UNIVERSITÉ DE NANTES
FACULTÉ DES SCIENCES ET DES TECHNIQUES

———

Sciences Pour l'Ingénieur, Géosciences, Architecture

Année 2012

N° attribué par la bibliothèque

Amélioration des approches Bayésiennes MCMC pour l'analyse régionale des crues.

———

THÈSE DE DOCTORAT
Discipline : Génie Civil
Spécialité : Hydrologie

Présentée
et soutenue publiquement par

Chi Cong NGUYEN

Le 13 juillet 2012, devant le jury ci-dessous

Président :	Christophe BOUVIER	LHSM, Montpellier
Rapporteurs :	Michel LANG	IRSTEA, Lyon
	Christophe BOUVIER	LHSM, Montpellier
Examinateurs :	Philippe NAVEAU	LSCE, Versailles
	Pietro BERNARDARA	EDF - R&D / LNHE, Chatou
	Remy GARCON	EDF - DTG, Grenoble

Directeur de thèse : Franck SCHOEFS GeM - Université de Nantes, Nantes

Co- directeur: Eric GAUME IFSTTAR, Nantes

Encadrant : Olivier PAYRASTRE IFSTTAR, Nantes

UNIVERSITY OF NANTES
FACULTY OF SCIENCE AND TECHNOLOGY

Sciences Pour l'Ingénieur, Géosciences, Architecture

Year 2012

N° assigned by the library

Improvement of Bayesian MCMC approaches for regional flood frequency analyses

DOCTORAL THESIS
Discipline: Civil Engineering
Specialty: Hydrology

Presented
and publicly supported by

Chi Cong NGUYEN

On July 13, 2012, before the jury below

President:	Christophe BOUVIER	LHSM, Montpellier
Reviewers:	Michel LANG	IRSTEA, Lyon
	Christophe BOUVIER	LHSM, Montpellier
Examiners:	Philippe NAVEAU	LSCE, Versailles
	Pietro BERNARDARA	EDF - R&D / LNHE, Chatou
	Remy GARCON	EDF - DTG, Grenoble

Thesis advisor: Franck SCHOEFS GeM - University of Nantes, Nantes

Co-advisor: Eric GAUME IFSTTAR, Nantes

Supervisor: Olivier PAYRASTRE IFSTTAR, Nantes

A gift for my mother. . .

Remerciements

Après trois ans de recherche à l'IFSTTAR, j'ai appris une chose : obtenir des résultats comme aujourd'hui n'aurait pas été possible sans le soutien et l'encouragement d'un grand nombre de personnes. Ainsi, je tiens à exprimer ma gratitude à un certain nombre de personnes qui au fil des ans ont contribué de diverses façons à la réalisation de cette thèse.

Tout d'abord, je tiens à exprimer mes remerciements chaleureux et sincères au Professeur François RESCHE, Ancien président de l'Université de Nantes, au Professeur Yves LECOINTE, Président de l'Université de Nantes, et au Professeur Van Ga BUI, ancien président de l'Université de Danang qui m'a présenté à l'IFSTTAR.

Ensuite, je suis profondément reconnaissant envers mon directeur de thèse, le Professeur Franck SCHOEFS et le Professeur Eric GAUME qui m'ont dirigé dans cette étude. Leur aide scientifique et leur conseils ont été d'une grande valeur dans cette étude. Leurs idéaux et concepts ont eu une influence remarquable sur toute ma carrière dans le domaine de l'hydrologie. Je tiens à exprimer ma profonde gratitude à mon encadrant Dr. Olivier PAYRASTRE pour sa contribution substantielle à ma thèse. Cette thèse n'aurait pas été possible sans sa direction. Il m'a aussi appris à rédiger des articles, à mieux m'organiser, à présenter mes pensées de manière plus claire dans ma thèse.

Par ailleurs, je remercie également les autres membres de mon comité de thèse : Mr Hervé ANDRIEU et Mr Benjamin RENARD qui ont apporté des commentaires, des supports et des conseils dans ma thèse. En outre, je tiens à remercier les rapporteurs, Mr Michel LANG et Mr Christophe BOUVIER, pour leur examen détaillé, la critique constructive et d'excellents conseils lors de la préparation de cette thèse. Au cours de ce travail au département de Géotechnique Eau et Risques, j'ai collaboré avec de nombreux collègues pour lesquels j'ai une grande estime, et je tiens à adresser mes plus vifs remerciements à tous ceux qui m'ont aidé dans mon travail.

Enfin, je dois des remerciements affectueux à ma femme et ma fille. Ils ont perdu beaucoup à cause de mes recherches à l'étranger. Sans leurs encouragements et leur compréhension, il aurait été impossible pour moi de finir ce travail. A côté, ma gratitude toute spéciale est due à mes parents et mon frère de soutien.

Acknowledgements

After three years of research at the Institut Français des Sciences et Techniques des Transports, de l'Aménagement et des Réseaux (IFSTTAR), I have learned one thing. I would not have achieved results of today, without the support and encouragement of a lot of people. Thus I would like to express my gratitude to a number of people who over the years have contributed in various ways to the completion of this thesis.

First, I wish to express my warm and sincere thanks to Professor François RESCHE, former president of University of Nantes, Professor Yves LECOINTE, president of University of Nantes, Professor Van Ga BUI, former president of University of Danang who introduced me to the IFSTTAR.

Next, I am deeply grateful to my advisors, Professor Franck SCHOEFS and Professor Eric GAUME who directed me in this study. Their kind support and guidance have been of great value in this study. Their ideals and concepts have had a remarkable influence on my entire career in the field of hydrology. I would like to express my deepest gratitude to my supervisor Dr. Olivier PAYRASTRE for his substantial contribution to my thesis. This thesis would not have been possible without his guidance. He also taught me how to write papers, how to organize and present my thoughts clearly in my thesis.

Furthermore, I also thank the other members of my thesis committee : Mr Hervé ANDRIEU and Mr Benjamin RENARD who provided comments, supports and advices in my thesis. In addition, I would like to thank the reviewers, Mr Michel LANG and Mr Christophe BOUVIER, for their detailed review, constructive criticism and excellent advice during the preparation of this thesis. During this work at the department of Geotechnics Water and Risks, I have collaborated with many colleagues for whom I have great regard, and I wish to extend my warmest thanks to all those who have helped me in my work.

Finally, I owe my loving thanks to my wife and my daughter. They have lost a lot due to my research abroad. Without their encouragement and understanding it would have been impossible for me to finish this work. Besides, my special gratitude is due to my parents and my brother support.

Contents

List of Figures

List of Tables

Notations

B_K	Bias of L_K^R
c	Coefficient
CI	Credibility intervals
$\mathbf{D}$	Regional sample
$f(Q)$	Density function
$F(Q)$	Cumulative probability function
$\hat{F}_{Q_i^{(100)}}$	Posterior distributions estimated of the 100-year
$\hat{F}$	Distribution estimated
ℓ	Likelihood function
$\ell(\mathbf{D}/\theta)$	Likelihood of the data set $\mathbf{D}$ given the parameter vector θ
ML	Maximum likelihood
N_{sim}	The number of simulation
n_i	Length of serie Q_i
$P(Q)$	Probability of Q
$p(\theta/\mathbf{D})$	Probability density of the parameter vector θ given the data set $\mathbf{D}$
$p(\theta)$	Prior distribution of θ
$p(\mathbf{D})$	Probability of the sample $\mathbf{D}$
Q_{ij}	Annual maximal discharges recorded j at gauged sites i (m^3/s)
Q_k	Discharge of ungauged extremes sites k (m^3/s)
Q^U	Upper discharge estimated (m^3/s)
Q^L	Lower discharge estimated (m^3/s)
$\hat{Q}$	Discharge estimated (m^3/s)
$\hat{Q}_{i,ML}^{(100)}$	Discharge estimated of the 100-year (m^3/s)
$\hat{Q}_{i,5}^{(100)}$	Discharge estimated of the 100-year corresponding 5% credibility intervals (m^3/s)
$\hat{Q}_{i,95}^{(100)}$	Discharge estimated of the 100-year corresponding 95% credibility intervals (m^3/s)
$\hat{Q}_{i,disp}^{(T)}$	Discharge estimated of the T-year corresponding dispersed distribution $\hat{F}_{Q_i^{(T)}}^{disp}$ (m^3/s)
$\hat{F}_{Q_i^{(T)}}^{disp}$	The posterior dispersed distributions estimated of the T-year
q	Discharge of regional growth curve ($m^3/s/km^2$)
$\hat{q}_{ML}^{(100)}$	Discharge estimated of the regional growth curve ($m^3/s/km^2$)
S_i	Catchment area of gauged sites i (km^2)

S_k	Catchment area of ungauged sites k (km^2)
s_i	Total number of site i of the region
t^i, t_3^i, t_4^i	L-coefficient of variation, $L_{skewness}$, and $L_{kurtosis}$ at site i
t^R, t_3^R, t_4^R	The weighted regional average, $L_{skewness}$, and $L_{kurtosis}$ of region
t_2^R	The regional average L_{CV}
H_1	The measure of the Hosking and Wallis homogeneity test
M	The number of MCMC chains
m	The value to retrieve a uniform distribution of $\hat{F}^{disp}_{Q_i^{(T)}}(Q_i^{(T)})$
N	The number of iterations of the MCMC algorithms
V	The weighted standard deviation of the at-site sample or observed dispersion
Z^{DIST}	The goodness-of-fit measure of distribution
δ	Coefficient of variation
μ	Index flood relation
φ_V	Mean of the N_{sim} values of V
$\varphi_{L_{CA}}$	The mean of L_{CA}
$\varphi_{L_{CV}}$	The mean of L_{CV}
σ_V	Standard deviation of the N_{sim} values of V
σ_K	Standard deviation of L_K^R
$\sigma_{L_{CA}}$	The standard deviations of L_{CA}
$\sigma_{L_{CV}}$	The standard deviations of L_{CV}
β	Parameter of the index-flood
$\xi; \alpha; \kappa$	Three-parameter of GEV
θ	Parameter
ρ_0	The coefficient of correlation in a bivariate normal distribution
Φ_i	The values of characteristics of the local growth curve in the regression techniques
$\lambda_1, \lambda_2, \lambda_3, \lambda_4$	4-parameter of L-Moments

Glossaire

ARI	Average recurrence interval
CCA	Canonical correlation analysis
CV	Coefficient of variation
CROI	Catchment region of influence
FFA	Flood frequency analysis
GEV	Generalized extreme value distribution
GROI	Geographical region of influence
GLS	Generalized least squares
HROI	Hybrid region of influence
LN	Lognormal distribution
MCMC	Markov chain Monte Carlo
MOR	Method of residuals
NQR	Normalized quantile regression
OLS	Ordinary least squares
POT	Peaks over a threshold
RFFA	Regional flood frequency analysis
ROI	Region of influence
RGC	Regional growth curve
WMW	Wilcoxon-Mann-Whitney test

Thèse en Français - Résumé

Préface

Ce résumé en Français reprend la structure du mémoire de thèse et présente les principaux résultats obtenus. Les différents chapitres y sont repris un à un, et les principales expressions mathématiques, ainsi que les figures et tableaux, illustrant les aspects essentiels du travail mené, sont tous repris et commentés.

0.1 Introduction

En hydrologie, un grand nombre d'études ont porté sur l'amélioration de l'estimation des quantiles de crues (débit maximum ou tout autre caractéristique de période de retour donnée). Ce estimation repose généralement sur des séries de crues maximales annuelles mesurées ou de valeurs dépassant un seuil. Cependant, les séries de données locales s'avèrent souvent trop courtes pour fournir des estimations fiables des quantiles de périodes de retour intermédiaires (typiquement 50 à 1000 ans). Sans informations supplémentaires pour affiner la forme de la distribution statistique, les résultats de l'analyse fréquentielle conduite s'avèrent généralement très incertains.

Pour cette raison, les hydrologues ont tenté de faire usage de sources d'information alternatives. Plusieurs approches développées consistent en une "extension temporelle" du jeu de données, basée sur l'intégration des données historiques et préhistoriques (Hosking et al. 1985b, Hosking and Wallis 1986a;b, Stedinger and Tasker 1986, Cohn and Stadinger 1987, Gary and Stedinger 1987, Sutcliffe 1987, Minghui and Stedinger 1989, Sheffer et al. 2003, Reis et al. 2005, Werritty et al. 2006, Ribatet et al. 2007b; 2009, Neppel et al. 2010, Payrastre et al. 2011). D'autres approches couramment utilisées pour enrichir les jeux de données disponibles pour l'inférence statistique correspondent à une "extension spatiale", et consistent en la fusion de données considérées comme statistiquement homogènes pour construire un large échantillon régional de données (Hosking and Wallis 1997, Charles and Stedinger 1999, Ouarda et al. 2001, Kjeldsen et al. 2002, Merz and Bloschl 2003, Seidou et al. 2006, Ribatet et al. 2007a, Norbiato et al. 2007, Wallis et al. 2007, Kjeldsen and Jones 2009). La plupart de ces approches sont basées sur l'hypothèse du "simple scaling" qui suppose que les distributions statistiques des débits de crues ont une forme identique au sein d'une région homogène, et diffèrent simplement par un facteur d'échelle spécifique à chaque site : "l'index de crue" Dalrymple (1960). Pour permettre cette hypothèse, les régions considérées sont constituées de bassins versants qui se ressemblent climatologiquement et

qui ont des mécanismes de production de crues semblables. Ils sont susceptibles d'être géographiquement contigus. Les résultats sont bien valables à condition que les hypothèses sur lesquelles l'analyse régionale des crues se fonde soient vérifiées. Malgré l'extension des jeux de données permise par ces approches, les quantiles estimés peuvent cependant rester incertains, notamment pour les bassins versants relativement petits et exposés à des crues de grande variabilité inter-annuelle.

Face à ce constat, Gaume et al. (2010) a proposé une méthode permettant d'incorporer au sein d'une analyse régionale des crues, en complément des séries hydrométriques, des informations relatives aux crues extrêmes observées sur des bassins non-jaugés. Ces crues extrêmes, parce qu'elles sont généralement les plus fortes connues sur une période assez longue, constituent l'équivalent d'une longue série d'enregistrements supplémentaires, qui est susceptible d'améliorer grandement la précision d'estimation des quantiles de crue. L'approche proposée pour incorporer ces données est toujours basée sur le principe de l'index de crue Dalrymple (1960). Toutefois, si une approche classique consiste à estimer l'index de crue par la moyenne de l'échantillon observé sur chaque site (Hosking and Wallis 1997), le calcul de cette moyenne n'est pas possible sur des sites non-jaugés. L'approche proposée par Gaume et al. (2010) consiste par conséquent à calibrer une relation d'index de crue, qui est une fonction des caractéristiques du bassin versant, sa surface en l'occurrence, et qui permet donc d'évaluer l'index de crue sur n'importe quel site non-jaugé. Le recours à cette relation d'index de crue représente une contrainte supplémentaire d'homogénéité de la région considérée, sensée respecter la relation théorique. Le complément de paramétrisation que représente cette relation (un paramètre supplémentaire), combinée à l'hypothèse d'homogénéité de la relation au sein de la région considérée, est susceptible de limiter les performances de cette approche, malgré le fait que des données supplémentaires correspondant aux crues extrêmes peuvent être mobilisées.

