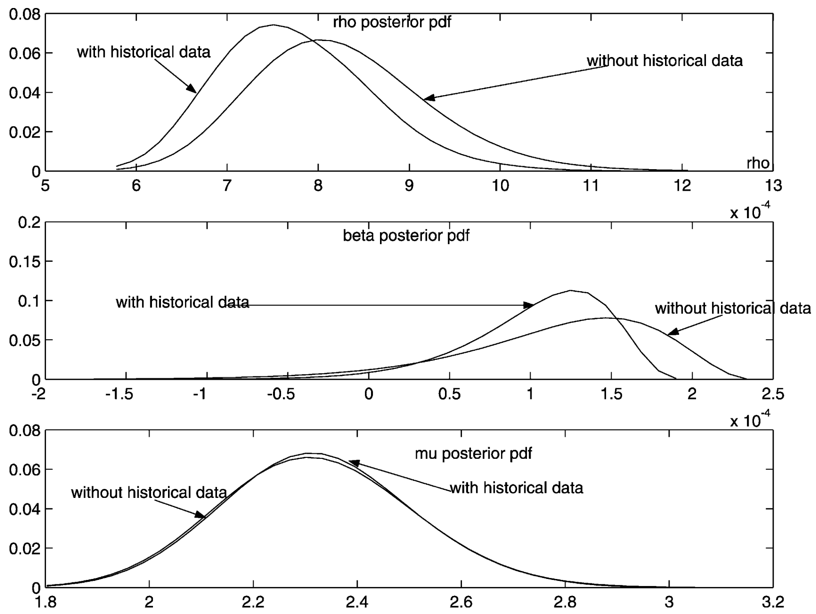

Fig. 3. Parameter posterior pdfs with and without historical data.

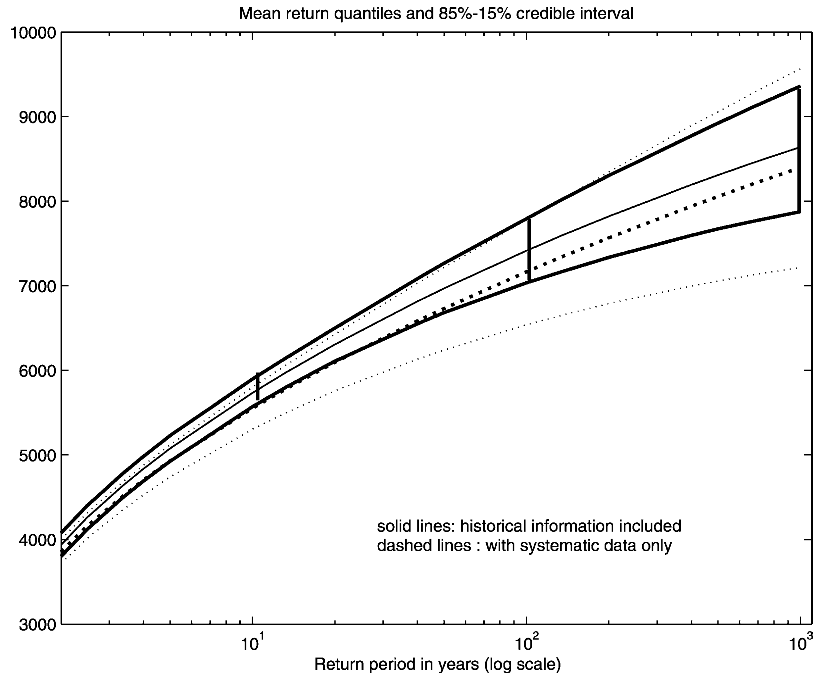

Fig. 4. Mean return quantiles and a 60% credibility interval with and without historical data.

