



Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations

Chi Cong Nguyen, Eric Gaume, Olivier Payrastre

► To cite this version:

Chi Cong Nguyen, Eric Gaume, Olivier Payrastre. Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations. *Journal of Hydrology*, 2014, 508, pp. 385-396. 10.1016/j.jhydrol.2013.09.058 . hal-00923359

HAL Id: hal-00923359

<https://hal.science/hal-00923359>

Submitted on 2 Jan 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Regional flood frequency analyses involving
2 extraordinary flood events at ungauged sites: further
3 developments and validations

4 C.C. Nguyen^{a,b}, E. Gaume^a, O. Payrastre^a

5 ^a*LUNAM Université, IFSTTAR, GERS, F-44344 Bouguenais, France*

6 ^b*Danang university, Vietnam*

7 **Abstract**

8 Flood frequency analyses are often based on recorded series at gauging
9 stations. However, the length of the available data sets is usually too short to
10 provide reliable estimates of extreme design floods. Hence, hydrologists have
11 tried to make use of alternative sources of information to enrich the datasets
12 used for the statistical inferences. Two main approaches were therefore pro-
13 posed. The first consists in extending the information in time, making use
14 of historical and paleoflood data. The second, spatial extension, consists
15 in merging statistically homogeneous data to build large regional data sam-
16 ples. Recently, a combination of the two techniques aiming at including
17 estimated extreme discharges at ungauged sites of a region in the regional
18 flood frequency analyses has been proposed. This paper presents a consoli-
19 dation of this approach and its comparison with the standard regional flood
20 frequency approach proposed by Hosking & Wallis. A modification of the
21 likelihood function is introduced to enable the simultaneous calibration of a
22 regional index flood relation and of the parameters of the regional growth
23 curve. Moreover, the efficiency of the proposed method is evaluated based

*Corresponding author. Tel.: +33(0)2 40 84 58 84, eric.gaume@ifsttar.fr

Preprint submitted to Journal of Hydrology

August 5, 2013

on a large number of Monte Carlo simulated data sets. This work confirms that extreme peak discharges estimated at ungauged sites may be of great value for the evaluation of large return period (typically over 100 years) flood quantiles. They should therefore not be neglected despite the uncertainties associated to these estimates.

Keywords:

Floods, regional analysis, statistics, Extremes , GEV

1. Introduction

Although a large number of statistical inference methods have been progressively developed, the question of estimating extreme design floods is still problematic due to the generally limited amount of available data. Continuous discharge series at gauged sites are generally too short to provide reliable estimates of extreme quantiles - typically the 100-year or higher return period quantiles (NERC, 2000). To cope with this difficulty, hydrologists have tried to complement the available data sets, either through a “temporal extension”, incorporating data on historical and paleofloods (Hosking and Wallis, 1986a,b; Stedinger and Cohn, 1986; Cohn and Stedinger, 1987; Gary and Stedinger, 1987; Sutcliffe, 1987; Minghui and Stedinger, 1989; Sheffer et al., 2003; Reis et al., 2005; Neppel et al., 2010; Payraastre et al., 2011), or through a “spatial extension”, merging data sets in regions considered as statistically homogeneous, “trading space for time” according to the words of Hosking & Wallis (Hosking and Wallis, 1997; Heinz and Stedinger, 1998; Charles and Stedinger, 1999; Ouarda et al., 2001; Kjeldsen et al., 2002; Merz and Blöschl, 2003; Seidou et al., 2006; Ribatet et al., 2007; Norbiato et al., 2007; Wallis

et al., 2007; Kjeldsen and Jones, 2009; Haddad and Rahman, 2011).

Recently, Gaume et al. (2010) observed that estimated extreme peak discharges at ungauged sites are often available, but never really used in flood frequency studies and proposed a method to incorporate such information in regional flood frequency analyses.

The proposed approach is based on the index flood principle (Dalrymple, 1960), assuming that, within a statistically homogeneous region, all local statistical distributions are identical apart from a site-specific scaling factor: the index flood. Usually, the index flood corresponds to the mean of the local series (Hosking and Wallis, 1997). The approach proposed by Gaume et al. (2010) is based on the calibration of an index flood relation linking the characteristics of the watersheds and the index flood value. Although this relationship represents an additional homogeneity requirement that may limit the extent of the region used for the statistical analysis, it also enables to estimate the index flood at ungauged sites, and thus to incorporate the corresponding ungauged extremes in the regional sample.

Based on several case studies, Gaume et al. (2010) showed the possible great value of such an approach, depending on the characteristics of available extreme flood inventories. The index flood relationship proposed, of the form S^β (S being the area of the watershed and β a parameter to be calibrated), appeared satisfactory in the test regions. The presented inference results were based on a Bayesian MCMC framework (Castellarin, 2005; Reis et al., 2005; Seidou et al., 2006; Ribatet et al., 2007; Castellarin et al., 2007; Payrastre et al., 2005, 2011) to adjust the regional growth curve with associated 90% credibility intervals. The results showed that the incorporation of ungauged

73 extremes could lead to a significant reduction of the width of the computed
74 credibility intervals.

75 In the initial version of the method (Gaume et al., 2010), the index flood
76 relation was adjusted prior to the calibration of the regional growth curve,
77 and the uncertainties associated with its calibration were not taken into
78 account. This led certainly to underestimate the credibility intervals and
79 over-rate the added value of the ungauged extremes and of the proposed
80 method. The effects of possible variations (heterogeneities) in the average
81 relation calibrated in a given region should also be considered for a fair
82 comparison with other statistical methods. This paper proposes an extension
83 of the initial method to account for uncertainties in the calibrated index flood
84 relation. It also tests the effect of possible regional variations in the average
85 relation on the efficiency of the proposed statistical inference approach.

86 The performances (i.e. widths and correctness of computed credibility
87 intervals) of the proposed approach and of the standard regional frequency
88 approach proposed by Hosking & Wallis (1997) are first compared in the
89 case where gauged data only are considered. The comparison is based on
90 samples generated through Monte Carlo simulations in order to be able to
91 verify the accuracy of the calculated credibility intervals and to introduce
92 controlled heterogeneities in the samples. In a second step, both approaches
93 are applied to the statistical analysis of a data set from the Ardèche region
94 in France composed of 168 records at 5 gauging stations and 18 estimated
95 ungauged extremes.

96 The paper is organised as follows: section 2 presents the basics and adap-
97 tations of the two regional flood frequency methods: Hosking & Wallis and

the proposed approach. The performances of the two approaches are compared based on simulated samples of random variables in section 3. In section 4, the methods are applied to the real-world case study. Conclusions are drawn in section 5.

2. Tested regional flood frequency analysis methods

2.1. The index flood hypothesis

The two approaches considered in this paper are based on the same fundamental simple scaling hypothesis or index-flood principle (Dalrymple, 1960): in a statistically homogeneous region, all the local annual maximum peak discharge distributions are supposed to be identical apart from a site-specific scaling factor. This hypothesis is summarized in equation 1:

$$Q_i(F) = \mu_i q(F) \quad (1)$$

Where F is the probability of non-exceedance, i is the index of the site ($i = 1, \dots, s$), s the total number of sites in the homogeneous region, $Q_i(F)$ is the discharge quantile, $q(F)$ is the regional dimensionless (i.e. reduced) quantile and μ_i is the index flood (or scaling factor).

The index flood may be any constant value proportional to the expectancy of the local distribution. Usually, when only data from gauged sites are considered, the index flood is estimated by the at-site sample mean (Hosking and Wallis, 1997; Castellarin, 2005; Castellarin et al., 2007). A regional flood frequency method where the index flood is computed as the average of the local series of annual maxima will be called hereafter method of Hosking & Wallis. This, even if a likelihood based Bayesian MCMC procedure rather

120 than a L-moment based procedure, as suggested by Hosking & Wallis (1997),
 121 is used to calibrate the parameters of the regional growth curve. Gaume et al.
 122 (2010) suggested an alternative approach to account for extreme discharge
 123 estimates that may be available at ungauged sites. An inventory of ungauged
 124 extremes may includes h extreme peak discharges Q_k ($k = 1, \dots, h$), each Q_k
 125 corresponding to the largest flood at site k during a period of length n_k . In
 126 order to include this additional information in the regional dataset, Gaume
 127 et al. (2010) proposed to use an index flood relation linking the index flood
 128 value to the catchment area S , since an average annual peak discharge can
 129 obviously not be computed at ungauged sites:

$$\mu_i = S_i^\beta \text{ and } \mu_k = S_k^\beta \quad (2)$$

130 Where S_i and S_k are the catchment areas at the corresponding sites, and
 131 β a coefficient to be calibrated. More complex relations based on various
 132 climatic and physio-geographic characteristics may be tested in the future,
 133 but at the price of an increased number of parameters to be calibrated. In
 134 the initial version of the method, the value of β was adjusted through a
 135 regression between the log tranform of the average annual peak discharges
 136 and the watershed areas at gauged sites.

137 It is proposed here to calibrate β along with the parameters of the regional
 138 growth curves using a modified version of likelihood as described below.

139 *2.2. Likelihood of the observed sample*

140 The inference approach applied herein is directly derived from Gaume
 141 et al. (2010) and inspired by numerous previous works (Reis et al., 2005;

Renard et al., 2006; Payraastre et al., 2011): i.e. based on the likelihood of the available data sets and a Bayesian MCMC algorithm for the estimation of the parameters of the growth curve and of their posterior distribution according to the observed data set.