Cette thèse a été l'occasion de tester et d'évaluer les performances de cette approche (nommée "approche proposée" par la suite), par comparaison à une approche d'analyse régionale plus conventionnelle, et ceci dans plusieurs contextes de disponibilité des données et d'hétérogénéité des régions considérées. Les travaux menés ont également permis d'intégrer des développements supplémentaires à l'approche initiale développée par Gaume et al. (2010). La suite de ce rapport présente une synthèse des résultats obtenus, organisée de la façon suivante :

Le deuxième chapitre présente d'abord tous les concepts de base de l'analyse fréquen-

tielle régionale des crues. Elle met notamment l'accent sur les différences entre l'approche proposée, basée sur une relation d'index de crue, et l'approche plus conventionnelle qui a servi de référence pour l'ensemble des travaux (**?**). Le détail des méthodes d'inférence développées dans la suite du mémoire est également présenté dans ce chapitre : inférence basée sur la calibration d'une distribution GEV, les formulations de la vraisemblance des observations, et enfin algorithme bayésien MCMC utilisé pour l'estimation des paramètres. Enfin, les questions soulevées et les propositions d'améliorations développées dans le reste de la thèse sont présentées en détail, toujours par référence à l'approche régionale de référence d'Hosking et Wallis.

Le chapitre trois présente les évolutions méthodologiques mises en œuvre sur l'approche proposée de façon à répondre au mieux aux questions soulevées. Ce chapitre présente ensuite une première évaluation comparative des deux approches régionales présentées dans la section 2. Cette évaluation est basée sur des simulations Monte Carlo. Elle permet de tester la fiabilité des deux approches dans différents contextes de disponibilité des données et de présence d'hétérogénéités au sein de la région considérée. Dans un premier temps, les deux méthodes sont comparées sur la base de jeux de données équivalents et limités à des séries jaugées, en examinant spécifiquement l'impact des hétérogénéités présentes au sein de la région considérée. L'évaluation menée porte sur la qualité de l'estimation des quantiles 100 ans et sur la fiabilité des intervalles de crédibilité calculés dans chaque cas. Elle permet de faire ressortir certains avantages ainsi que les principales limites des deux approches de régionalisation considérées. La sensibilité à la présence d'hétérogénéités est examinée sous plusieurs formes : les hétérogénéités associées à la relation d'index de crue et à la loi de probabilité régionale sont successivement considérées.

La chapitre quatre présente ensuite l'application des deux approches à deux études de cas : la région de l'Ardèche (168 séries jaugées sur 5 sites et 18 crues extrêmes non-jaugées) et la région du Var (249 séries jaugées sur 9 sites et 4 crues extrêmes non-jaugées) dans le sud-est de la France. Cette application permet d'illustrer les avantages et les limites de chaque méthode lors de l'application à des cas d'étude réels, et de confronter ces méthodes à des approches locales (pouvant mobiliser ou non des informations historiques). Elle permet également surtout d'illustrer un aspect supplémentaire : l'apport de l'information relative aux crues extrêmes recensées sur des sites non-jaugés, apport qui devient possible grâce aux évolutions méthodologiques introduites avec l'approche proposée. L'intérêt de la mobilisation de ces informations sur les crues extrêmes non-jaugées ressort assez nettement des cas d'étude présentés.

Enfin, la section cinq propose les conclusions et perspectives de ce travail.

0.2 Description des méthodes d'analyse régionales étudiées

0.2.1 Principe général des deux approches étudiées

Les deux approches considérées reposent toutes les deux sur le principe de l'index de crue Dalrymple (1960). Ce principe revient à considérer qu'au sein d'une région statistiquement homogène, l'ensemble des distributions statistiques des débits de pointes de crues sont identiques moyennant un facteur d'échelle spécifique à chaque site : l'index de crue μ_i. Ceci se résume par la relation suivante :

$$Q_i^{(T)} = \mu_i q^{(T)} \tag{1}$$

où i représente un site donné au sein de la région (i= 1,..,s), s est le nombre total de sites dans la région, $Q_i^{(T)}$ est le quantile de débit de pointe de période de retour T sur le site i, $q^{(T)}$ est le quantile correspondant de la distribution régionale réduite, et μ_i est l'index de crue du site i.

L'approche usuelle d'analyse régionale développée par Hosking and Wallis (1997) consiste à évaluer μ_i par la moyenne des observations effectuées sur chaque site i. Par construction, cette approche présente l'inconvénient d'empêcher la mobilisation des informations disponibles sur des sites non jaugés, pour lesquels la moyenne locale des observations ne peut être estimée. De façon à pouvoir valoriser ces informations supplémentaires, la méthodologie développée par Gaume et al. (2010) prévoit d'estimer l'index de crue μ_i de chaque site i constituant la région, à partir d'une relation du type :

$$\mu_i = S_i^{\beta} \tag{2}$$

Dans laquelle S_i représente la surface du bassin versant du site i, et β est un paramètre à déterminer.

Le recours à la relation (2) permet le calcul de μ_i y compris pour des sites non-jaugés, et par conséquent d'incorporer dans l'analyse les données relatives aux crues extrêmes observées sur ces sites. Dans les tests initiaux présentés par Gaume et al. (2010), la relation (2) était calibrée a priori à partir des sites jaugés présents dans la région, et sa validité vérifiée à partir d'un test de Wilcoxon, Mann et Withney. L'approche qui sera développée ici est sensiblement différente puisqu'elle permet une estimation de β conjointe avec celle

des autres paramètres de la distribution statistique régionale réduite, ce qui permet une prise en compte explicite des incertitudes inhérentes à la calibration de ce paramètre.

Dans la suite du document, les termes d'approche proposée et d'approche de référence (pour celle de Hosking et Wallis) seront utilisés.

0.2.2 Procédure d'inférence utilisée et formulation de la vraisemblance des données

La procédure d'inférence utilisée repose sur le calcul de la vraisemblance de l'échantillon régional de données, à partir d'un modèle statistique candidat correspondant ici à une distribution GEV de paramètres (ξ, α, κ).

Dans sa forme la plus générale, l'échantillon **D** de données régionales considéré inclut :

- des séries de valeurs maximales annuelles de débit de pointe $Q_{i,j}$ enregistrées sur s sites i=(1,..,s), chaque série incluant n_i années d'enregistrements (j=(1,..,n_i)),
- des valeurs extrêmes de dépit de pointe Q_k observées sur h sites non jaugés (k=(1,..,h)), chaque valeur Q_k constituant la plus forte valeur de débit observée sur une période de n_k années.

La formulation classique de la vraisemblance de l'échantillon **D** prend alors la forme suivante :

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

où f_θ et F_θ représentent respectivement la fonction densité de probabilité et la fonction de répartition de la distribution GEV (cf. equation.4 et 5), les valeurs Q/μ représentent les valeurs réduites (divisées par l'index de crue) des débits de pointe mesurés sur chaque site, et θ représente le vecteurs des paramètres à estimer. Ici, θ inclut les trois paramètres de la distribution GEV : θ=(ξ, α, κ), et le paramètre β dans le cas de l'approche proposée.

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

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

Dans le cas de l'approche de référence d'Hosking et Wallis (1997), les crues extrêmes Q_k ne pouvant pas être mobilisée, l'expression (3) de la vraisemblance se limite au premier terme $\prod_{i=1}^{s}\left[\prod_{j=1}^{n_i} f_\theta\left(\frac{Q_{i,j}}{\mu_i(\theta)}\right)\right]$.

0.2.3 Estimation des paramètres à l'aide d'un algorithme bayésien MCMC

L'estimation des paramètres θ du modèle est réalisée en utilisant une approche bayésienne MCMC (Monte Carlo Markov Chain) désormais relativement répandue pour des applications en hydrologie (Reis et al. 2005, Renard et al. 2006, Neppel et al. 2010, Gaume et al. 2010, Payrastre et al. 2011), et dont le principe est succinctement rappelé ici. Cette approche repose directement sur l'application du théorème de Bayes, qui fournit une estimation de la distribution de probabilité des paramètres θ conditionnellement aux observations $\mathbf{D}$:

$$p\left(\theta \mid \mathbf{D}\right) = \frac{\ell\left(\mathbf{D} \mid \theta\right) p\left(\theta\right)}{p\left(\mathbf{D}\right)} \tag{6}$$

Dans l'expression (6) $p\left(\theta\right)$ représente l'information disponible a priori sur la distribution des paramètres, et $p\left(\mathbf{D}\right)$ est la probabilité du jeu d'observations $\mathbf{D}$, qui est une constante et n'influe donc pas sur le résultat obtenu. Ici aucune information a priori n'a été considérée. Une distribution $p\left(\theta\right)$ uniforme a donc été retenue.

L'estimation de $p\left(\theta \mid \mathbf{D}\right)$ est réalisée en appliquant un algorithme MCMC, qui permet d'échantillonner un grand nombre de vecteurs de paramètres θ dont la densité de probabilité est $p\left(\theta \mid \mathbf{D}\right)$. A partir du jeu de paramètres θ échantillonné, la distribution de probabilité des quantiles de débits peut être estimée en chaque site au sein de la région. Ce qui permet de calculer des intervalles de crédibilité à 90% sur chacun de ces quantiles.

0.2.4 Premiers tests réalisés par Gaume et questions soulevées

Les résultats présentés par Gaume et al. (2010) ont été obtenus en appliquant l'approche proposée, décrite ci-avant, à plusieurs études de cas (régions du Gard et de l'Ardèche dans le sud-est France, et régions de Slovaquie). La méthode d'inférence appliquée correspond à l'approche Bayésienne MCMC présentée au paragraphe précédent. Dans chaque cas une analyse de sensibilité à la nature des données mobilisées est conduite (avec ou sans information issues des sites non jaugés). Ces quelques études de cas montrent clairement que la mobilisation des informations sur les crues extrêmes identifiées sur des sites non jaugés peut conduire à une réduction très importante de la largeur des intervalles de crédibilité calculés sur les quantiles de périodes de retour supérieures à 100 ans.

Si ces premiers résultats s'avèrent très prometteurs, ils ont néanmoins été produits à partir de relations d'index de crue calibrées a priori dans chacune des régions considérées : les intervalles de crédibilité calculés ne tiennent donc pas compte de l'incertitude d'estimation du paramètre de cette relation. Par ailleurs, ces intervalles reposent sur une hypothèse d'homogénéité parfaite des régions considérées : cette homogénéité supposée concerne à la fois relation d'index de crue calibrée, mais également la forme des distributions locales de débits de pointe qui est supposée semblable sur l'ensemble des sites jaugés et non jaugés au sein des régions considérées. Sur ce dernier point (forme des distributions locales), une mesure d'hétérogénéité est proposée par **?**, mais cette vérification se limite aux sites jaugés.

Certaines questions importantes, car susceptibles de remettre en cause les conclusions initiales, restent donc en suspens à l'issue de ce premier travail d'évaluation. Ces questions peuvent se résumer de la façon suivante :

- d'une façon générale, quel est le degré de validité des intervalles de crédibilité calculés et dans quelle mesure ces intervalles représentent ils l'incertitude réelle ?
- quelle incertitude supplémentaire doit être attribuée à la calibration de la relation d'index de crue, qui représente un paramètre supplémentaire par comparaison à l'approche régionale de référence ?
- quel est l'impact des hétérogénéités inévitables de la relation d'index de crue dans la région considérée ? Si cet impact est important, il peut conduire à limiter fortement l'étendue des régions considérées avec l'approche proposée.
- quel est impact d'hétérogénéités de la forme des distributions locales au sein de la région considérée, sachant que cette l'hypothèse d'homogénéité de ces distributions ne peut être testée que sur la base des séries jaugées et est ensuite appliquée aux sites non-jaugés ? Cette hypothèse nécessite une délimitation soigneuse de la région, fondée par exemple sur la proximité spatiale et / ou sur les caractéristiques des bassins versants, et peut, là encore, conduire à limiter l'étendue de la région considérée.

Les résultats produits dans le cadre de la thèse ont pour objectif de répondre partiellement aux questions soulevées ci-dessus. De façon à pouvoir estimer l'impact de l'hétérogénéité des régions considérées, ces résultats reposent assez largement sur des simulations, dont le principe sera présenté dans la section suivante. Ils conduisent finalement, à partir de nouvelles études de cas présentées en section 4, à mieux illustrer les avantages et les

limites possibles de l'incorporation de l'information sur les crues extrêmes non jaugées, à partir de l'approche proposée par Gaume et al. (2010).

0.3 Evaluation des deux approches à partir de simulations

Cette section présente dans un premier temps les évolutions méthodologiques introduites de façon à permettre une meilleure évaluation des intervalles de crédibilité associés à l'approche proposée : ces évolutions permettent une évaluation commune des paramètres de la relation d'index de crue et de la distribution GEV régionale réduite, au sein de l'algorithme MCMC. L'introduction d'un paramètre supplémentaire à estimer soulève également la question de la convergence des algorithmes MCMC mis en œuvre : un paragraphe est donc consacré à la vérification de cette convergence.

Dans un deuxième temps, la section présente la méthode d'évaluation comparative des deux approches d'analyse régionale décrites dans la section 2, développée à partir de simulations et en considérant des jeux de données équivalents pour les deux méthodes. Cette évaluation, outre la simple comparaison des deux approches, permet d'avoir un premier aperçu de l'impact possible des hétérogénéités de la région considérée sur les résultats fournis par chacune des deux méthodes.

Enfin, les premiers résultats de simulation obtenus sont présentés et commentés, dans un premier temps en considérant des possible régions parfaitement homogènes, puis en mettant l'accent sur les impacts de la présence d'hétérogénéités dans les régions étudiées.

0.3.1 Adaptation des formulations de vraisemblance

Cas de l'approche proposée

L'estimation commune du paramètre β de la relation d'index de crue, et des paramètres (ξ, α, κ) de la loi GEV représentant la forme des distributions locales des débits, peut être mise en œuvre assez simplement, mais nécessite une adaptation de l'expression de la vraisemblance présentée dans l'équation (3). La formulation qui a été utilisée dans la suite du travail présenté ici est la suivante :

$$\ell(\mathbf{D} \mid \theta) = \prod_{i=1}^{s}\left[\prod_{j=1}^{n_i}\left[F_\theta\left(1.01 * \frac{Q_{i,j}}{\mu_i(\theta)}\right) - F_\theta\left(0.99 * \frac{Q_{i,j}}{\mu_i(\theta)}\right)\right]\right] \times \prod_{k=1}^{h}\left[F_\theta\left(1.01 * \frac{Q_k}{\mu_k(\theta)}\right) - F_\theta\left(0.99 * \frac{Q_k}{\mu_k(\theta)}\right)\right] \times \prod_{k=1}^{h}\left[F_\theta\left(\frac{Q_k}{\mu_k(\theta)}\right)\right]^{(n_k-1)} \quad (7)$$

Dans cette expression, le jeu de paramètres θ est constitué des 4 paramètres précités (β, ξ, α, κ),et les valeurs d'index de crue μ_i sur chaque site ne sont plus fixées a priori puisque dépendantes du paramètre β à estimer.

