HAL

archives-ouvertes.fr

Statistical analysis of extreme events in a nonstationary context via a Bayesian framework. Case study with peak-over-threshold data

B. Renard, M. Lang, P. Bois

▶ To cite this version:

HAL Id: hal-00452224

https://hal.archives-ouvertes.fr/hal-00452224

Submitted on 1 Feb 2010

Statistical analysis of extreme events in a non-stationary context via a Bayesian framework. Case study with peak-over-threshold data.

B. Renard[1], M. Lang[1], P. Bois[2]

(1) *Cemagref Centre de Lyon, U.R. Hydrologie-Hydraulique, 3 bis Quai Chauveau, CP 220, 69336 Lyon cedex 09, France.*

telephone:33 4 72 20 87 72
fax:33 4 78 47 78 75

e-mail : renard@lyon.cemagref.fr

(2) *LTHE, Laboratoire d'études des Transferts en Hydrologie et Environnement, BP53, 38041 Grenoble cedex 09, France.*

Abstract

Statistical analysis of extremes currently assumes that data arise from a stationary process, although such an hypothesis is not easily assessable and should therefore be considered as an uncertainty. The aim of this paper is to describe a Bayesian framework for this purpose, considering several probabilistic models (stationary, step-change and linear trend models) and four extreme values distributions (exponential, generalized Pareto, Gumbel and GEV). Prior distributions are specified by using regional prior knowledge about quantiles. Posterior distributions are used to estimate parameters, quantify the probability of models and derive a realistic frequency analysis, which takes into account estimation, distribution and stationarity uncertainties. MCMC methods are needed for this purpose, and are described in the article. Finally, an application to a POT discharge series is presented, with an analysis of both occurrence process and peak distribution.

Keywords

Bayesian analysis; extreme events; stationarity; uncertainties; climate change; frequency analysis.

1. Introduction

Since Gumbel's precursory work (Gumbel, 1958), a great number of methods have been developed in the statistics of extreme events studies (see Katz et al. (2002), for a recent review). Among them, Bayesian analysis recently received attention, thanks to advances in computational and numerical tools (Coles and Powell, 1996; Coles and Tawn, 1996; Coles and Pericchi, 2003; Coles et al., 2003; Diebolt et al., 2003; Parent and Bernier, 2003).

In most cases, these methods assume that data are independent and stationary, which means that the statistical properties of the studied process are invariant with respect to time. Nevertheless, this assumption sometimes appears very doubtful. For example, with discharges series, changes in land use can modify the floods and droughts regimes. More generally, in a global climate change context, hydro-meteorological data could exhibit trends or step-changes, as has been demonstrated for global temperature, as an example (IPCC, 2001). Strupczewski et al. (2001) therefore derived a likelihood-based approach to model trends in the distribution parameters, including a criterion for the "best" model choice. Cunderlick and Burn (2003) used the same kind of method in order to include data from several sites and to estimate quantiles from pooled frequency analysis. An alternative of interest was proposed by Perreault et al. (2000a; 2000b), by quantifying the probability of each model in a Bayesian framework, and combining their results with appropriate weightings. In this manner, model choice uncertainty can be taken into account, which seems to be a very pragmatic approach as it is impossible to verify what the "true" model is – and even if such a model exists. However, Perreault et al. made an assumption of normality, which prevents this methodology from being applied to extreme events data.

The aim of this article is to derive a Bayesian framework for extreme events analysis, based on a comparison of models. In section 2, we describe the successive steps of the Bayesian approach. The probabilistic models and prior distributions used in this paper are described in sections 3 and 4. Particular attention will be paid to parameter estimation, detection of changes and frequency analysis (section 5). The particular case of the exponential distribution for peak over threshold (POT) or inter-arrival time data is discussed in section 6. Finally, we present an application to hydrological data, with exponential and generalized Pareto distributions (section 7), before drawing some conclusions and discussing potential improvements (section 8).

2. An overview of the Bayesian approach

A complete and detailed presentation of Bayesian analysis and its applications can be found in Berger (1985) or in Gelman et al. (1995). Figure 1 just shows the general principle of the method. As a first step, probabilistic models for the data must be chosen. In the context of the potential changes described here, these models consist of including trends or step changes in the distribution parameters. Given an independent observations vector X, it is then possible to compute the likelihood $p(X|\theta)$ relating to each model, as is usually done in more standard approaches. The fundamental difference of Bayesian analysis concerns the parameters vector θ: it is not viewed as an unknown but constant quantity, which has to be estimated, but as a random vector, whose distribution can be computed from observations. For this purpose, a prior distribution $\pi(\theta)$ is assumed for θ. This distribution reflects the prior knowledge about the parameters, and must be assessed without using the observations. In some cases, the lack of information about a phenomenon leads to the use of so-called non-informative priors. We won't consider this particular case in this paper, thus assuming that prior information is available, however vague it may be. This is not a drastic restriction, as environmental variables can often be connected with not totally unknown physical processes.

Given the likelihood and the prior distribution, Bayes theorem provides the parameters posterior distribution:

$$p(\boldsymbol{\theta}|X) = \frac{\pi(\boldsymbol{\theta})p(X|\boldsymbol{\theta})}{\int \pi(\boldsymbol{\theta})p(X|\boldsymbol{\theta})d\boldsymbol{\theta}} \qquad (1)$$

The posterior distribution can be viewed as an updating of the prior information about $\boldsymbol{\theta}$, thanks to the observations. It provides a more complete estimation than the classical point estimates. However, as soon as the dimension of $\boldsymbol{\theta}$ becomes large, the use of this distribution is no longer straightforward, and specific computational methods are needed. The next sections describe the successive steps to achieve a complete analysis framework.

3. Models for extreme values analysis

Table 1 summarizes the probability density functions (pdf) considered in this paper. These distributions arise from theoretical results, thus applying to a wide range of situations involving extreme values. Two main sampling methods are concerned. Firstly, the block maximum consists of selecting the maximum value within a block. Normally, a one year block is chosen, leading to the annual maxima (AM) series. Secondly, peak over threshold (POT) data consists of selecting all the events exceeding a high threshold (see Rosbjerg and Madsen (2004) for an up-to-date review of literature). The extreme value theorem (Fisher and Tippett, 1928) shows that when the block size tends to infinity, the limit distribution, if it exists, of the block maximum series computed from *iid* variables is a GEV distribution. Similarly, when the threshold value increases, the limit distribution of a POT series can be approximated with a Generalized Pareto distribution (Pickands, 1975). In real-world problems, the data from which the extreme variables are extracted may be too complex to meet the requirements of the above theorems. As an example, daily discharges or rainfalls are neither independent nor identically distributed in case of non-stationarity. Moreover, the block size is usually equal to 365 days, which might be insufficient to apply the asymptotic arguments of theorems' proofs. However, these distributions have been extensively applied in studies involving extreme values (e.g. floods, heavy rainfalls, wave heights), and have demonstrated their usefulness in most hydro-meteorological applications.

Moreover, if a Poisson process is assumed for the occurrence of POTs, it can be shown that a Poisson + Exponential model leads to a Gumbel distribution for the annual maxima, whereas Poisson + Generalized Pareto similarly leads to the GEV distribution (Lang et al., 1999). Finally, the inter-arrivals time between two events can be modelled with a one-parameter exponential distribution when a Poisson process is assumed for the occurrences (Lang, 1999).

With our notations, σ is the threshold value, which is supposed to be known, and μ, λ, and ξ are the location, scale and shape parameters. The two-parameter distributions (Exponential and Gumbel) can be considered as limiting cases of the corresponding three-parameter ones (Generalized Pareto and GEV), when $\xi \to 0$.

In order to take into account non-stationary data, a step change or a linear trend can be assumed for the parameters, as formalized in table 2. t denotes time, in a discrete (AM) or continuous (POT) way. Notice that the shape parameter ξ is always assumed to be constant (although the same kind of changes could theoretically be modelled), because of the difficulty of precisely estimating this parameter. Moreover, other kinds of changes could be modelled, such as a linear trend on the log-transformed scale parameter (Katz et al., 2002), non-simultaneous step changes in the location and the scale parameters, multiple step changes or polynomial trends. Although the methodology presented here can take into account such kind of models, it always has to be kept in mind that complex models can improve the description

of a phenomenon, but can dramatically decrease the accuracy of the estimations or increase the numerical difficulties.

Finally, for each of the 12 models presented in table 2, the likelihood of an observed vector $X = (x_1, ..., x_n)$ can be computed as follows:

$$p(X \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} p(x_i \mid \boldsymbol{\theta}) \qquad (2)$$

4. Specifying priors

Scientists usually have to estimate some environmental characteristics at one site of interest, without any data being available at this particular location. Standard statistical analysis is then impossible, and alternative methods have been developed for this purpose. Regionalization can thus be used to transfer knowledge from gauged to ungauged sites (GREHYS, 1996; Javelle et al., 2002; Madsen et al., 2002). A more physical approach consists of deducing some estimates from underlying physical processes. As an example, in the hydrology field, a large number of methods have been used to relate stream flow to explanatory variables, such as rainfall, topography of the watershed, soil properties, etc (see Reed (1999) for an example in UK).