Considering the regional sample \mathbf{D} described above, including both (i) s series of gauged annual maximum discharges $Q_{i,j}$, $j = 1, \dots, n_i$ the index of the year, and (ii) the h estimated largest peak discharges Q_k over n_k years at h ungauged sites, the standard expression of the likelihood of the regional sample \mathbf{D} would be the following:

$$\ell(\mathbf{D} \mid \theta) = \prod_{i=1}^s \left[\prod_{j=1}^{n_i} f_{\theta} \left(\frac{Q_{i,j}}{\mu_i} \right) \right] \prod_{k=1}^h \left[f_{\theta} \left(\frac{Q_k}{\mu_k} \right) \right] \prod_{k=1}^h \left[F_{\theta} \left(\frac{Q_k}{\mu_k} \right) \right]^{(n_k-1)} \quad (3)$$

Where f_{θ} and F_{θ} are respectively the probability density function and the cumulative probability function of the selected statistical distribution for the regional growth curve, and θ corresponds to the vector of parameters to be estimated. The GEV distribution, often used to describe peak discharge growth curves (Lu and Stedinger, 1992; Stedinger and Lu, 1995; Coles and Powell, 1996; Coles and Tawn, 1996; Heinz and Stedinger, 1998; Seidou et al., 2006), was selected here (Eq. 4 and 5). The vector θ comprises the position, scale and shape parameters (ξ , α , κ) of the GEV distribution.

$$F_{\theta}(Q) = \exp \left[- \left(1 - \frac{\kappa(Q - \xi)}{\alpha} \right)^{1/\kappa} \right]_{\alpha > 0} \quad (4)$$

$$f_{\theta}(Q) = \frac{1}{\alpha} \left(1 - \frac{\kappa(Q - \xi)}{\alpha}\right)^{1/\kappa - 1} \exp \left[- \left(1 - \frac{\kappa(Q - \xi)}{\alpha}\right)^{1/\kappa} \right]_{\alpha > 0} \quad (5)$$

159 In equation 3, the first term corresponds to the probability of the gauged
 160 series. It is the only necessary term if continuous series of measured annual
 161 maximum discharges are used. The second term is the probability of the
 162 ungauged extremes. The third complementary term is the probability asso-
 163 ciated to the fact that the ungauged extreme value has not been exceeded
 164 during the remaining $(n_k - 1)$ years at each ungauged site. The index values
 165 μ can be estimated before the calibrating the regional growth curve : method
 166 of Hosking & Wallis and initial version of the proposed approach. Equation 2
 167 can be also directly introduced in the formulation of the likelihood (eq. 3)
 168 adding one parameter, β , to be calibrated. This replacement necessitates
 169 a slight but determining modification of the expression of likelihood (sec-
 170 tion 2.4). Depending on the option, the set of parameters to be calibrated
 171 will count either 3 or 4 parameters.

172 *2.3. Bayesian Monte Carlo Markov Chain algorithm*

173 The Bayesian Monte Carlo Markov Chain procedure is now relatively
 174 common for hydrological applications (Reis et al., 2005; Renard et al., 2006;
 175 Seidou et al., 2006; Ribatet et al., 2007; Neppel et al., 2010; Gaume et al.,
 176 2010; Payraastre et al., 2011; Viglione et al., 2013) and will only be briefly
 177 presented here. Part of the algorithms used is included in the R software
 178 library nsRFA (Viglione, 2013). Recalling the Bayes' theorem, the likelihood

179 of the sample given the parameters of the statistical model $\ell(\mathbf{D} | \theta)$ can be
 180 related to the likelihood or density of probability of the parameters given the
 181 sample $p(\theta | \mathbf{D})$ (posterior distribution):

$$p(\theta | \mathbf{D}) = \frac{\ell(\mathbf{D} | \theta) p(\theta)}{p(\mathbf{D})} \quad (6)$$

182 Where $p(\theta)$ is the so called prior distribution of θ , which summarizes
 183 any prior or alternative knowledge on the distribution of θ , and $p(\mathbf{D})$ is the
 184 probability of the sample \mathbf{D} which is unknown. When prior information on
 185 the distribution of θ does not exist, then $p(\theta)$ is often taken as uniform.
 186 It is the case here, which implies that $p(\theta | \mathbf{D})$ is proportional to $\ell(\mathbf{D} | \theta)$.
 187 The statistical model being chosen, it is possible to compute the probability
 188 density function of its parameters according to the observed data sample.

189 $p(\theta | \mathbf{D})$ being known, or more precisely $\ell(\mathbf{D} | \theta)$ which is proportional to
 190 it according to our hypotheses (uniform prior), parameter sets θ will be sam-
 191 pled according to $p(\theta | \mathbf{D})$ to build their posterior distributions and compute
 192 the corresponding credibility limits for the discharge quantiles. The MCMC
 193 algorithms, combining random walk Monte Carlo methods with Markov
 194 chains, are a class of algorithms for the efficient sampling from multivari-
 195 ate random distributions (Tanner, 1996; Robert and Casella, 2004).

196 Figure 1 illustrates the Bayes MCMC inference procedure: posterior dis-
 197 tributions of two of the estimated parameters for 3 different Markov chains
 198 run in parallel.

199 2.4. Modification of the likelihood formulation

200 A modification of the values of the index floods μ_i and μ_k , corresponding
 201 to a change of scale of the analyzed discharges, will also affect the values

202 of the optimal parameters ξ and α , the variable $(Q/\mu - \xi)/\alpha$ remaining
 203 unchanged. The absolute value of the cumulative density function (eq. 4) will
 204 be unaffected by such a change of variable, but not the absolute value of the
 205 density function, due to the factor $1/\alpha$ in front of the expression of the density
 206 (eq. 5). Any method based on the maximization of likelihood expression 3,
 207 will therefore tend to minimize the value of α , and hence to maximize the
 208 values of the indices $\mu_i = S_i^\beta$ if they are calibrated. Therefore, the use
 209 of likelihood expression 3 in an inference procedure where the parameter β
 210 is adjusted will inevitably lead to biased estimates of β (over-estimations)
 211 and of the parameters of the regional growth curve which are linked to β .
 212 This is confirmed by numerical tests as illustrated in Figure 1. To avoid
 213 this disruptive effect, one possibility consists in including only cumulative
 214 densities in the likelihood expression (eq. 7).

$$\begin{aligned}
 \ell(\mathbf{D} \mid \theta) = & \prod_{i=1}^s \left[\prod_{j=1}^{n_i} \left[F_\theta \left(\frac{Q_{i,j}^U}{S_i^\beta} \right) - F_\theta \left(\frac{Q_{i,j}^L}{S_i^\beta} \right) \right] \right] \\
 & \prod_{k=1}^h \left[F_\theta \left(\frac{Q_k^U}{S_k^\beta} \right) - F_\theta \left(\frac{Q_k^L}{S_k^\beta} \right) \right] \prod_{k=1}^h \left[F_\theta \left(\frac{Q_k^U}{S_k^\beta} \right) \right]^{(n_k-1)} \quad (7)
 \end{aligned}$$

215 Where the exponents U and L indicate the upper and lower estimate for
 216 the considered discharge values (upper and lower guesses for an estimated
 217 ungauged extreme or known uncertainty range for measured discharges). It
 218 has been verified that equations 7 and 3 provide very close results if the
 219 difference between $Q_{i,j}^U$ and $Q_{i,j}^L$ is small. In the rest of the manuscript $Q_{i,j}^U$
 220 will be set equal to $1.01 Q_{i,j}$ and $Q_{i,j}^L$ to $0.99 Q_{i,j}$.

221 Figure 1 illustrates the type of results obtained through the application of

222 the MCMC algorithm using either expressions 3 or 7 of the likelihood, when
 223 the index flood relation is directly introduced in the likelihood formulation
 224 and the parameter β is calibrated along with the parameters of the regional
 225 growth curve. The simulated regional data sample is described in section 3.
 226 These results show the tremendous under-estimation of the parameter α of
 227 the regional growth curve when equation 3, containing density functions,
 228 is used. The assessment seems correct with equation 7, even if affected
 229 by uncertainties, due to the finite number of records (i.e. to the limited
 230 information content of the data set).

231 *2.5. MCMC convergence diagnosis*

232 Like for any other optimization method, convergence of the MCMC al-
 233 gorithm towards the posterior distribution of the model parameters may
 234 sometimes be difficult to reach, especially when the number of parameters to
 235 be calibrated is increased. For the inference on a single data set, convergence
 236 problems may be solved by tuning the MCMC algorithm: better selection
 237 of the parameter starting values, adjustment of the Monte Carlo Markov
 238 chain random search controlling factors. The tests presented hereafter im-
 239 plied repeated applications of the MCMC algorithm on a large number of
 240 data samples. It was therefore necessary to strengthen the algorithm to en-
 241 sure rapid convergence in most cases, to verify automatically the convergence
 242 of the MCMC algorithm for each run and to avoid introducing results corre-
 243 sponding to insufficient convergence or convergence failures in the analysis.
 244 Among the large number of methods developed for convergence monitoring
 245 of MCMC algorithms (Cowles and Carlin, 1996; Salaheddine et al., 2006), the
 246 popular Gelman & Rubin test has been selected (Gelman and Rubin, 1992).

247 It consists in running several Markov chains in parallel and in computing
248 the evolution during the iterations of the square root of the ratio between
249 average within-chain and average between-chains variance of the likelihood.
250 This ratio should ideally be equal to 1 if convergence is achieved.

251 The test was applied herein with two chains and a threshold value of
252 1.05. Figure 2 shows the evolution of the Gelman & Rubin criterion with
253 the number of iterations when MCMC algorithms are run on synthetic data
254 samples presented in section 3. To achieve fast convergence, a first MCMC
255 chain was systematically run with a limited number of iterations (30000)
256 to explore the shape of likelihood function in the vicinity of the parameter
257 values maximizing the likelihood and evaluate the covariance matrix of the
258 parameters (important controlling factor of the MCMC chain). Figure 2
259 shows that the convergence is rapidly achieved with the reference approach,
260 when it appears to be more difficult with the proposed approach. According
261 to the conducted tests, the maximum iteration number was set to 100 000
262 and the various MCMC outputs computed on the basis on the last 30 000
263 iterations. This number of iterations was sufficient to ensure the convergence
264 of the MCMC algorithm for all the simulations presented in this paper.

265 **3. First test of the two methods**

266 The two regionalisation approaches (standard and proposed) have been
267 first tested for the statistical analysis of simulated random data sets to eval-
268 uate their performances, before their application to a real-world case study
269 (statistical analysis of data samples available in the Ardèche region, South of
270 France). A special emphasis has been put on (i) the validation the credibility

intervals computed with the MCMC procedure and (ii) the analysis of the sensitivity to regional variability in the index flood relation.

3.1. Method

(i) Monte Carlo simulation procedure.

The test set is composed of 1000 simulated random samples drawn from a regional GEV distribution ($\xi=3.34$, $\alpha=2.24$, $\kappa=-0.16$) and with an index flood relation ($\beta=0.76$) adjusted on the Ardèche set analysed in section 4. The samples have the same characteristics as the Ardèche regional gauged sample: catchment areas and record lengths, see table 1.

Both regionalisation approaches are then applied to each simulated dataset, including the estimation of the posterior distribution of parameters thanks to the Bayesian MCMC procedure. The cases for which convergence of the MCMC algorithm is not achieved, according to the Gelman & Rubin test, are discarded. The posterior distribution of the 100-year quantile is then built, and the corresponding maximum likelihood value $\hat{Q}_i^{ML}(0.99)$ (mode of this posterior distribution since no prior information is considered) and 90% centered credibility interval $[\hat{Q}_i^5(0.99), \hat{Q}_i^{95}(0.99)]$ are obtained for each site and each dataset. The final result is a set of 1000 (or less if some MCMC chains did not converge which did not happen here) of maximum likelihood quantiles $\hat{Q}_i^{ML}(0.99)$ and corresponding 90% credibility intervals. They will be shown as box plots.