On peut également remarquer dans les deux premiers termes de l'expression (7), que la densité de probabilité $f_\theta\left(\frac{Q}{\mu}\right)$ usuellement utilisée pour représenter des valeurs discrètes de débit a été remplacée ici par la probabilité d'observer une valeur dans un intervalle de $\pm 1\%$ autour de $\frac{Q}{\mu}$: $F_\theta\left(1.01 * \frac{Q}{\mu}\right) - F_\theta\left(0.09 * \frac{Q}{\mu}\right)$. Cette modification s'est avérée nécessaire car le rapport $\frac{Q}{\mu}$ étant décroissant en fonction du paramètre β, la densité de probabilité $f_\theta\left(\frac{Q}{\mu}\right)$ croît en fonction de β, ce qui perturbe fortement le fonctionnement de l'algorithme MCMC. La valeur de la densité de probabilité n'est pas indépendante de l'ordre de grandeur des débits réduits. L'algorithme d'optimisation MCMC conduit donc à une estimation biaisée du paramètre β (surestimation afin de maximiser la valeur des densités de probabilité) et par conséquent des autres paramètres calés. Afin d'éviter cet effet, la vraisemblance ne doit comporter que des fonctions de répartition. A titre d'illustration, la figure 1 compare les résultats de l'estimation des paramètres, selon l'expression de vraisemblance utilisée (expression classique (3) ou modifiée (7)). Cette figure illustre très clairement le biais d'estimation important de ce paramètre et par conséquent des autres paramètres lorsque la vraisemblance repose sur l'expression (3).

Cas de l'approche de référence

Pour l'application de l'approche de référence d'Hosking (1997), de façon à conserver des formulations de vraisemblance homogènes entre les deux approches, la formulation de vraisemblance qui a été utilisée par la suite correspond au premier terme de l'équation (7). Néanmoins, dans ce dernier, cas le jeu de paramètres à estimer se limite aux 3 paramètres de la distribution GEV(ξ, α, κ), et les valeurs d'index de crue (estimées à partir des moyennes locales) sont fixes :

$$\ell(\mathbf{D} \mid \theta) = \prod_{i=1}^{s}\left[\prod_{j=1}^{n_i}\left[F_\theta\left(1.01 * \frac{Q_{i,j}}{\mu_i(\theta)}\right) - F_\theta\left(0.99 * \frac{Q_{i,j}}{\mu_i(\theta)}\right)\right]\right] \quad (8)$$

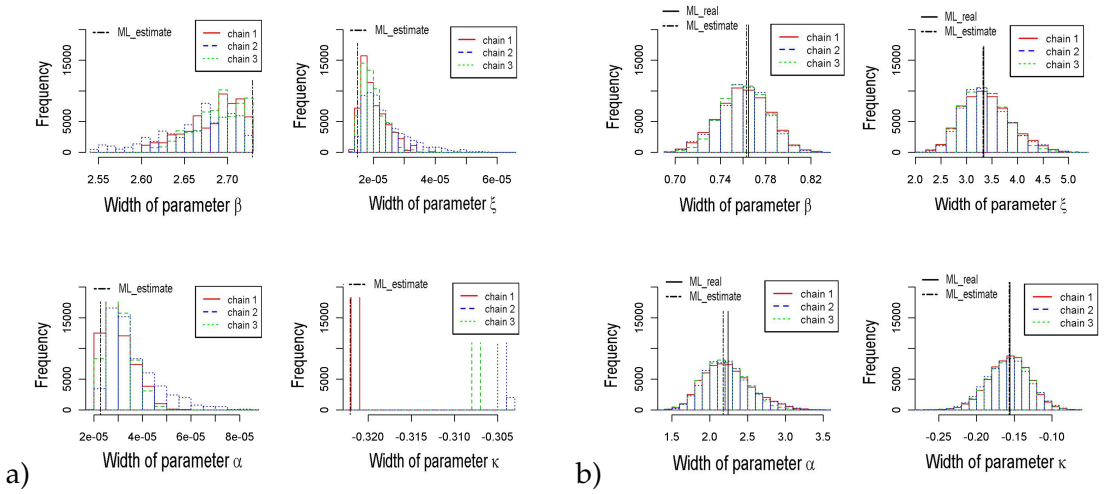


FIGURE 1 – *Comparaison des résultats d'estimation du paramètre θ par l'algorithme MCMC, à partir d'un jeu de données identiques, et suivant l'expression de vraisemblance utilisée : a) expression (3), et b) expression (7).*

Les résultats de l'inférence statistique basée sur l'expression 7 (ou 8) et sur l'expression 3 sont extrêmement proches dans le cas de l'approche de référence.

0.3.2 Vérification de la convergence des algorithmes MCMC

Le recours à un algorithme MCMC pour l'estimation des paramètres θ, nécessite de vérifier avec attention la bonne convergence de l'algorithme, et ce à chaque application. Si cette convergence posait rarement problème dans le cas de l'approche de référence, où seuls les trois paramètres de la distribution GEV sont à estimer, le passage à l'estimation conjointe de 4 paramètres avec l'approche proposée est susceptible de poser des difficultés de convergence bien plus importantes.

De façon à pouvoir vérifier de façon systématique cette convergence, le test de convergence de Gelman and Rubin (1992) a été sélectionné. Ce test fait partie des plus couramment utilisés, parmi les nombreux tests par ailleurs développés dans la littérature. Ce diagnostic est appliqué à chacun des paramètres estimés, et basé sur une analyse de variance de plusieurs chaines de Markov conduites en parallèle. Le test résulte dans le calcul d'un coefficient **R** (rapport entre la variance des valeurs des paramètres entre chaînes et intra-chaînes), qui prend des valeurs significativement supérieures à 1 lorsqu'une absence de convergence significative est détectée, et des valeurs proches de 1 lorsque la convergence semble satisfaisante. Par la suite, une valeur seuil de 1.05 a été considérée comme représentative d'un niveau de convergence satisfaisant.

La figure 2 présente l'évolution des valeurs du coefficient **R**, pour des algorithmes MCMC appliqués à un même jeu de données, et en fonction du nombre d'itérations effectuées (de 40 000 à 300 000). Cette figure confirme d'une part que la convergence est rapidement atteinte dans le cas de l'approche de référence (3 paramètres à estimer). Dans le cas de l'approche proposée, elle parait plus difficile à atteindre notamment pour un nombre d'itérations inférieur à 60 000. En conséquence, dans la suite des résultats présentés, le nombre d'itérations des algorithmes appliqués a été fixé à 60 000, et le test de Gelman and Rubin (1992) appliqué de façon systématique de façon à éliminer les cas de convergence insatisfaisante.

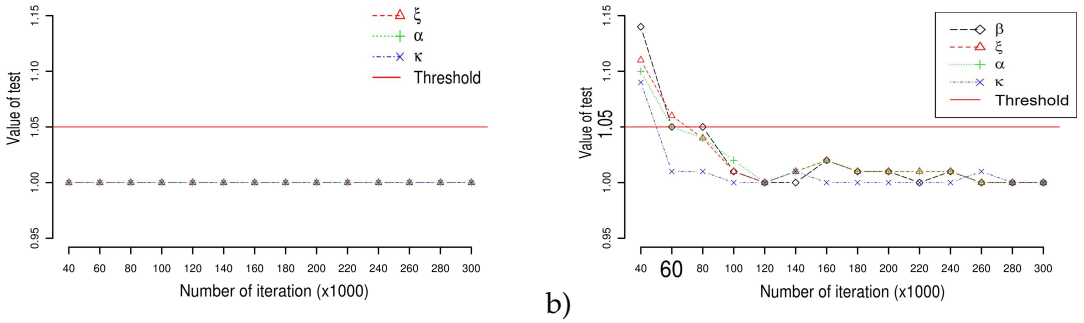


FIGURE 2 – *Résultats du diagnostic de convergence de Gelman et Rubin's appliqué à l'algorithme MCMC : a) approche de référence ; et b) approche proposée. Ces résultats correspondent aux séries jaugées de la région de l'Ardèche*

0.3.3 Principe général des simulations Monte Carlo

Cas d'échantillons homogènes

La procédure de simulations Monte Carlo développée pour l'évaluation comparative des deux approches régionales (référence et proposée) consiste à générer 1000 jeux de données régionales dont les caractéristiques (nombre et longueur des séries, surfaces des bassins correspondantes) sont définies au préalable et peuvent par exemple correspondre à une étude de cas à traiter. Les différentes séries de débits sont générées à partir d'une distribution régionale réduite et d'une relation d'index de crue toutes deux définies au préalable, par exemple ajustées à partir des données réelles disponibles. La forme de la distribution régionale considérée correspond ici systématiquement à une distribution GEV de paramètres (ξ_0, α_0, κ_0). La relation d'index de crue utilisée est celle présentée dans la section 2, avec un paramètre β_0.

Le cas de simulation le plus simple correspond à une région homogène. Dans ce cas,

les échantillons de données générés respectent parfaitement la forme de la distribution régionale des débits réduits, et de la relation d'index de crue (voir Figure 3). La procédure de simulation est alors la suivante :

- une série de débits réduits $q_{i,j}$ est générée pour chaque site i de la région considérée. La longueur n_i (j=[1,..,n_i]) de chacune de ces séries est celle fixée au préalable. Les débits réduits $q_{i,j}$ sont directement déduits de l'expression de la fonction de répartition de la distribution GEV :

$$q_{i,j} = \xi_0 + \frac{\alpha_0}{\kappa_0}\left[1 - \left(-log\left(F_j\right)\right)^{\kappa_0}\right] \tag{9}$$

  Les probabilités de non dépassement F_j étant générées par tirage aléatoire dans une loi uniforme sur la plage [0,1],
- la série de chaque site i est ensuite remise à la bonne échelle en utilisant la relation d'index de crue :

$$Q_{i,j} = q_{i,j} * S_i^{\beta_0} \tag{10}$$

Les deux approches d'analyse régionale ont ensuite appliquées à chacun des 1000 échantillons générés. La distribution a posteriori des paramètres est estimée dans chaque cas en appliquant l'algorithme Bayésiens MCMC, dont la bonne convergence est vérifiée à partir du test de Gelman and Rubin (1992). Les résultats d'estimation sont ensuite comparés aux véritables distributions statistiques parentes des échantillons (notamment aux valeurs des quantiles "vrais") : cette étape d'analyse des résultats sera présentée en détails dans le paragraphes 0.3.4.

Introduction d'hétérogénéités dans la relation d'index de crue

L'hétérogénéité de la région considérée autour de la relation d'index de crue théorique a été introduite de la façon suivante dans les simulations effectuées : pour chaque site et chaque échantillon simulé, la valeur théorique $\mu_i = S_i^{\beta}$ de l'index de crue du site i a été remplacée par une valeur aléatoire tirée dans une distribution log-normale de moyenne S_i^{β} et d'écart type $\delta * S_i^{\beta}$, δ étant successivement pris égal à 0.1 et 0.3. La figure 4 présente les fluctuations correspondantes des valeurs d'index de crue μ_i, et leurs répercussion sur les moyennes des échantillons générés sur chaque site.

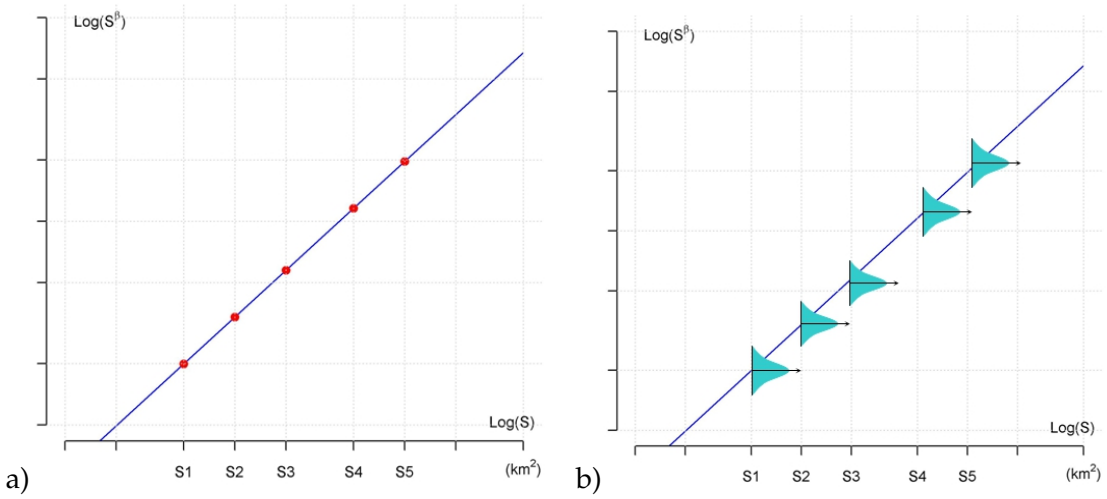


FIGURE 3 – *Relation d'index de crue utilisée pour la génération d'échantillons régionaux simulés : a) cas d'une région homogène, b) introduction d'hétérogénéités dans la relation d'index de crue à partir d'une distribution log-normale.*

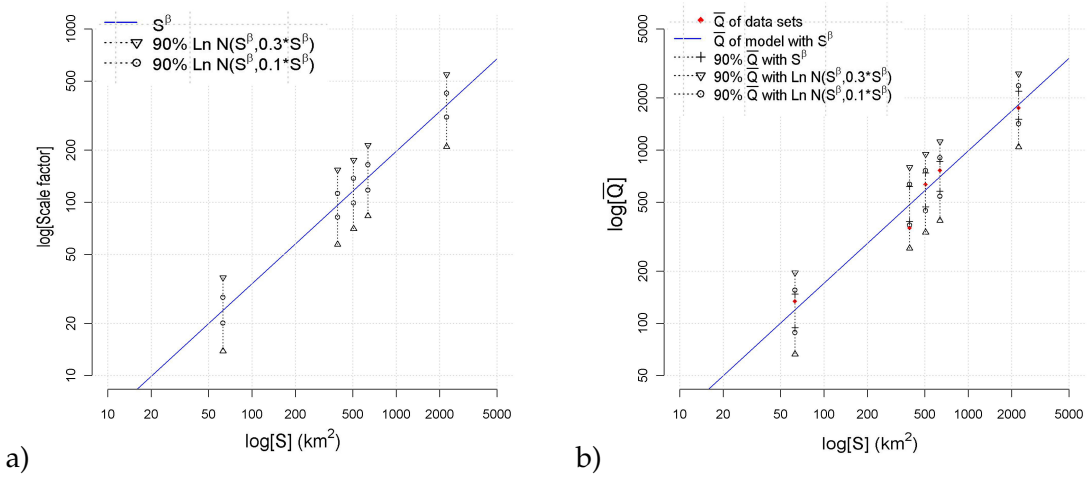


FIGURE 4 – *Dispersion des a) valeurs d'index de crue μ_i, et b) moyennes des échantillons simulés, associée à la génération aléatoire des valeurs de μ_i à partir d'une distribution log-normale $LN(S_i^\beta, \delta * S_i^\beta)$. La simulation présentée (nombre de sites, surfaces de bassins, longueurs des séries) correspond à la région de l'Ardèche.*

Introduction d'hétérogénéités dans la forme des distributions locales

? ont élaboré un test d'homogénéité des distributions locales au sein d'une région fondé sur les différences des valeurs des ratios de L-moments L_{CV} et $L_{skewness}$ estimées sur chacun des sites de la région. Pour une série de débits triés par ordre croissant, (Q_1, Q_2,..,Q_n), les expressions des quatre premiers L-moments sont les suivantes, i étant le rang des données dans l'ordre établi :

$$\lambda_1 = \frac{1}{C_n^1} \sum_{i=1}^{n} Q_i \tag{11}$$

$$\lambda_2 = \frac{1}{2C_n^2} \sum_{i=1}^{n} \left(C_{i-1}^1 - C_{n-1}^1 \right) Q_i \tag{12}$$

$$\lambda_3 = \frac{1}{3C_n^3} \sum_{i=1}^{n} \left(C_{i-1}^2 - 2C_{i-1}^1 C_{n-1}^1 + C_{n-1}^2 \right) Q_i \tag{13}$$

$$\lambda_4 = \frac{1}{4C_n^4} \sum_{i=1}^{n} \left(C_{i-1}^3 - 3C_{i-1}^2 C_{n-1}^1 + 3C_{i-1}^1 C_{n-1}^2 - C_{n-1}^3 \right) Q_i \tag{14}$$

Les ratios des L-moments se définissent de la façon suivante : t_2=λ_2/λ_1 est le L_{CV} (L coefficient de variation) ; $t_3 = \lambda_3/\lambda_2$ est appelé $L_{skewness}$; et $t_4 = \lambda_4/\lambda_2$ est appelé $L_{kurtosis}$.