When specifying priors, the situation is the same, as data must not be used at this step. The non-exhaustive above list of examples shows that regional knowledge is often available. Such information just has to be transferred to the probabilistic model parameters. In extreme values analysis, a variable of current concern is the p-quantile q_p, that is the value which is exceeded with probability $(1\text{-}p)$. Coles and Tawn (1996) thus proposed a method for producing prior distributions from these quantiles. Let $p(x|\boldsymbol{\theta})$ be one of the stationary pdf (models M_0), with dimension k ($\boldsymbol{\theta} \in \mathbb{R}^k$). Let $p_1 < ... < p_k$, and $q_{p_1} < ... < q_{p_k}$ be the corresponding quantiles, which can be expressed as $\boldsymbol{\theta}$ and p functions. As an example, with the $M_0^{\exp}$ model, $q_p = \sigma - \lambda \log(1-p)$. As a natural order exists between $q_{p_1}, ..., q_{p_k}$, thus leading to strongly non-independent values, prior production is done using the following modified quantiles:

$$\tilde{q}_1 = q_{p_1}, \ \tilde{q}_2 = q_{p_2} - q_{p_1}, ..., \tilde{q}_k = q_{p_k} - q_{p_{k-1}} \qquad (3)$$

Let $G(y|a, b)$ be the pdf of a gamma distribution with parameters a and b:

$$G(y \mid a,b) = \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by} 1_{\{t>0\}}(y) \qquad (4)$$

where $\Gamma(.)$ denotes the gamma function and $1_\Omega(y)$ is the indicator function of the Ω set ($1_\Omega(y) = 1$ if $y \in \Omega$, 0 otherwise). The $(\tilde{q}_i)_{i=1,...,k}$ are supposed to follow independent gamma distributions:

$$\pi(\tilde{q}_1, ..., \tilde{q}_k) = G(\tilde{q}_1 \mid a_1, b_1)...G(\tilde{q}_k \mid a_k, b_k) \qquad (5)$$

The gamma distribution is used because of its flexibility, but alternative choices are possible. The hypothesis of independence is used here for convenience. Nevertheless, this is not a drastic restriction, as the joint prior distribution must reflect diffuse knowledge, which will be updated in the following step. Moreover, structural dependence between parameters will be preserved when transferring prior knowledge from the quantiles space to the parameters space.

The prior distribution is finally computed by replacing the $\tilde{q}_i$ by their expression in terms of $\boldsymbol{\theta}$, and multiplying by the Jacobian of the transformation $\boldsymbol{\theta} \mapsto (\tilde{q}_1, ..., \tilde{q}_k)$:

$$\pi(\boldsymbol{\theta}) = G(\tilde{q}_1(\boldsymbol{\theta}) \mid a_1, b_1) ... G(\tilde{q}_k(\boldsymbol{\theta}) \mid a_k, b_k) \mid \boldsymbol{J} \mid \tag{6}$$

where $\boldsymbol{J}$ is the Jacobian matrix of the above transformation, and the couples (a_i, b_i) are called the hyperparameters of the model, which have to be fixed at a particular value, reflecting the prior knowledge about the modified quantiles $(\tilde{q}_i)_{i=1,...,k}$. Tables 3a and 3b show the $q_p(\boldsymbol{\theta})$ mathematical expressions in each model M_0, M_1 and M_2, and the corresponding Jacobian matrixes. The determinants of these matrixes can be computed numerically with any mathematical software.

In the case of a stationary model M_0, prior specification is thus related to at most three modified quantiles $(\tilde{q}_1, \tilde{q}_2, \tilde{q}_3)$. Prior specification can be achieved in the same manner with non-stationary models M_1 and M_2, despite an increase in the $\boldsymbol{\theta}$ dimension. In the step change model (M_1), a discrete distribution $\pi(\tau)$ has to be assumed for the change point τ (uniform on $1,...,n-1$ for instance). Let r be an integer such as $1 \leq r < k$. Prior specifications are the following:

$$\begin{cases} \text{Before change, } (\tilde{q}_i^{(1)})_{i=1,...,r} \text{ follow independent gamma distributions.} \\ \text{After change, } (\tilde{q}_i^{(2)})_{i=1,...,k-r} \text{ follow independent gamma distributions.} \end{cases}$$

Assuming independence between the change point τ and the $\tilde{q}_i$, prior distribution is then:

$$\begin{aligned} \pi(\boldsymbol{\theta}) = &G(\tilde{q}_1^{(1)}(\boldsymbol{\theta}) \mid a_1^{(1)}, b_1^{(1)}) ... G(\tilde{q}_r^{(1)}(\boldsymbol{\theta}) \mid a_r^{(1)}, b_r^{(1)}) \\ &\times G(\tilde{q}_1^{(2)}(\boldsymbol{\theta}) \mid a_1^{(2)}, b_1^{(2)}) ... G(\tilde{q}_{k-r}^{(2)}(\boldsymbol{\theta}) \mid a_{k-r}^{(2)}, b_{k-r}^{(2)}) \pi(\tau) \mid \boldsymbol{J} \mid \end{aligned} \tag{7}$$

If no prior knowledge about the existence of a change is available, the same specifications for quantiles with same probability before and after change can be chosen:

$$(a_i^{(1)}, b_i^{(1)}) = (a_i^{(2)}, b_i^{(2)}), \forall i = 1,..., \min(r, k-r) \tag{8}$$

If a trend is assumed in the parameters (M_2), let $t_1 < t_2$ be two time points, and $1 \leq r < k$. Prior specifications are then:

$$\begin{cases} \text{At } t = t_1, (\tilde{q}_i(t_1))_{i=1,...,r} \text{ follow independent gamma distributions.} \\ \text{At } t = t_2, (\tilde{q}_i(t_2))_{i=1,...,k-r} \text{ follow independent gamma distributions.} \end{cases}$$

The prior distribution can thus be written:

$$\begin{aligned} \pi(\boldsymbol{\theta}) = &G(\tilde{q}_1(\boldsymbol{\theta}(t_1)) \mid a_1^{(1)}, b_1^{(1)}) ... G(\tilde{q}_r(\boldsymbol{\theta}(t_1)) \mid a_r^{(1)}, b_r^{(1)}) \\ &\times G(\tilde{q}_1(\boldsymbol{\theta}(t_2)) \mid a_1^{(2)}, b_1^{(2)}) ... G(\tilde{q}_{k-r}(\boldsymbol{\theta}(t_2)) \mid a_{k-r}^{(2)}, b_{k-r}^{(2)}) \mid \boldsymbol{J} \mid \end{aligned} \tag{9}$$

As previously explained in the M_1 case, equation (8) can be used to complete the prior specifications.

At last, the approach proposed by Coles and Tawn (1996) was exposed here because regional or empirical knowledge is often expressed in terms of quantiles values, rather than parameters values. However, alternative prior distributions can be used without additional difficulties within the presented framework.

5. Computations on posterior distributions

The methods described in sections 3 and 4 allow Bayes theorem to be used to compute the posterior distribution $p(\theta|X)$. However, the integral of the denominator of equation (1) is generally intractable, $p(\theta|X)$ is thus only known up to a constant of proportionality:

$$p(\theta \mid X) \propto p(X \mid \theta)\pi(\theta) \tag{10}$$

Moreover, the high dimensionality of $p(\theta|X)$ makes this function difficult to handle in practice. Next sections present some methods to overcome this drawback. Readers interested in further exploration of posterior distributions than those described below can refer to Tanner (1996).

5.1. Parameters estimation via marginal densities

The $p(\theta|X)$ j^{th} marginal is defined by:

$$p(\theta_j \mid X) = \int p(\theta_1, ..., \theta_k \mid X)d\theta_1 ...d\theta_{j-1}d\theta_{j+1}...d\theta_k \tag{11}$$

Once again, this integration is often intractable. No analytical expression of the posterior marginal is then available. An alternative is to simulate a sample from the joint posterior distribution $p(\theta|X)$. Two main methods can be used for this purpose: the Metropolis-Hastings algorithm (Metropolis and Ulam, 1949; Metropolis et al., 1953; Hastings, 1970) and the griddy Gibbs sampling (Ritter and Tanner, 1992). The latter is conceptually easier to understand, and is thus presented in detail in Annex 1, although the former method may be more efficient in some situations.

Given a sample of N vectors $(\theta^{(i)})_{i=1,...,N} = (\theta_1^{(i)},...,\theta_k^{(i)})_{i=1,...,N}$, a sample from the jth marginal is simply obtained by considering $(\theta_j^{(i)})_{i=1,...,N}$. Estimation can then be made by means of this sample: the mean or the median gives a point estimate, and a confidence interval can be calculated by computing lower and upper empirical quantiles. Notice that the values $(\theta_j^{(i)})_{i=1,...,N}$ are not independent, a large sample size is thus required to avoid any bias due to autocorrelation. A sensitivity analysis could be necessary to determine a critical sample size, above which estimates stay stable.

5.2. Detection of changes

We applied the methodology described by Perreault at al. (2000b), based on Bayes factors (Kass and Raftery, 1995), to detect changes in environmental series. Let us define the marginal distribution of X as follow:

$$p(X \mid M) = \int p(X, \theta \mid M)d\theta \tag{12}$$

A simple calculation shows that $p(X|M)$ is the denominator of the Bayes theorem applied to a particular model M:

$$p(X \mid M) = \int p(X \mid \theta, M)\pi(\theta \mid M)d\theta \tag{13}$$

As stated previously, this quantity can generally not be explicitly computed. Chib (1995) proposed a simple method for estimating the marginal distribution of X, based on griddy Gibbs sampling. A detailed description of this approach is made in Annex 2. Let $\Omega = \{M_1, ..., M_q\} = \Omega_1 \cup \Omega_1^c$, where the exponent c denotes the complement of a set, be a

partition of the collection of models considered. As an example, $\Omega = \left\{ M_0^{\exp}, M_1^{\exp}, M_2^{\exp}, M_0^{gp}, M_1^{gp}, M_2^{gp} \right\} = \left\{ M_0^{\exp}, M_0^{gp} \right\} \cup \left\{ M_1^{\exp}, M_1^{gp}, M_2^{\exp}, M_2^{gp} \right\}$ could be a partition for POT data models. More generally, the partition $\Omega_1 \cup \Omega_1^c$ can take the form "stationary vs non stationary models", or "2 parameters vs 3 parameters models". Assume that $\pi(.)$ is a prior probability distribution on Ω, the posterior probability of a model M_j can then be computed with the Bayes theorem:

$$p(M_j \mid X) = \frac{p(X \mid M_j)\pi(M_j)}{\sum_{i=1}^{q} p(X \mid M_i)\pi(M_i)} \tag{14}$$

The Bayes factor between models M_i and M_j is then defined by:

$$B_{i,j} = \left(\frac{p(M_i \mid X)}{p(M_j \mid X)} \right) \Bigg/ \left(\frac{\pi(M_i)}{\pi(M_j)} \right) = \frac{p(X \mid M_i)}{p(X \mid M_j)} \tag{15}$$

A composite Bayes factor between Ω_1 et Ω_1^c can also be computed:

$$B_{\Omega_1, \Omega_1^c} = \left(\frac{\sum_{M \in \Omega_1} p(M \mid X)}{\sum_{M \in \Omega_1^c} p(M \mid X)} \right) \Bigg/ \left(\frac{\sum_{M \in \Omega_1} \pi(M)}{\sum_{M \in \Omega_1^c} \pi(M)} \right) \tag{16}$$

The Bayes factor provides a measurement of the pertinence of a hypothesis compared to another. It can thus be compared to classical likelihood ratios. However, model likelihood is not computed by replacing the parameters by their ML estimates, but by integrating on the prior distribution. More generally, this approach is not a classical test procedure. The Bayes factor value depends on prior specifications in addition to sampling variance. Thus, it cannot be interpreted as a risk in terms of limit error percentage when resampling size tends to infinity, but rather in terms of a false decision cost (see Perreault (2000) for further explanations). Classically, the two error types are assumed to have the same cost. The Bayes factor is then compared to 1, with greater certainty for high or near zero values (Kass and Raftery, 1995).

