

A regional Bayesian POT model for flood frequency analysis

Mathieu Ribatet · Eric Sauquet · Jean-Michel Grésillon ·
Taha B. M. J. Ouarda

Published online: 9 August 2006
© Springer-Verlag 2006

Abstract Flood frequency analysis is usually based on the fitting of an extreme value distribution to the local streamflow series. However, when the local data series is short, frequency analysis results become unreliable. Regional frequency analysis is a convenient way to reduce the estimation uncertainty. In this work, we propose a regional Bayesian model for short record length sites. This model is less restrictive than the index flood model while preserving the formalism of “homogeneous regions”. The performance of the proposed model is assessed on a set of gauging stations in France. The accuracy of quantile estimates as a function of the degree of homogeneity of the pooling group is also analysed. The results indicate that the regional Bayesian model outperforms the index flood model and local estimators. Furthermore, it seems that working with relatively large and homogeneous regions may lead to more accurate results than working with smaller and highly homogeneous regions.

Keywords Regional frequency analysis · Bayesian inference · Index flood · L -moments · Markov Chain Monte Carlo

1 Introduction

Flood frequency analysis is essential in preliminary studies to define the design flood. Methods for estimating design flow usually consist of fitting one of the distributions given by the extreme value theory to a sample of flood events. If modelling exceedance over a threshold is of interest, a theoretical justification (Fisher and Tippett 1928; Balkema and Haan 1974; Pickands 1975) exists for the use of the Generalized Pareto (GP) distribution.

$$F(x) = 1 - \left[1 + \frac{\zeta(x - \mu)}{\sigma} \right]^{-1/\zeta} \quad (1)$$

where $1 + \zeta(x - \mu)/\sigma > 0, \sigma > 0$. μ, σ and ζ are the location, scale and shape parameters. This distribution is defined for $\zeta \neq 0$, and can be derived by continuity in the case $\zeta = 0$, corresponding to the Exponential case:

$$F(x) = 1 - \exp\left(-\frac{x - \mu}{\sigma}\right) \quad (2)$$

A comprehensive review of the Extreme Value Theory is given by Embrechts et al. (1997) and Coles (2001).

However, frequency analysis can lead to unreliable flood quantiles when little data is available at the site of interest. A convenient way to improve estimates of flood statistics is to incorporate data from other gauged locations in the estimation procedures. This approach is widely applied in hydrology and is known as regional flood frequency analysis (RFFA). One of the most popular and simple approaches favoured by engineers

M. Ribatet (✉) · T. B. M. J. Ouarda
INRS-ETE, University of Québec, 490, de la Couronne
Québec, QC, Quebec, Canada G1K 9A9
e-mail: Mathieu.Ribatet@ete.inrs.ca

E. Sauquet · J.-M. Grésillon · M. Ribatet
Cemagref, 3 bis quai Chauveau CP 220,
69336 Lyon Cedex 09, France

is the index flood method (Dalrymple 1960). The standard procedure allows: (a) the delineation of homogeneous regions, i.e. a set of sites which behave—hydrologically and/or statistically—in the same way; (b) the derivation of a regional flood frequency distribution; (c) and the estimation of the parameters and quantiles at the site of interest.

Regions are collections of gauged basins with similar site characteristics related to the flood magnitude. The pooled stations are not necessarily close to the site of interest. Forming homogeneous regions can be achieved in various ways. Regions were first established geographically. More recent work promoted the use of geographically non-contiguous regions (Burn 1990; GREHYS 1996). Recent research has defined the concept of “region of influence” (Acreman and Wiltshire 1989). Other techniques can be used such as Artificial Neural Networks to identify groups of stations (Hall et al. 2002).

The index flood model assumes that flood distributions at all sites within a region are identical, up to a scale factor. The index flood approach is not exempt from critics as its application requires strong assumptions. One major implicit assumption, noticed by Gupta et al. (1994), is that the coefficient of variation of peak flows has to be constant across the region. This fundamental property seems not to be verified in practice (Robinson and Sivapalan 1997) and not physically justified (Katz et al. 2002).

The assumptions of the index flood model often need to be relaxed to suit the observations. For this purpose, Gabriele and Arnell (1991) proposed a hierarchical approach to RFFA. The skewness is still supposed to be constant over the whole region, but the coefficient of variation and the mean annual flood can vary slightly from one subregion to another. However, the two authors underlined the practical difficulty of delineating these subregions.

In the index flood model, each observation from any site within the region has the same weight. However, it seems not optimal as, obviously, the most valuable information comes from the target site. Indeed target data—even short—are the only ones which are “really” distributed like the target site.

We suggest that it would be better to use a Bayesian approach that encompasses the classical index flood model and uses all the data in a more efficient manner. In summary, the proposed Bayesian approach differs from the index flood model as it: (a) uses the at-site information in a more efficient way since this approach distinguishes the target site data and the regional data and (b) does not impose a purely deterministic relationship between sites within the region.

The main goal of this article is to test the efficiency and robustness of the regional Bayesian model developed when dealing with short record length series. For this purpose, classical frequency analysis i.e. local and traditional RFFA will be compared to the suggested regional Bayesian approach. Section 2 presents a brief summary of the classical index flood model. Relevant theoretical aspects of Bayesian theory are introduced and applied to flood modelling in a RFFA context in Sect. 3. Section 4 describes the data set used to illustrate the method. Section 5 describes the procedure used to elicit the prior distribution. Section 6 outlines the weaknesses and strengths of each approach on a typical homogeneous region. Finally Sect. 7 presents an analysis of the effect of homogeneity level on quantile estimation.

2 The index flood model

The index flood method states that flood frequency distributions within a particular region are supposed to be identical when divided by a scale factor—namely the index flood. Mathematically, this assumption is expressed as:

$$Q^{(S)} = C^{(S)} Q^{(R)} \quad (3)$$

where $Q^{(S)}$ is the quantile function at site S , $C^{(S)}$ the index flood at site S and $Q^{(R)}$ the regional quantile function i.e. the dimensionless quantile function valid across the homogeneous region.

Equation 3 is the core of the model and leads to strong constraints concerning at-site distribution parameters. Consequently, the shape parameter is the same throughout the homogeneous region, whereas the location and scale parameters have simple scaling behaviour—see Appendix.

Equation 3 is supposed to be satisfied if all sites are hydrologically and/or statistically similar. Therefore, one of the main aspects of this approach is to identify a homogeneous region which includes the target site.

Similarity in basin characteristics is necessary but not sufficient to ensure the homogeneity of the region in terms of the flood peaks statistics. Hosking and Wallis (1993, 1997) suggested a heterogeneity measurement H_1 to assess whether a region is “acceptably homogeneous” ($H_1 < 1$), “probably heterogeneous” ($1 \leq H_1 < 2$) or “definitively heterogeneous” ($H_1 \geq 2$). Note that the case $H_1 \leq 0$ seems to detect correlations between sites within the region.