La mesure d'homogénéité d'Hosking et Wallis est basée sur le calcul de l'écart type pondéré V des valeurs de L_{CV} calculées sur chaque site, pondérées par la longueur de chaque série (cf. section 2.3.3 pour l'expression complète dans la chapitre 2). Les fluctuations de V sont ensuite estimées à partir de 500 échantillons de caractéristiques similaires, simulées à partir d'une distribution kappa ajustée aux données réelles (méthode des L-moments). La distribution des valeurs de V obtenue est généralement proche d'une distribution normale et est caractérisée par sa moyenne φ_V et son écart-type σ_V. La mesure d'hétérogénéité se fait par le calcul du critère H_1 :

$$H_1 = \frac{(V - \varphi_V)}{\sigma_V} \tag{15}$$

La région pouvant être jugée homogène si H_1<1, et hétérogène si $H_1 \geq 3$ (section 2.3.3 dans la chapitre 2).

L'introduction d'hétérogénéités dans les distributions locales a également été réalisée ici en se basant sur les L-moments : des valeurs de L_{CV} et $L_{skewness}$ ont été générées aléatoirement, et ont été considérées comme représentatives de la forme de la distribution locale pour chacun des sites de la région considérée. Cette procédure nécessite trois étapes supplémentaires par rapport à la simulation de régions homogènes présentée au paragraphe 0.3.3 :

- **(i)** l'évaluation de gammes de fluctuations raisonnables des valeurs de L_{CV} et $L_{skewness}$, et le choix d'une distribution aléatoire représentant ces fluctuations,

– **(ii)** l'incorporation d'hétérogénéités, à partir de valeurs de L_{CV} et $L_{skewness}$ générées aléatoirement, dans les distributions utilisées pour générer les échantillons locaux de débits réduits,
– **(iii)** l'application du test d'homogénéité d'Hosking et Wallis décrit ci-dessus à chaque échantillon généré, permettant de filtrer en partie de l'hétérogénéité produite.

L'étape **(i)** a été réalisée en simulant 10 000 séries d'observations locales à partir de la distribution "moyenne", représentative d'une région homogène. Les fluctuations de valeurs de L_{CV} et $L_{skewness}$ estimées sur ces échantillons, dues à la variabilité d'échantillonnage, ont ensuite été observées, pour des échantillons incluant successivement n= 40 puis n= 20 observations, de façon à obtenir deux niveaux de dispersion différents des valeurs de L-moments. La figure 5 illustre la dispersion des 10 000 couples (L_{CV}, $L_{skewness}$) obtenus dans chaque cas. Cette figure laisse apparaitre la nette corrélation existante entre les deux valeurs de L-moments. L'ajustement d'une distribution normale bivariée à chacun de ces deux nuages de points permet finalement de générer aléatoirement des couples (L_{CV}, $L_{skewness}$) ayant des fluctuations équivalentes, comme en témoigne la figure 6.

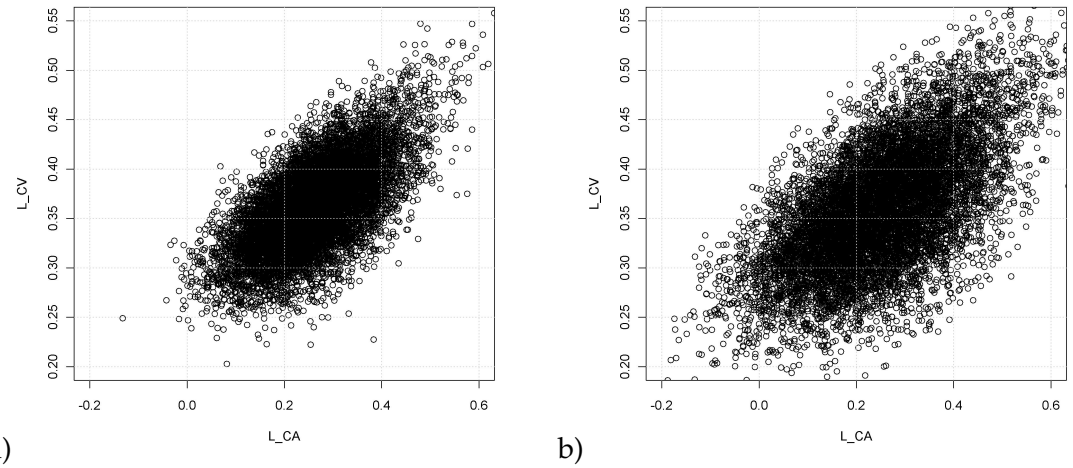
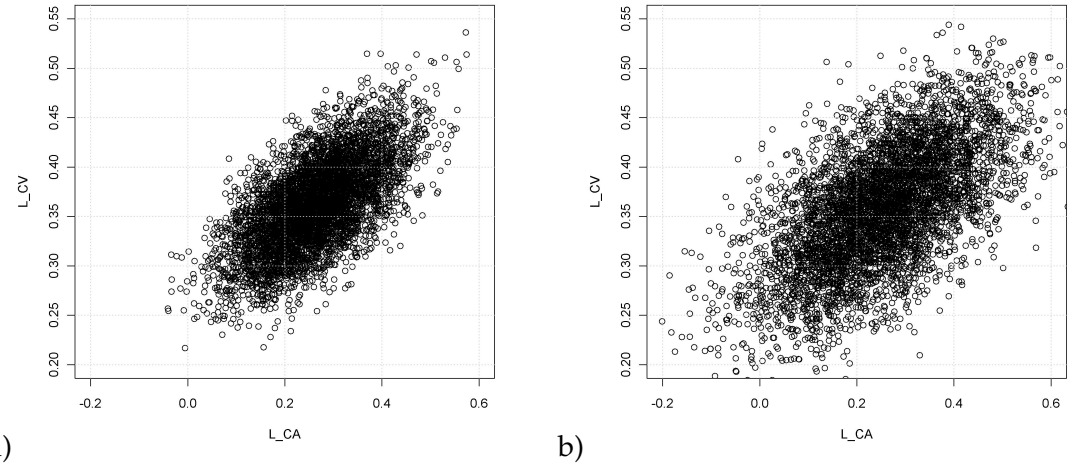


FIGURE 5 – *Dispersion des valeurs de L_{CV} et $L_{skewness}$ estimées sur 10 000 échantillons simulés d'une longueur de : a) n= 40 ans, et b) n= 20 ans. Les échantillons ont été simulés à partir d'une distribution GEV donc les paramètres ξ_0= 3.34, α_0= 2.24, et κ_0= -0.16, correspondent à ceux identifiés pour l'étude de cas de l'Ardèche.*

L'étape **(ii)** a ensuite été réalisée en générant aléatoirement, pour chaque série locale à simuler, un couple de valeurs de (L_{CV}, $L_{skewness}$) à partir de la distribution normale bivariée ajustée à l'étape **(i)**. Les paramètres de la distribution GEV appliquée pour générer la série ont été directement déduits des valeurs de L_{CV} et L_{CA} ou $L_{skewness}$ obtenues.



a) b)

FIGURE 6 – *Dispersion des valeurs L_{CV} et $L_{skewness}$ simulées à partir de distributions normales bivariées ajustées sur les données présentées sur la figure 5. Chaque nuage de points inclut 5000 valeurs. a) distribution normale bivariée ajustée sur les échantillons de durée n= 40 ans (Fig 5.a avec les paramètres de la loi normale bivariée de moyenne $\varphi_{L_{CA}}$= 0.267, $\varphi_{L_{CV}}$= 0.363, d'écart-type $\sigma_{L_{CA}}$= 0.076, $\sigma_{L_{CV}}$= 0.035, et coefficient de corrélation ρ_0= 0.668), et b) distribution normale bivariée ajustée sur les échantillons de durée n= 20 ans (Fig 5.b avec $\varphi_{L_{CA}}$= 0.259, $\varphi_{L_{CV}}$= 0.361, $\sigma_{L_{CA}}$= 0.109, $\sigma_{L_{CV}}$= 0.051, et ρ_0= 0.660).*

Les valeurs des paramètres (ξ_i, α_i, κ_i) de la distribution GEV utilisée pour simuler les séries de débits réduits sur le site i ont ensuite été directement estimés sur la base des valeurs de L_{CV-i} et $L_{skewness-i}$, en considérant une valeur attendue moyenne φ_{q_i} identique à celle de la courbe de référence :

$$\varphi_{q_i} = \xi_0 + \frac{\alpha_0 \left(\Gamma \left(1+\kappa_0\right) - 1\right)}{\kappa_0} \tag{16}$$

Il n'existe pas de solution explicite pour estimer κ_i, à l'aide de (φ_{q_i}, L_{CV-i}, L_{CA-i}) mais les approximations suivantes ont une précision supérieure à $9 * 10^{-4}$ pour -0.5<L_{CA-i}<0.5

$$\kappa_i = \quad 7.8590 * c + 2.9554 * c^2 \text{ avec } c = \frac{2}{3 + L_{CA-i}} - \frac{log2}{log3} \tag{17}$$

Les autres paramètres sont donnés par :

$$\alpha_i = \frac{\varphi_{q_i} * L_{CV-i} * \kappa_i}{\left(1 - 2^{\kappa_i}\right) * \Gamma \left(1+\kappa_i\right)} \tag{18}$$

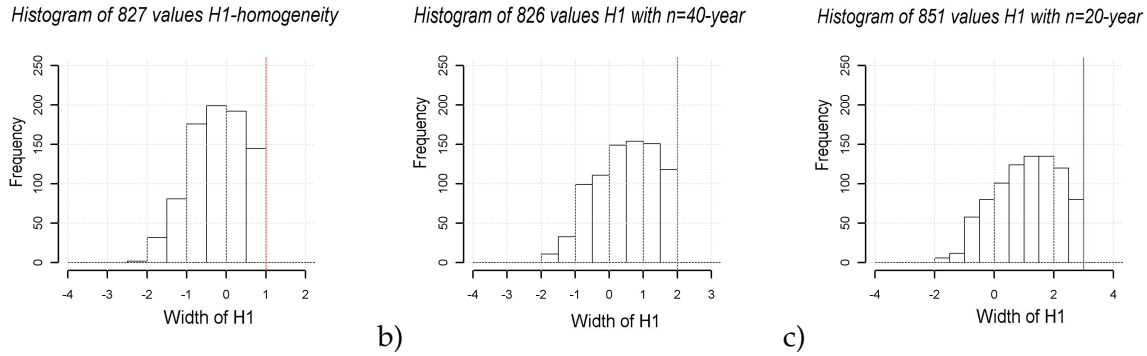
$$\xi_i = \varphi_{q_i} - \frac{\alpha_i * \left(1 - \Gamma \left(1+\kappa_i\right)\right)}{\kappa_i} \tag{19}$$

Finalement, la production de séries de débits réduits au site i est effectuée de la même manière que dans le cas de régions homogènes mais est, dans le cas présent, basée sur les valeurs des paramètres ξ_i, α_i, κ_i :

$$q_{i,j} = \xi_i + \frac{\alpha_i}{\kappa_i}\left[1 - \left(-log\left(F_{i,j}\right)\right)^{\kappa_i}\right] \tag{20}$$

Il est important de remarquer que les hétérogénéités ainsi introduites ne peuvent être que partiellement filtrées par la mesure d'homogénéité H_1 proposée par Hosking et Wallis (1993), puisqu'elles peuvent être expliquées par la variabilité d'échantillonnage. De façon à effectuer des simulations aussi réalistes que possible, cette mesure a été systématiquement calculée pour chacun des échantillons de données régionales simulés (étape **iii**). Sur la base des résultats obtenus, les trois situations d'hétérogénéité suivantes ont finalement été prises en compte (cf. Figure 7) :

- **Cas 1 :** ce cas correspond à des jeux de données régionales homogènes (échantillons générés à partir d'une distribution GEV unique). Dans ce cas, seuls les échantillons avec une valeur $H_1 \leq 1$ ont été sélectionnés,
- **Cas 2 :** ce cas correspond à un niveau d'hétérogénéité limitée (fluctuations des valeurs de L_{CV} et $L_{skewness}$ correspondant à des échantillons de n= 40 ans) : les échantillons ayant une valeur $H_1 \leq 2$ ont été sélectionnés ,
- **Cas 3 :** ce cas correspond à des données très hétérogènes (fluctuations des valeurs de L_{CV} et $L_{skewness}$ correspondant à des échantillons de n= 20 ans), dans ce cas les échantillons ayant une valeur $H_1 \leq 3$ ont été conservés.



a) b) c)

FIGURE 7 – *Histogramme des valeurs de H_1 des échantillons régionaux sélectionnés, pour les trois niveaux d'hétérogénéité considérés : a)* ***Cas 1*** *: échantillons homogènes, 827/1000 échantillons sélectionnés en fonction d'une seuil $H_1 \leq 1$, b)* ***Cas 2*** *: échantillons avec hétérogénéité limitée, 826/1000 échantillons choisis en fonction d'un seuil $H_1 \leq 2$, et c)* ***Cas 3*** *: échantillons avec une forte hétérogénéité, 851/1000 échantillons choisis en fonction d'un seuil $H_1 \leq 3$.*

Introduction d'hétérogénéités dans la relation d'index de crue et dans les formes de distributions locales

La procédure de simulation présentée ici inclut simultanément l'incorporation d'hétérogénéités sur la forme des distributions locales et sur la relation d'index de crue. Cette procédure a été appliquée aux deux études de cas présentées dans la section 0.4, de façon à avoir un aperçu des impacts possibles des hétérogénéités supposées au sein de chacune de ces régions.

La procédure de simulations reste similaire à ce qui a été décrit précédemment. Elle peut être résumée de la façon suivante :

- étalonnage d'une distribution normale bivariée qui sera utilisée pour l'introduction des hétérogénéités sur les formes de distributions locales,
- la génération de la série des débits réduits sur chaque site de la région, en intégrant l'hétérogénéité sur les courbes de croissance utilisées sur chaque site,
- remise à l'échelle de chaque série locale réduite avec introduction d'hétérogénéité dans la relation d'index de crue (δ= 0.1),
- application du test d'homogénéité d'Hosking et Wallis et suppression de tous les échantillons identifiés comme hétérogènes, suivant le principe décrit au paragraphe 0.3.3.

0.3.4 Modalités d'analyse des résultats de simulations

Les deux approches de régionalisation décrites dans la section 0.2 ont été appliquées à chaque échantillon de données simulé, en vérifiant dans chaque cas la bonne convergence des algorithmes MCMC. Les résultats obtenus ont dans chaque cas été interprétés selon les modalités décrites ci-après.

Qualité d'estimation des quantiles

Le mode de la distribution a posteriori (ou maximum de vraisemblance) sur le quantile centennal $\hat{Q}_{i,ML}^{(T)}$ (T=100) et l'intervalle de crédibilité à 90% correspondant $[\hat{Q}_{i,5}^{(T)}, \hat{Q}_{i,95}^{(T)}]$ ont été calculés pour chaque site au sein de la région considérée. Le résultat final est un jeu de 1000 valeurs (ou un peu moins lorsque le test d'homogénéité d'Hosking et Wallis est appliqué) de débits centennaux estimés avec dans chaque cas l'intervalle de crédibilité à 90% associé. La qualité des estimations obtenues est vérifiée en examinant la dispersion

des valeurs estimées autour de la valeur vraie du quantile centennal de chaque site $Q_i^{(T)}$ (cf. Figure 4 par exemple).