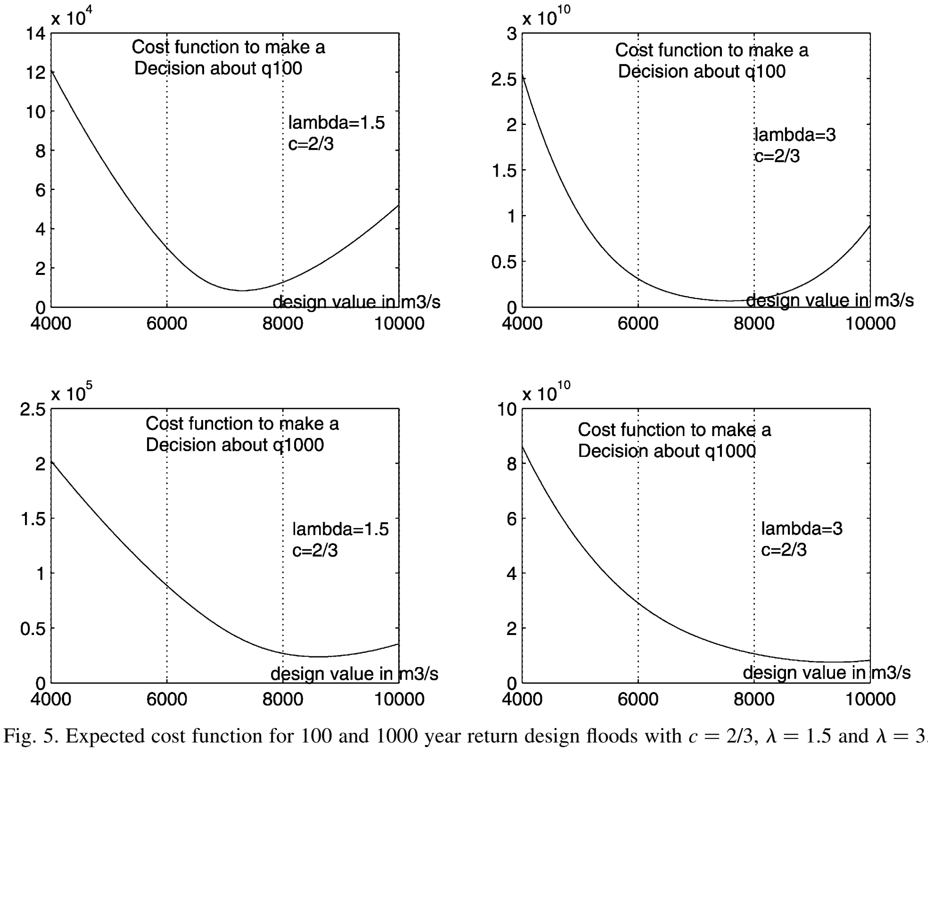

Fig. 5. Expected cost function for 100 and 1000 year return design floods with $c = 2/3$, $\lambda = 1.5$ and $\lambda = 3$.

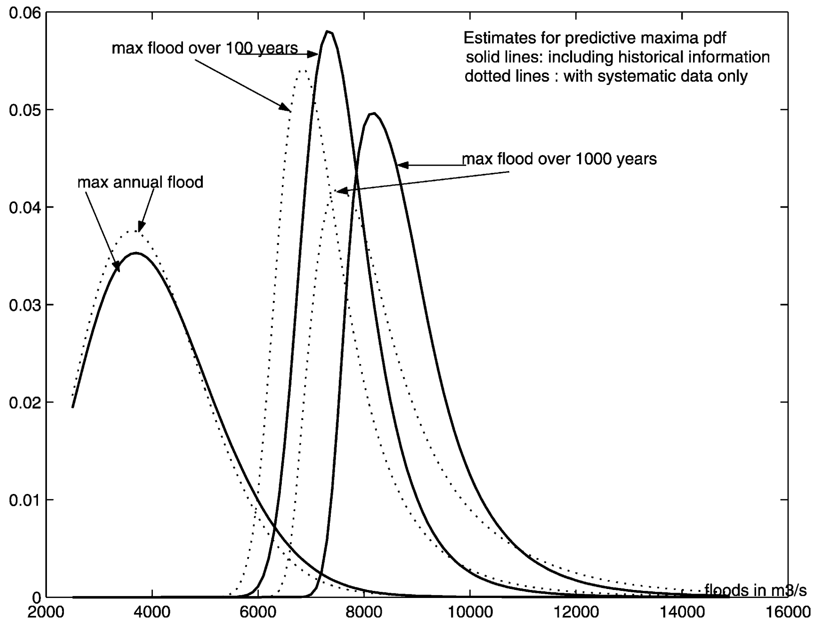

Fig. 6. Predictive maximum flood pdf over periods of 1, 100 and 1000 years with and without historical data.

historical evidence. These curves are the predictive distributions of the maximum flood, i.e. the pdf for the maximum flood on a given period P after integrating out the uncertainty over the parameters. They are less diffuse when historical information is incorporated into the analysis.

7.3. Imprecise historical data

In the following application, imprecise historical values are represented by a normal variable centered on the recorded reference $\bar{y}_i$ with a standard deviation σ and truncated to the interval $[y_i^{\min} = 6000\ \mathrm{m}^3/\mathrm{s}, y_j^{\max} = 8500\ \mathrm{m}^3/\mathrm{s}]$. For the 12 historical records, the corresponding loop is added into the data augmentation algorithm. The width of the 90% credible interval for the 1000 year return percentile reaches 3700 m^3/s with $\sigma = 1000\ \mathrm{m}^3/\mathrm{s}$ and 3060 m^3/s with $\sigma = 500 \times \mathrm{m}^3/\mathrm{s}$, which are intermediate states between the ranges with precise historical data and no historical data. As expected, this width increases with the variance of the error of measurement, but also depends on the information carried by the truncations boundaries.