Once the region satisfies the homogeneity test of Hosking and Wallis (1993, 1997), the regional flood

frequency distribution and the related at-site distribution is computed in a classical way. That is, by fitting the regional distribution to the weighted mean of sample L -moments. Details for computing heterogeneity statistics, regional flood frequency and at-site distribution can be found in (Hosking and Wallis 1993, 1997).

By definition of the index flood model, it can be seen that any parameters derived from each sample have the same weight. Giving equal weights to all site observations is debatable since the most relevant information is certainly the target site one. The relevance of the target site information is obvious as this is the only one which is “really” distributed like the target site. Thus, in this approach the available information is not efficiently used.

3 Regional Bayesian model

The Bayesian concepts have already been applied with success to the regional frequency analysis of extreme rainfalls (Coles and Pericchi 2003) and floods (Madsen and Rosbjerg 1997; Northrop 2004). Regional information is not used to build a regional distribution but to specify a kind of “suspicion” about the target site distribution. This is easily achieved in the Bayesian framework through the so-called prior distribution.

The main goal of Bayesian inference is to compute the posterior distribution. The posterior distribution $\pi(\theta | x)$ is given by the Bayes Theorem:

$$\pi(\theta|x) = \frac{\pi(\theta)\pi(\theta; x)}{\int_{\Theta} \pi(\theta)\pi(\theta; x)d\theta} \propto \pi(\theta)\pi(\theta; x) \tag{4}$$

where θ is the vector of parameters of the distribution to be fitted, Θ is the space parameter, $\pi(\theta; x)$ is the likelihood function, x is the vector of observation and $\pi(\theta)$ is the prior distribution.

In theory, the posterior distribution is entirely known but is often insolvable—because of the integral. One of the solutions is to fix a prior model which leads to an analytical—or semi-analytical—posterior distribution and which allows the posterior distribution to be computed more easily (Parent and Bernier 2003). Nevertheless, the most convenient way is to implement Markov Chain Monte Carlo (MCMC) techniques to sample the posterior distribution. This approach avoids using a purely artificial prior model with no theoretical and/or physical justifications.

For our application, the likelihood function corresponds to the GP distribution as peaks over a threshold are of interest. From Eq. 4, if the prior distribution is known, posterior distribution can be computed—up to a constant. The next section describes how to define the prior distribution.

3.1 Prior distribution

The prior model is usually a multivariate distribution which must represent beliefs about the distribution of the parameters i.e. μ , σ and ξ prior to having any information about the data.

As the proposed model is fully parametric, the prior distribution $\pi(\theta)$ is a multivariate distribution entirely defined by its hyper parameters. In our case study, the marginal prior distributions were supposed to be independent lognormal for both location and scale parameters and normal for the shape parameter. Thus,

$$\pi(\theta) \propto J \exp \left[(\theta' - \gamma)^T \Sigma^{-1} (\theta' - \gamma) \right] \tag{5}$$

where γ , Σ are hyper parameters, $\theta' = (\log \mu, \log \sigma, \xi)$ and J is the Jacobian of the transformation from θ' to θ , namely $J = 1/\mu \sigma$. $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ is the mean vector, Σ is the covariance matrix. As marginal priors are supposed to be independent, Σ is a 3–3 diagonal matrix with diagonal elements d_1, d_2, d_3 .

3.2 Estimation of the hyper parameters

Hyper parameters are defined through the index flood concept. Consider all sites of a region except the target site—say the j th site. A set of pseudo target site parameters can be computed:

$$\tilde{\mu}^{(i)} = \mu_*^{(i)} C^{(j)} \tag{6}$$

$$\tilde{\sigma}^{(i)} = \sigma_*^{(i)} C^{(j)} \tag{7}$$

$$\tilde{\xi}^{(i)} = \xi_*^{(i)} \tag{8}$$

for all $i \neq j$, where $C^{(j)}$ is the at-site Index Flood and $\mu_*^{(i)}$, $\sigma_*^{(i)}$, $\xi_*^{(i)}$ are respectively the location, scale and shape at-site parameter estimates from the rescaled sample.

Under the hypothesis of the index flood model, pseudo parameters $(\tilde{\mu}^{(i)}, \tilde{\sigma}^{(i)}, \tilde{\xi}^{(i)})$ for $i \neq j$ are expected to be similar to the target site distribution parameters. Note that, information from the target site sample is not used to elicit the prior distribution. Thus, $C^{(j)}$ in Eqs. 6 and 7 must be estimated without use of the j th site sample.

From these pseudo parameters, hyper parameters can be computed:

$$\gamma_1 = \frac{1}{N-1} \sum_{i \neq j} \log \tilde{\mu}^{(i)}, \quad d_1 = \frac{1}{N-1} \sum_{i \neq j} \text{Var} [\log \tilde{\mu}^{(i)}] \quad (9)$$

$$\gamma_2 = \frac{1}{N-1} \sum_{i \neq j} \log \tilde{\sigma}^{(i)}, \quad d_2 = \frac{1}{N-1} \sum_{i \neq j} \text{Var} [\log \tilde{\sigma}^{(i)}] \quad (10)$$

$$\gamma_3 = \frac{1}{N-1} \sum_{i \neq j} \tilde{\xi}^{(i)}, \quad d_3 = \frac{1}{N-2} \sum_{i \neq j} (\tilde{\xi}^{(i)} - \gamma_3)^2 \quad (11)$$

It is important at this step to incorporate the uncertainties on the elicitation of the prior distribution. Indeed, it may avoid problems related to misleading information resulting from a region not so homogeneous and moderating a “suspicion” that may be too true.

For this purpose, two types of uncertainties are taken into account: one from parameter estimation, and the other from the target site index flood estimation. Thus, hyper parameters γ_1 and γ_2 are estimated differently than γ_3 , as pseudo parameters for location and scale parameters depend on the target site index flood. Under the hypothesis of independence between $C^{(i)}$ and $\mu_*^{(i)}, \sigma_*^{(i)}$ the variance terms in Eqs. 9 and 10 are computed according these two types of uncertainties:

$$\text{Var} [\log \tilde{\mu}^{(i)}] = \text{Var} [\log C^{(i)}] + \text{Var} [\log \mu_*^{(i)}] \quad (12)$$

$$\text{Var} [\log \tilde{\sigma}^{(i)}] = \text{Var} [\log C^{(i)}] + \text{Var} [\log \sigma_*^{(i)}] \quad (13)$$

The independence assumption between $C^{(i)}$ and $\mu_*^{(i)}, \sigma_*^{(i)}$ is not too restrictive as the target site index flood $C^{(i)}$ is estimated independently from $\mu_*^{(i)}, \sigma_*^{(i)}$.

Note that $\text{Var} [\log \mu_*^{(i)}]$ are estimated using Fisher Information and the Delta method. Estimation of $\text{Var} [C^{(i)}]$ is a special case and depends on the method for estimating the at-site index flood. Nevertheless, it is always possible to carry out an estimation of this variance, at least through standard errors.