(ii) Reliability of the computed credibility intervals.

Monte Carlo simulations enable to test the relevance of the posterior distributions of the discharge quantile $\hat{Q}_i(0.99)$ obtained through the Bayesian

295 MCMC procedure. The real quantile value $Q_i(0.99)$ is known, and its prob-
 296 ability of non-exceedance $\hat{F}(Q_i(0.99))$ can be computed according to the
 297 computed posterior distribution of the discharge quantile $\hat{Q}_i(0.99)$ for each
 298 simulated sample. If these posterior distributions of the estimated discharge
 299 quantiles $\hat{Q}_i(0.99)$ are unbiased - i.e. reflect the correct density of prob-
 300 ability of the real quantile $Q_i(0.99)$, the values of $\hat{F}(Q_i(0.99))$ estimated
 301 for the 1000 samples should be uniformly distributed over $[0, 1]$. Figure 3
 302 illustrates how the distributions of $\hat{F}(Q_i(0.99))$ should be interpreted. For
 303 instance in case (c), the real quantile value has an estimated probability of
 304 non-exceedance close to 0 or 1 for too many cases : the real quantile lies too
 305 often at the margins of the computed posterior distribution for the estimated
 306 quantiles, sign that the variance of the computed distribution is too low and
 307 that the computed intervals are too narrow. Likewise, case (d) illustrates the
 308 case of an overestimation of the variance of the posterior distributions and
 309 resulting too large credibility intervals. Case (b) reveals an underestimation
 310 of the quantiles.

311 **(iii)** Introduction of regional variability in the index flood relation.
 312 The Monte Carlo simulations have been repeated with the introduction of a
 313 regional variability in the index flood relation to test the robustness of the
 314 new proposed method. For each generated sample, the index flood value
 315 at site i , $\mu_i = S_i^\beta$ was replaced by a random value drawn from a lognormal
 316 distribution with mean S_i^β and standard deviation $\delta * S_i^\beta$, δ being successively
 317 set to 0.1 and 0.3. Statistical tests indicate that values of δ equal to 0 or 0.1
 318 are compatible with the gauged sample available in the Ardèche region.

319 3.2. Results

320 (i) Case of homogeneous data sets

321 Figure 4 presents the inference results based on 1000 simulated samples,
 322 without introduction of regional variability in the index flood relation. The
 323 fluctuations of the maximum likelihood estimates $\hat{Q}_i^{ML}(0.99)$ (respectively
 324 $\hat{q}^{ML}(0.99)$ for the regional growth curve), and associated 90% credibility
 325 upper and lower bounds $[\hat{Q}_i^5(0.99), \hat{Q}_i^{95}(0.99)]$, are presented both, for the
 326 estimates corresponding to each gauged watershed area of the region and for
 327 the regional growth curve. In order to facilitate comparisons, all estimations
 328 have been divided by the real quantile value $Q_i(0.99)$ (or $q(0.99)$). Therefore
 329 the position of the real quantile is 1 for all the box plots.

330 Figure 4.a and b appear overall very similar except for the regional growth
 331 curves: the maximum likelihood estimates and the width of computed cred-
 332 ibility intervals fluctuate in similar ranges, particularly for the larger catch-
 333 ment areas. The uncertainties increase logically in the proposed method for
 334 smaller catchment areas, because the sensitivity of the index flood to the
 335 value of the parameter β is increasing with the watershed area according to
 336 equation 2. Conversely, the smallest areas have a lower weight in the de-
 337 termination of the index flood relation and as a consequence the calibrated
 338 index flood values are less accurate for lower surfaces.

339 The regional growth curves have different significations in both methods:
 340 distribution of the reduced discharges (values divided by the expectancy of
 341 the distribution) in the method of Hosking & Wallis and distribution of the
 342 discharges for a 1 km^2 watershed in the proposed approach. This is largely in
 343 the extrapolation range of the index flood relation for low values of watershed

344 areas and explains the large uncertainties for the estimated quantiles for the
 345 regional growth curve in the proposed approach.

346 An in depth analysis of both figures reveals nevertheless some differences
 347 of the two regionalisation approaches: the median values of the maximum
 348 likelihood and 5% and 95% credibility bounds for the 100-year quantile are
 349 slightly lower when the standard Hosking & Wallis method is used, but the
 350 uncertainties affecting these estimates (widths of the box plots) appear larger
 351 except for the smallest considered watershed area (63 km^2). The median
 352 maximum likelihood estimated value lower than 1 for the standard method
 353 suggests also that it is slightly biased.

354 The reliability tests conducted on the computed credibility intervals indi-
 355 cate furthermore that, while the intervals seem correctly estimated with the
 356 proposed method (Fig. 5.e and f) with the exception of the smallest watershed
 357 area where a tendency to the overestimation of the quantile appears. The
 358 credibility intervals computed with the standard method appear too narrow
 359 (Fig. 5.a and b). A complementary test presented in appendix Appendix A
 360 indicates that a 20% to 40% underestimation affects the computed widths of
 361 credibility intervals in this case study: a correction factor m equal to 1.2 to
 362 1.4 should be applied to the standard-deviation of the posterior distributions
 363 of $\hat{Q}_i(0.99)$ to obtain a uniform distribution of $\hat{F}(Q_i(0.99))$ (Fig. 5.c and
 364 d). This under-estimation of the variance of the posterior distributions of
 365 $\hat{Q}_i(0.99)$ for the standard method is logical since the uncertainty associated
 366 with the estimation of the local sample means is not accounted for in the ex-
 367 pression of the likelihood. Its effect on the quantile estimations can therefore
 368 not be evaluated by the Bayesian MCMC procedure.

369 The origins of the observed limited biases - tendency to underestimate
370 the quantile on average for the standard method and to overestimate the
371 quantile for the smaller areas for the proposed approach - are difficult to
372 depict.

373 These results are overall extremely satisfactory. When the regional sam-
374 ple corresponds to the proposed statistical model (regional growth curve of
375 the GEV type and power-law linking the index flood and the area of the
376 watershed), the proposed approach appears consistent and accurate. The
377 parameters of the regional statistical model seem correctly retrieved from
378 samples with only a slight bias in the estimation of the parameter β af-
379 fecting the estimates of the discharge quantiles of the watersheds with the
380 smallest areas. In the standard method, an additional source of uncertainty
381 seems to be introduced by the computation of local sample means, estimates
382 of the local sample expectancies, leading to an unexplained minor but de-
383 tectable underestimation of the quantiles and to an increase of the estimation
384 uncertainties of these quantiles if compared to the proposed method. The
385 credibility intervals computed through the MCMC algorithm underestimate
386 in this case the uncertainties associated to the calculated quantiles.

387 The additional parameter β appears to be a factor of stabilization rather
388 than a source of further uncertainties in a flood frequency analysis. Fixing
389 an index flood relation seems to help to filter out the variability of the local
390 sample means due to sampling fluctuation. This is an unexpected result.
391 The proposed approach, initially designed to introduce ungauged extremes in
392 flood frequency analysis, is as efficient, if not more efficient than the standard
393 approach when applied to gauged series. To be really able to draw such a

394 conclusion, it is necessary to consider that the regional data set may not
 395 follow perfectly the simple suggested index flood relation. How does the
 396 method resist to some regional variability in the index flood relation? It is
 397 the focus of the next section.

398 **(ii)** Impact of some regional variability in the index flood relation

399 Without surprise, the introduction of variability in the index flood rela-
 400 tion increases both, the estimated credibility intervals and the uncertainties
 401 affecting the estimates of the maximum likelihood, 5% and 95% credibility
 402 bound estimates (Fig. 6).

403 The effect remains nevertheless limited when the variability is relatively
 404 low (Fig. 6.a). The results obtained for $\delta = 0.1$ do not differ much from the
 405 results obtained in the ideal case ($\delta = 0$ in Fig. 4.b). The maximum likelihood
 406 estimate is still unbiased and it could be verified that the credibility intervals
 407 were also only slightly biased. Uncertainties increase drastically when the
 408 regional variability is increased (Fig. 6.b), but still with a limited bias, on
 409 average, on the maximum likelihood estimate of the quantile (overestimation)
 410 and some bias appearing in the estimated credibility intervals. Of course, the
 411 biases on the individual local estimates can be significant as shown by the
 412 widths of the box plots.

413 These results finally suggest that the proposed approach may resist to a
 414 certain level of heterogeneity of the proposed index flood relationship in the
 415 considered region. Important heterogeneities should nevertheless be avoided
 416 through a careful selection of the regional data set based on the plot of
 417 the local sample means versus watershed areas (Gaume et al., 2010). This
 418 poses an additional constraint to the delineation of homogeneous regions if
 419 compared to the method of Hosking & Wallis, limiting the size of the regional
 420 gauged sets, but opening the possibility to value the ungauged extremes.
 421 This constraint should be kept in mind for a fair comparison of both methods.

422 **4. Application to a real-world example with ungauged extremes:** 423 **the Ardèche area in France**

424 *4.1. The Ardèche data set*

425 The Ardèche case study had already been selected by Gaume et al. (2010)
 426 for the initial tests of the regionalisation method involving ungauged ex-
 427 tremes. It has been again used here to illustrate the changes introduced
 428 by the new inference procedure including the calibration of the index flood
 429 relationship.

430 The Ardèche region is located in the southeast of Massif Central in south-
 431 ern France (Fig. 7). It is one of the areas in Europe exposed to the most
 432 frequent and severe flash floods (Gaume et al., 2009). The region includes six
 433 main rivers: the Ardèche River 2380 km^2 , Ouvèze River 120 km^2 , Eyrieux
 434 River 860 km^2 , Doux River 630 km^2 , Ay River 110 km^2 and Cance River
 435 380 km^2 . Five stream gauges provide accurate, continuous and long series
 436 of discharges in the region. The gauged dataset of annual maximum peak

437 discharges counts 168 records. The situation of the 5 gauging stations is
438 reported in figure ?? and the characteristics of corresponding samples are
439 presented in table 1.