5.3. Frequency analysis.

Section 5.1 provides a way of simulating a sample for the $p(\theta|X)$ distribution, which relates to estimation uncertainty. Section 5.2 describes the computing of posterior probabilities of the models: as a collection of models cannot be considered totally exhaustive, a part of the modelling uncertainty is thus controlled. These two kinds of uncertainty can be taken into account in frequency analysis, by the following procedure:

Repeat N times:
1. Random choice of a model M^*, with probabilities $p(M_1|X),...,p(M_k|X)$.
2. Random choice of a parameter vector θ^* from the sample of $p(\theta|X, M^*)$.
3. Computation of $q_p(t)$, the p-quantile in model M^* with parameter θ^* at time t.

The quantiles collection obtained can be considered as a sample of the quantile distribution at time t, including estimation and modelling uncertainties. It can thus be used to provide a point estimate, a confidence interval or even a distribution estimate for $q_p(t)$. If data

stationarity is questionable, the function $t \mapsto q_P(t)$ provides a description of the possible change in terms of quantile drift with time.

6. A particular case: the exponential distribution

Up to this point, the methods used in this paper have been described as generally as possible. The framework is thus still applicable with other models. Nevertheless, distributions belonging to the exponential family can be treated in a simpler way, with analytical results being available without using MCMC methods. In the next sections, posterior marginal distributions and the marginal distribution of observations are computed for exponential models, leading to substantial simplifications. Notice that Parent and Bernier (2003) also succeeded in deriving a MCMC-free method in the case of a stationary generalized Pareto distribution.

6.1. Stationary model

The likelihood for $M_0^{\exp}$ model is:

$$p(X \mid \theta) = \prod_{i=1}^{n} p(x_i \mid \theta) = \frac{1}{\lambda^n} \exp\left(-\frac{n(\bar{X} - \sigma)}{\lambda}\right) \tag{17}$$

Viewed as a function of λ, this likelihood is proportional to an inverse gamma pdf $IG(y|a,b)$:

$$IG(y \mid a,b) = \frac{b^a}{\Gamma(a)} \frac{1}{y^{a+1}} \exp(-\frac{b}{y}) \mathbb{1}_{\{t>0\}}(y) \tag{18}$$

This particularity is known as the conjugacy property. Thus, if $\pi(\lambda)$ is supposed to follow an inverse gamma distribution with parameters a and b, the posterior distribution of λ is:

$$\begin{aligned} p(\lambda \mid X) &\propto \pi(\lambda)p(X \mid \lambda) \\ &= \frac{b^a}{\Gamma(a)} \frac{1}{\lambda^{n+a+1}} \exp\left(-\frac{b+n(\bar{X}-\sigma)}{\lambda}\right) \\ &= f(\lambda) \end{aligned} \tag{19}$$

Up to proportionality, λ posterior distribution is still an inverse gamma distribution, with updated parameters:

$$\begin{aligned} a' &= a+n \\ b' &= b+n(\bar{X}-\sigma) \end{aligned} \tag{20}$$

The marginal distribution of observations is finally simply obtained by integrating $f(\lambda)$:

$$\begin{aligned} p(X \mid M_0^{\exp}) &= \int f(\lambda)d\lambda \\ &= \frac{b^a}{\Gamma(a)} \int \frac{1}{\lambda^{a'+1}} \exp\left(-\frac{b'}{\lambda}\right) d\lambda \\ &= \frac{b^a}{\Gamma(a)} \frac{\Gamma(a')}{b'^{a'}} \end{aligned} \tag{21}$$

6.2. Step change model

Following the table 2 notations, the $M_1^{\exp}$ likelihood is:

$$p(\boldsymbol{X}\,|\,\boldsymbol{\theta}) = \left(\frac{1}{\lambda_1}\right)^{\tau}\left(\frac{1}{\lambda_2}\right)^{n-\tau}\exp\left(-\sum_{i=1}^{\tau}\frac{x_i-\sigma}{\lambda_1}\right)\exp\left(-\sum_{i=\tau+1}^{n}\frac{x_i-\sigma}{\lambda_2}\right) \qquad (22)$$

For fixed τ, the likelihood is proportional to the product of two inverse gamma distributions. Assuming independence between the change point and λ_i values, the prior distribution is:

$$\pi(\boldsymbol{\theta}) = IG(\lambda_1\,|\,a_1,b_1\,)IG(\lambda_2\,|\,a_2,b_2\,)\pi(\tau) \qquad (23),$$

where $\pi(\tau)$ is any discrete distribution on $1,\ldots,n\text{-}1$. Once again, a simple calculation shows that, for a fixed τ the posterior distribution is still proportional to a product of inverse gamma densities:

$$p(\boldsymbol{\theta}\,|\,\boldsymbol{X}) \propto IG(\lambda_1\,|\,a_1,b_1\,)IG(\lambda_2\,|\,a_2,b_2\,)\pi(\tau)p(\boldsymbol{X}\,|\,\boldsymbol{\theta})$$

$$= \frac{b_1^{a_1}}{\Gamma(a_1)}\frac{b_2^{a_2}}{\Gamma(a_2)}IG(\lambda_1\,|\,a_1',b_1'\,)IG(\lambda_2\,|\,a_2',b_2'\,)\pi(\tau) \qquad (24)$$

$$= f(\lambda_1,\lambda_2,\tau)$$

with:

$$a_1' = a_1 + \tau$$

$$a_2' = a_2 + n - \tau$$

$$b_1' = b_1 + \sum_{i=1}^{\tau}(x_i - \sigma) \qquad (25)$$

$$b_2' = b_2 + \sum_{i=\tau+1}^{n}(x_i - \sigma)$$

The marginal posterior distribution can directly be computed by integration from the joint posterior distribution:

$$p(\tau \mid \boldsymbol{X}) \propto \int f(\lambda_1, \lambda_2, \tau) d\lambda_1 d\lambda_2$$

$$\propto \frac{\Gamma(a_1')\Gamma(a_2')}{(b_1')^{a_1'}(b_2')^{a_2'}} \pi(\tau)$$

$$p(\lambda_1 \mid \boldsymbol{X}) \propto \int f(\lambda_1, \lambda_2, \tau) d\lambda_2 d\tau$$

$$\propto \sum_{\tau=1}^{n-1} \frac{\Gamma(a_2')}{(b_2')^{a_2'}} \pi(\tau) \left(\frac{1}{\lambda_1}\right)^{a_1'+1} \exp\left(-\frac{b_1'}{\lambda_1}\right) \tag{26}$$

$$p(\lambda_2 \mid \boldsymbol{X}) \propto \int f(\lambda_1, \lambda_2, \tau) d\lambda_1 d\tau$$

$$\propto \sum_{\tau=1}^{n-1} \frac{\Gamma(a_1')}{(b_1')^{a_1'}} \pi(\tau) \left(\frac{1}{\lambda_2}\right)^{a_2'+1} \exp\left(-\frac{b_2'}{\lambda_2}\right)$$

Finally, the marginal distribution of observations is:

$$p(\boldsymbol{X} \mid M_1^{\exp}) = \int f(\lambda_1, \lambda_2, \tau) d\lambda_1 d\lambda_2 d\tau$$

$$= \sum_{\tau=1}^{n-1} \frac{\Gamma(a_1')\Gamma(a_2')}{\Gamma(a_1)\Gamma(a_2)} \frac{(b_1)^{a_1}(b_2)^{a_2}}{(b_1')^{a_1'}(b_2')^{a_2'}} \pi(\tau) \tag{27}$$

6.3. Trend model

The likelihood can be written as:

$$p(\boldsymbol{X} \mid \boldsymbol{\theta}) = \left(\frac{1}{\lambda_0}\right)^n \prod_{i=1}^{n} \left(\frac{1}{1+\lambda_1 t_i}\right) \exp\left(-\frac{1}{\lambda_0} \sum_{i=1}^{n} \frac{x_i - \sigma}{1+\lambda_1 t_i}\right) \tag{28}$$