3.3 Specificities of the proposed prior model

The construction of the prior distribution with regional information was already suggested by Northrop (2004). Nevertheless, the location parameter—or equivalently the threshold in the GP case—was assumed to be known. Yet, from a theoretical point of view, the

location parameter cannot be known prior to having any information from the target site sample. Northrop (2004) developed a similar approach based on the index flood but uncertainty associated with the scale factor prediction was not considered. The prior distribution was elicited directly from the distribution of the “pseudo target site” estimates $(\tilde{\mu}^{(i)}, \tilde{\sigma}^{(i)}, \tilde{\xi}^{(i)})$. In this perspective, “pseudo target site” estimates are viewed as constant and not as random variables. When dealing with sites with a long record, uncertainties on parameter distributions are low. On the contrary, this has a much greater impact for the index flood, as uncertainties are much larger since the target site index flood is estimated without use of at-site data, even with long record length sites. Note that if the prior distribution is overly accurate, estimation and credibility intervals are influenced. For these reasons and unlike the approach proposed by Northrop (2004), the target site index flood in the proposed methodology is considered to be a random variable and not a constant.

Thus, our prior distribution is not too falsely “tight fit”. But it reflects “real” beliefs about target site behaviour without any use of target site sample.

Madsen and Rosbjerg (1997) and Fill and Stedinger (1998) both presented a regional empirical Bayesian estimator. Both models used conjugate families for prior distributions. However, even if conjugate families are convenient devices, they should not only be used just because computations are easier. In their approaches, prior distributions are elicited with quantile regression on relevant physiographic characteristics.

Our approach differs from the two previous empirical Bayesian approaches (i.e. the target site sample is not used to elicit priors) and respects in that way absolutely the Bayesian theory. Moreover, conjugate priors are not considered, but priors are suited to the data. For example, the lognormal distribution for both location and shape parameters is justified by a physical and theoretical lower boundary as: (a) discharge data are naturally non-negative; so the location parameter should also be non-negative; (b) the scale parameter is strictly positive by definition of the GP distribution.

This prior model is quite different from that proposed by Coles and Tawn (1996) who introduced a lognormal prior distribution only for the scale parameter. Note that it is possible to work with return levels (Coles and Tawn 1996) or return periods (Crowder 1992) instead of working with distribution parameters. However, regional information is suitable for working directly with distribution parameters. For other studies, such prior models could be of interest if “suspicion” is based on return levels or return periods.

4 Data description

Streamflow data were collected at 48 gauging stations in an area reaching from 45°N to 47°N and from 3°E to 8°E. The selection of the gauging sites was initially based on the 22 regions into which France is divided for the implementation of the Water Framework Directive (Wasson et al. 2004). Seven regions cover the area under study. These regions were manually delineated taking into account the spatial pattern of mean annual rainfall, altitude and underlying geology. All these variables might influence flood generation processes. Therefore this division is considered as a preliminary guide for pooling stations. According to Hosking and Wallis (1997), pre-regions were slightly altered by identifying discordant sites while maximising the number of sites within the region and meeting the heterogeneity test. Finally, a set of 14 stations was selected for this study. The heterogeneity statistic for this group is $H_1 = 0.17 < 1$. Consequently, the region is considered to be “acceptably homogeneous”.

The dataset includes seven tributaries to the Loire River and seven gauging stations located in the French part of the Rhône basin (Fig. 1; Table 1). The record length of the instantaneous discharge time series ranges from a minimum of 22 years to a maximum of 37 years, with a mean value of 32 years. The drainage areas vary from 32 to 792 km². Moreover, most of the gauging stations monitored were on first-order stream catchments i.e. all but two pairs of catchments are un-nested. The large majority of floods in the region occur during autumn and winter and are caused by heavy rainfalls.

Partial duration flood series were extracted from the time series for each station. Figure 2 illustrates time series for stations U4505010, U4635010 and V3015010

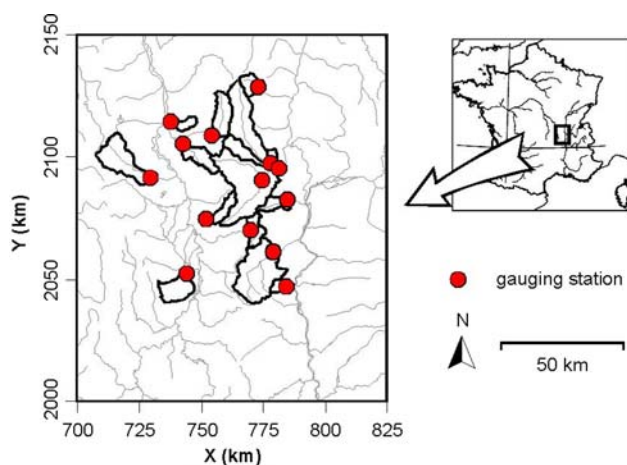


Fig. 1 Location of the gauging stations within the studied area

and their associated thresholds. Threshold levels were selected to extract on average about two events per year while meeting the criteria of independence between floods (Lang et al. 1999).

Three stations U4505010, U4635010 and V3015010 were of particular interest because of their extended record length of 37 years. The time series of these three sites are displayed in Fig. 2.

In this case study, the scale factor was set to correspond to the 1-year return flood quantile—or equivalently the quantile associated with probability of non-exceedance of 0.5. Thus, our choice for the index flood is close to the sample median which was the reference in Robson and Reed (1999) but differs from Hosking and Wallis (1997) where the sample mean was used. This particular choice for the index flood is not unintentional as estimating the quantile with a probability of non-exceedance of 0.5 is more robust than estimating the sample mean. Analysing the influence of index flood selection is beyond the scope of this work. The main point is to keep the same index flood throughout the case study to compare approaches on the same basis.

5 Elicitation of the prior distribution

To estimate the target site index flood, the most popular way is to develop an empirical formula that relates the flow statistic to geomorphological, land-use and climatic descriptors. This relationship is usually established by multivariate regression procedures. In our case study, we consider a simple model for which only one explanatory variable is introduced in the regression analysis: the drainage area. The power form model is adopted:

$$C = aA^b \quad (14)$$

where A is the catchment area. Parameters a and b are computed through ordinary least square procedures on logarithmically transformed data.