440 Using the large set of documentary sources available in the region, un-
441 gauged largest peak discharge estimates over the last 50 years could be re-
442 trieved at 14 additional sites (Fig. 8). Similarly, information on the 50 years
443 preceding the gauging period could be found at 4 of the gauged sites. The
444 set of ungauged extremes is finally composed of 18 maximum values over 50
445 years, that is equivalent to 900 censored records, six times more than the
446 regional gauged data set (Table 2). Note that the selected set of ungauged
447 extremes has been reduced if compared to the set used by Gaume et al.
448 (2010). A minimum distance between sites on the same stream has for in-
449 stance been imposed to avoid redundancy (see appendix Appendix B for a
450 short discussion about the sampling of extremes).

451 The homogeneity of the growth curves within the region has been tested
452 using the Hosking & Wallis heterogeneity measure (Hosking and Wallis,
453 1997). The H_1 value based on the 5 gauged series is equal to 2.7, which
454 suggests that the region is possibly heterogeneous. Homogeneity has never-
455 theless been assumed to proceed with the computations. The influence of
456 regional heterogeneities in the growth curves on the inference results will be
457 studied in a future publication.

458 Provided the preceding assumption, this case study finally offers an inter-
459 esting opportunity to apply and compare the approaches described above, in
460 two different contexts: (i) regional dataset limited to gauged sites, and (ii)
461 availability of information on the largest discharge at ungauged sites.

4.2. Results and discussion

Figure 8 and table 3 summarize the results obtained for the Ardèche case study. They should be interpreted at the light of the conclusions drawn in the previous section. The presented results correspond to the two extreme gauged catchment surfaces in the region: 63 km^2 (Borne watershed at Puylaurens), and 2240 km^2 (close to the outlet of the entire Ardèche watershed at Saint Martin).

The two methods - Hosking & Wallis and the proposed approach - lead to close results when applied to the gauged data set (Table 3, cases a and b). The widths of the estimated credibility intervals are slightly larger in the case of the proposed approach, but these intervals are underestimated by about 20% in the reference approach according the previous results. The underestimation of the quantiles is not systematic when the reference approach is used as illustrated by table 3. The inference results are depending on the characteristics of the sample.

The main advantage of the proposed approach lies in the possibility to incorporate information on ungauged extremes in the analysis. In the Ardèche case, the ungauged extremes appear consistent with the gauge series and the GEV distribution (Fig. 8). Their incorporation in the statistical inference leads to a clear reduction of the credibility intervals that are divided by a factor of about 2. This confirms the possible great value of the proposed approach in regions for which information on extremes floods at ungauged sites is available. The assumption of homogeneity of the index flood relationship limits the number of gauged sites that can be included in the homogeneous region. But, the extremes at ungauged sites represent a potentially large ad-

ditional source of information (equivalent to 900 records in the present case), that may reduce significantly the uncertainties attached to estimated high return period quantiles.

Figure 9 compares the modelled relation between average discharges and watershed areas and the observed locally estimated mean discharges at the light of the proposed Lognormal model to reproduce regional variability in the index flood relationship. This illustrates the consistency of the proposed index flood relation with the observed data. The second smallest watershed, the Eyrieux, appears to have relatively smaller empirical mean discharge and could be possibly regarded as non-consistent with the other sites as far as the index flood relation is considered. This low mean value may nevertheless be attributed to sampling fluctuation but may likewise be explained by the fact that this site is located further North and less influenced by the Mediterranean climate. This last comment illustrates that any regional flood frequency method is based on homogeneity hypotheses that can hardly be verified. It is therefore certainly sound to test the robustness of the obtained inference results to different hypotheses and region delineations. In this perspective, the proposed approach enriches the available panel of methods and possible hypotheses.

Finally, the same Monte Carlo simulations as the ones used in section 3 were conducted with generated samples including 18 ungauged sites where only the highest discharge over 50 years is known to verify if the conclusions drawn on the real-world example can be generalized (Fig. 10.d). In fact, the results of an analysis conducted on one single data set may be attributed to sampling variability. On average, the credibility intervals appear

512 to be divided by two on a sample of 1000 simulated datasets when the "un-
 513 gauged" extremes are taken into account, which is the order of magnitude
 514 obtained for the Ardèche data set. 100-year quantile estimates based on
 515 the L-moments (standard method suggested by Hosking & Wallis) have also
 516 been computed for sake of verification (Fig. 10.a). L-moment and maximum
 517 likelihood based estimates appear to have similar performances on this data
 518 sample (figures 10.a and b). The L-moment estimates exhibit a slightly lower
 519 dispersion (L-moment are less sensitive to sampling variability) but a higher
 520 bias.

521 **5. Conclusions**

522 The approach initially proposed by Gaume et al. (2010), aiming at incor-
 523 porating information on ungauged extremes in regional flood frequency anal-
 524 yses, has been consolidated and thoroughly tested. The proposed method
 525 and algorithm, after some necessary adjustments, in particular of the formu-
 526 lation of the likelihood, appear to be able to estimate accurately quantiles
 527 and their uncertainty bounds on the basis of a regional sample combining
 528 continuous gauged series and ungauged censored data. The introduction of
 529 an index flood relationship and of an additional parameter to be calibrated,
 530 proved to be a stabilization factor rather than a factor of complexity and
 531 a source of uncertainty. The method shows a relative robustness to some
 532 regional variability in the index flood relation. The additional homogeneity
 533 constraint limits the possible extent of homogeneous regions if compared to
 534 the reference approach, but this is compensated by the additional informa-
 535 tion brought by the estimated ungauged extremes. In the case of the Ardèche

536 region the ungauged data set represents 900 censored recording station-years:
537 six times the amount of the available regional gauged series. Its incorpora-
538 tion in the regional flood frequency analysis led to a division by two of the
539 estimated 100-year quantile 90% credibility intervals.

540 Uncertainties on the extreme peak discharge estimates were not consid-
541 ered herein and the ungauged extreme dataset that could be retrieved in the
542 Ardèche region may appear as particularly rich if compared to what could
543 reasonably be obtained in other regions. Previous works, conducted on the
544 usefulness of historic records in flood frequency studies, indicated that the
545 accuracy level of the estimated extreme discharges and the number of ex-
546 tremes for which an estimated discharge is available, have little influence on
547 the inference results when the historic period of record is fixed Payrastré
548 et al. (2011). It remains to be confirmed that the same conclusion holds for
549 ungauged extremes when included in regional flood frequency analyses.

550 In the same line of thoughts, the possible variability in the growth curves
551 in a region considered as homogeneous has not been considered yet. Its
552 influence could be evaluated in the same manner as the influence of variability
553 in the index flood relation has been considered in section 3. It may modify
554 partly the judgment on the relative efficiency of the various tested regional
555 flood frequency analysis methods.

556 Finally, it is important to keep the conclusion of section 4 in mind. All
557 the regional flood frequency methods are based on homogeneity hypotheses
558 that can hardly be verified. It is therefore certainly sound to test the ro-
559 bustness of the obtained inference results to different hypotheses and region
560 delineations: to combine various methods. In this perspective, the proposed

561 approach does not replace but enriches the available panel of methods and
562 possible hypotheses. It may lead, as shown herein, to quantile estimates that
563 are theoretically equally or even more accurate than the estimates obtained
564 with standard methods. Its main constraint may also be considered as an
565 advantage: the two necessary homogeneity hypotheses limit the extent of
566 homogeneous regions that may be, on the other hand, hydrologically more
567 consistent and less questionable than in standard regional flood frequency
568 analyses.

569 **Acknowledgements**

570 This work has benefited from the support of the French Ministry of foreign
571 affairs.

Appendix A. Method for the evaluation of the underestimation factor for the computed credibility intervals

In many cases, the proposed reliability test for the computed credibility intervals presented in section 3.1 indicates that the intervals are too narrow (Fig. 5). This is due the fact that some sources of uncertainties are not accounted for in the bayesian MCMC framework used to compute these intervals. The following procedure is proposed to evaluate the underestimation factor m of the computed intervals. Based on the hypothesis that the shapes of the posterior distribution of the discharge quantiles $\hat{Q}_i(0.99)$ are correct and that only their dispersion is underestimated, the procedure to evaluate the underestimation factor m is the following.

The distance between the real quantile value $Q_i(0.99)$ and the median value of the computed posterior distribution $Me(\hat{Q}_i(0.99))$ is artificially reduced by the factor m : the probability of non-exceedance $\hat{F}_m(Q_i(0.99))$ (eq. A.1) is computed rather than the probability $\hat{F}(Q_i(0.99))$ based on the MCMC results for each simulated sample.

$$\hat{F}_m(Q_i(0.99)) = \hat{F}\left(Me(\hat{Q}_i(0.99)) + 1/m \left[Q_i(0.99) - Me(\hat{Q}_i(0.99))\right]\right) \quad (\text{A.1})$$

The factor m leading to the distribution of $\hat{F}_m(Q_i(0.99))$ closest to a uniform distribution for the 1000 simulated samples, gives an estimate of the underestimation factor of the credibility intervals.

Appendix B. Sampling ungauged extremes in a region

The regional samples of ungauged extremes must be carefully built to avoid introducing bias into the statistical analysis. At least, the extreme data collation method must seek to meet three conditions: (i) be based on an explicit sampling strategy that can be translated into the likelihood formulation, (ii) be comprehensive for each selected location, (iii) limit possible dependencies between records and samples. The next three sections explain how these three conditions were accounted for to build the Ardèche regional sample of ungauged extreme floods.

Appendix B.1. Explicit and systematic sampling method for regional extremes

Isolated extreme discharge values are often reported but can not be directly integrated in regional flood frequency analyses if their representativeness is unknown. Over which period of time can it be asserted that a larger

flood event did not occur at the same site ? And, even more crucial, these values have been reported because they appear as extreme, i.e. this means implicitly that they seem large if compared the known flood events observed in neighbouring comparable catchments. Over how many years and for how many similar catchments (station) could it be finally asserted that the considered extreme discharge value - or rescaled discharge value - has not been exceeded ? Large uncertainties concerning the record length (station.years) corresponding to each extreme may drastically reduce their added value in regional flood frequency analyses as illustrated by the Slovakian test case in Gaume et al. (2010). An inventory of regional extremes should therefore be based, if possible, on a systematic data mining procedure. The various sites in a region where information on extremes may be available should be determined a priori: typically significant towns along rivers where large floods may have induced damages and may have been reported. An inventory of past large floods should be realized at each of these sites based on local memory and archives and it must at least be verified that some given threshold values have or have not been exceeded at the considered sites in the recent past: non-exceedance is a statistically valuable information.