For fixed λ_1, this likelihood viewed as a λ_0 function is proportional to an inverse gamma pdf. Unfortunately, no conjugacy property appears for the λ_1 parameter. Any prior distribution can thus be chosen for λ_1, for instance a normal distribution with parameters m and s. The joint prior distribution is then:

$$\pi(\boldsymbol{\theta}) = IG(\lambda_0 \mid a,b) N(\lambda_1 \mid m,s) \tag{29},$$

leading to the following posterior distribution:

$$p(\boldsymbol{\theta} \mid \boldsymbol{X}) \propto p(\boldsymbol{X} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta})$$

$$= \left(\frac{1}{\lambda_0}\right)^{n+a+1} \frac{b^a}{\Gamma(a)} \frac{1}{s\sqrt{2\pi}} \prod_{i=1}^{n} \left(\frac{1}{1+\lambda_1 t_i}\right) \exp\left(-\frac{1}{\lambda_0}\left(b + \sum_{i=1}^{n} \frac{x_i - \sigma}{1+\lambda_1 t_i}\right) - \frac{(\lambda_1 - m)^2}{2s^2}\right) \tag{30}$$

$$= f(\lambda_0, \lambda_1)$$

The f function can be explicitly integrated with respect to λ_0, thus giving the marginal posterior distribution of λ_1:

$$p(\lambda_1 \mid X) \propto \int f(\lambda_0, \lambda_1) d\lambda_0$$

$$= \frac{b^a}{\Gamma(a)} \frac{\Gamma(n+a)}{\left(b + \sum_{i=1}^{n} \frac{x_i - \sigma}{1 + \lambda_1 t_i}\right)^{n+a}} \frac{1}{s\sqrt{2\pi}} \prod_{i=1}^{n} \left(\frac{1}{1 + \lambda_1 t_i}\right) \exp\left(-\frac{(\lambda_1 - m)^2}{2s^2}\right) \quad (31)$$

$$= f_1(\lambda_1)$$

The marginal posterior distribution of λ_0 and the marginal distribution of observations can be computed by 1-D numerical integration:

$$p(\lambda_0 \mid X) \propto \int f(\lambda_0, \lambda_1) d\lambda_1$$

$$p(X \mid M_2^{\exp}) = \int f(\lambda_0, \lambda_1) d\lambda_0 d\lambda_1 \quad (32)$$

$$= \int f_1(\lambda_1) d\lambda_1$$

7. Application

The Bayesian method was applied to a stream flow series from the Drôme river at Luc en Diois (194 km^2, South East of France). Daily stream flows were available between the years 1907 and 2003. We extracted a series of independent peaks above a 23 m^3.s^{-1} threshold, leading to a sample of 93 values. Over-threshold values and inter-arrivals series were studied (figure 2). The former was studied with the six models $M_0^{\exp}, M_1^{\exp}, M_2^{\exp}, M_0^{gp}, M_1^{gp}$ and M_2^{gp}, the latter with $M_0^{\exp}, M_1^{\exp}$ and $M_2^{\exp}$.

7.1. Prior specification

The first step consisted of specifying a discrete prior distribution on the models collection we considered. In order to avoid any bias in the hypothesis we wished to evaluate (stationary *vs.* non-stationary and Exponential *vs.* Generalized Pareto), the following probabilities were chosen:

$$\begin{cases} \pi(M_0^{\exp}) = \pi(M_0^{gp}) = 0.25 \\ \pi(M_1^{\exp}) = \pi(M_2^{\exp}) = \pi(M_1^{gp}) = \pi(M_2^{gp}) = 0.125 \end{cases}$$

for the over-threshold series, and

$$\begin{cases} \pi(M_0^{\exp}) = 0.5 \\ \pi(M_1^{\exp}) = \pi(M_2^{\exp}) = 0.25 \end{cases}$$

for the inter-arrivals series. In this manner, each hypothesis had an equal prior probability.

The second step consisted of evaluating our prior knowledge on quantiles, as explained in section 4. We applied the methodology proposed by Galéa and Prudhomme (1997) and Javelle et al. (1999), known as the QdF (flow-duration-frequency) models. As a first step, we estimated the instantaneous discharge quantile with probability 0.9 by using the Crupedix formula (CTGREF et al., 1980-1982):

$$q_{0.9} = S^{0.8} \left(\frac{r_{0.9}}{80} \right)^2 R \qquad (33)$$

where S is the watershed area, $r_{0.9}$ is the daily rainfall quantile with probability 0.9, and R is a regional index. The Socose duration D (CTGREF et al., 1980-1982), which characterizes the flood duration, was then computed:

$$\log(D) = -0.69 + 0.32 \log(S) + 2.2 \left(\frac{\overline{r}}{r_{0.9}.\overline{t}} \right)^{1/2} \qquad (34)$$

where $\overline{r}$ is the mean annual rainfall and $\overline{t}$ the mean annual temperature. All the parameters needed for the calculation of $q_{0.9}$ and D are available on maps. The QdF methodology can then be used to deduce flood quantiles with other durations or frequencies, by comparison with three reference data sets, representative of the three main groups of hydrological flood regimes in France. The reference data set suitable for the Drôme river had already been chosen by Prudhomme (1995).

We thus obtained 0.9- and 0.99-quantiles of daily stream flows, leading to $\hat{\tilde{q}}_1$ and $\hat{\tilde{q}}_2$ values as defined in section 4. In order to fit a two-parameters prior distribution to each modified quantile, an indication about their variability was also needed. In the Crupedix method, a 90% confidence interval is $[q/2; 2q]$. Because of the lack of information about the final uncertainty in our prior specification process, we decided to use the same shape of confidence intervals for the modified quantiles. We thus evaluated the hyperparameters a and b of the Gamma distribution, so that the mean prior value of $\tilde{q}_i$ was equal to $\hat{\tilde{q}}_i$ and the 0.95-quantile of the gamma distribution was equal to $2 \times \hat{\tilde{q}}_i$. In the exponential case, the simplified procedure described in section 6 was applied. The estimates $\hat{\lambda} = \dfrac{\sigma - \hat{\tilde{q}}_1}{\log(1 - p_1)}$ and

$\hat{\lambda}' = \dfrac{\sigma - 2 \times \hat{\tilde{q}}_1}{\log(1 - p_1)}$ were computed, and the hyperparameters a and b of the Inverse Gamma distribution were evaluated so that the mean was equal to $\hat{\lambda}$ and the 0.95-quantile was equal to $\hat{\lambda}'$. Finally, the same value of the a hyperparameter was chosen for the prior distribution relative to inter-arrivals (leading to the same coefficient of variation), and the b hyperparameter was computed to obtain a mean value of 1 for the λ parameter (leading to one event per year on average). Equation (8) was used in models M_1 and M_2. Table 4 gives the results of the prior specification process.

7.2. Results

Table 5 summarizes the models' posterior probabilities and associated Bayes factors. According to the scale proposed by Kass and Raftery (1995), evidence for non-stationarity of the inter-arrivals series appears to be positive, a linear trend on the exponential distribution being the most probable model. Figure 3 shows the process mean for each of the three models. For this purpose, parameter estimates are the means of marginal posterior distributions (figure 4), except for the change point in model M_1^{exp}, which is the mode of the discrete posterior distribution.

No clear evidence appears concerning the stationarity of the over-threshold series, with a Bayes factor near to 1. Conversely, the generalized Pareto distribution seems to be most efficient, as shown by figure 5 for models M_0^{exp} and M_0^{gp}. Surprisingly, the model M_1^{gp} appears to be the more probable, but does not provide an accurate estimate of the change-point (Figure 6). One possible explanation could be the sensitivity of Chib approach to the starting point: the accuracy of the method may be lower when no clear isolated mode appears in the posterior distribution of τ. Nevertheless, the posterior distribution of the scale parameter after change is very close to that of the scale parameter in model M_0^{gp} (Figure 7). Consequently, results of the frequency analysis should also be very close for these two models.

Figure 8 shows the result of such an analysis. In order to take into account changes in the inter-arrivals process Y (which is supposed to be independent from the over-thresholds X), the procedure presented in section 5.3 was modified as follows:

Repeat N times:
1. Random choice of a model M^* for the over-thresholds, with probabilities $p(M_1|X),...,p(M_k|X)$.
2. Random choice of a parameter vector θ^* from the sample of $p(\theta|X, M^*)$.
3. Random choice of a model M^+ for the inter-arrivals, with probabilities $p(M_1|Y),...,p(M_k|Y)$.
4. Random choice of a parameter vector θ^+ from the sample of $p(\theta|Y, M^+)$.
5. Computation of $\mu(t)$ $(=1/\lambda(t))$, the mean number of events per year in model M^+ at time t, with parameter θ^+.
6. Computation of the probability relative to a return period T at time t: $p = 1-1/T\mu(t)$.
7. Computation of $q_P(t)$, the p-quantile in model M^* with parameter θ^* at time t.

Flood quantiles have decreased by about 19% between years 1907 and 2003. Figure 9 focuses on the trajectory of the 10-year return flood. The curve shape in the early years is due to the step-change model: once the posterior probability of a step change has become too weak, this model assumes that the process will stay in the second state, thus behaving like a stationary model. The curve is thus smoothed after such a date. Prediction of quantile future values must be regarded with caution, because we neglect the uncertainty due to change persistency in the future. Nevertheless, such a figure emphasizes the ambiguity relative to the use of "return periods" for characterizing frequency in a non-stationary context. In such a case, it would be better to use the 0.9-quantile, which has a 0.1 probability of being exceeded in the next year, this value being different each year.

8. Conclusion and discussion

The aim of this article was to derive a Bayesian framework to deal with data for which stationarity is questionable. The method presented here can take into account modelling uncertainties, in order to obtain a more realistic quantification of the results accuracy. This should result in wider confidence intervals, compared to methods ignoring the modelling uncertainties. Conversely, prior specifications can be used to inject physical or regional knowledge about a phenomenon, which may increase the accuracy of estimates. Coles and Powell (1996) thus noted that prior spatial information can be very useful for estimating scale and shape parameters of a GEV distribution, even with few data, whereas the location parameter is mostly site-specific and data-dependent. More information about the advantages and drawbacks of Bayesian methods, compared to likelihood-based ones, in extreme values analysis are needed. More precisely, future studies could compare the accuracy of estimates, the robustness to outliers, and investigate the influence of prior specifications (or misspecifications).

Although we focused on 12 particular models, including two sampling methods, two kinds of distributions and three change possibilities, the framework can easily be used in other contexts, like polynomial changes, multiple step changes, seasonal distributions mixture, use of covariates, etc. Some hypotheses we made could also be relaxed by means of appropriate models. Firstly, we assumed that data arise from an independent sample, which can sometimes be very doubtful. ARMA models are then good candidates to overcome this difficulty (see Ray and Tsay (2002) for instance). Secondly, we assumed any change to be final. Thus, when forecasting a future quantile value, the uncertainty due to change persistency is not taken into account. In the step-change case, Hidden Markov Chains can be used for this purpose (Robert et al., 2000; Perreault and Fortin, 2003). Nevertheless, most of the published approaches assumed data normality. Some improvements are thus still foreseeable in these topics.