However, more sophisticated models could be carried out. Nevertheless, for our case study, observations demonstrate that Eq. 14 is a good parameterisation for estimating the index flood. Figures 3 and 4 illustrate the efficiency of the regressive and prior model for site U4505010 for which:

$$\hat{C} = 0.12A^{1.01} \quad (R^2 = 0.94) \quad (15)$$

Regional information was incorporated into the prior distribution through the index flood model. Moreover, uncertainties in the prior distribution were

Table 1 Characteristics of the stations in the homogeneous region

| Code | Station | Area (km ²) | X (km) | Y (km) | Record |
|----------|------------------------------------|-------------------------|--------|----------|-----------|
| K0624510 | The Bonson river at St. Marcellin | 104 | 744.72 | 2,053.90 | 1971–2003 |
| K0663310 | The Coise river at Larajasse | 61 | 770.67 | 2,072.11 | 1971–2003 |
| K0704510 | The Toranche river at St. Cyr | 62.3 | 752.63 | 2,076.68 | 1977–2003 |
| K0813020 | The Aix river at St. Germain Laval | 193 | 729.48 | 2,093.71 | 1973–2002 |
| K0943010 | The Rhins river at Amplepuis | 114 | 754.52 | 2,111.10 | 1973–2003 |
| K0974010 | The Gand river at Neaux | 85 | 743.45 | 2,107.75 | 1972–2003 |
| K1004510 | The Rhodon river at Perreux | 32 | 738.40 | 2,116.64 | 1973–2003 |
| U4505010 | The Ardieres river at Beaujeu | 54.5 | 773.67 | 2,130.75 | 1969–2003 |
| U4624010 | The Azergues river at Chatillon | 336 | 779.07 | 2,099.72 | 1970–2003 |
| U4635010 | The Brevenne river at Sain Bel | 219 | 775.90 | 2,092.57 | 1969–2003 |
| U4644010 | The Azergues river at Lozanne | 792 | 782.56 | 2,098.09 | 1981–2003 |
| V3015010 | The Yzeron river at Craponne | 48 | 785.47 | 2,084.50 | 1969–2003 |
| V3114010 | The Gier river at Rive de Gier | 319 | 780.54 | 2,062.67 | 1981–2003 |
| V3315010 | The Valencize river at Chavanay | 36 | 786.54 | 2,048.60 | 1978–2003 |

incorporated. Thus, the prior information is, on one hand, not too falsely accurate and on the other hand, informative enough because of the supposed homogeneity of stations.

6 Performance of the Bayesian model on a homogeneous region

When making classical inference on small samples, the uncertainties may be too large. If an extremal event or too many “regular events” in this short record period are present, the estimation can be affected. It could lead to a dramatic overestimation or underestimation of quantiles corresponding to different return levels. A perfect model is expected: (a) to perform well enough even with small samples; (b) to be robust enough when an extreme event occurs in the sample; (c) to be robust enough when too many “regular” events occur in the sample.

In this section, three different models will be applied. For this purpose, the three stations U4505010, U4635010 and V3015010 were selected to assess the robustness and efficiency of the local, regional and Bayesian regional models. These three different approaches correspond to: (a) local: fit the GP distribution to the peaks over threshold data with the maximum likelihood estimator (MLE), unbiased probability weighted moments (PWU) and the biased probability weighted moments (PWB); (b) regional (REG): fit a regional GP distribution as described in section 2 and obtain the target site distribution; (c) regional Bayesian (BAY): elicit the prior density from regional information, then compute the posterior density through MCMC techniques. As an illustration of MCMC output, Fig. 5 displays the prior and posterior marginal densities for the GP parameters of the proposed model. Marginal posterior distribution obtained from an uninformative prior model are also displayed. That is with the same prior model but with a large variance i.e. $d_i = 1,000$, $i = 1, \dots, 3$.

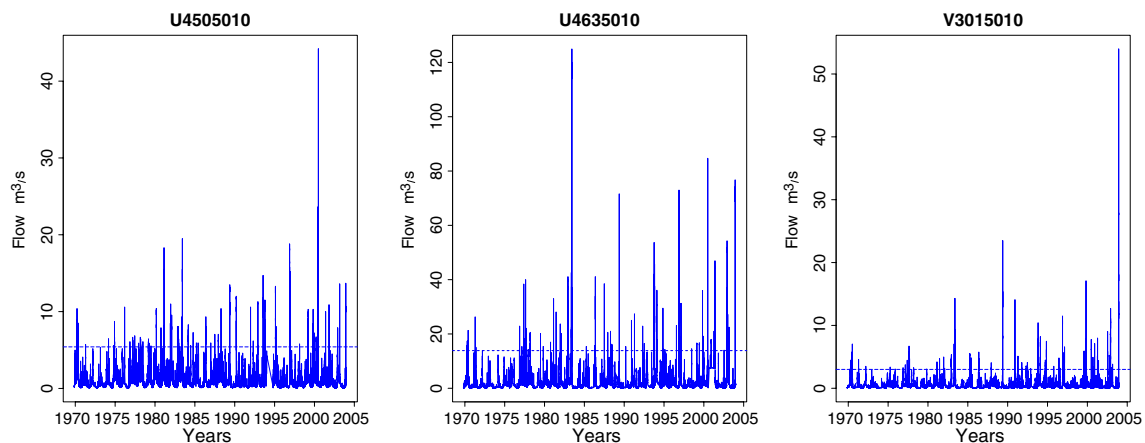
**Fig. 2** Times series for sites U4505010, U4635010 and V3015010 and thresholds associated

Fig. 3 Histograms of pseudo target site estimates of location, scale and shape parameters for site U4505010

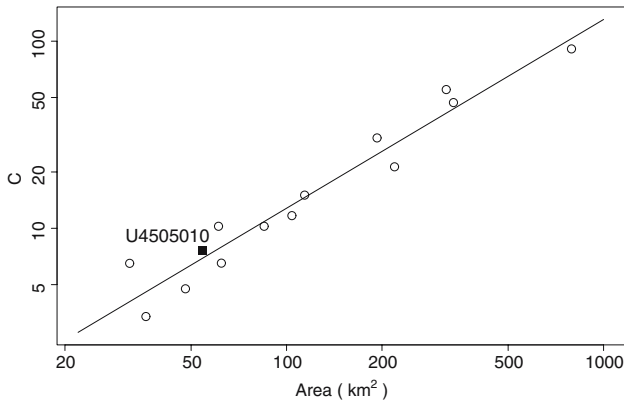
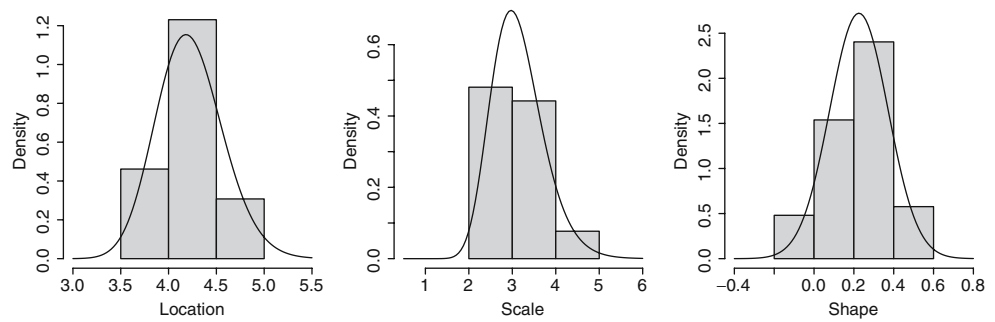


Fig. 4 Regression on the basin area for estimating the at-site index flow for station U4505010

Figure 5 shows the relevance of regional information as the proposed prior model is clearly more accurate than an analysis directly from data. Moreover, for the proposed model and even with only 5 years record length, marginal posterior densities are more accurate than marginal prior densities—except for the shape parameter. Thus, combination of regional and target site information at two different stages is worthwhile, even when only few data are available. The location parameter is a special case as the modes of both the marginal prior and posterior densities seem to be significantly dissimilar.