Appendix B.2. Comprehensiveness

Collation of extreme events over a given period of time provide censored data sets: only discharges exceeding a given threshold, that depends on the site and the considered period, are documented. Flood events exceeding the threshold should be exhaustively listed at each site over the considered past period for a correct statistical inference. To be on the safe side, it is generally preferable to limit the considered past period of time and to raise the thresh-

old value to ensure exhaustiveness, even if this limits the information content of the inventories of large or extreme floods. For this reason, the collation period has been limited in the Ardèche to the last 50 years (typical depth of local memory). Likewise, the inventory has been limited to the largest known flood event at each ungauged site. This last choice facilitates also the computation of the plotting positions for the ungauged extremes, which can become tricky: the return periods of the ungauged extremes reported on the cumulated distribution plots are estimated through Monte Carlo simulations (Gaume et al., 2010).

Appendix B.3. Independency

Likelihood formulations and the computation of the credibility intervals are based on the assumption of independence between the various recorded maximum annual peak discharges. But, extreme peak discharges induced by the same rainfall event sometimes occur at the same date at nearby ungauged sites, suggesting at least weak dependencies in the inventory of ungauged extremes. To prevent any effect of these dependencies on the results and especially to limit the risk of underestimation of the credibility intervals due to an implicate overrating of the information content of the ungauged extreme inventory neglecting dependencies, the data set had to be refined. When the largest reported flood occurred at the same date at close ungauged sites, one randomly selected site has been kept in the final ungauged data set. Keeping the largest value would have introduced a sampling bias. The refinement of the Ardèche ungauged extreme set led to eliminate more than half of the records initially used by Gaume et al. (2010), hence resulting in a slight increase only of the computed credibility intervals.

- Castellarin, A., 2005. Probabilistic behavior of a regional envelope curve. *Water Resources Research* 41, W06018.
- Castellarin, A., Vogel, R., Matalas, N., 2007. Probabilistic envelope curves for design flood estimation at ungauged sites. *Water Resources Research* 43, W04406.
- Charles, N., Stedinger, J., 1999. Development of regional regression relationship with censored data. *Water Resources Research* 35(4), 775–784.
- Cohn, T., Stedinger, J., 1987. Use of historical information in a maximum likelihood framework. *Journal of Hydrology* 96 (1-4), 215–223.
- Coles, S., Powell, E., 1996. Bayesian methods in extreme value modelling. A review and new developments. *Int. Stat. Rev* 43, 1–48.
- Coles, S., Tawn, J., 1996. A bayesian analysis of extreme rainfall data. *Appl. Stat* 45, 463–478.
- Cowles, M., Carlin, B., 1996. Markov Chain Monte Carlo convergence diagnostics: a comparative review. *Statist-Assoc* 91, 883–904.
- Dalrymple, T., 1960. Flood frequency analyses. *Water Supply: Geological Survey, Reston, Virginia, USA* , 1543–A.
- Gary, D., Stedinger, J., 1987. Regional regression of flood characteristics employing historical information. *Journal of Hydrology* 96, 255–264.
- Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaskovicova, L., Blöschl, G., Borga, M., Dumitrescu, A., Daliakopoulos, I., Garcia, J., Irimescu, A., Kohnova, S., Koutroulis, A., Marchi, L.,

- Matreata, S., Medina, V., Preciso, E., Sempere-Torres, D., Stancalie, G., Szolgay, J., Tsanis, J., Velasco, D., Viglione, A., 2009. A compilation of data on European flash floods. *Journal of Hydrology* 367, 70–80.
- Gaume, E., Gaal, L., Viglione, A., Szolgay, J., Kohnova, S., Blöschl, G., 2010. Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. *Journal of Hydrology* doi:10.1016/j.jhydrol.2010.01.008.
- Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* 7, 457–511.
- Haddad, K., Rahman, A., 2011. Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework: Quantile Regression vs. Parameter Regression Technique. *Journal of Hydrologic Engineering* 920, doi.org/10.1061/(ASCE)HE.1943-5584.0000395, 1–6.
- Heinz, D., Stedinger, J., 1998. Using regional regression within index flood procedures and an empirical Bayesian estimator. *Journal of Hydrology* 210, 128–145.
- Hosking, J., Wallis, J., 1986a. Paleoflood hydrology and flood frequency analysis. *Water Resources Research* 22, 543–550.
- Hosking, J., Wallis, J., 1986b. The value of historical data in flood frequency analysis. *Water Resources Research* 22, 1606–1612.
- Hosking, J., Wallis, J., 1997. Regional frequency analysis: An approach Based on L-Moments. Cambridge University Press, London, UK.

- Kjeldsen, T., Jones, D., 2009. A formal statistical model for pooled analysis of extreme. *Hydrology Research* 40(5), 465–480.
- Kjeldsen, T., Smithers, J., Schulze, R., 2002. Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method. *Journal of Hydrology* 255, 194–211.
- Lu, L., Stedinger, J., 1992. Sampling variance of normalized GEV/PWM quantile estimators and regional homogeneity test. *Journal of Hydrology* 138, 223–245.
- Merz, R., Blöschl, G., 2003. A process typology of regional floods. *Water Resources Research* 39(12), 1340.
- Minghui, J., Stedinger, J., 1989. Flood frequency analysis with regional and historical information. *Water Resources Research* 25, 925–936.
- Natural Environment Research Council, 2000. Flood estimation Handbook. NERC - Center for Ecology & Hydrology. Wallingford, UK.
- Neppel, L., Renard, B., Lang, M., Ayral, P., Coeur, D., Gaume, E., Jacob, N., Payrastre, O., Pobanz, K., Vinet, F., 2010. Flood frequency analysis using historical data: accounting for random and systematic errors. *Hydrological Sciences Journal* 55, 192–208.
- Norbiato, D., Borga, M., Sangati, M., Zanon, F., 2007. Regional frequency analysis of extreme precipitation in the eastern italian Alps and the August 29, 2003 flash flood. *Journal of Hydrology* 345, 149–166.

- Ouarda, T., Girard, C., Cavadias, G., Bobee, B., 2001. Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology* 254, 157–173.
- Payrastre, O., Gaume, E., Andrieu, H., 2005. Use of historical data to assess the occurrence of floods in small watersheds in the French Mediterranean area. *Advances in Geosciences* 2, 1–8.
- Payrastre, O., Gaume, E., Andrieu, H., 2011. Usefulness of historical information for flood frequency analyses: Developments based on a case study. *Water Resources Research* 47, W08511.
- Reis, D., Stedinger, J., Martins, E., 2005. Bayesian GLS regression with application to LP3 regional skew estimation. *Water Resources Research* 41, W10419.
- Renard, B., Garreta, V., Lang, M., 2006. An application of Bayesian analysis and Markov Chain Monte Carlo methods to the estimation of a regional trend in annual maxima. *Water Resources Research* 42(12), W12422.
- Ribatet, M., Sauquet, E., Gresillon, J., Ouarda, T., 2007. Usefulness of the reversible jump Markov Chain Monte Carlo in regional flood frequency analysis. *Water Resources Research* 43, W08403.
- Robert, C., Casella, G., 2004. Monte Carlo statistical methods. Springer, New York. 2nd ed.
- Salaheddine, A., Anne-Catherine, F., Bernard, B., 2006. Comparison of methodologies to assess the convergence of Markov Chain Monte Carlo methods. *Computational statistics and data analysis* 50, 2685–2701.

- Seidou, O., Ouarda, T., Barbet, M., Bruneau, P., Bobe, B., 2006. A parametric Bayesian combination of local and regional information in flood frequency analysis. *Water Resources Research* 42, W11408.
- Sheffer, N.A., Enzel, Y., Benito, G., Grodek, T., Poart, N., Lang, M., Naulet, R., Coeur, D., 2003. Paleofloods and historical floods of the Ardche river, France. *Water Resources Research* 39(12), 1379.
- Stedinger, J., Cohn, T., 1986. Flood frequency analysis with historical and paleoflood information. *Water Resources Research* 22, 785–793.
- Stedinger, J., Lu, L., 1995. Appraisal of regional and index flood quantile estimators. *Stochastic Hydrology and Hydraulics* 9, 49–75.
- Sutcliffe, J., 1987. The use of historical records in flood frequency analysis. *Journal of Hydrology* 96, 159–171.
- Tanner, M., 1996. Tools for statistical inference: Methods for the Exploration of Posterior, Distributions and Likelihood Functions. Springer, Evanston, USA.
- Viglione, A., 2013. Non-supervised regional frequency analysis R library. <http://cran.r-project.org/web/packages/nsRFA/>.
- Viglione, A., Merz, R., Salinas, J.L., Blöschl, G., 2013. Flood frequency hydrology: 3.A Bayesian analysis. *Water Resources Research* 49, WR010782, doi:10.1029/2011WR010782.
- Wallis, J.R., Schaefer, M., Barker, B., Taylor, G., 2007. Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-

hour durations for Washington State. *Hydrology and Earth System Sciences* (11), 415–442.

List of Tables

1	The sample size and the surface of the gauging stations in the regional Ardèche.	38
2	Detail of the extreme floods reported to be the highest ones in the last 50 years (or last 50 years before the gauged period)	44
3	Comparison the credibility intervals (CI) at St. Laurent and St.Martin gauging stations to three cases corresponding to return periods $T=100$ years.	45

List of Figures

1	Posterior distributions for the parameters α and β . Calibration of a regional growth curve and an index flood relation, based on a simulated sample of 1000 records: $\alpha = 2.24$, $\xi = 3.34$, $k = -0.16$ and $\beta = 0.76$. (a) totally biased result with the standard likelihood expression 3 and (b) correct result with likelihood expression 7.	37
2	Gelman & Rubin’s convergence diagnosis with the regional data sets of the Ardèche: a) reference approach; and b) proposed approach.	38

3	Posterior distributions of $\hat{F}(Q_i(0.99))$: (a) perfect credibility intervals, (b) tendency to over-estimate the quantile, (c) too narrow estimated credibility intervals, (d) too large estimated credibility intervals.	39
4	Estimates of the 100-year quantile through Bayesian MCMC regional flood frequency inferences conducted on 1000 different samples: (a) Method of Hosking & Wallis and (b) proposed method. Box plots of the 1000 maximum likelihood and 5% and 95% credibility bounds for each considered watershed area. All the values have been divided by the real 100-year quantiles $\hat{Q}_i(0.99)$	40
5	Distributions of the values of $\hat{F}(Q_i(0.99))$ estimated for the 1000 simulated samples: (a,b) Method of Hosking & Wallis, (c,d) Hosking & Wallis with correction factor m for the posterior distributions and (e,f) proposed method.	41
6	Results of Bayesian MCMC regional flood frequency inferences conducted on 1000 samples with the proposed method when a regional variability in the index flood relation is introduced: (a) $\delta = 0.1$, (b) $\delta = 0.3$	42
7	The Ardèche region, location of the 5 gauged sites (triangles) and the 18 sites where ungauged extremes were retrieved (dots).	43