Vérification de la validité des intervalles de crédibilité à 90%

Cette vérification a été effectuée en calculant la position du quantile "vrai" centennal $Q_i^{(T)}$ sein de chacune des distributions a posteriori estimées : $\hat{F}_{Q_i^{(T)}}\left(Q_i^{(T)}\right)$. Si les distributions a posteriori $\hat{F}_{Q_i^{(T)}}$ sont correctement estimées, les valeurs de $\hat{F}_{Q_i^{(T)}}\left(Q_i^{(T)}\right)$ obtenues doit être distribuées uniformément sur les 1000 simulations effectuées.

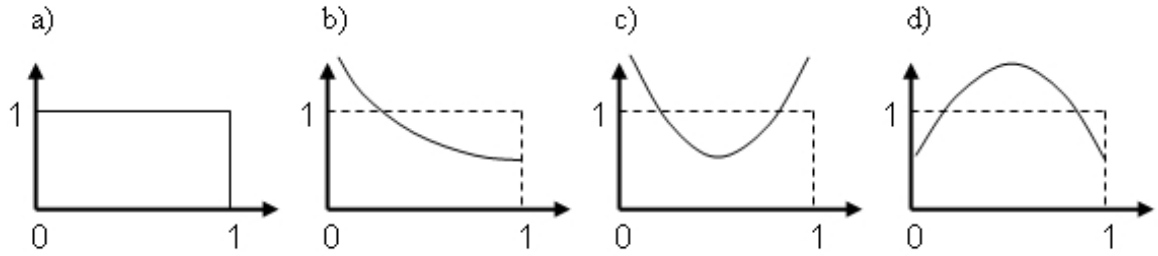


FIGURE 8 – *Exemples de distributions possibles des valeurs de* $\hat{F}_{Q_i^{(T)}}\left(Q_i^{(T)}\right)$*, et interprétations correspondantes : a) estimation parfaite, b) estimation biaisée du quantile (surestimation), c) limites de crédibilité estimées trop rapprochées, et d) limites de crédibilité estimées trop éloignées.*

La figure 8 présente quelques exemples des distributions de $\hat{F}_{Q_i^{(T)}}\left(Q_i^{(T)}\right)$ qui peuvent être obtenues ainsi que l'interprétation correspondante en termes de qualité de l'estimation.

Quelques-uns des résultats obtenus montrent une tendance à sous-estimer la variance des distributions a posteriori (cf. Figure 8.c) : limites de crédibilité estimées trop rapprochées). Dans ces cas, des distributions plus dispersées $\hat{F}^{disp}_{Q_i^{(T)}}$ ont été générées, en gardant la même médiane, mais avec un écart type multiplié par un facteur m=(1.2 ; 1.4 ; 1.6 ; 1.8). Ceci est obtenu en appliquant la formule suivante pour chaque valeur estimée $Q_i^{(T)}$:

$$\hat{Q}^{(T)}_{i,disp} = m * \left(\hat{Q}_i^{(T)} - median\left(\hat{Q}_i^{(T)}\right)\right) + median\left(\hat{Q}_i^{(T)}\right) \quad (21)$$

La valeur de m permettant d'aboutir à une distribution uniforme des valeurs $\hat{F}^{disp}_{Q_i^{(T)}}\left(Q_i^{(T)}\right)$ permet d'avoir une idée du niveau de sous-estimation de la largeur des intervalles de crédibilité.

0.3.5 Résultats des simulations

Cas de régions homogènes

Les simulations présentées ici correspondent à une région dont les caractéristiques (longueur des séries, nombre de sites et surfaces associées) sont celles rencontrées lors de l'étude de cas de l'Ardèche. Les séries simulées incluent donc 168 années-stations réparties sur 5 sites. Elles ont été générées à partir d'une distribution GEV régionale et d'une relation d'index de crue toutes deux ajustées à partir des données réelles disponibles dans la région de l'Ardèche : ξ= 3.34, α= 2.24, κ= -0.16 pour la distribution régionale réduite, β= 0.76 la relation d'index de crue.

Les figure 9 et figure 10 synthétisent les résultats des simulations réalisées sans introduire d'hétérogénéités dans les séries régionales générées. Ces figures révèlent la similitude globale des estimations obtenues entre les deux approches d'analyse régionale (référence et proposée), en particulier pour les surfaces de bassins versant les plus grandes : les fluctuations des valeurs de quantiles estimées $\hat{Q}_{i,ML}^{(T)}$ sont équivalentes, et les largeurs des intervalles de crédibilité sont proches. En revanche, les tendances observées pour la distribution régionale réduite (RGC) s'avèrent très différentes. Le débit centennal estimé est moins dispersé que les estimations locales et les intervalles de crédibilité sont relativement resserrés dans le cas de l'approche de référence d'Hosking et Wallis. Avec l'approche proposée, la valeur estimée est au contraire très fluctuante et les intervalles de crédibilité nettement plus larges.

Ces différences ont plusieurs explications. Dans le cas de l'approche proposée, l'équation (2) montre que la valeur du paramètre β influe sur l'ordre de grandeur de l'index de crue μ_i et donc sur celui des débits régionaux réduits $q_{i,j}$. De ce fait, la calibration de la relation d'index de crue induit une incertitude sur l'ordre de grandeur de la distribution régionale réduite estimée (qui correspond ici à une surface de 1 km^2). Cet effet est ici accentué par l'absence d'observations disponibles pour des surfaces proches de 1 km^2, la gamme d'observations couvrant la plage de surfaces 63-2240 km^2. L'incertitude sur l'ordre de grandeur des débits régionaux réduits est partiellement compensée lors du retour aux estimations locales car les débits réduits sont alors remultipliés par les index de crue μ_i estimés. Toutefois, l'équation (2) montre également que les valeurs de μ_i sont plus sensibles au paramètre β pour les grands bassins : autrement dit, les données correspondant aux bassins de grandes surfaces ont plus de poids dans la détermination de β. Ceci explique que l'effet d'incertitude observé sur la loi régionale réduite reste partiellement présent pour les petites surfaces (voir le cas du bassin de 63 km^2). En ce qui concerne l'approche de ré-

férence, on peut remarquer, contrairement à l'approche proposée, que les fluctuations du débit centennal sont moins importantes pour la distribution régionale réduite que pour les estimations locales. Ceci peut être attribué aux erreurs réalisées lors de l'estimation de l'index de crue sur chaque site à partir de la moyenne des observations locales, erreurs qui se compensent vraisemblablement au sein de l'échantillon régional réduit.

Les figure 11 et figure 12 présentent l'évaluation des distributions a posteriori calculées $\hat{F}_{Q_i^{(T)}}$. L'approche de référence (Figure 11) conduit globalement à sous-estimer la largeur des intervalles de crédibilité calculés. Cet effet de sous-estimation n'est pas présent pour la distribution régionale réduite (pour laquelle les largeurs d'intervalles sont plutôt surestimées), mais il est très net pour les différents sites jaugés. Cette sous-estimation est là encore liée à l'incertitude d'estimation de l'index de crue (espérance de chaque série estimée par sa moyenne), source d'incertitude qui n'est pas prise en compte dans le calcul des intervalles. Le calcul des distributions $\hat{F}^{disp}_{Q_i^{(T)}}$ avec m= 1.2 (Figures 11.d et 11.e) montre que cette sous-estimation est de l'ordre de 20% dans le cas présenté. Toutefois, cet effet est d'autant plus marqué que les séries concernées sont de faible longueur (voir figure 8 dans la paragraphe 0.3.4). La figure 12, pour sa part, montre qu'avec l'approche proposée l'effet d'incertitude observé pour les petites surfaces s'accompagne d'un biais d'estimation significatif (tendance à la sous-estimation du quantile réel). En revanche, avec cette approche les intervalles de crédibilité sont estimés de façon tout à fait satisfaisante dès que les surfaces considérées augmentent.

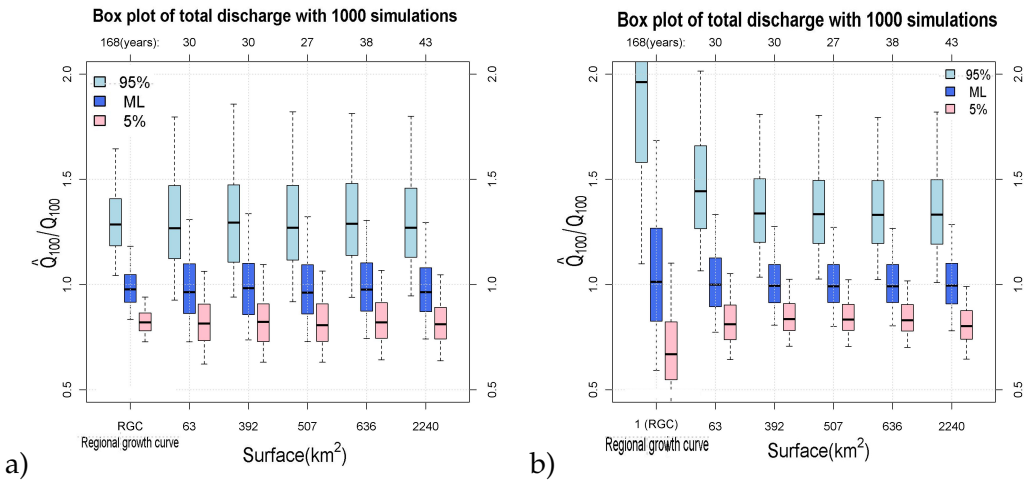


FIGURE 9 – *Dispersion des valeurs du débit centennal estimé $\hat{Q}^{(100)}_{i,ML}$ (respectivement $\hat{q}^{(100)}_{ML}$ pour la distribution régionale réduite ou courbes de croissance régionale), et des bornes des intervalles de crédibilité à 90% $[\hat{Q}^{(100)}_{i,5}, \hat{Q}^{(100)}_{i,95}]$, calculés à partir des 1000 échantillons régionaux simulés : a) approche de référence d'Hosking et Wallis, et b) approche proposée. L'ensemble des valeurs présentées sont divisées par la valeur réelle du quantile centennal $Q^{(100)}_i$ (ou $q^{(100)}$).*

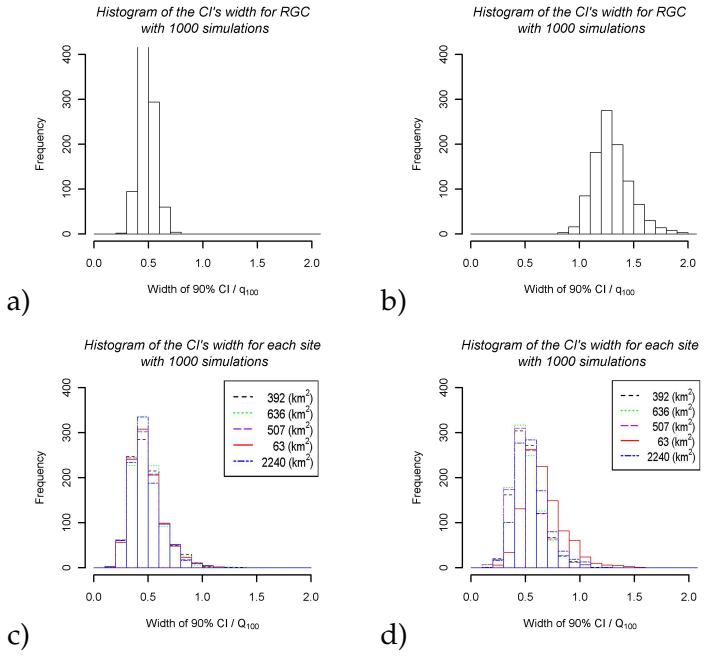


FIGURE 10 – *Largeurs des intervalles de crédibilité à 90% $[\hat{Q}_{i,5}^{(100)}, \hat{Q}_{i,95}^{(100)}]$ calculés à partir des 1000 échantillons régionaux simulés : a) Distribution régionale réduite pour l'approche de référence, b) Distribution régionale réduite avec l'approche proposée c) estimations locales pour l'approche de référence d) estimations locales pour l'approche proposée. L'ensemble des valeurs présentées ont été divisées par la valeur vraie du quantile centennal $Q_i^{(100)}$ (ou $q^{(100)}$).*

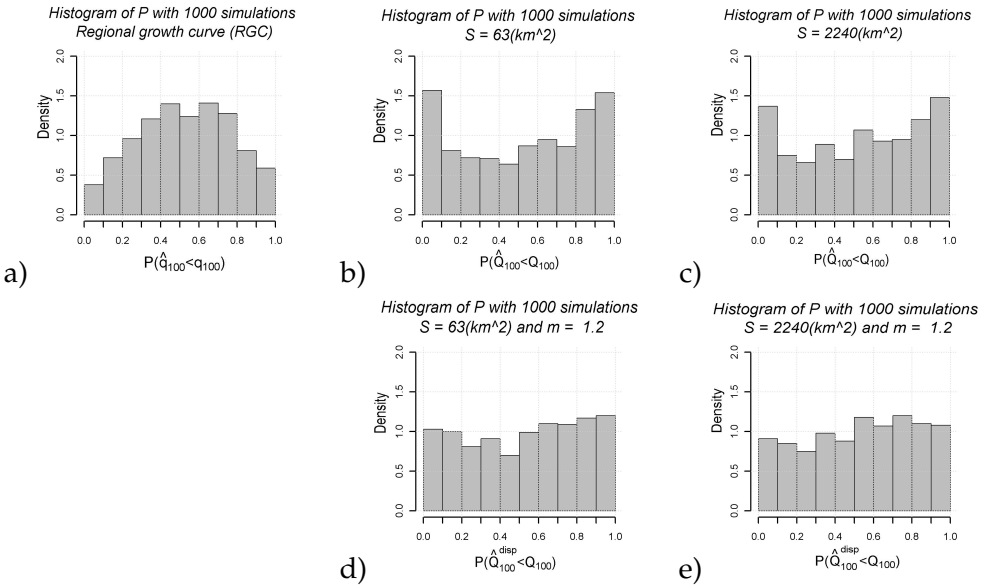


FIGURE 11 – *Distribution des valeurs de $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ (ou $\hat{F}_{q^{(100)}}(q^{(100)})$ estimées pour les 1000 échantillons simulés, dans le cas de l'approche de référence d'Hosking et Wallis : a) distribution régionale réduite, b) S= 63 km^2, c) S= 2240 km^2, et distributions corrigées $\hat{F}^{disp}_{Q_i^{(100)}}(Q_i^{(100)})$ associées : d) S= 63 km^2 avec m= 1.2, e) S= 2240 km^2 avec m= 1.2*

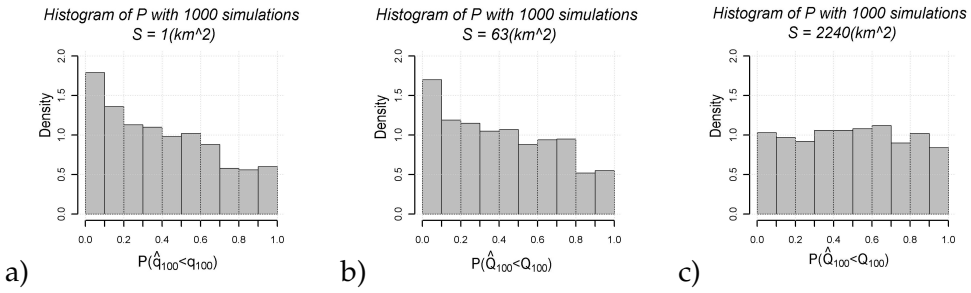


FIGURE 12 – *Distribution des valeurs de* $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ *(ou* $\hat{F}_{q^{(100)}}(q^{(100)})$ *dans le cas de la distribution régionale réduite) estimées pour les 1000 échantillons simulés, dans le cas de l'approche proposée : a) distribution régionale réduite (S= 1 km*^{2}*), b) S= 63 km*^{2}*, c) S= 2240 km*^{2}*.*

Influence de la longueur des séries considérées

Ce paragraphe présente une simulation complémentaire à la précédente et vise à mieux illustrer l'influence de la longueur des séries locales d'enregistrements sein de la région considérée. La taille des basins versants et la forme des distributions locales considérés sont identiques au cas présenté précédemment, mais le nombre d'observations disponibles sur chaque site a été fixé à 88 pour le bassin de 63 km^2, et respectivement à 40, 25, 10, et 5 observations pour les bassins de 392, 507, 636, et 2240 km^2. Le nombre total d'observations disponibles dans l'échantillon régional (168) est donc le même que dans le cas initial. Néanmoins, dans ce cas le faible effectif des observations présentes sur les basins de grande surface a un double effet : il limite la précision d'estimation de la moyenne sur les sites concernés, et il limite également le poids des bassins de grande surface dans la calibration de la relation d'index de crue.