8. Discussion and concluding remarks

8.1. Frequentist versus Bayesian approaches for flood analysis

On the basis of the systematic records and a point process model with Poisson occurrence for floods exceeding a threshold and Weibull pdf for their magnitudes, Miquel (1984) performed an asymptotic Bayesian estimation for the same case study. Table 4 gives confidence intervals for the 1000 year return design floods with the historical sample. Since that date, it seems that most flood analyses were based on the frequentist North American approach. We did not follow this avenue of thought for the reasons

Table 4
Asymptotic Bayesian credible Intervals for Garonne quantiles (Miquel, 1984) around the Max likelihood estimates (normal approximation)

	Estimate	Sup 95%	Inf 95%
q_{10}	5510	5940	5090
q_{100}	6810	7550	6070
q_{1000}	8730	9370	7550

presented in the introduction. The Garonne example illustrates our arguments:

• Traditional frequentist interpretation of probability can be misleading: for extreme values, the conceptual point of view of repeating experiments may appear paradoxical and unrealistic. In addition, usual frequentist criteria such as unbiasedness or minimum sampling variance are worthless for operational purposes. On the contrary, most decision makers would not give the same weight to over and underestimation, which should be formalized by the analysist. Eq. (12) gives an illustrative cost function to meet such requirements. As a consequence, the expected cost curves from Eq. (13) on Fig. 5 are not symmetrical.

• With a locally uniform prior (as taken in this case), maximum likelihood estimates are asymptotically identical to the mode of the posterior pdf. However, asymptotic normal conditions to be met by ML estimates are not fulfilled here, even with 151 records: Fig. 2 highlights that percentile posterior pdfs are highly skewed. This illustrates that some traditional floods frequency analyses (Ouarda et al., 1998) can be far from coherent when based on asymptotic confidence intervals without validation.

• Model uncertainty can be illustrated by the previous Bayesian inference. Compared with the Exponential model ($\beta = 0$), the generalized POT, here letting β free to vary, increases the range of the 90% credible interval by 180% and lowers by 22% the estimation of the 1000 years percentile when systematic data are used. Conversely, Monte Carlo simulations conditioned upon fixed ML values of the model parameters completely miss the point of model uncertainties. Extreme values frequentist studies with historical data or not, such as in Martins and Stedinger (2001a,b), would therefore ignore parameter (and model) uncertainty and only partially test the properties of old or new estimates. Such model uncertainty may, as in the Garonne case, represent a great part of the unknown, even when restricting the possible models to the GEV family. Consequently, most traditional flood frequency studies give the end user (reasoning with decisional bets) a fallacious feeling of overconfidence in the results derived this way, however, computer intensive the simulation experiment may be.

• The managers rather interpret probabilities in a subjective manner. On the other hand, the frequentist setting is mostly adopted by hydrologists. This may lead to costly misunderstandings until both agree to work and to communicate in the same conceptual framework.

• Missing floods are naturally interpreted as latent variables and the data augmentation algorithm is an essential part of the estimation process. The concept of latent (or hidden) variables extends to many other models such as the binomial censored data model of Stedinger and Cohn (1986) in which the only information used is the number of exceedances above a threshold.

• To incorporate past historical data into the analysis, strong additional modeling assumptions must be made. The most stringent one is the stationarity of the hydraulic regime, which states that the three parameters of the POT model have not changed since 1770. Due to the limiting number of historical events, designing alternate models getting rid of this stationarity assumption is an open challenge since only very few additional parameters can be considered to describe a drift with time or the influence of covariates.

9. Conclusions

The following conclusions have been attained:

- The Garonne case study highlights that conditional model structures are conveniently handled within the Bayesian perspective. The data augmentation algorithms not simply a ‘nice Bayesian trick’, but it unties the implicit mathematical constraints of tractability that curbed the creativity of the analyst. We believe that the released conceptual effort should now be reinvested with much profit into the worthwhile task of modeling.
- The data augmentation algorithm has a direct hydrological interpretation. For a practitioner it seems quite natural to fill the gaps left by the missing values. Bayesian methodology provides the probability-based guidelines to put this idea into practice. Recent advances in graphical representation of the model properties and conditional thinking (Spiegelhalter et al., 1996) can further improve the communication between scientists and end users.
- Historical evidence, even imprecise, provides highly valuable information to reduce model uncertainty. On the Garonne example, the design

value and its credible interval are notably changed when incorporating historical evidence in the study, even if this information is not very precise.