As the main goal of this work is to compare models on small samples, efficiency will be evaluated on sub-

samples from the original data. Local maximum likelihood estimation on the whole sample will be used as a benchmark to assess the performance of each model. This particular case will be denoted THEO in the following sections. The choice of MLE estimate as a benchmark value is reasonable because of its theoretical motivation and asymptotic efficiency. Moreover, the MLE approach allows the profile confidence intervals to be calculated. This is a key point as these profile confidence intervals are often more accurate than those based on the Delta Method and Fisher Information (Coles 2001).

Furthermore, as interpretation on quantile estimates is more natural than for distribution parameter estimates, the analysis will focus on quantiles corresponding to return period 2, 5, 10 and 20 years. Benchmark values for these quantiles—and their associated 90% profile likelihood confidence intervals are given in Table 2. Benchmark values with return periods greater than 20 years will be considered unreliable—as uncertainties on these quantiles are too large with only 37 years of record.

Moreover, for such return periods, benchmark values are quite similar to those obtained with PWM estimates—with a mean bias of 0.89%. So, the performance of each model is not too much affected by the choice of the MLE estimator for benchmark values.

Different frequency curves for site V3015010 with only the last 15 years of records are displayed in Fig. 6. Let us focus on the highest discharge observation. The

Fig. 5 Proposed prior (dashed line), proposed posterior (solid line) and posterior from an uninformative prior (dotted line) marginal densities for GP parameters. Site U4505010 with 5 years record length. Vertical lines denotes benchmark values

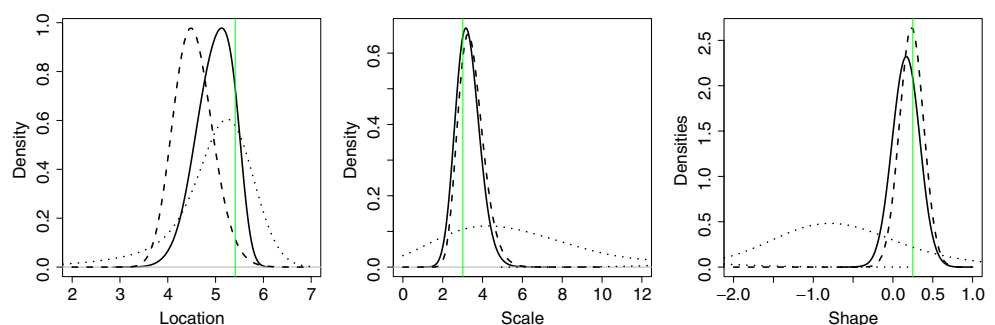


Table 2 Benchmark values for 2, 5, 10 and 20 years quantiles and the associated 90% profile likelihood confidence intervals in brackets

| Station | Q_2 | Q_5 | Q_{10} | Q_{20} |
|----------|-------------------|-------------------|-------------------|--------------------|
| U4505010 | 10.8 (10.1, 11.7) | 15.3 (13.9, 17.4) | 19.5 (17.2, 23.4) | 24.4 (20.6, 31.5) |
| U4635010 | 33.0 (30.0, 36.5) | 52.2 (45.5, 62.5) | 72.2 (60.2, 95.4) | 98.9 (69.2, 200.5) |
| V3015010 | 7.5 (6.9, 8.3) | 11.7 (10.4, 13.7) | 15.9 (13.6, 19.9) | 21.3 (17.3, 28.8) |

return period related to this event is very high for the REG approach. All the other models lead to significantly lower return periods. This flood event is extreme at a regional scale but not in a local context. This underestimation is due to the misuse of the target site sample to establish the regional distribution. On the other hand, the regional Bayesian model performs well for all return periods. Indeed, Fig. 6 indicates that the return level curve is very similar to that of the benchmark. This is quite logical as it adds up the advantage of efficiently using the target site sample and a good “suspicion” on the overall behaviour of the flood peak distribution thanks to the so-called prior distribution. Local approaches suggest a very heavy tail as the extremal event that occurred in 2004 (see Fig. 2, right panel) was in the last sequence of 15 years of records.

As one of the main goals of a RFFA procedure is to deal with small samples, the target site sample was truncated to obtain shorter periods of records of m years, $m \in \{5, 10, 15, 20, 25, 30, 37\}$. The robustness and efficiency of the methods to converge on the parameters of the target site distribution are measured. For this purpose, quantile estimates corresponding to return periods of 2, 5, 10 and 20 years—corresponding to non-exceedance probabilities of 0.75, 0.9, 0.95 and 0.975, respectively—are used. The evolution of quantile estimates as a function of the record length period is presented in Fig. 7. The figure only takes into

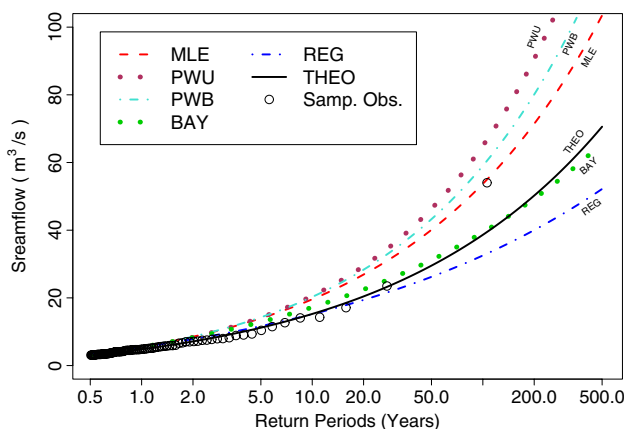
account the first m years; that is, for example, estimates related to the 5 year record length correspond to the period 1969–1973.

Because of the extreme event observed in 1983 (see Fig. 2, middle panel), benchmark values for local and REG approaches were systematically underestimated. This result shows that: on one hand, for small samples classical inference like MLE, PWB and PWU are too responsive if too many “regular” events occurred. On the other hand, for the index flood model, underestimation of quantiles is related to the underestimation of the scale factor $C^{(S)}$ in Eq. 3 because of these “regular” events. Only the Bayesian model performs well enough even with record lengths lower than 15 years.

A monotonic increase of the design flood estimates with the sample length can be noticed in Fig. 7. This behaviour is easily explained by Fig. 2, middle panel. Indeed, only the last part of the time series shows really extreme events. As the record length increases, many more extreme events occur leading to higher estimates. The Bayesian approach is the only one which does not really present this monotonic behaviour.

Moreover, the Bayesian approach is by far the most robust and accurate model as, on the whole range of record length, and for all benchmark values, estimation lies in the 90% profile likelihood confidence interval. This is not true with any other model. The advantage of incorporating regional information within a Bayesian framework is certainly to define a “restricted space” to which the distribution parameters belong. Thus, the impact of a very extremal event—or conversely too many low-level events—should be regarded as an extreme event related to this “restricted space”.