Les résultats obtenus sont présentés sur les figures 3.13 and 3.14. Ils conferment l'analyse issue des tests initiaux. Avec l'approche de référence, la dispersion des estimations locales $\hat{Q}_{i,ML}^{(100)}$ et des intervalles de crédibilité associés $[\hat{Q}_{i,5}^{(100)}, \hat{Q}_{i,95}^{(100)}]$ augmente avec la diminution de la longueur des séries disponibles, en raison de la plus grande incertitude d'estimation de la moyenne locale. En parallèle, la sous-estimation de la dispersion des distributions a posteriori $\hat{F}_{Q_i^{(100)}}$ augmente fortement.

Avec l'approche proposée, la dispersion des estimations $\hat{Q}_{i,ML}^{(100)}$ et des intervalles de crédibilité $[\hat{Q}_{i,5}^{(100)}, \hat{Q}_{i,95}^{(100)}]$ tend à diminuer pour les petites surfaces, et augmente pour les surfaces les plus importantes car l'estimation du paramètre β et des valeurs d'index de crue correspondante est moins dépendante des observations présentes sur ces bassins de grande surface.

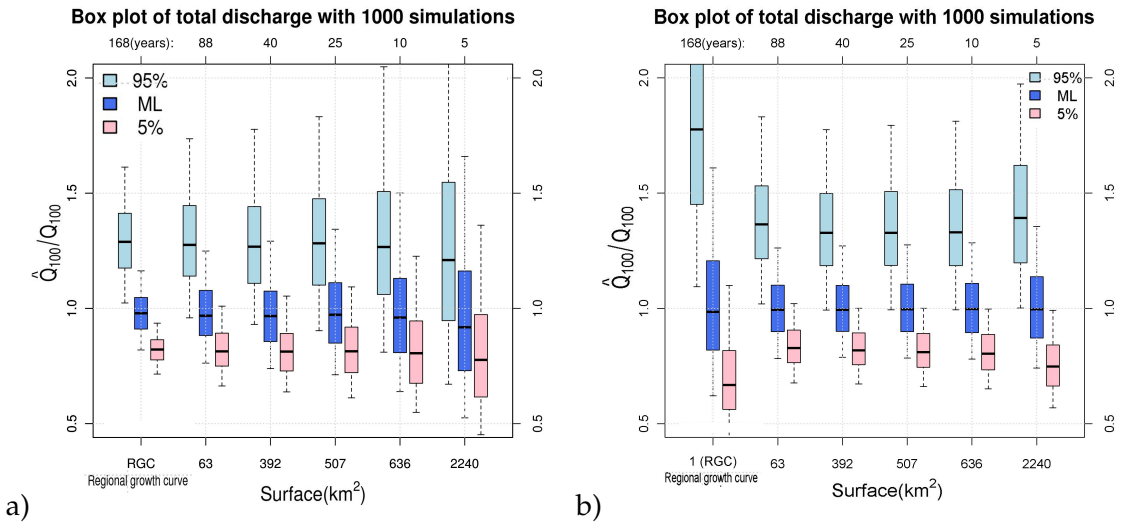


FIGURE 13 – *Dispersion des valeurs du débit centennal estimé $\hat{Q}_{i,ML}^{(100)}$ (respectivement $\hat{q}_{ML}^{(100)}$ pour la distribution régionale réduite), et des bornes des intervalles de crédibilité à 90% [$\hat{Q}_{i,5}^{(100)}$, $\hat{Q}_{i,95}^{(100)}$], calculés à partir des 1000 échantillons régionaux simulés et pour des longueurs d'enregistrements locales différentes de celles présentées sur la figure 9 (88, 40, 20, 10, et 5 ans) : a) approche de référence d'Hosking et Wallis, et b) approche proposée. L'ensemble des valeurs présentées sont divisées par la valeur réelle du quantile centennal $Q_i^{(100)}$ (ou $q^{(100)}$).*

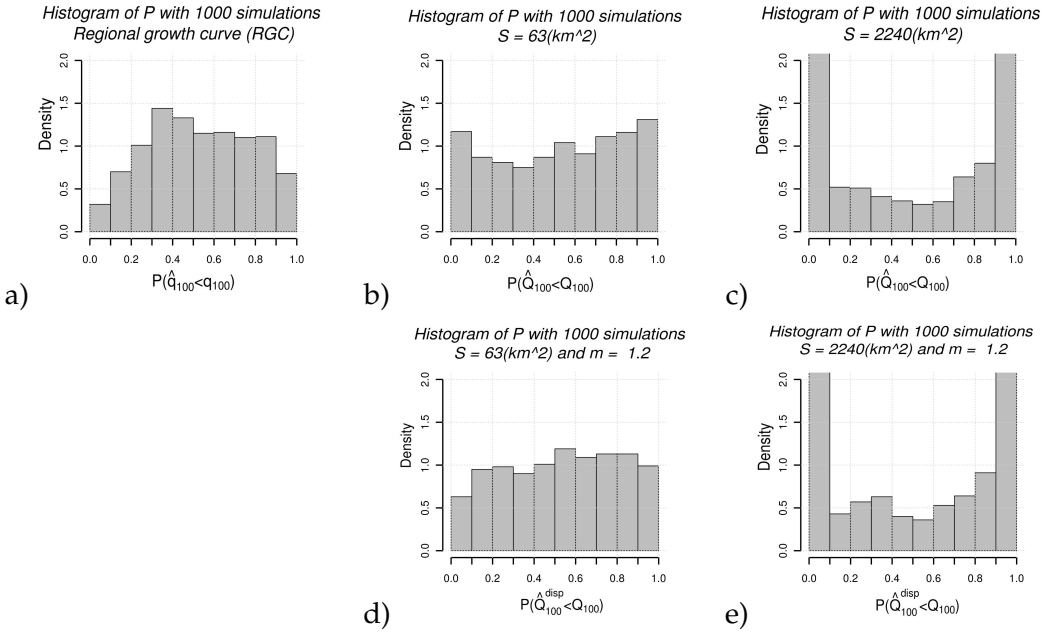


FIGURE 14 – *Distribution des valeurs de $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ (ou $\hat{F}_{q^{(100)}}(q^{(100)})$ estimées pour les 1000 échantillons simulés présentés sur la figure 13, dans le cas de l'approche de référence d'Hosking et Wallis : a) distribution régionale réduite, b) S= 63 km^2, c) S= 2240 km^2, et distributions corrigées $\hat{F}^{disp}_{Q_i^{(100)}}(Q_i^{(100)})$ associées : d) S= 63 km^2 avec m= 1.2, e) S= 2240 km^2 avec m= 1.2*

Cas d'hétérogénéités dans la relation d'index de crue

La figure 15 illustre l'effet d'une hétérogénéité des séries simulées autour de la relation d'index de crue théorique. Les simulations présentées ici correspondent à une région dont les caractéristiques sont identiques à celle présentée au paragraphe 0.4.2 (la région de l'Ardèche).

Les résultats fournis par l'approche de référence, qui ne repose pas sur cette relation d'index de crue, restent logiquement inchangés. Avec l'approche proposée, cette hétérogénéité a pour effet un élargissement général des intervalles de crédibilité qui s'accompagne d'un effet de biais plus marqué (tendance nette à la surestimation des valeurs des quantiles touchant cette fois l'ensemble des surfaces). Si ces effets restent faibles dans le cas d'une hétérogénéité modérée (δ= 0.1), ils deviennent beaucoup plus importants en cas d'hétérogénéité forte (δ= 0.3).

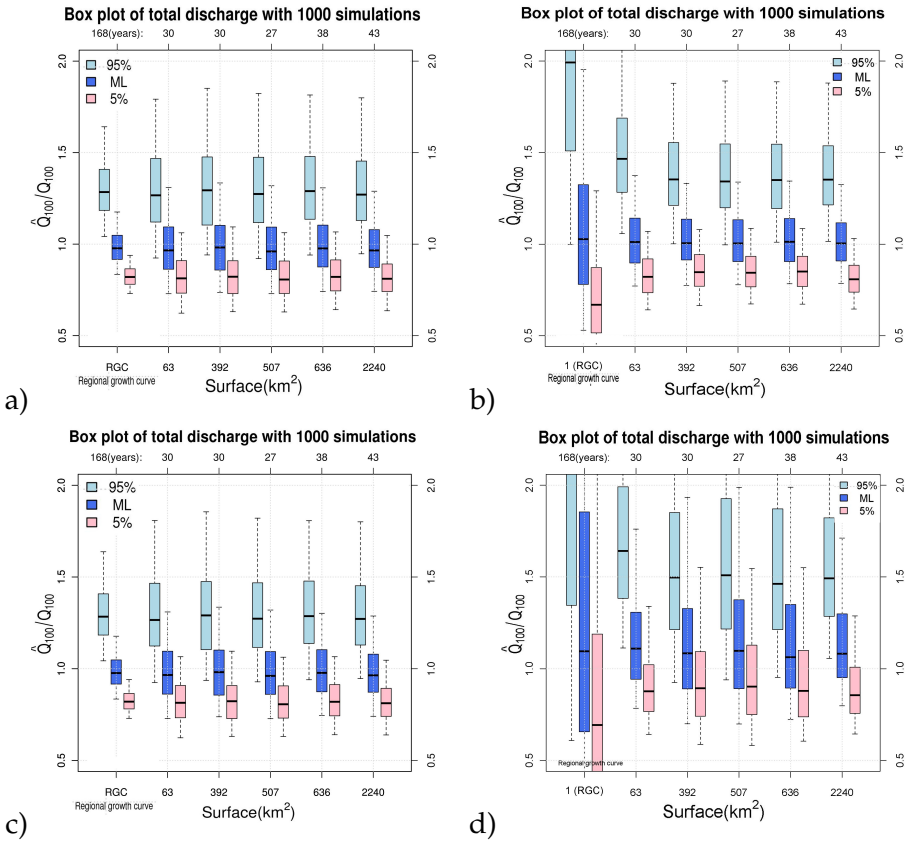


FIGURE 15 – *Dispersion des valeurs du débit centennal estimé $\hat{Q}_{i,ML}^{(100)}$ (respectivement $\hat{q}_{ML}^{(100)}$ pour la distribution régionale réduite), et des bornes des intervalles de crédibilité à 90% $[\hat{Q}_{i,5}^{(100)}, \hat{Q}_{i,95}^{(100)}]$, calculés à partir de 1000 échantillons régionaux simulés et présentant une hétérogénéité de la relation d'index de crue : a) approche de référence (Hosking1997) avec δ= 0.1, b) approche proposée avec δ= 0.1, c) approche de référence (Hosking1997) avec δ= 0.3, d) approche proposée avec δ= 0.3. L'ensemble des valeurs présentées sont divisées par la valeur réelle du quantile centennal $Q_i^{(100)}$ (ou $q^{(100)}$).*

Cas d'hétérogénéités dans les courbes de croissances locales

Les caractéristiques de la région considérée ici diffère de ce qui a été présenté aux paragraphes précédents : la région inclut ici 5 bassin versants jaugés d'une surface identique de 1 km^2, avec dans chaque cas une longueur d'enregistrements de 30 ans (soit 150 ans au total pour les 5 sites considérés). Dans ce cas précis, l'approche proposée ne nécessite pas (et ne permet pas) la calibration de la relation d'index de crue, les séries constituant la région étant déjà ramenées à une surface unitaire de 1 km^2. Les résultats présentés ici correspondent donc simplement à l'estimation de la courbe de croissance régionale (paramètres ξ, α, et κ), et s'affranchissent des incertitudes associées à l'estimation de la relation d'index de crue. Pour l'approche de référence d'Hosking et Wallis, chaque série générée est divisée par sa moyenne pour être correctement redimensionnée et pour permettre à la distribution régionale correspondante d'être estimée.

La figure 16 présente la dispersion des quantiles estimés $\hat{q}_{i,ML}^{(100)}$ et des intervalles de crédibilité associés $[\hat{q}_{i,5}^{(100)}, \hat{q}_{i,95}^{(100)}]$ obtenus avec les deux approches régionales (approche de référence et approche proposée) et dans les trois cas d'hétérogénéité des distributions locales présentés au paragraphe 0.3.3 :

- Le **cas 1** correspond à une région homogène, avec sélection des échantillons ayant une valeur $H_1 \leq 1$ (827/1000),
- Le **cas 2** correspond à une hétérogénéité limitée (dispersion des L_{CV} et $L_{skewness}$ correspondant à des séries d'une durée de 40 ans), avec sélection des échantillons ayant une valeur $H_1 \leq 2$ (826/1000),
- Enfin, le **cas 3** correspond à des données très hétérogènes (dispersion des L_{CV} et $L_{skewness}$ correspondant à des séries d'une durée de 20 ans), et à un seuil $H_1 \leq 3$ pour les échantillons sélectionnés (851/1000).

Ces résultats illustrent de façon très claire les effets significatifs de la présence d'hétérogénéités dans la forme des distributions locales constituant la région : **(i)** les quantiles estimés deviennent plus dispersés autour du quantile réel, **(ii)** la dispersion affecte aussi les limites de crédibilité $[\hat{q}_{i,5}^{(100)}, \hat{q}_{i,95}^{(100)}]$ et la largeur des intervalles de crédibilité augmente. Ces résultats restent relativement similaires quelle que soit l'approche de régionalisation adoptée.

La figure 17 montre par ailleurs que la dispersion des distributions a posteriori estimées (largeur des intervalles de crédibilité) devient rapidement sous-estimée de façon

importante, y compris pour un niveau d'hétérogénéité intermédiaire (**cas 2**). D'après la figure 18 qui présente les distributions de valeurs de $\hat{F}^{disp}_{Q_i^{(100)}}(Q_i^{(100)})$, la sous-estimation atteint, avec l'approche proposée, environ 60% dans le **cas 2** et 80% dans le **cas 3**. Elle est encore plus importante avec l'approche de référence : Ceci est tout à fait logique puisque cette approche de référence reste affectée par l'estimation de la moyenne de chaque série.