Acknowledgements

We thank Jean Noel Bacro, for his seminal ideas about GEV models. We are indebted to Michel Lang and Bernard Bobée for long-lasting discussions vividly challenging us to show that Bayesian analysis could be of any use in hydrology. We also wish to acknowledge the help from two anonymous referees and the editor.

References

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis, Springer, New York.

Berger, J.O., 1999. Bayesian analysis: a look at today and thoughts of tomorrow. ISDS discussion paper 99-30 (available at http://www.stat.duke.edu).

Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory, Wiley, Londres.

Bernier, J., 1967. Les méthodes Bayésiennes en hydrologie. International Hydrology Symposium, Fort Collins, USA.

Bernier, J., Parent, E., 2002. Encoding prior knowledge to improve risk analysis of extreme hydrological events via POT modeling, submitted for publication.

Box, G., Tiao, G., 1973. Bayesian Inference in Statistical Analysis, Addison-Wesley, reading, MA.

Coles, S.G., 2001. An Introduction to Statistical Modeling of Extreme Values, Springer, New York.

Coles, S.G., Powell, E.A., 1996. Bayesian methods in extreme value modelling: a review and new developments. Int. Statist. Rev. 64, 119–136.

Coles, S.G., Tawn, J.A., 1996. A Bayesian analysis of extreme rainfall data. Appl. Statist. 45 (4), 463–478.

Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds (with discussion). J. R. Stat. Soc., B 52, 393–442.

Gelfand, A., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85, 398–409.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis, Chapman & Hall, Londres.

Hosking, J.R.M., 1985. Maximum likelihood estimation of the parameters of the generalized extreme value distribution. Appl. Stat. 34, 301–310.

Hyfran, 2000. Hydrological Frequency Analysis Software package. Chaire d'hydrologie statistique INRS Eau, software (available at http://www.inrs-eau.uquebec.ca/activites/groupes/chaire_hydrol).

Kuczera, G., 1999. Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. Water Resour. Res. 35 (5), 1551–1557.

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

Martins, E.S., Clarke, R.T., 1993. Likelihood-based confidence intervals for estimating floods with given return periods. J. Hydrol. 147, 61–81.

Martins, E.S., Stedinger, J.R., 2001. Generalized maximum likelihood Pareto–Poisson estimators for partial duration series. Water Resour. Res. 37 (10), 2551–2558.

Martins, E.S., Stedinger, J.R., 2001. Historical information in a generalized maximum likelihood framework with partial duration and annual maximum series. Water Resour. Res. 37 (10), 2559–2568.

Miquel, J., 1984. Guide pratique d'estimation des probabilités de crues, Eyrolles, Paris.

NERC, 1975. Flood study report. National Environment Research Council, London.

Ouarda, T.B.M., Rasmussen, P.F., Bobée, B., Bernier, J., 1998. Utilisation de l'information historique en analyse hydrologique frequentielle. Rev. Sciences de l'Eau, no special, 41–49.

Pardé, M., 1935. Le régime de la Garonne. Revue Géographique des Pyrénées et Sud-Ouest.

Pickands, J., 1975. Statistical inference using extreme order statistics. Ann. Stat. 3, 11–130.

Raiffa, H., Schlaifer, R., 1961. Applied Statistical Decision Theory, Harvard University Press, Cambridge, MA.

Rasmussen, P.F., Ashkar, F., Rosbjerg, D., Bobée, B., 1994. The POT method for flood estimation: a review. In: Hipel, Kluwer (Eds.), Stoch. Stat. Methods Hydrol Envir. Engng, Kluwer, Dordrecht.

Robert, C.P., Casella, G., 1998. Monte Carlo Statistical Methods, Springer, Berlin.

Smith, R.L., 1985. Maximum Likelihood estimation in a class of nonregular cases. Biometrica 72, 67–90.

Smith, R.L., 1986. Extreme value theory based on the r largest annual events. J. Hydrol. 86, 27–43.

Spiegelhalter, D.J., Thomas, A., Best, N.G., 1996. Computation on Bayesian graphical models. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 5, Oxford University Press, Oxford, pp. 407–425.

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

Tanner, M.A., 1996. Tools for Statistical Inference: Methods for the Exploration of Posterior Distribution and Likelihood Functions, Springer, New York.

Tribus, M., 1972. Décision rationnelle dans l'incertain, Masson, Paris.

USWRC, 1992. Guidelines for determining flood frequency. US Water Resources Council, Bulletin, vol. 17.B. Hydrology Subcommittee, Office for Water Data Coordination, Washington, DC.

Wang, Q.J., 1991. The POT model described by the generalized Pareto distribution with Poisson arrival rate. J. Hydrol. 129, 263–280.