Finally, one of the more promising tasks would be to derive a Bayesian multivariate method for extreme values analysis. More precisely, in a global climate change context, at-site analysis is not satisfactory. Unfortunately, generalization of existing methods (Perreault et al., 2000c) is not straightforward for extreme values distributions. The use of copulas (Favre et al., 2004) might be considered for a restricted number of locations. Adaptation of the flood-duration-frequency approach (Javelle et al., 2002) in a non-stationary context is also one of the possibilities. At last, hierarchical Bayesian models, including trends or step-changes parameters, may be useful for regional analysis of extremes (Cooley, 2005; Cooley et al., 2005), as well as for measurement uncertainties handling (Reis and Stedinger, 2005).

Acknowledgements

Bibliography

Berger, J. O. 1985. Statistical decision theory and bayesian analysis. Springer-Verlag. 617 p. New York.

Chib, S. 1995. Marginal Likelihood From the Gibbs Output. Journal of American Statistical Association **90**:1313-1321.

Coles, S., and L. Pericchi. 2003. Anticipating catastrophes through extreme value modelling. Journal of the Royal Statistical Society Series C-Applied Statistics **52**:405-416.

Coles, S., L. R. Pericchi, and S. Sisson. 2003. A fully probabilistic approach to extreme rainfall modelling. Journal of Hydrology **273**:35-50.

Coles, S. G., and E. A. Powell. 1996. Bayesian methods in extreme value modelling: A review and new developments. International Statistical Review **64**:119-136.

Coles, S. G., and J. A. Tawn. 1996. A Bayesian analysis of extreme rainfall data. Applied Statistics-Journal of the Royal Statistical Society Series C **45**:463-478.

Cooley, D. 2005. Statistical Analysis of Extremes Motivated by Weather and Climate Studies: Applied and Theoretical Advances. University of Colorado. 122 p.

Cooley, D., D. Nychka, and P. Naveau. 2005. A Spatial Bayesian Hierarchical Model for a Precipitation Return Levels Map. *in*. Extreme Value Analysis, Gothenburg, Sweden.

CTGREF, Srae, and S. H. Diame. 1980-1982. Synthèse nationale sur les crues des petits bassins versants. Fascicule 2 : la méthode Socose ; Information Technique n° 38-2 (Juin 1980) ; Fascicule 3 : la méthode Crupedix.

Cunderlik, J. M., and D. H. Burn. 2003. Non-stationary pooled frequency analysis. Journal of Hydrology **276**:210-223.

Diebolt, J., M. A. El-Aroui, M. Garrido, and S. Girard. 2003. Quasi-conjugate Bayes estimates for GPD parameters and application to heavy tails modelling. Rapport de recherche INRIA. 29 p.

Favre, A. C., S. El Adlouni, L. Perreault, N. Thiemonge, and B. Bobee. 2004. Multivariate hydrological frequency analysis using copulas. Water Resources Research **40**.

Fisher, R. A., and L. H. Tippett. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Cambridge Phil. Soc. **24**.

Galéa, G., and C. Prudhomme. 1997. Notions de base et concepts utiles pour la compréhension de la modélisation synthétique des régimes de crue des bassins versants au sens des modèles QdF. Revue des Sciences de l'Eau **1**:83-101.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. Bayesian data analysis. C. Hall. 526 p. Chapman & Hall.

GREHYS. 1996. Presentation and review of some methods for regional flood frequency analysis. Journal of Hydrology **186**:63-84.

Gumbel, E. J. 1958. Statistics of extremes. 375 p. Columbia University Press, New York.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**:97-109.

IPCC. 2001. Climate change 2001: synthesis report. Cambridge University Press. 408 p, Cambridge.

Javelle, P., J. M. Grésillon, and G. Galéa. 1999. Discharge-duration-Frequency curves modeling for floods and scale invariance. Comptes Rendus de l'Académie des Sciences, Sciences de la terre et des planètes. **329**:39-44.

Javelle, P., T. Ouarda, M. Lang, B. Bobee, G. Galéa, and J. M. Grésillon. 2002. Development of regional flood-duration-frequency curves based on the index-flood method. Journal of Hydrology **258**:249-259.

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

Katz, R. W., M. B. Parlange, and P. Naveau. 2002. Statistics of extremes in hydrology. Advances in Water Resources **25**:1287-1304.

Lang, M. 1999. Theoretical discussion and Monte-Carlo simulations for a Negative Binomial process paradox. Stochastic Environmental Research and Risk Assessment **13**:183-200.

Lang, M., T. B. M. J. Ouarda, and B. Bobée. 1999. Towards operational guidelines for over-threshold modeling. Journal of Hydrology **225**:103-117.

Madsen, H., P. S. Mikkelsen, D. Rosbjerg, and H. P. 2002. Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series statistics. Water Resources Research **38**:1239.

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

Metropolis, N., and S. Ulam. 1949. The Monte Carlo method. Journal of the American Statistical Association **44**:335-341.

Parent, E., and J. Bernier. 2003. Encoding prior experts judgments to improve risk analysis of extreme hydrological events via POT modeling. Journal of Hydrology **283**:1-18.

Perreault, L. 2000. Analyse bayésienne rétrospective d'une rupture dans les séquences de variables aléatoires hydrologiques. ENGREF, INRS-Eau. 200 p.

Perreault, L., J. Bernier, B. Bobee, and E. Parent. 2000a. Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. Journal of Hydrology **235**:221-241.

Perreault, L., J. Bernier, B. Bobee, and E. Parent. 2000b. Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. Journal of Hydrology **235**:242-263.

Perreault, L., and V. Fortin. 2003. Mixture and Hidden Markov models for peak flow analysis. *in*. Seizièmes entretiens du centre Jacques Cartier, Lyon, France.

Perreault, L., E. Parent, J. Bernier, B. Bobee, and M. Slivitzky. 2000c. Retrospective multivariate Bayesian change-point analysis: A simultaneous single change in the mean of several hydrological sequences. Stochastic Environmental Research and Risk Assessment **14**:243-261.

Pickands, J. 1975. Statistical inference using extreme order statistics. Annals of Statistics **3**:119-131.

Prudhomme, C. 1995. Modèles synthétiques des connaissances en hydrologie. Université Montpellier II, CEMAGREF Lyon. 400 p, Montpellier.

Ray, B. K., and R. S. Tsay. 2002. Bayesian methods for change-point detection in long-range dependent processes. Journal of Time Series Analysis **23**:687-705.

Reed, D. W. 1999. Flood Estimation Handbook. Volume 1: Overview. I. o. Hydrology. 108 p. Wallingford.

Reis, D. S., and J. R. Stedinger. 2005. Bayesian MCMC flood frequency analysis with historical information. Journal of Hydrology **313**:97-116.

Ritter, C., and M. A. Tanner. 1992. Facilitating the Gibbs Sampler - the Gibbs Stopper and the Griddy-Gibbs Sampler. Journal of the American Statistical Association **87**:861-868.

Robert, C. P., T. Ryden, and D. M. Titterington. 2000. Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. Journal of the Royal Statistical Society Series B-Statistical Methodology **62**:57-75.

Rosbjerg, D., and H. Madsen. 2004. Advanced approaches in PDS/POT modelling of extreme hydrological events. *in*. Hydrology : Science & Practice for the 21st Century, London. B. H. Society. 217-221.

Strupczewski, W. G., and Z. Kaczmarek. 2001. Non-stationary approach to at-site flood frequency modelling II. Weighted least squares estimation. Journal of Hydrology **248**:143-151.

Strupczewski, W. G., V. P. Singh, and W. Feluch. 2001. Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation. Journal of Hydrology **248**:123-142.

Tanner, M. A. 1996. Tools for Statistical Inference. Springer-Verlag. 208 p. New York.

Annex 1. Griddy Gibbs sampling

The aim of this method is to simulate a sample from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{X})$. The Gibbs sampling algorithm can be written as follow:

- Choose a starting value $\boldsymbol{\theta}^{(0)} = \left(\theta_1^{(0)},...,\theta_k^{(0)}\right)$, and set $j = 0$.
- Repeat N times :
  - $j = j+1$;
  - Sample $\theta_1^{(j)} \sim p\left(\theta_1 \mid \theta_2^{(j-1)},...,\theta_k^{(j-1)}, \boldsymbol{X}\right)$
  - Sample $\theta_2^{(j)} \sim p\left(\theta_2 \mid \theta_1^{(j)},\theta_3^{(j-1)},...,\theta_k^{(j-1)}, \boldsymbol{X}\right)$
  - …
  - Sample $\theta_q^{(j)} \sim p\left(\theta_q \mid \theta_1^{(j)}, ...,\theta_{q-1}^{(j)},\theta_{q+1}^{(j-1)},...,\theta_k^{(j-1)}, \boldsymbol{X}\right)$
  - ……
  - Sample $\theta_k^{(j)} \sim p\left(\theta_k \mid \theta_1^{(j)},...,\theta_{k-1}^{(j)}, \boldsymbol{X}\right)$
- end

The vectors series ($\boldsymbol{\theta}^{(j)}$) converges to the target posterior distribution as j tends to infinity. In order to decrease the influence of the starting point, the first m iterations are usually deleted, and inference is made with the last $N\text{-}m$ iterations. Sensitivity analysis may be necessary to determine acceptable values for m and N (usually, at least a few thousands of iterations are used for m and N). More generally, this method presents the same drawbacks as other iterative simulation techniques. Convergence thus has to be monitored, by choosing several starting points or computing convergence indices. Gelman et al. (1995) or Tanner (1996) provide some guidelines for improving the numerical simulations efficiency.