The gain of accuracy in the target site from using regional information is clearly established in Sect. 6 (Figs. 6 and 7). The Bayesian approach seems to be robust even with small samples while being accurate with a larger sample. The poor performance of the REG model is related to a bad selection of sites within the “homogeneous” region being considered and estimates may have been more accurate if “better” regions were considered. Unfortunately, building up such a region is difficult because of the purely deterministic

**Fig. 6** Comparison of frequency curves for site V3015010 with 15 last years recording

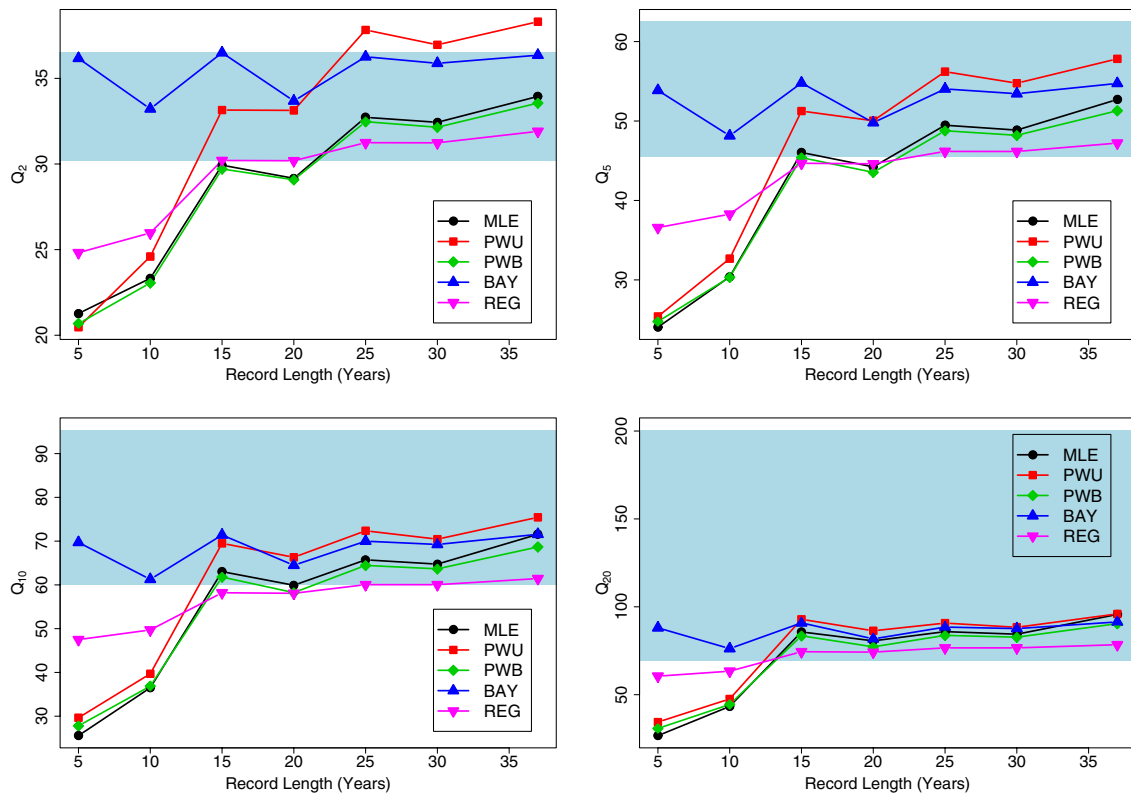


Fig. 7 Evolution of Q_2, Q_5, Q_{10}, Q_{20} estimates as the size increases for the site U4635010 and 90% profile likelihood confidence interval for the benchmark values—light blue area

relation (3). As the Bayesian approach relaxes the REG model, the search for more homogeneous regions could be ineffective. The goal of the next section is to measure the potential gain, for the Bayesian model, against homogeneity property.

7 Effect of degree of heterogeneity on quantile estimation

As indicated in the previous section, we focus now on the impact of the degree of homogeneity of the region. For this purpose, we consider four different regions—denoted He^+, He, Ho and Ho^+ , which correspond to increasingly homogeneous regions according to the test of Hosking and Wallis (1997). The Ho region corresponds to the region analysed in the previous section and described in Table 1. All regions have 14 sites except for the most homogeneous one Ho^+ which contains only 8 stations. He and He^+ regions are derived from Ho . One to five sites are withdrawn and replaced by other stations to obtain a higher heterogeneity value. The Ho^+ region is a sub-region of Ho . Heterogeneity statistics for these regions are summarised in Table 3.

To evaluate the influence of degree of homogeneity of a region on quantile estimation, models are assessed using two performance criteria: the Normalised Bias (NBIAS) and the Normalised Root Mean Square Error (NRMSE). These indices are defined as follows:

$$NBIAS = \frac{1}{k} \sum_{i=1}^k \frac{\hat{Q}_i - Q}{Q} \tag{16}$$

$$NRMSE = \sqrt{\frac{1}{k} \sum_{i=1}^k \left(\frac{\hat{Q}_i - Q}{Q} \right)^2} \tag{17}$$

where k is the number of estimates of Q and \hat{Q}_i is the i th estimate of the benchmark value Q . To compute these two indices, we fit all models on all trimmed

Table 3 Heterogeneity statistics for the four regions considered—statistics in brackets are obtained with the scale factor taken to be the 1-year quantile corresponding to a non-exceedance probability of 0.5

| Region | He^+ | He | Ho | Ho^+ |
|--------|-------------|-------------|---------------|---------------|
| H_1 | 7.11 (6.83) | 1.35 (1.37) | 0.17 (0.08) | -0.60 (-0.67) |
| H_2 | 3.46 (3.38) | 1.00 (1.03) | 0.41 (0.33) | -1.28 (-1.31) |
| H_3 | 1.40 (1.45) | 0.30 (0.28) | -0.09 (-0.14) | -1.14 (-1.18) |

periods of size m years— $m \in \{5, 10, 15, 20, 25, 30, 37\}$. Moreover, the overall performance of each model is evaluated using a rank score. This technique was already used to compare different models in Shu and Burn (2004).

To calculate the rank score, the p models are ordered using their performance indices—1 corresponding to the best model and p to the worst. For each model, the scores for the different criteria are summed to obtain the overall rank score R_O for the model. For convenience, the overall rank score R_O is standardised in such a way that it lies within the interval $[0, 1]$:

$$R_S = \frac{pq - R_O}{pq - q} \quad (18)$$

where p is the number of models being considered, and q the number of indices. A standardised rank score close to 1 is associated with a model with a good performance, and 0 with that of poor performance.

Three quantiles are of particular interest Q_5 , Q_{10} and Q_{20} —i.e. associated with a probability of non-exceedance of 0.9, 0.95 and 0.975, respectively. *NRMSE*, *NBIAS* and the standardised rank score for station U4635010 and a record length of 5 years are illustrated in Table 4. Notations for different models in this table consist of one lowercase letter referring to the Bayesian approach b or the Regional Index Flood r and the code for the degree of the homogeneity of the region. Only the MLE model does not use these codes as it is completely independent of the homogeneity level.