8	Results of the Bayesian MCMC regional flood frequency inferences conducted on the Ardèche dataset. Results provided for the Ardèche at Saint Martin (2240 km^2): (a) Method of Hosking & Wallis and (b) proposed approach including ungauged extremes. Statistical distributions corresponding to the maximum likelihood (continuous line) and 5% and 95% credibility bounds (dotted lines). Gauged values (crosses) and ungauged extremes (dots).	46
9	Calibrated relation between average discharges and watershed areas, sample mean discharges and variation ranges corresponding to the tested lognormal random models.	47
10	Maximum likelihood estimate and 5% and 95% bounds of the computed credibility intervals for 4 tested methods on 1000 simulated samples having the same characteristics as the Ardèche data sample (St Martin case): (a) method of Hosking and Wallis with L-moments, (b) method of Hosking and Wallis used in this paper, (c) proposed method applied on gauged series only, (d) proposed method with 18 ungauged extremes. .	48

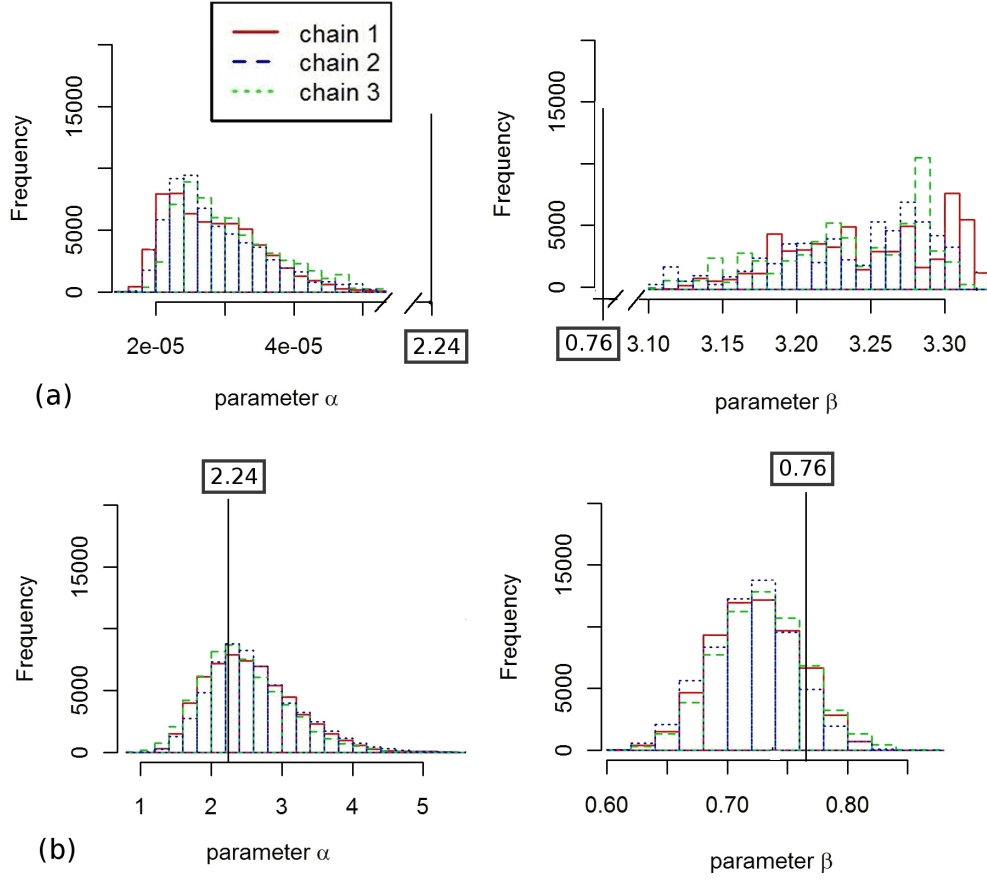


Figure 1: Posterior distributions for the parameters α and β . Calibration of a regional growth curve and an index flood relation, based on a simulated sample of 1000 records: $\alpha = 2.24$, $\xi = 3.34$, $k = -0.16$ and $\beta = 0.76$. (a) totally biased result with the standard likelihood expression 3 and (b) correct result with likelihood expression 7.

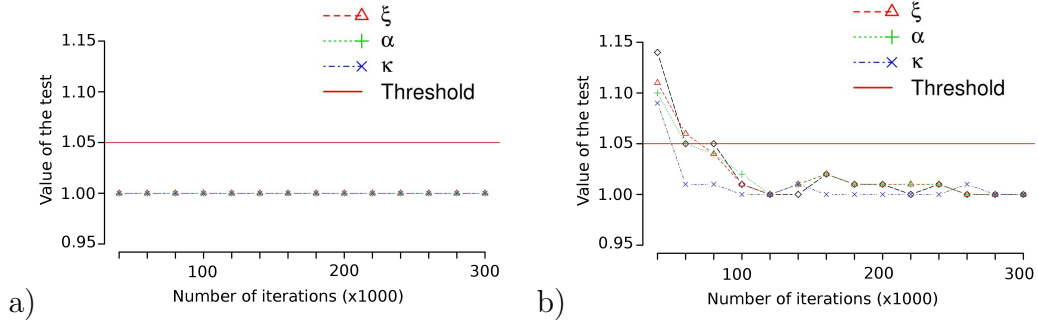


Figure 2: Gelman & Rubin's convergence diagnosis with the regional data sets of the Ardèche: a) reference approach; and b) proposed approach.

Table 1: The sample size and the surface of the gauging stations in the regional Ardèche.

Station	River name	Period of data	N(year)	S (km^2)
Beauvene	Eyrieux	1969-1998	30	392
Vogüe	Ardèche	1966-2003	38	636
Chambonas	Chassezac	1971-1997	27	507
St.Laurent	Borne	1968-1997	30	63
St.Martin	Ardèche	1963-2005	43	2240

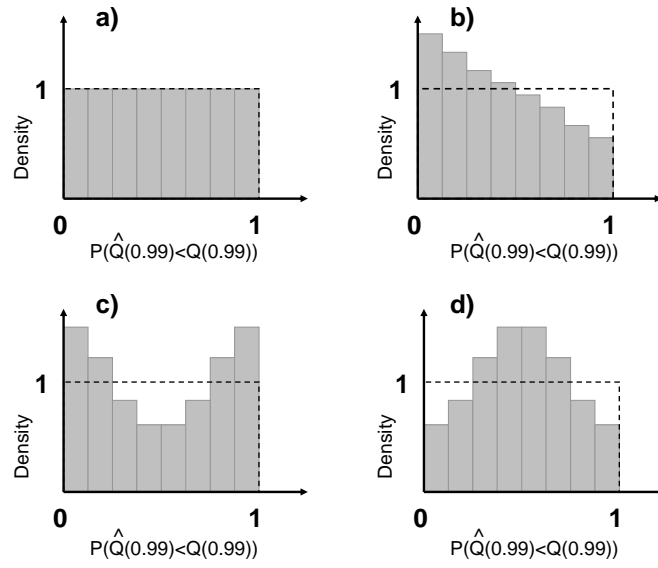


Figure 3: Posterior distributions of $\hat{F}(Q_i(0.99))$: (a) perfect credibility intervals, (b) tendency to over-estimate the quantile, (c) too narrow estimated credibility intervals, (d) too large estimated credibility intervals.

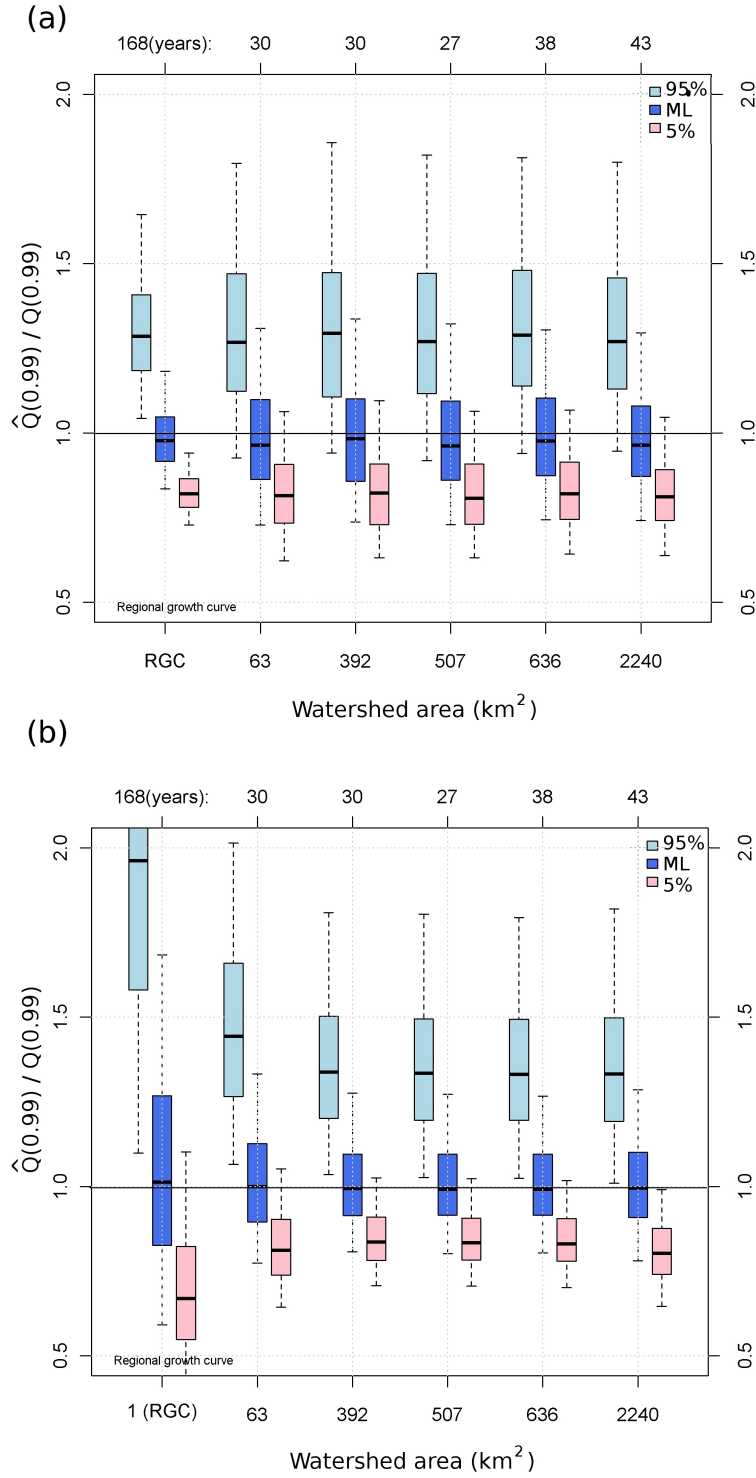


Figure 4: Estimates of the 100-year quantile through Bayesian MCMC regional flood frequency inferences conducted on 1000 different samples: (a) Method of Hosking & Wallis and (b) proposed method. Box plots of the 1000 maximum likelihood and 5% and 95% credibility bounds for each considered watershed area. All the values have been divided