Unfortunately, the Gibbs sampling algorithm can usually not be used in this raw version, because it involves being able to sample from the full conditional densities $p\left(\theta_q \mid \theta_1^{(j)}, ...,\theta_{q-1}^{(j)},\theta_{q+1}^{(j-1)},...,\theta_k^{(j-1)}, \boldsymbol{X}\right)$. Ritter and Tanner (1992) proposed the use of a discrete approximation of the cumulative density functions (cdf) of these distributions:

- Choose a grid of points $y_1,\dots, y_p$
- Evaluate $p(\theta_i \mid \theta_{1,\dots}, \theta_{i-1}, \theta_{i+1,\dots}, \theta_q, \boldsymbol{X})$ on this grid, to obtain $w_1,\dots, w_p$
- Compute the cumulative sums of $w_1,\dots, w_p$ to obtain an approximation of the CDF.
- Sample u from a uniform distribution on [0, 1].
- Transform u by the inverse of the approximate CDF.

This algorithm must be added at each step of the Gibbs iteration. If the full conditional densities are known only up to proportionality, $p\left(\theta_q \mid \theta_1^{(j)}, ...,\theta_{q-1}^{(j)},\theta_{q+1}^{(j-1)},...,\theta_k^{(j-1)}, \boldsymbol{X}\right)$ can be replaced by $f\left(\theta_1^{(j)}, ...,\theta_{q-1}^{(j)},y,\theta_{q+1}^{(j-1)},...,\theta_k^{(j-1)}\right)$, the product of prior and likelihood, and the cumulative sums of $w_1,\dots, w_p$ must then be divided by the total sum to provide the CDF approximation. Inversion of the CDF can be made with a linear interpolation between two grid points. The grid choice is the most important issue of this technique: it has to be broad enough to cover the range of the distributions, and fine enough to ensure a sufficient accuracy, keeping in mind that this supplementary step is computing-time expensive. Some improvements of the method are described in Ritter and Tanner (1992).

Annex 2. Chib method

The aim of this method is to compute the marginal distribution of the observations, which is the normalizing constant of the Bayes theorem:

$$p(X \mid M) = \frac{\pi(\boldsymbol{\theta}) p(X \mid \boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid X)} = \frac{f(\boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid X)} \tag{A1}$$

This relationship being true for any vector $\boldsymbol{\theta}$, let us consider a particular $\boldsymbol{\theta}^* = (\theta_1^*, ..., \theta_k^*)$. $f(\boldsymbol{\theta}^*)$ is directly computable, which is not the case of the denominator. Consider the following relationship:

$$p(\theta_1^*, ..., \theta_k^* \mid X) = p(\theta_1^* \mid X) p(\theta_2^* \mid \theta_1^*, X) ... p(\theta_q^* \mid \theta_1^*, \theta_2^*, ..., \theta_{q-1}^*, X)$$
$$... \times p(\theta_k^* \mid \theta_1^*, ..., \theta_{k-1}^*, X) \tag{A2}$$

The first term can be evaluated thanks to the sample of the first marginal distribution, by using a Gaussian kernel as an example. The last term can be computed by 1-D numerical integration:

$$p(\theta_k^* \mid \theta_1^*, ..., \theta_{k-1}^*, X) = \frac{f(\boldsymbol{\theta}^*)}{\int f(\theta_1^*, ..., \theta_{k-1}^*, \theta_k) d\theta_k} \tag{A3}$$

The griddy Gibbs sampling can be used to compute intermediary terms. $p(\theta_q^* \mid \theta_1^*, \theta_2^*, ..., \theta_{q-1}^*, X)$ is indeed the first marginal of the distribution $p(\theta_q, \theta_{q+1}, ..., \theta_k \mid \theta_1^*, \theta_2^*, ..., \theta_{q-1}^*, X)$, evaluated at θ_q^*. The griddy Gibbs algorithm can thus be applied to the non-normalized posterior density with the first q-1 components being fixed, that is $f(\theta_1^*, \theta_2^*, ..., \theta_{q-1}^*, \theta_q, \theta_{q+1}, ..., \theta_k)$.

Although this approach is theoretically valid for any value $\boldsymbol{\theta}^*$ with non-zero posterior probability, Chib recommends the use of a high-density point to increase the method accuracy.

List of tables

List of figures

	POT data	Block Maximum data
2-parameters distributions	**Exponential** $$p(x\mid\boldsymbol{\theta}) = \frac{1}{\lambda}\exp(-\frac{x-\sigma}{\lambda})$$ $$\boldsymbol{\theta} = (\sigma,\lambda)$$ σ known (threshold value) $$\lambda > 0$$	**Gumbel** $$p(x\mid\boldsymbol{\theta}) = \frac{1}{\lambda}\exp\left(-\frac{x-\mu}{\lambda} - \exp\left(-\frac{x-\mu}{\lambda}\right)\right)$$ $$\boldsymbol{\theta} = (\mu,\lambda)$$ $$\mu > 0$$ $$\lambda > 0$$
3-parameters distributions	**Generalized Pareto** $$p(x\mid\boldsymbol{\theta}) = \frac{1}{\lambda}\left(1-\frac{\xi(x-\sigma)}{\lambda}\right)^{\frac{1}{\xi}-1}$$ $$\boldsymbol{\theta} = (\sigma,\lambda,\xi)$$ σ known (threshold value) $$\lambda > 0$$ $$\xi \neq 0$$ $$1 - \frac{\xi(x-\sigma)}{\lambda} > 0$$	**Generalized Extreme Value (GEV)** $$p(x\mid\boldsymbol{\theta}) = \frac{1}{\lambda}\left(1-\frac{\xi(x-\mu)}{\lambda}\right)^{\frac{1}{\xi}-1}\exp\left(-\left(1-\frac{\xi(x-\mu)}{\lambda}\right)^{\frac{1}{\xi}}\right)$$ $$\boldsymbol{\theta} = (\mu,\lambda,\xi)$$ $$\mu > 0$$ $$\lambda > 0$$ $$\xi \neq 0$$ $$1 - \frac{\xi(x-\mu)}{\lambda} > 0$$

Table 1

Sampling method	Distribution	Model		
		stationary	step-change	linear trend
POT	Exponential	M_0^{exp} $\theta(t)=\theta$	M_1^{exp} $\lambda(t) = \begin{cases} \lambda_1 & \text{if } t \le \tau \\ \lambda_2 & \text{if } t > \tau \end{cases}$	M_2^{exp} $\lambda(t) = \lambda_0 (1 + \lambda_1 t)$
	Generalized Pareto	M_0^{gp} $\theta(t)=\theta$	M_1^{gp} $(\lambda(t), \xi(t)) = \begin{cases} (\lambda_1, \xi) & \text{if } t \le \tau \\ (\lambda_2, \xi) & \text{if } t > \tau \end{cases}$	M_2^{gp} $\lambda(t) = \lambda_0 (1 + \lambda_1 t)$ $\xi(t) = \xi$
Block Maximum	Gumbel	M_0^{gu} $\theta(t)=\theta$	M_1^{gu} $(\mu(t), \lambda(t)) = \begin{cases} (\mu_1, \lambda_1) & \text{if } t \le \tau \\ (\mu_2, \lambda_2) & \text{if } t > \tau \end{cases}$	M_2^{gu} $\mu(t) = \mu_0 (1 + \mu_1 t)$ $\lambda(t) = \lambda_0 (1 + \lambda_1 t)$
	GEV	M_0^{gev} $\theta(t)=\theta$	M_1^{gev} $(\mu(t), \lambda(t), \xi(t)) = \begin{cases} (\mu_1, \lambda_1, \xi) & \text{if } t \le \\ (\mu_2, \lambda_2, \xi) & \text{if } t > \end{cases}$	M_2^{gev} $\mu(t) = \mu_0 (1 + \mu_1 t)$ $\lambda(t) = \lambda_0 (1 + \lambda_1 t)$ $\xi(t) = \xi$

Table 2

	stationary	step-change	linear trend
		Model	

(a)

M_0^{exp}

$$q_p(t) = \sigma - \lambda \log(1-p)$$

Specification on $\tilde{q}_1$

$$J = -\log(1-p_1)$$

M_1^{exp}

$$q_p^{(i)}(t) = \sigma - \lambda_i \, \log(1-p)$$

Specification on $(\tilde{q}_1^{(1)}, \tilde{q}_1^{(2)})$

$$\boldsymbol{J} = \begin{pmatrix} -\log(1-p_1) & 0 \\ 0 & -\log(1-p_1) \end{pmatrix}$$

M_2^{exp}

$$q_p(t) = \sigma - \lambda_0(1+\lambda_1 t)\log(1-p)$$

Specification on $(\tilde{q}_1(t_1), \tilde{q}_1(t_2))$

$$\boldsymbol{J} = \begin{pmatrix} -(1+\lambda_1 t_1)\log(1-p_1) & -\lambda_0 t_1 \, \log(1-p_1) \\ -(1+\lambda_1 t_2)\log(1-p_1) & -\lambda_0 t_2 \, \log(1-p_1) \end{pmatrix}$$

(b)

M_0^{gp}

$$q_p(t) = \sigma + \frac{\lambda}{\xi}(1-(1-p)^\xi)$$

$$y_i = (1-p_i)^\xi$$

Specification on $(\tilde{q}_1, \tilde{q}_2)$

$$\boldsymbol{J} = \begin{pmatrix} \frac{1-y_1}{\xi} & \frac{\lambda}{\xi^2}\big(y_1(1-\log(y_1))-1\big) \\ \frac{y_1-y_2}{\xi} & \frac{\lambda}{\xi^2}\big(y_1(\log(y_1)-1)-y_2(\log(y_2)-1)\big) \end{pmatrix}$$

M_1^{gp}

$$q_p^{(i)}(t) = \sigma + \frac{\lambda_i}{\xi}(1-(1-p)^\xi)$$

Specification on $(\tilde{q}_1^{(1)}, \tilde{q}_2^{(1)}, \tilde{q}_1^{(2)})$

$$\boldsymbol{J} = \begin{pmatrix} \frac{1-y_1}{\xi} & 0 & \frac{\lambda_1}{\xi^2}\big(y_1(1-\log(y_1))-1\big) \\ \frac{y_1-y_2}{\xi} & 0 & \frac{\lambda_1}{\xi^2}\big(y_1(\log(y_1)-1)-y_2(\log(y_2)-1)\big) \\ 0 & \frac{1-y_1}{\xi} & \frac{\lambda_2}{\xi^2}\big(y_1(1-\log(y_1))-1\big) \end{pmatrix}$$