The results from Table 4 demonstrate that the Bayesian model performs quite well independently of the region being considered. However, this model seems to perform even better when applied to an “acceptably homogeneous” or “probably heterogeneous” region. For the Ho^+ region, the Bayesian approach performs poorly. This may be explained by the fact that the prior distribution is too informative and probably not consistent with the target site sample.

This comment does not apply to the good overall performance of the REG model on this region. Indeed, as the prior distribution is elicited using Eqs. 6, 7, 8, and the scale factor $C^{(j)}$ is estimated without any use of the target site sample, this can lead to a misleading prior distribution while the REG model performs well. The bad estimation of the scale factor is less important with a more heterogeneous region as the prior information is less informative, thus the Bayesian model performance is not highly affected.

On the other side, the overall rank score of the REG model increases with the degree of homogeneity of the region. Yet, the overall rank score for the REG model never exceeds the value of 0.6—reached for the Ho^+ region. This value remains much lower than the best rank score for the Bayesian model—i.e. 0.85. These results corroborate the superiority of the Bayesian approach.

From Table 4, two conclusions can be established. On one hand, for small samples, the Bayesian approach is the most efficient model. On the other hand, the results seem to indicate that there is no need to keep increasing the homogeneity of the region as it increases the risk of being too confident in the “homogeneous region” without increasing significantly the efficiency of the model.

These results are in line with similar results obtained for stations U4505010 and V3015010, except for the poor behaviour of the Bayesian model on the Ho^+ region. Indeed, for the other stations, the Bayesian model remains more efficient than the REG model within the Ho^+ region. However, its overall rank score remains stable through out the different regions -it He, Ho and Ho^+ . The “risk” of dealing with a too homogeneous region such as Ho^+ is also corroborated, as the overall rank score for the index flood model for station U4505010 decreases dramatically to 0.06. Thus, the index flood for the Ho^+ region performs quite well for stations U4635010 and V3015010, while very surprisingly badly with U4505010.

Table 4 Estimation of *NRMSE* and *NBIAS* for station U4635010 with a record length of 5 years

| Model | NRMSE | | | NBIAS | | | Rank score |
|---------|-------|----------|----------|--------|----------|----------|------------|
| | Q_5 | Q_{10} | Q_{20} | Q_5 | Q_{10} | Q_{20} | |
| MLE | 0.33 | 0.34 | 0.39 | 0.01 | − 0.09 | − 0.18 | 0.26 |
| bHe^+ | 0.16 | 0.13 | 0.18 | 0.09 | − 0.02 | − 0.13 | 0.65 |
| rHe^+ | 0.27 | 0.30 | 0.37 | − 0.12 | − 0.22 | − 0.31 | 0.18 |
| BHe | 0.10 | 0.07 | 0.11 | 0.08 | 0.00 | − 0.09 | 0.85 |
| RHe | 0.27 | 0.26 | 0.28 | − 0.03 | − 0.10 | − 0.17 | 0.43 |
| bHo | 0.14 | 0.09 | 0.08 | 0.12 | 0.05 | − 0.02 | 0.76 |
| rHo | 0.27 | 0.26 | 0.27 | 0.01 | − 0.06 | − 0.12 | 0.58 |
| bHo^+ | 0.29 | 0.28 | 0.25 | 0.29 | 0.27 | 0.25 | 0.19 |
| rHo^+ | 0.28 | 0.27 | 0.26 | 0.02 | − 0.01 | − 0.04 | 0.60 |

In Fig. 8, the changes in the overall rank score as a function of the record length is illustrated for station V3015010. The left panel corresponds to the REG method, while the right one stands for the Bayesian approach. The MLE scores are also presented in both panels. Fig. 8 indicates that the changes in the overall rank score are more stable for regional models, i.e. REG and Bayesian models, than for MLE. Furthermore, the benefit of increasing the degree of homogeneity of the region is more relevant for the REG model than for the Bayesian model. Nevertheless, the worst Bayesian rank score is always quite close to the best REG rank score. This seems to indicate the superiority of the Bayesian approach. This last point is corroborated with the results corresponding to stations U4505010 and U4635010 except for the bHo^+ model for station U4635010 because of the bad estimation of the scale factor $C^{(j)}$, as mentioned earlier. The effect of bad estimation of the target site Index Flood on prior and thereby on posterior distributions is depicted in Fig. 9.

From Fig. 9, it is overwhelmingly clear that the prior model is not appropriate—particularly for the location parameter. The prior for the shape parameter is not too false as it does not depend on the target site index flood estimate.

As the record length increases, the MLE model becomes increasingly efficient. In particular, for record lengths greater than 15 years, it is more effective than rHe^+ , rHe and rHo models. On one hand, for record lengths smaller than 15 years, MLE is always less efficient than Bayesian approaches and even significantly for bHe , bHo and bHo^+ models. This is quite logical as Bayesian estimation can be looked at as a restrictive maximum likelihood estimator - restriction being defined by the prior distribution. So, under the hypothesis that the prior distribution is well-defined,

the “restrictive estimator” is unbiased and has a smaller variance. On the other hand, for record lengths greater than 15 years, MLE, bHe and bHo seems to be similar.

8 Conclusion

A framework for performing a regional Bayesian frequency analysis for partially gauged stations is presented. The proposed model has the advantage of being less restrictive than the most widely used regional model, that is the index flood. Several case studies from French sites were analysed to illustrate the superiority of the Bayesian approach in comparison to the traditional index flood and to local approaches. The influence of the homogeneity level of the pooling group on quantile estimates was also considered. The results demonstrate that working with quite large and homogeneous regions rather than small and strongly homogeneous regions is more efficient. Further work can focus on the regional estimation of other characteristics of the flood hydrograph. For instance, a regional Bayesian model can focus on Flood Duration Frequency.

All statistical analysis was carried out in the R Development Core Team (2005) framework. For this purpose, two packages were contributed to this software within the framework of the present research work. These two packages integrate the tools that were developed to carry out the modelling effort presented in this paper. The first one POT performs statistical inference on peaks over thresholds, while the second one, RFA, contains several tools to carry out a regional frequency analysis. These two packages are available, free of charge, at the web site <http://www.R-project.org>, section CRAN, Packages.