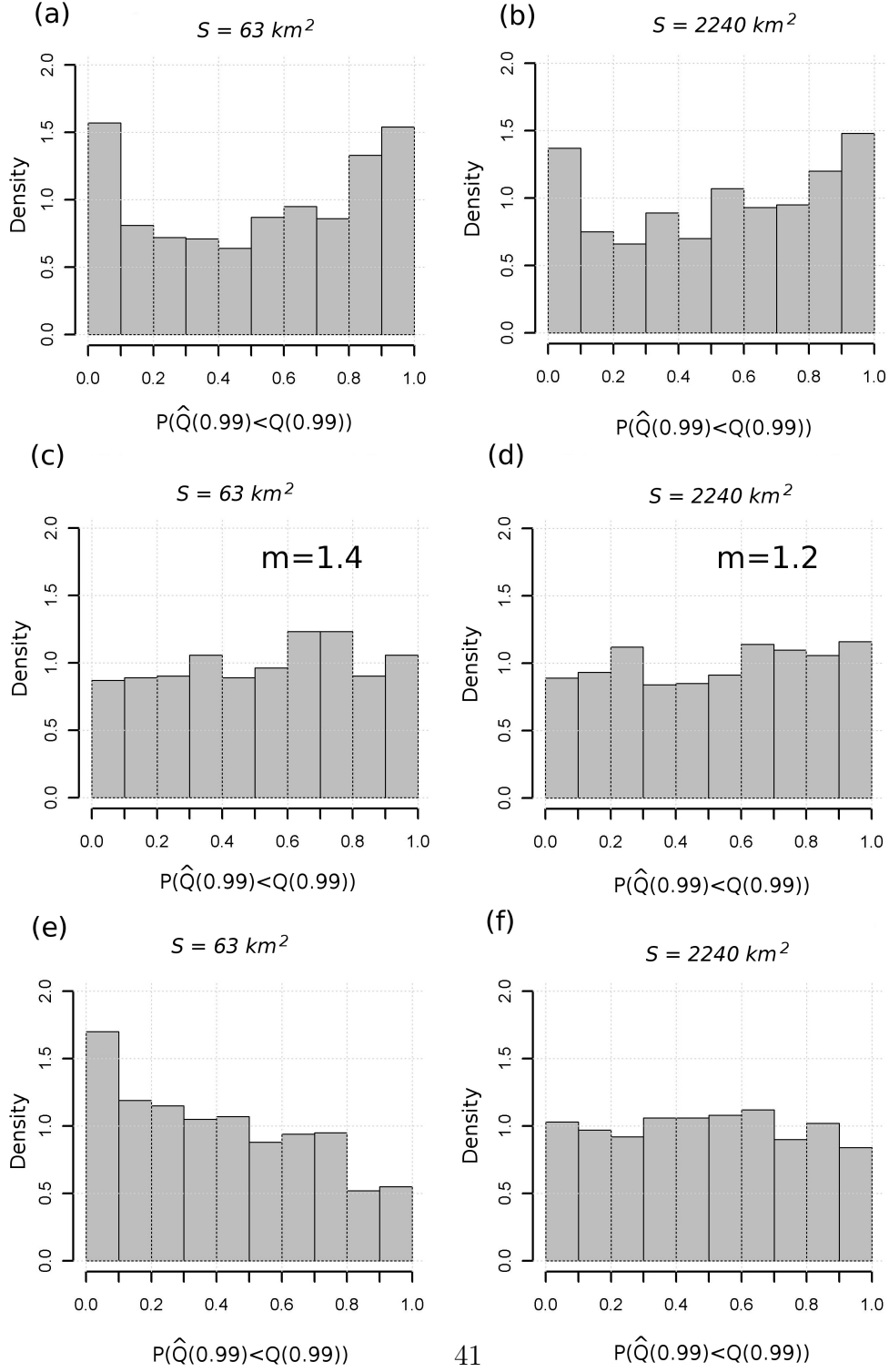


Figure 5: Distributions of the values of $\hat{F}(Q_i(0.99))$ estimated for the 1000 simulated samples: (a,b) Method of Hosking & Wallis, (c,d) Hosking & Wallis with correction factor m for the posterior distributions and (e,f) proposed method.

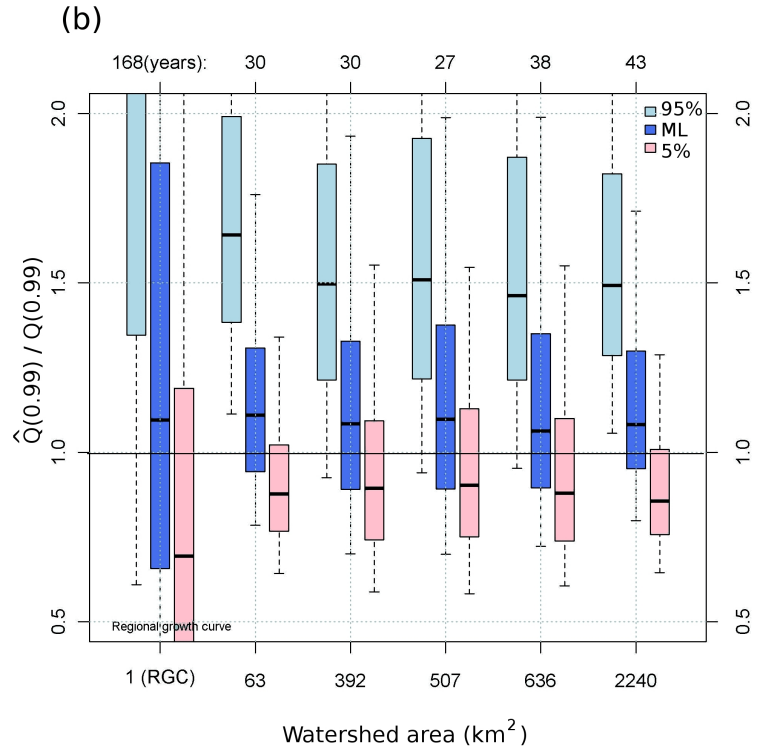
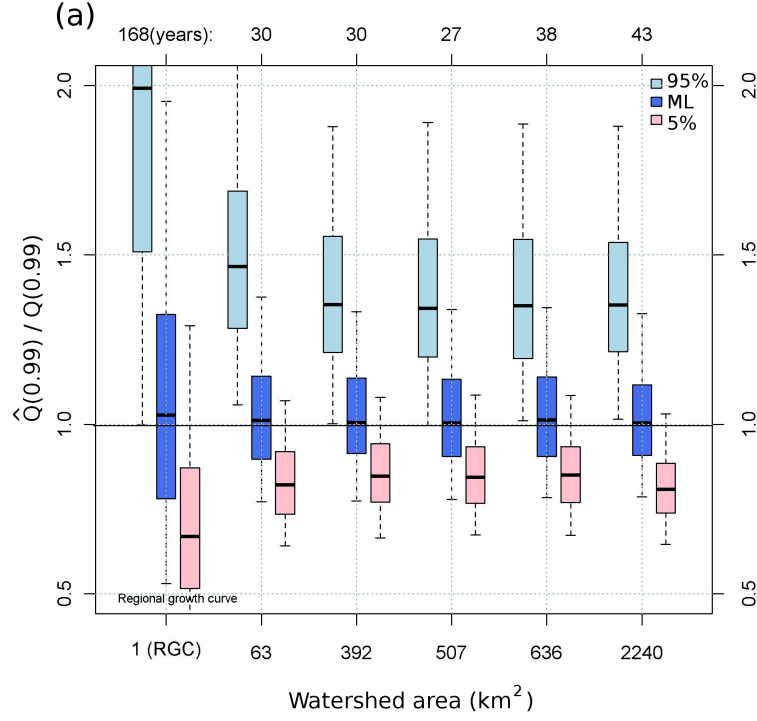


Figure 6: Results of Bayesian MCMC regional flood frequency inferences conducted on 1000 samples with the proposed method when a regional variability in the index flood relation is introduced: (a) $\delta = 0.1$, (b) $\delta = 0.3$.

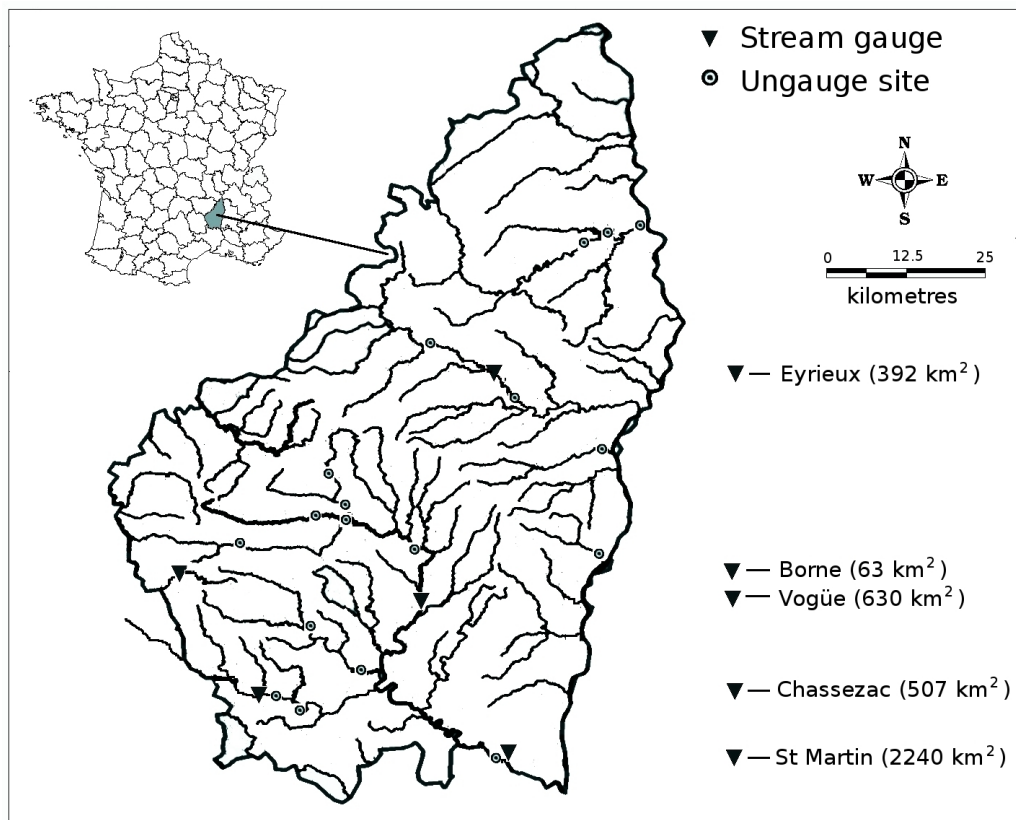


Figure 7: The Ardèche region, location of the 5 gauged sites (triangles) and the 18 sites where ungauged extremes were retrieved (dots).