M_2^{gp}

$$q_p(t) = \sigma + \frac{\lambda_0(1+\lambda_1 t)}{\xi}(1-(1-p)^\xi)$$

Specification on $(\tilde{q}_1(t_1), \tilde{q}_2(t_1), \tilde{q}_1(t_2))$

$$\boldsymbol{J} = \begin{pmatrix} \frac{(1+\lambda_1 t_1)(1-y_1)}{\xi} & \frac{\lambda_0 t_1(1-y_1)}{\xi} & \frac{\lambda_0(1+\lambda_1 t_1)}{\xi^2}\big(y_1(1-\log(y_1))-1\big) \\ \frac{(1+\lambda_1 t_1)(y_1-y_2)}{\xi} & \frac{\lambda_0 t_1(y_1-y_2)}{\xi} & \frac{\lambda_0(1+\lambda_1 t_1)}{\xi^2}\big(y_1(\log(y_1)-1)-y_2(\log(y_2)-1)\big) \\ \frac{(1+\lambda_1 t_2)(1-y_1)}{\xi} & \frac{\lambda_0 t_2(1-y_1)}{\xi} & \frac{\lambda_0(1+\lambda_1 t_2)}{\xi^2}\big(y_1(1-\log(y_1))-1\big) \end{pmatrix}$$

(c)

M_0^{gu}

$$q_p(t) = \mu - \lambda \log(-\log(p))$$

Specification on $(\tilde{q}_1, \tilde{q}_2)$

$$\boldsymbol{J} = \begin{pmatrix} 1 & -\log(-\log(p_1)) \\ 0 & \log\left(\frac{\log(p_1)}{\log(p_2)}\right) \end{pmatrix}$$

M_1^{gu}

$$q_p^{(i)}(t) = \mu_i - \lambda_i \, \log(-\log(p))$$

Specification on $(\tilde{q}_1^{(1)}, \tilde{q}_2^{(1)}, \tilde{q}_1^{(2)}, \tilde{q}_2^{(2)})$

$$\boldsymbol{J} = \begin{pmatrix} 1 & -\log(-\log(p_1)) & 0 & 0 \\ 0 & \log\left(\frac{\log(p_1)}{\log(p_2)}\right) & 0 & 0 \\ 0 & 0 & 1 & -\log(-\log(p_1)) \\ 0 & 0 & 0 & \log\left(\frac{\log(p_1)}{\log(p_2)}\right) \end{pmatrix}$$

M_2^{gu}

$$q_p(t) = \mu_0(1+\mu_1 t) - \lambda_0(1+\lambda_1 t)\log(-\log(p))$$

Specification on $(\tilde{q}_1(t_1), \tilde{q}_2(t_1), \tilde{q}_1(t_2), \tilde{q}_2(t_2))$

$$\boldsymbol{J} = \begin{pmatrix} 1+\mu_1 t_1 & \mu_0 t_1 & -(1+\lambda_1 t_1)\log(-\log(p_1)) & -\lambda_0 t_1 \, \log(-\log(p_1)) \\ 0 & 0 & (1+\lambda_1 t_1)\log\left(\frac{\log(p_1)}{\log(p_2)}\right) & \lambda_0 t_1 \, \log\left(\frac{\log(p_1)}{\log(p_2)}\right) \\ 1+\mu_1 t_2 & \mu_0 t_2 & -(1+\lambda_1 t_2)\log(-\log(p_1)) & -\lambda_0 t_2 \log(-\log(p_1)) \\ 0 & 0 & (1+\lambda_1 t_2)\log\left(\frac{\log(p_1)}{\log(p_2)}\right) & \lambda_0 t_2 \log\left(\frac{\log(p_1)}{\log(p_2)}\right) \end{pmatrix}$$

Table 3a

$$\mathbf{M_0^{gev}}$$

$$q_p(t) = \mu + \frac{\lambda}{\xi}\left(1 - (-\log(p))^\xi\right)$$

Specification on $(\tilde{q}_1, \tilde{q}_2, \tilde{q}_3)$

$$y_i = -\log(p_i)$$

stationary

$$J = \begin{pmatrix} 1 & \dfrac{1 - y_1^\xi}{\xi} & \dfrac{\lambda}{\xi^2}\left(y_1^\xi(1 - \log(y_1^\xi)) - 1\right) \\[3mm] 0 & \dfrac{y_1^\xi - y_2^\xi}{\xi} & \dfrac{\lambda}{\xi^2}\left(y_1^\xi(\log(y_1^\xi) - 1) - y_2^\xi(\log(y_2^\xi) - 1)\right) \\[3mm] 0 & \dfrac{y_2^\xi - y_3^\xi}{\xi} & \dfrac{\lambda}{\xi^2}\left(y_2^\xi(\log(y_2^\xi) - 1) - y_3^\xi(\log(y_3^\xi) - 1)\right) \end{pmatrix}$$

$$\mathbf{M_1^{gev}}$$

$$q_p^{(i)}(t) = \mu_i + \frac{\lambda_i}{\xi}\left(1 - (-\log(p))^\xi\right)$$

Specification on $(\tilde{q}_1^{(1)}, \tilde{q}_2^{(1)}, \tilde{q}_3^{(1)}, \tilde{q}_1^{(2)}, \tilde{q}_2^{(2)})$

Step change

$$J = \begin{pmatrix} 1 & 0 & \dfrac{1 - y_1^\xi}{\xi} & 0 & \dfrac{\lambda_1}{\xi^2}\left(y_1^\xi(1 - \log(y_1^\xi)) - 1\right) \\[3mm] 0 & 0 & \dfrac{y_1^\xi - y_2^\xi}{\xi} & 0 & \dfrac{\lambda_1}{\xi^2}\left(y_1^\xi(\log(y_1^\xi) - 1) - y_2^\xi(\log(y_2^\xi) - 1)\right) \\[3mm] 0 & 0 & \dfrac{y_2^\xi - y_3^\xi}{\xi} & 0 & \dfrac{\lambda_1}{\xi^2}\left(y_2^\xi(\log(y_2^\xi) - 1) - y_3^\xi(\log(y_3^\xi) - 1)\right) \\[3mm] 0 & 1 & 0 & \dfrac{1 - y_1^\xi}{\xi} & \dfrac{\lambda_2}{\xi^2}\left(y_1^\xi(1 - \log(y_1^\xi)) - 1\right) \\[3mm] 0 & 0 & 0 & \dfrac{y_1^\xi - y_2^\xi}{\xi} & \dfrac{\lambda_2}{\xi^2}\left(y_1^\xi(\log(y_1^\xi) - 1) - y_2^\xi(\log(y_2^\xi) - 1)\right) \end{pmatrix}$$

$$\mathbf{M_2^{gev}}$$

$$q_p(t) = \mu_0(1 + \mu_1 t) + \frac{\lambda_0(1 + \lambda_1 t)}{\xi}\left(1 - (-\log(p))^\xi\right)$$

Specification on $(\tilde{q}_1(t_1), \tilde{q}_2(t_1), \tilde{q}_3(t_1), \tilde{q}_1(t_2), \tilde{q}_2(t_2))$

linear trend

$$J = \begin{pmatrix} 1 + \mu_1 t_1 & \mu_0 t_1 & \dfrac{(1 + \lambda_1 t_1)(1 - y_1^\xi)}{\xi} & \dfrac{\lambda_0 t_1(1 - y_1^\xi)}{\xi} & \dfrac{\lambda_0(1 + \lambda_1 t_1)}{\xi^2}\left(y_1^\xi(1 - \log(y_1^\xi)) - 1\right) \\[3mm] 0 & 0 & \dfrac{(1 + \lambda_1 t_1)(y_1^\xi - y_2^\xi)}{\xi} & \dfrac{\lambda_0 t_1(y_1^\xi - y_2^\xi)}{\xi} & \dfrac{\lambda_0(1 + \lambda_1 t_1)}{\xi^2}\left(y_1^\xi(\log(y_1^\xi) - 1) - y_2^\xi(\log(y_2^\xi) - 1)\right) \\[3mm] 0 & 0 & \dfrac{(1 + \lambda_1 t_1)(y_2^\xi - y_3^\xi)}{\xi} & \dfrac{\lambda_0 t_1(y_2^\xi - y_3^\xi)}{\xi} & \dfrac{\lambda_0(1 + \lambda_1 t_1)}{\xi^2}\left(y_2^\xi(\log(y_2^\xi) - 1) - y_3^\xi(\log(y_3^\xi) - 1)\right) \\[3mm] 1 + \mu_1 t_2 & \mu_0 t_2 & \dfrac{(1 + \lambda_1 t_2)(1 - y_1^\xi)}{\xi} & \dfrac{\lambda_0 t_2(1 - y_1^\xi)}{\xi} & \dfrac{\lambda_0(1 + \lambda_1 t_2)}{\xi^2}\left(y_1^\xi(1 - \log(y_1^\xi)) - 1\right) \\[3mm] 0 & 0 & \dfrac{(1 + \lambda_1 t_2)(y_1^\xi - y_2^\xi)}{\xi} & \dfrac{\lambda_0 t_2(y_1^\xi - y_2^\xi)}{\xi} & \dfrac{\lambda_0(1 + \lambda_1 t_2)}{\xi^2}\left(y_1^\xi(\log(y_1^\xi) - 1) - y_2^\xi(\log(y_2^\xi) - 1)\right) \end{pmatrix}$$