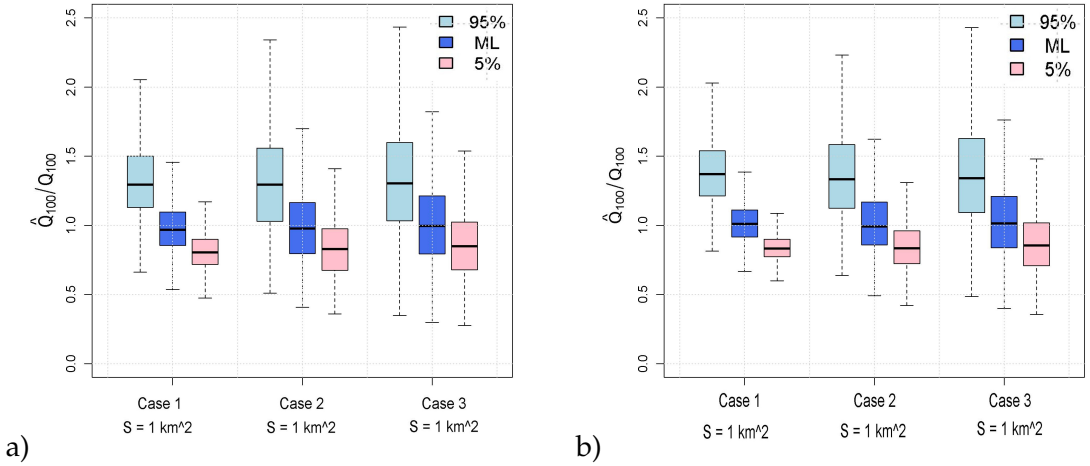


FIGURE 16 – *Dispersion des valeurs du débit centennal estimé $\hat{q}_{i,ML}^{(100)}$, et des bornes des intervalles de crédibilité à 90% $[\hat{q}_{i,5}^{(100)}, \hat{q}_{i,95}^{(100)}]$, pour plusieurs niveaux d'hétérogénéité des distributions locales constituant la région : a) approche de référence, et b) approche proposée. Toutes les valeurs présentées ont été divisées par la valeur réelle du quantile centennal $q_i^{(100)}$*

0.3.6 Conclusions partielles

Cette section a permis de comparer les performances respectives des deux approches d'analyse régionale des crues décrites dans la section 0.2 (approche de référence d'Hosking et approche proposée), dans un contexte où les données impliquées dans l'analyse sont identiques et se limitent à des séries jaugées aux stations hydrométriques.

Une procédure de simulations Monte Carlo a été mise en place de façon à **(i)** comparer la précision d'estimations des quantiles de crue obtenues avec les deux approches, et **(ii)** évaluer la pertinence des intervalles de crédibilité à 90% obtenus avec l'approche Bayésienne MCMC, et **(iii)** évaluer la sensibilité de ces résultats à la présence d'hétérogénéités dans les régions considérées. Deux types d'hétérogénéités ont été distingués : hétérogénéité autour de la relation d'index de crue théorique, et hétérogénéité de la forme des distributions locales.

Les résultats de simulations obtenus permettent d'identifier de façon très claire les fai-

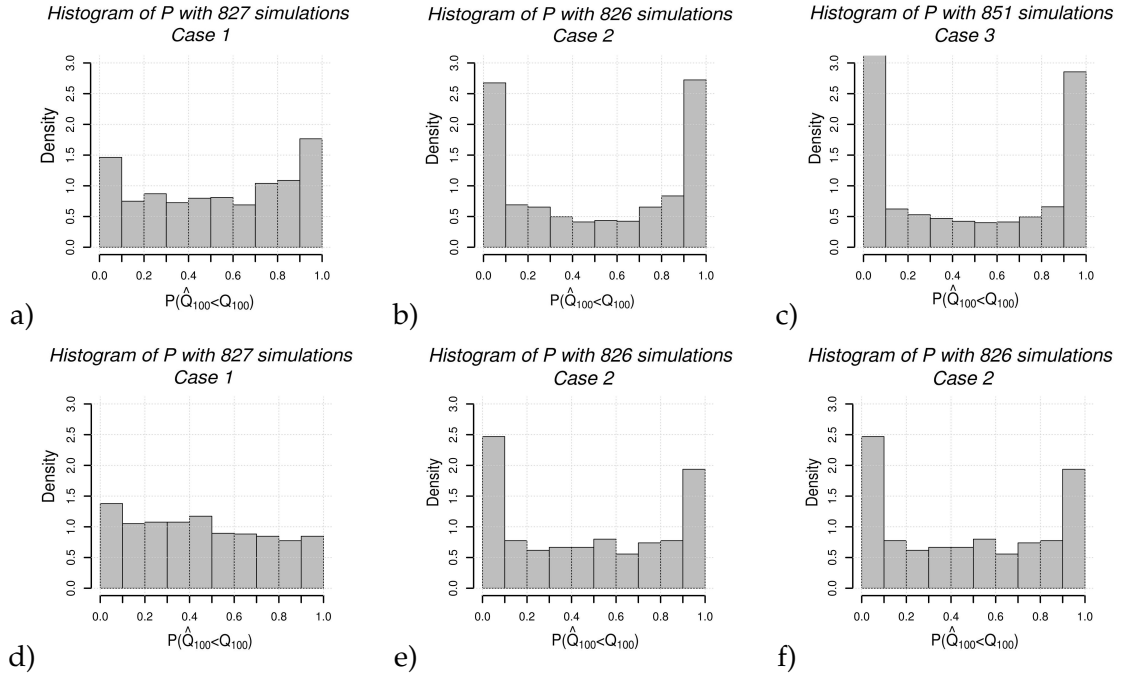


FIGURE 17 – *Distribution des valeurs de* $\hat{F}_{q_i^{(100)}}(q_i^{(100)})$ *obtenues pour les trois cas d'hétérogénéité des distributions locales considérés : a) approche de référence -* ***cas 1****, b) approche de référence -* ***cas 2****, c) approche de référence -* ***cas 3****, d) approche proposée -* ***cas 1****, e) approche proposée -* ***cas 2****, et f) approche proposée -* ***cas 3****.*

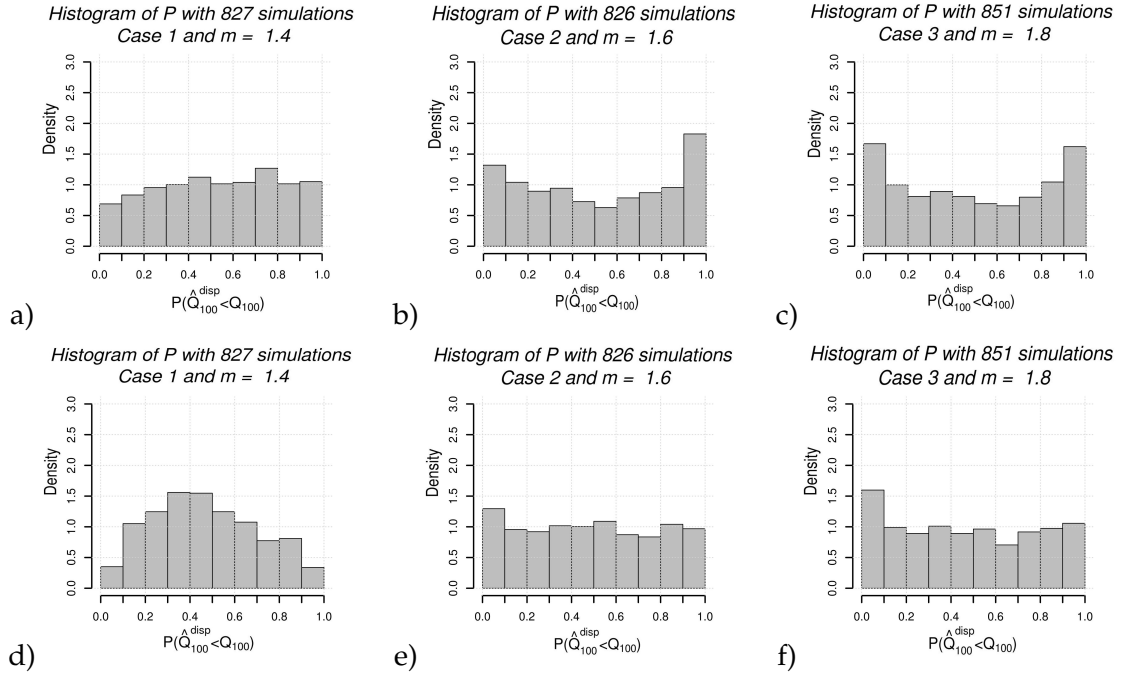


FIGURE 18 – *Distribution des valeurs de* $\hat{F}^{disp}_{q_i^{(100)}}(q_i^{(100)})$ *calculées dans les trios cas d'hétérogénéité considérés : a) approche de référence -* ***cas 1*** *avec m= 1.4, b) approche de référence -* ***cas 2*** *avec m= 1.6, c) approche de référence -* ***cas 3*** *avec m= 1.8, d) approche proposée -* ***cas 1*** *avec m= 1.4, e) approche proposée -* ***cas 2*** *avec m= 1.6, et f) approche proposée -* ***cas 3*** *avec m= 1.8.*

blesses de chacune des deux approches étudiées. L'approche de référence d'Hosking et Wallis montre une réelle sensibilité à la longueur des séries locales disponibles : des séries trop courtes (moins de 40 ans d'observation) génèrent une incertitude sur l'estimation de l'index de crue (moyenne des observations). Cette incertitude se répercute directement sur la précision d'estimations des quantiles et génère également une sous-estimation de largeur des intervalles de crédibilité calculés. De son côté, l'approche proposée montre une certaine sensibilité à la surface des bassins versants considérés : les bassins de petite surface ont moins de poids dans la détermination de la relation d'index de crue, ce qui a pour effet d'augmenter l'incertitude d'estimation pour ces bassins. Cette incertitude est assortie d'un biais d'estimation (tendance à la surestimation des quantiles), qui touche également les intervalles de crédibilité. Cet effet reste néanmoins modéré dans la gamme des surfaces jaugées, et devient nettement plus marquée pour les très petites surfaces (par exemple pour une surface de 1 km^2 correspondant à la distribution régionale réduite). Dès que les surfaces des bassins deviennent plus significatives (plus de 100 km^2 dans le cas présenté), cet effet disparait presque totalement et les intervalles de crédibilité s'avèrent alors parfaitement estimés. Finalement, dans les cas présentés ici où les séries locales considérées ont une longueur d'environ 30 ans en moyenne, le recours à la relation d'index de crue, apparait comme un facteur de stabilisation plus qu'un facteur d'incertitude, et semble conduire à une meilleure qualité d'estimation. C'est un résultat inattendu et qui plaide en faveur de l'approche proposée même lorsque seules des données mesurées sur des sites jaugés sont analysées.

L'introduction d'hétérogénéités dans la relation d'index de crue tend d'une façon générale à dégrader la qualité d'estimation obtenue. En ce qui concerne les hétérogénéités autour de la relation d'index de crue, de façon assez logique l'approche de référence ne se montre pas sensible à ce type d'hétérogénéité, alors que l'approche proposée peut être significativement affectée : augmentation de l'incertitude d'estimation des quantiles et apparition d'un biais (surestimation). Ces effets restent néanmoins limités si l'hétérogénéité est modérée (cas d'un écart-type de 10% autour de la relation théorique).

L'introduction d'hétérogénéité dans la forme des distributions locales apparait nettement plus problématique. Les deux approches sont alors affectées avec des effets similaires : nette dégradation de la qualité des estimations, associée à une importante sous-estimation de la largeur des intervalles de crédibilité (dispersion des distributions a posteriori). La question de l'homogénéité des distributions locales apparait comme une limite potentielle

importante des deux approches régionales présentées, car cette homogénéité ne peut jamais être garantie dans les études de cas réelles.

Enfin, il est également important de rappeler que les résultats présentés ici reposent sur une autre hypothèse qui ne peut pas jamais être vérifiée dans la réalité : celle d'une parfaite adéquation entre les observations aléatoires étudiées et la distribution statistique proposée pour représenter ces observations (ici une distribution GEV). Les résultats produits dans ce section restent par conséquent des résultats théoriques, et c'est pourquoi il s'avère important de les mettre à l'épreuve d'études de cas réelles.

0.4 Application à des études de cas

0.4.1 Introduction

Dans dans la section qui précède, les deux méthodes d'analyse régionale étudiées ont été comparées à partir de données simulées correspondant à des séries d'enregistrements de stations hydrométriques. Ce nouveau section vise désormais à évaluer les deux méthodes dans un contexte d'études de cas réelles, et pour lesquelles des données relatives à des crues extrêmes, parfois recensées sur des sites non-jaugés, sont disponibles. Les deux méthodes ne seront donc plus évaluées ici sur la base de jeux de données identiques, mais à partir des données réellement mobilisables dans chaque cas.

Deux études de cas seront présentées : elles correspondent à la région de l'Ardèche (déjà étudiée par Gaume et al. (2010)), et à celle du Var, dans le sud de la France. Dans chaque cas, une partie de l'information disponible sur les crues extrêmes correspond à des crues historiques identifiées sur les sites jaugés de la région (crue ayant précédé la période d'enregistrement continu). L'analyse menée a été volontairement ciblée sur ces sites, ce qui a permis de comparer les deux approches d'analyse régionale qui nous intéressent à une approche purement locale mobilisant ces données historiques. Dans chaque cas, après une délimitation attentive des contours des régions pouvant être considérées comme suffisamment homogènes et la sélection des données correspondantes, l'analyse a été menée de la façon suivante :

- dans un premier temps les méthodes ont été appliquées aux différents jeux de données disponibles de façon à déterminer une gamme de valeurs raisonnables des paramètres du modèle statistique à appliquer pour les simulations,
- sur la base du modèle statistique déterminé, des simulations ont été conduites suivant le principe présenté au paragraphe 0.3.3, en considérant des jeux de données de

caractéristiques identiques à ceux réellement disponibles. Ces simulations ont permis d'évaluer les performances théoriques de chaque approche au regard de la nature des données disponibles, et également d'évaluer la pertinence des intervalles de crédibilité évalués à l'aide de l'approche Bayésienne MCMC,

– les résultats fournis par les différentes approches à partir des jeux de données réels ont ensuite été analysés au regard des résultats de simulations qui précèdent.

La suite de cette section est organisée en deux parties : dans un premier temps les jeux de données et le travail de délimitation et d'analyse d'homogénéité des deux régions étudiées est présenté, dans une deuxième partie, les résultats de simulations et de l'application des différents approches aux données réelles sont présentés.

0.4.2 Présentation des deux études de cas

La région de l'Ardèche :

Cette étude de cas a déjà été utilisée par Gaume et al. (2010) pour les tests initiaux de la méthode proposée. Elle offre l'avantage de disposer d'un inventaire de crues extrêmes réalisé dans le cadre du programme Européen Hydrate Gaume et al. (2009), et présente par conséquent une opportunité intéressante de comparer les deux méthodes d'analyse régionale étudiées ici.