Table 2: Detail of the extreme floods reported to be the highest ones in the last 50 years (or last 50 years before the gauged period)

Location	River name	Date	Q (m^3/s)	S (km^2)
Pont de Rolandy	Ardèche	22/09/1992	1150	150
Aubenas	Ardèche	26/09/1992	2200	480
Sauze St.Martin	Ardèche	30/09/1958	4500	2240
Rosieres	Beaune	04/10/1958	1820	210
Joyeuse	Beaune	30/09/1958	1000	100
Chambon	Borne	30/09/1958	100	11
Vans	Boudaric	03/11/1989	130	6
Burzet	Bourges	22/09/1992	350	47
Chambonas	Chassezac	21/09/1980	3360	510
Dorne	Dorne	03/08/1963	630	78
Lamastre	Doux	03/08/1963	970	242
Pont de Cesar	Doux	03/08/1963	1500	635
Barrage des Collanges	Eyrieux	03/08/1963	1685	343
Meyras	Fontoliere	22/09/1992	900	130
Meysee	Lavezon	30/09/1960	500	56
Rieut ord	Loire	01/09/1992	444	62
Pouzin	Ouveze	10/08/1967	700	140
Saliousse	Saliousse	21/09/1980	300	61

Table 3: Comparison the credibility intervals (CI) at St. Laurent and St.Martin gauging stations to three cases corresponding to return periods T=100 years.

Surface (km^2)	N_{cont} (years)	N_{hist} (years)	Case	$\hat{Q}_i^5(0.99)$ (m^3/s)	$\hat{Q}_i^{ML}(0.99)$ (m^3/s)	$\hat{Q}_i^{95}(0.99)$ (m^3/s)
63	168	0	(a)	423.8	520.9	712.5
	168	0	(b)	382.3	479.5	682.1
	166	900	(c)	442.8	500.4	574.2
2240	168	0	(a)	5538.6	6806.8	9310.8
	168	0	(b)	5892.4	7602.2	10747.6
	166	900	(c)	5477.3	6560.0	7847.1

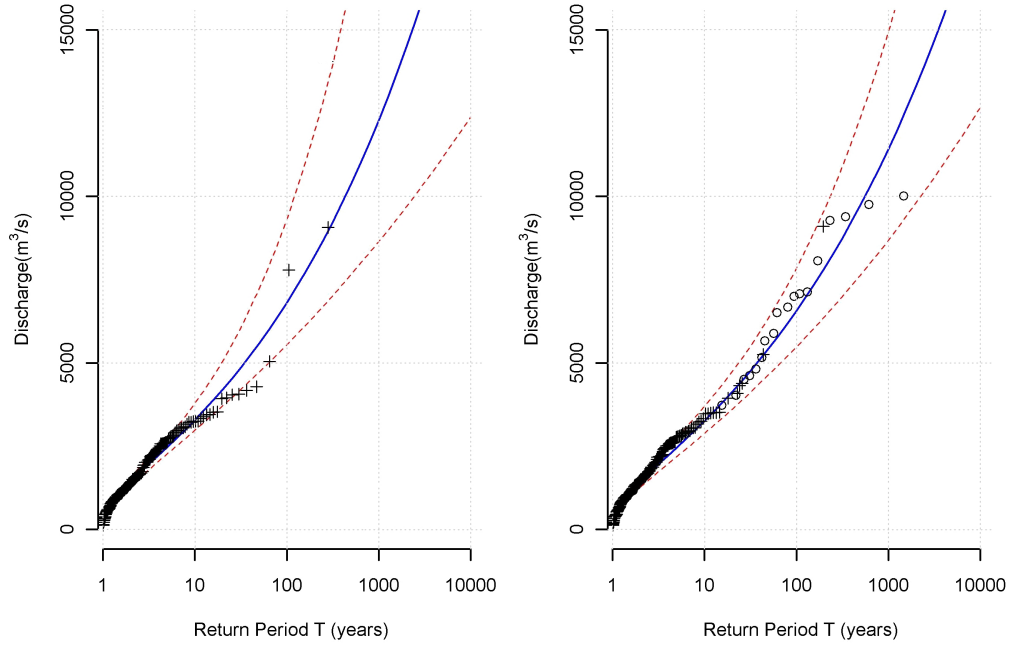


Figure 8: Results of the Bayesian MCMC regional flood frequency inferences conducted on the Ardèche dataset. Results provided for the Ardèche at Saint Martin (2240 km^2): (a) Method of Hosking & Wallis and (b) proposed approach including ungauged extremes. Statistical distributions corresponding to the maximum likelihood (continuous line) and 5% and 95% credibility bounds (dotted lines). Gauged values (crosses) and ungauged extremes (dots).

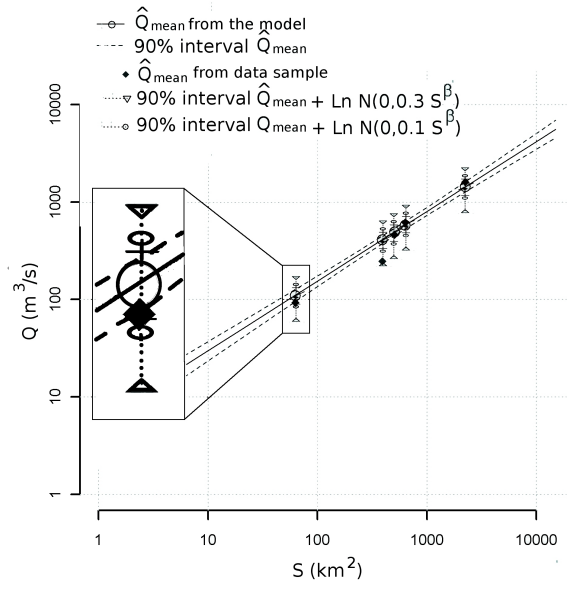


Figure 9: Calibrated relation between average discharges and watershed areas, sample mean discharges and variation ranges corresponding to the tested lognormal random models.

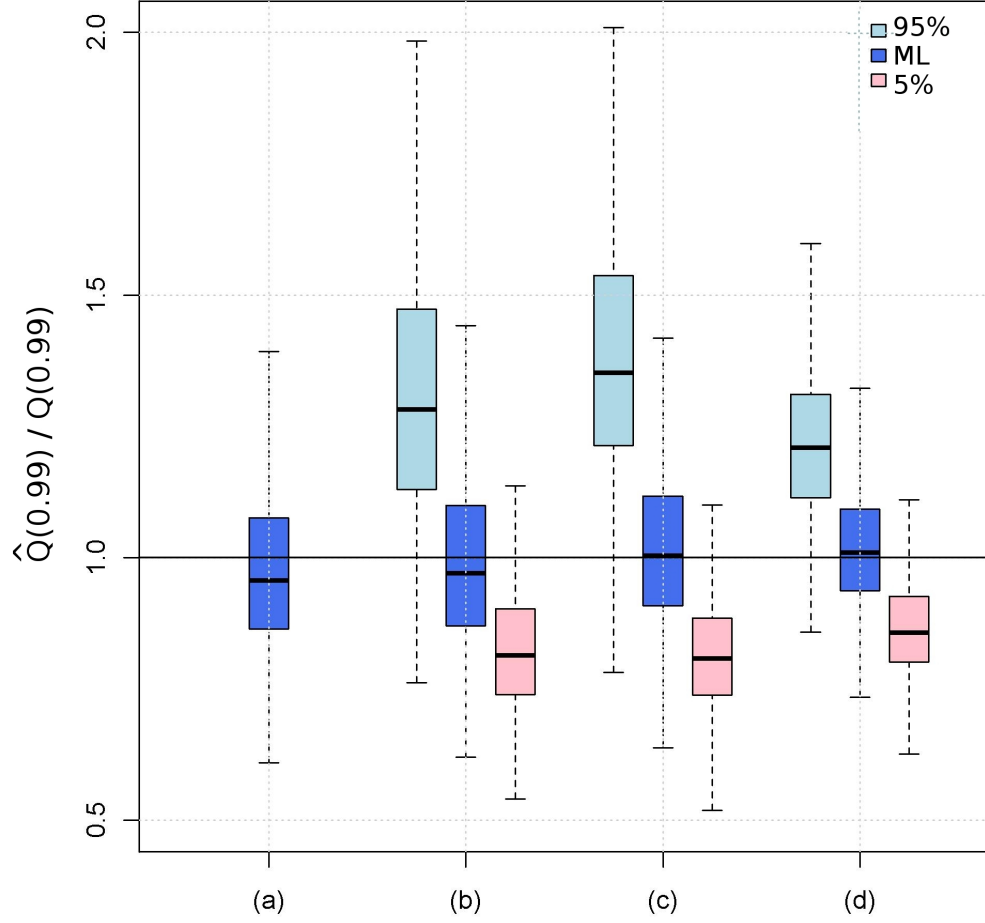


Figure 10: Maximum likelihood estimate and 5% and 95% bounds of the computed credibility intervals for 4 tested methods on 1000 simulated samples having the same characteristics as the Ardèche data sample (St Martin case): (a) method of Hosking and Wallis with L-moments, (b) method of Hosking and Wallis used in this paper, (c) proposed method applied on gauged series only, (d) proposed method with 18 ungauged extremes.