Table 3b

	Peaks				Inter-arrivals		
	Exponential	Generalized Pareto		Change point	Exponential	Change point	Trend coefficient
Parameters	$\hat{\lambda}$	$\hat{\tilde{q}}_1$	$\hat{\tilde{q}}_2$	$\hat{\tau}$	$\hat{\lambda}$	$\hat{\tau}$	$\hat{\lambda}_1$
	16.025	60 m^3.s^{-1}	42 m^3.s^{-1}	no prior information	1	no prior information	no prior information
Hyper-parameters	$IG(\lambda\,|\,a,b)$ a=2.5 b=24	$G(\tilde{q}_1\,|\,a,b)$ a=3.6 b=3.6/60	$G(\tilde{q}_2\,|\,a,b)$ a=3.6 b=3.6/42	uniform on $[1,\dots,n\text{-}1]$	$IG(\lambda\,|\,a,b)$ a=2.5 b=1.5	uniform on $[1,\dots,n\text{-}1]$	$N(\lambda_1\,|\,m,s)$ m=0 s=0.05

Table 4

	$M_0^{\exp}$	$M_1^{\exp}$	$M_2^{\exp}$	M_0^{gp}	M_1^{gp}	M_2^{gp}	non-stationary *vs.* stationary	Generalized Pareto *vs.* exponential
Inter-arrivals	0.16	0.21	0.63	/	/	/	5.4	/
Peaks	0.15	0.04	0.01	0.28	0.5	0.02	1.3	4.1

Table 5
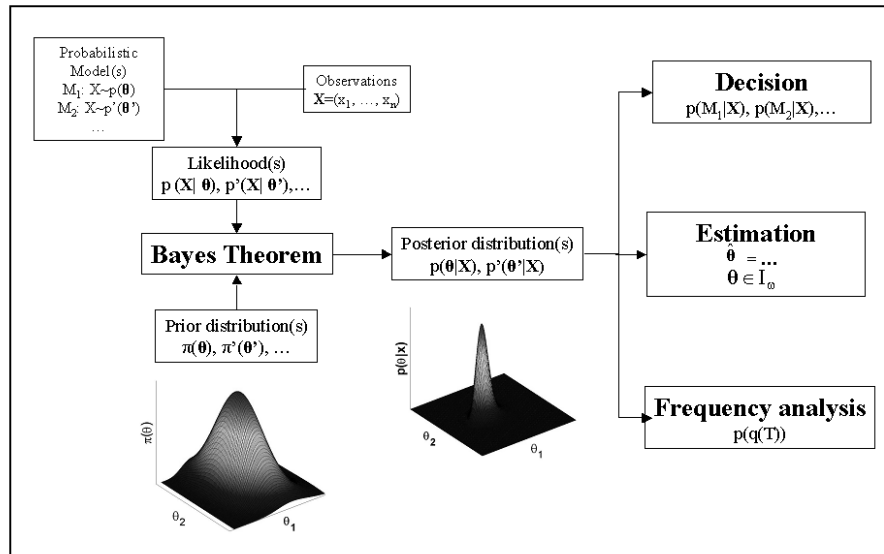
Figure 1
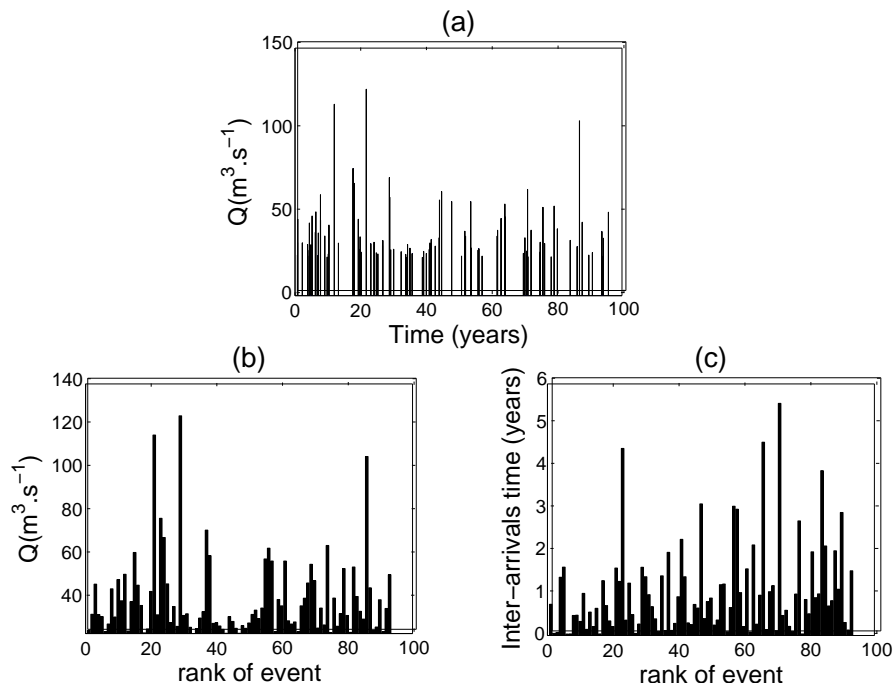
Figure 2
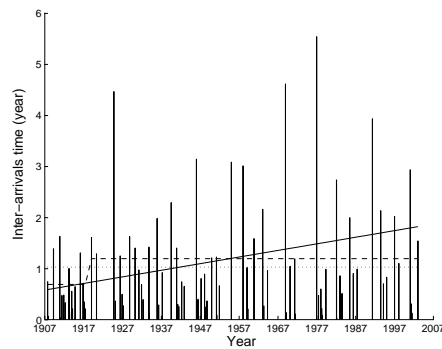
Figure 3
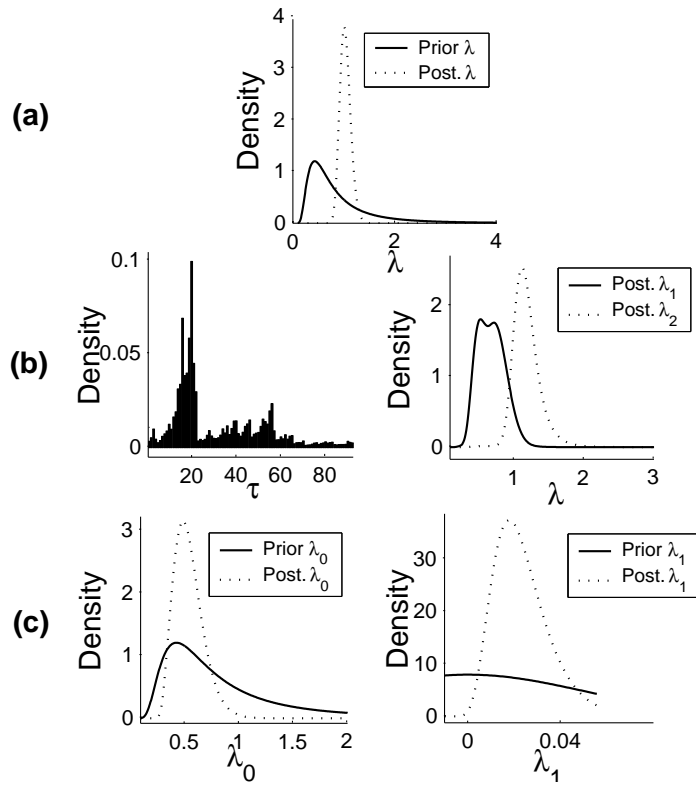
Figure 4
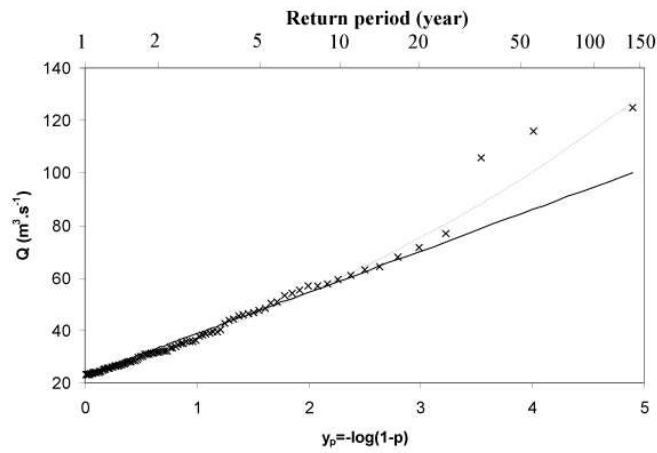
Figure 5
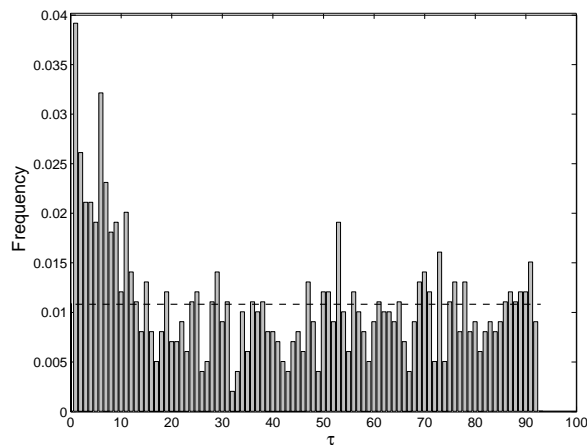
Figure 6
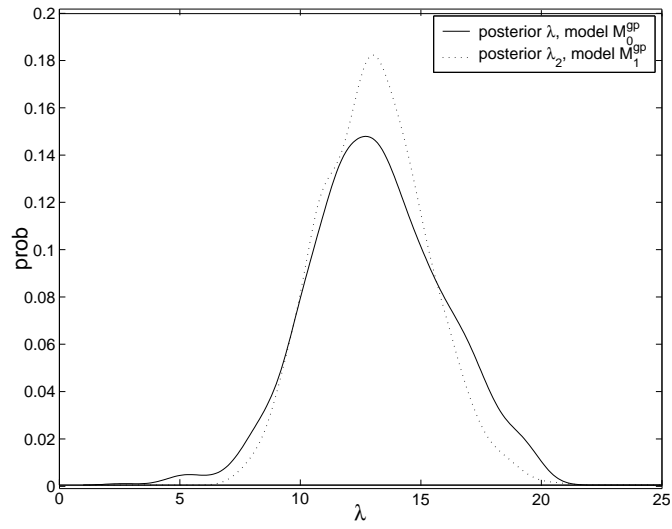
Figure 7

Figure 8

Figure 9