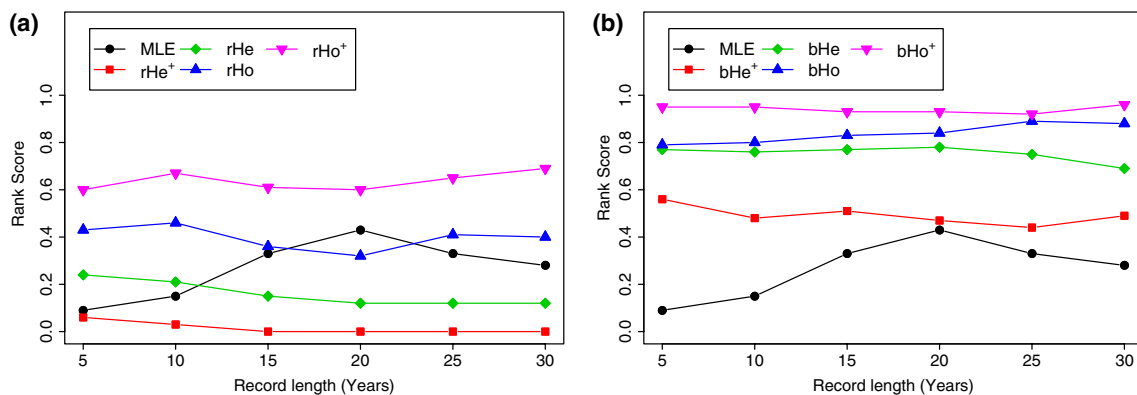
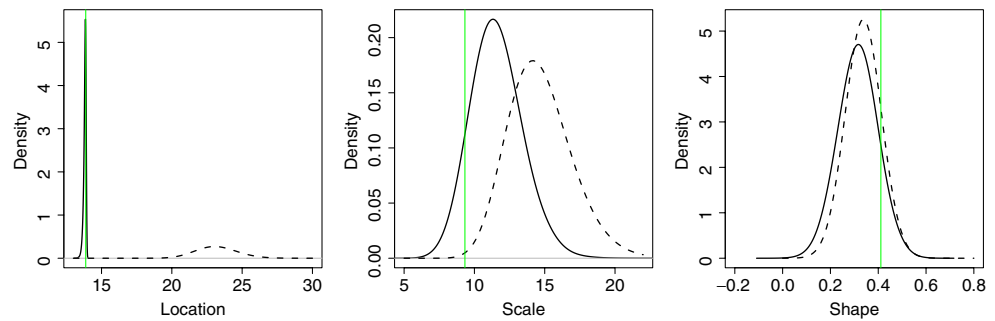


Fig. 8 Score evolution as a function of record length for Station V3015010. a REG scores, b BAY scores

Fig. 9 Effect of bad estimation of target site index flood on marginal prior and posterior densities. Site U4635010 with 10 years record length



Acknowledgements The authors wish to thank the DIREN Rhône-Alpes for providing data. The authors are also very grateful to the two referees for their constructive remarks which improved the document.

9 Appendix: Properties of the index flood on GP parameters

We provide in this appendix the proof for the following theorem:

Theorem 1 *Let X be a random variable GP distributed. So X has the cumulative distribution function defined by:*

$$F(x) = 1 - \left[1 + \frac{\xi(x - \mu)}{\sigma} \right]^{-1/\xi}$$

Let $Y = CX$ where $C \in \mathbb{R}_^+$. Then, Y is also GP distributed with parameters $(\mu C, \sigma C, \xi)$.*

Proof Let X be a r.v. GP distributed with parameters (μ, σ, ξ) and $Y = CX$ where $C \in \mathbb{R}_*^+$. Then:

$$\begin{aligned} \Pr[Y \leq y] &= \Pr\left[X \leq \frac{y}{C}\right] = 1 - \left[1 + \frac{\xi(yC - \mu)}{\sigma} \right]^{-1/\xi} \\ &= 1 - \left[1 + \frac{\xi(y - \mu C)}{\sigma C} \right]^{-1/\xi} \end{aligned}$$

So, Y is also GP distributed with parameters $(\mu C, \sigma C, \xi)$. The proof for the GEV case can be established in the same way.

References

- Acreman M, Wiltshire S (1989) The regions are dead: long live the regions. In: FRIENDS, vol 187. International association of hydrological sciences, Washington, pp 175–188
- Balkema A, de Haan L (1974) Residual life time at great age. *Ann Probab* 2:792–804
- Burn DH (1990) Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Res* 26(10):2257–2265
- Coles S (2001) An introduction to statistical modelling of extreme values. Springer Series in Statistics. Springer, London
- Coles S, Pericchi L (2003) Anticipating catastrophes through extreme value modelling. *J R Stat Soc C* 52(4):405–416
- Coles S, Tawn J (1996) A bayesian analysis of extreme rainfall data. *J R Stat Soc Ser C Appl Stat* 45(4):463–478
- Crowder M (1992) Bayesian priors based on a parameter transformation using the distribution function. *Ann Inst Stat Math* 44(3):405–416
- Dalrymple T (1960) Flood frequency analysis. US Geol Surv Water Supply 1543A
- Development Core Team R (2005) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>. ISBN 3-900051-07-0
- Embrechts P, Klüppelberg C, Mikosch T (1997) Modelling extremal events for insurance and finance. Springer, Berlin Heidelberg New York
- Fill HD, Stedinger JR (1998) Using regional regression within index flood procedures and an empirical bayesian estimator. *J Hydrol* 210(1–4):128–145
- Fisher RA, Tippett LH (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc Cam Phil Soc* 24:180–190
- Gabriele S, Arnell N (1991) A hierarchical approach to regional flood frequency analysis. *Water Resources Res* 27(6):1281–1289
- GREHYS (1996) Inter-comparison of regional flood frequency procedures for Canadian rivers. *J Hydrol* 186(1–4):85–103
- Gupta VK, Mesa OJ, Dawdy DR (1994) Multiscaling theory of flood peaks: Regional quantile analysis. *Water Resources Res* 30(12):3405–3421
- Hall MJ, Minns AW, Ashrafuzzaman AKM (2002) The application of data mining techniques for the regionalisation of hydrological variables. *Hydrol Earth Syst Sci* 6(4):685–694
- Hosking JRM, Wallis JR (1993) Some statistics useful in regional frequency analysis. *Water Resources Res* 29(2):271–281
- Hosking JRM, Wallis JR (1997) Regional frequency analysis. Cambridge University Press, Cambridge
- Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. *Adv Water Resources* 25(8–12):1287–1304
- Lang M, Ouarda TBMJ, B. Bobée B (1999) Towards operational guidelines for over-threshold modeling. *J Hydrol* 225(3–4):103–117
- Madsen H, Rosbjerg D (1997) Generalized least squares and empirical bayes estimation in regional partial duration series index-flood modeling. *Water Resources Res* 33(4):771–781
- Northrop PJ (2004) Likelihood-based approaches to flood frequency estimation. *J Hydrol* 292(1–4):96–113

- Parent E, Bernier J (2003) Encoding prior experts judgments to improve risk analysis of extreme hydrological events via pot modeling. *J Hydrol* 283(1–4):1–18
- Pickands J III (1975) Statistical inference using extreme order statistics. *Ann Statist* 3:119–131
- Robinson JS, Sivapalan M (1997) An investigation into the physical causes of scaling and heterogeneity of regional flood frequency. *Water Resources Res* 33(5):1045–1059
- Robson AJ, Reed DW (1999) Flood estimation handbook, vol 3. Institute of Hydrology, Wallingford
- Shu C, Burn DH (2004) Artificial neural network ensembles and their application in pooled flood frequency analysis. *Water Resources Res* 40(9):W09301
- Wasson JG, Chandesris A, Pella H, Blanc L (2004) Hydro-eco-regions: a functional approach of river typology for the European water framework directive. *Ingénieries EAT in French* 40:3–10