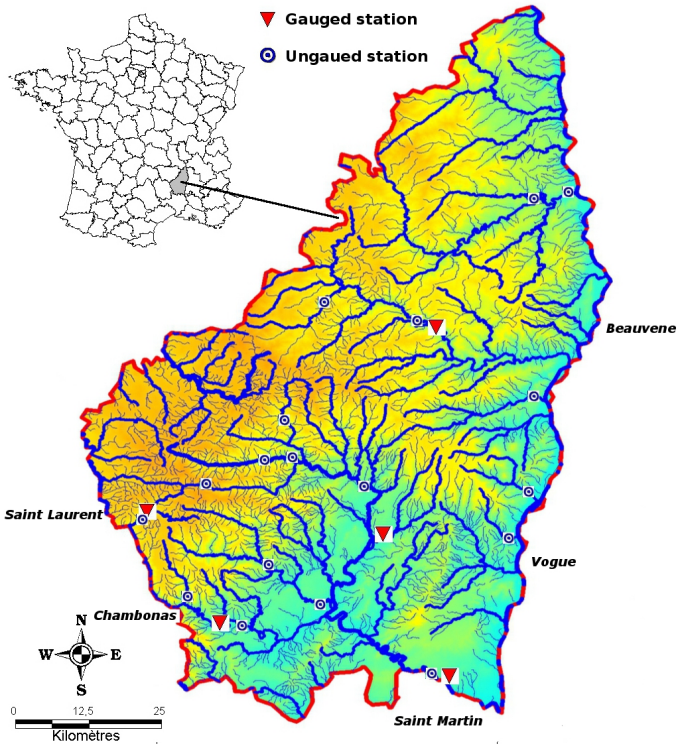


FIGURE 19 – *Localisation des sites constituent la région Ardèche étudiée.*

(i) séries hydrométriques disponibles

Le jeu de données régionales constitué repose sur 5 stations hydrométriques qui ont été sélectionnées pour la qualité et la continuité de leurs enregistrements, ainsi que la longueur des séries correspondantes. Ces stations se situent sur les bassins versants principaux de l'Ardèche (2380 km^2) et de l'Eyrieux (860 km^2), leur situation est indiquée sur la figure 19. L'échantillon régional ainsi constitué inclut 168 années-station d'enregistrements, réparties sur les 5 sites. Les surfaces des bassins versants des sites jaugés varient de 63 à 2240 km^2, et les longueurs des séries d'enregistrements locaux sont relativement homogènes, de 27 à 43 ans (cf. Tableau 1).

TABLE 1 – *Caractéristiques des séries hydrométriques sélectionnées dans la région de l'Ardèche.*

Station	**Rivière**	**Période couverte**	**NP** (années)	**S** (km^2)
Beauvene	Eyrieux	1969-1998	30	392
Vogue	Ardèche	1966-2003	38	636
Chambonas	Chassezac	1971-1997	27	507
Saint Laurent	Borne	1968-1997	30	63
Saint Martin	Ardèche	1963-2005	43	2240

(ii) Homogénéité des distributions locales

L'homogénéité des distributions au sein de cette région a été testée en utilisant la mesure d'hétérogénéité définie par Hosking and Wallis (1997). La valeur H_1 obtenue est de 2.7, ce qui suggère que la région est "peut être hétérogène". Toutefois, les distributions empiriques des séries locales disponibles ne paraissent pas très significativement différentes, en dehors de celle de Saint Laurent qui présente une valeur forte pouvant être considérée comme un horsain à ce stade de l'analyse et de Beauvène sur l'Eyrieux (station située la plus au nord) présentant des débits normés plus faibles sans que l'écart par rapport aux autres séries soit réellement très important. En conséquence, l'hypothèse d'homogénéité des distributions des 5 sites jaugés constituant la région à été retenue pour la suite de l'analyse.

(iii) Homogénéité de la relation d'index de crue

En ce qui concerne l'homogénéité de la relation d'index de crue, la comparaison des moyennes des séries jaugées à celles issues du meilleur modèle statistique utilisant cette relation d'index de crue (voir Figure 21), montre que les écarts constatés sont dans l'ensemble explicables par une simple fluctuation d'échantillonnage. Seul le site de l'Eyrieux s'écarte de façon un peu plus significative de la relation moyenne. Pour cette raison, une

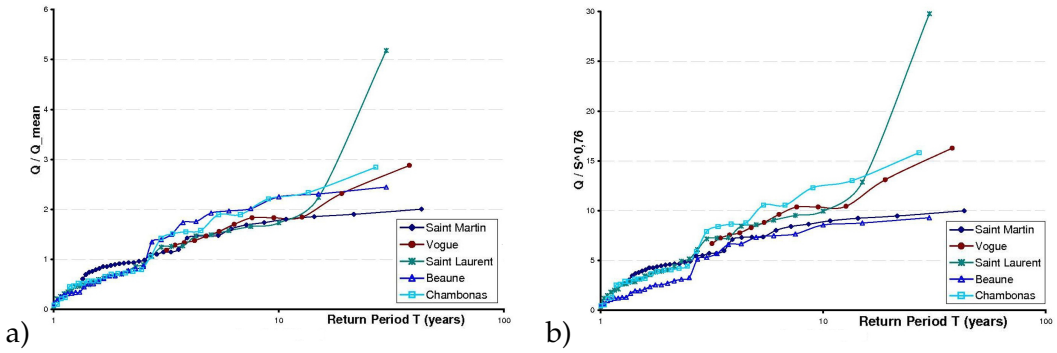


FIGURE 20 – *Distributions empiriques des 5 séries jaugées de la région Ardèche. a) valeurs des débits divisées par la moyenne de chaque série, et b) valeurs des débits normées en utilisant la relation d'index de crue* $S^{0.76}$.

hétérogénéité modérée de la région a été considérée dans les simulations présentées au paragraphe 0.3.5.

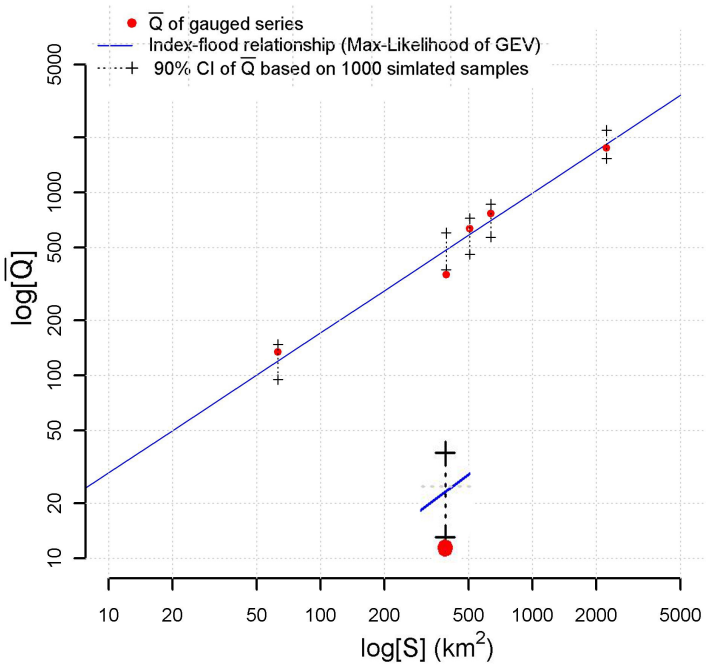


FIGURE 21 – *Comparaison des moyennes des séries jaugées et des moyennes associées de 1000 échantillons simulés (même surface et longueur d'observation) à partir du modèle régional optimisé (distribution GEV et la relation d'index de crue) : les intervalles correspondent aux fluctuations à 90% des valeurs des moyennes des échantillons simulés.*

(iv) Données disponibles sur les crues extrêmes

Les séries hydrométriques de cette région sont complétées par un jeu de 18 crues extrêmes, chacune de ces crues pouvant être considérée comme la plus forte observée sur les

50 dernières années. Parmi ces crues, 4 correspondent à des sites jaugés, et correspondant dans ce cas à la plus forte crue connue sur les 50 années précédant les enregistrements hydrométriques. L'ensemble de ces données sont synthétisées dans le tableau 2.

TABLE 2 – *Détail des crues extrêmes de la région Ardèche connues pour être les plus fortes observées sur les 50 dernières années (ou les 50 années ayant précédé la période d'enregistrement hydrométriques).*

Location	**Rivière**	**Date**	**Q** (m^3/s)	**S** (km^2)	**Période associée**
Pont de Rolandy	Ardèche	22/09/1992	1150	150	50
Aubenas	Ardèche	26/09/1992	2200	480	50
Sauze St.Martin	**Ardèche**	**30/09/1958**	**4500**	**2240**	**50**
Rosieres	Beaune	04/10/1958	1820	210	50
Joyeuse	Beaune	30/09/1958	1000	100	50
Chambon	**Borne**	**30/09/1958**	**100**	**11**	**50**
Vans	Boudaric	03/11/1989	130	6	50
Burzet	Bourges	22/09/1992	350	47	50
Chambonas	**Chassezac**	**21/09/1980**	**3360**	**510**	**50**
Dorne	Dorne	1963	630	78	50
Lamastre	Doux	03/08/1963	970	242	50
Pont de Cesar	Doux	03/08/1963	1500	635	50
Barrage des Collanges	**Eyrieux**	**1963**	**1685**	**343**	**50**
Meyras	Fontoliere	22/09/1992	900	130	50
Meysee	Lavezon	30/09/1960	500	56	50
Rieut ord	Loire	01/09/1992	444	62	50
Pouzin	Ouveze	10/08/1967	700	140	50
Saliouse	Saliouse	1980	300	61	50

La région du Var :

La région du Var est également située dans le sud-est de la France. Cette région comprend cinq cours d'eau côtiers principaux : Argens, Gapeau, Giscle, Siagnole, et Grenouiller. Parmi eux, l'Argens a de loin le plus grand bassin versant, bassin qui inclut plusieurs affluents notables : le Caramy, l'Issole, le Cauron, le Reyran, l'Aille, et la Nartuby. Cette région présente la particularité d'avoir été touchée en 2010 par une crue de forte intensité,

dépassant nettement l'ensemble des crues enregistrées dans les séries hydrométriques, et dont la période de retour s'avère de ce fait délicate à évaluer.

(i) séries hydrométriques disponibles

Les séries hydrométriques disponibles dans la région comprennent 472 observations réparties sur 17 sites jaugés. Bien que certaines de ces séries aient une longueur très limitée (Cogolin, Cabasse, Carcès), la plupart d'entre elles offrent une durée d'observation d'environ 30 ans (voir Tableau 3). Les tailles de bassins versants varient de 44 à 2530 km^2.

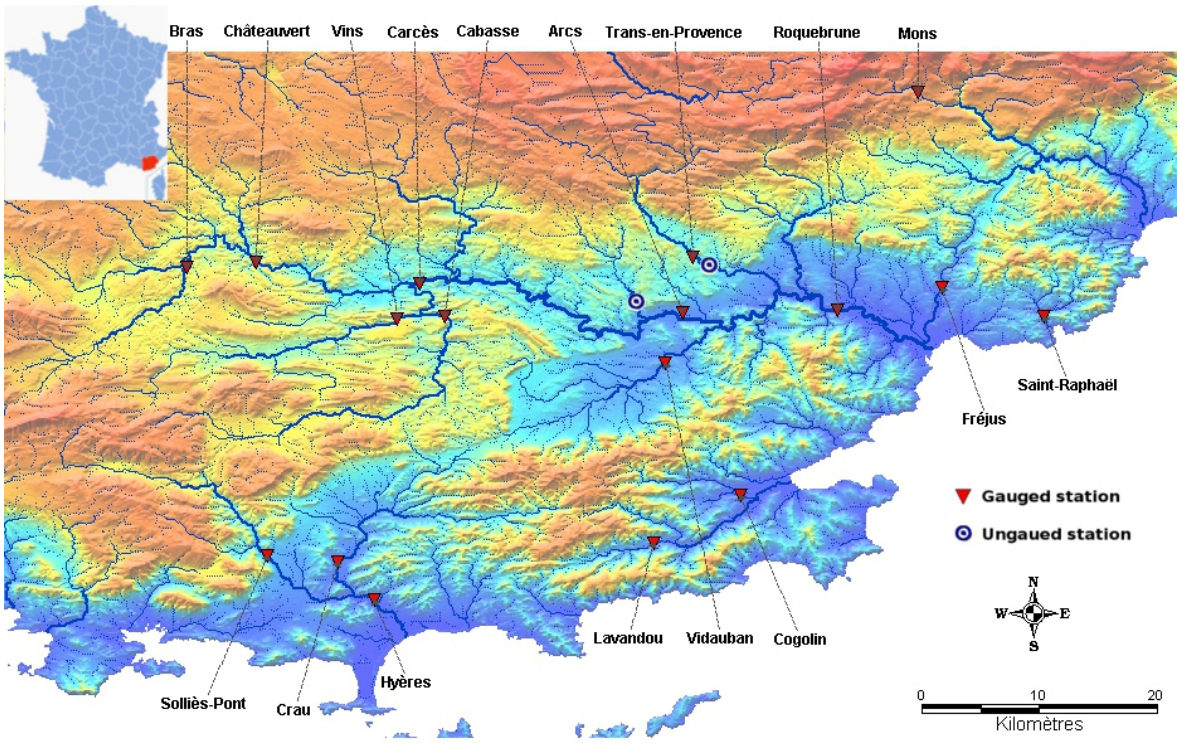


FIGURE 22 – *Localisation des stations hydrométriques présentes dans la région du Var étudiée.*

(ii) Homogénéité des distributions locales

La mesure d'hétérogénéité définie par Hosking and Wallis (1997) indique que la région est hétérogène lorsque les 17 sites jaugés sont tous pris en compte (H_1= 3.78). Par ailleurs, la forme des distributions empiriques locales, non présentée ici, confirme cette hétérogénéité. Sur la base de ce résultat, la région a donc été divisée en deux parties : **(i)** la partie nord, incluant 9 sites jaugés et **(ii)** la partie côtière, incluant 8 sites jaugées. Dans la partie côtière, la valeur H_1 indique toujours une hétérogénéité significative (H_1= 3). La partie nord de la région semble en revanche présenter un niveau d'homogénéité acceptable (H_1= 1.5). Par conséquent, seule cette partie nord de la région a été examinée ci-après.

(iii) Homogénéité de la relation d'index de crue

La figure 23.a montre que deux des sites présents dans la partie nord s'écartent très si-

TABLE 3 – *Caractéristiques des séries hydrométriques disponibles dans la région du Var.*

Région	**Station**	**Rivière**	**Période couverte**	**NP** (années)	**S** (km^2)
partie nord	Bars	Cauron	1975-2003	29	154
	Chateauvert	Argens	1972-2003	32	485
	Vins	Caramy	1972-2003	32	215
	Cabasse	Issole	1975-1988	13	223
	Carces	Argens	1989-2000	12	1181
	Arcs	Argens	1969-2003	35	1730
	Trans-en-Provence	Nartuby	1970-2003	34	190
	Roquebrune	Argens	1972-2003	32	2530
	Mons	Siagnole	1980-2009	30	87
partie côtière	Sollies-Pont	Gapeau	1968-2003	36	169
	Crau	Real Martin	1984-2003	20	277
	Hyeres	Gapeau	1971-2003	33	517
	Lavandou	Mole	1969-2000	32	44.4
	Vidauban	Aille	1971-2003	33	229
	Cogolin	Giscle	1985-2003	16	65.8
	Frejus	Reyran	1970-1991	22	71
	Saint Raphael	Grenouiller	1970-2000	31	48

gnificativement de la relation d'index de crue calibrée : les moyennes des échantillons correspondants, en effet, ne semblent pas être expliquées par une simple fluctuation d'échantillonnage. Les sites concernés sont Bras (154 km^2) et Chateauvert (485 km^2) qui correspondent à la partie amont (nord-ouest) du bassin versant de l'Argens. Par conséquent, deux régions différentes ont été distinguées en fonction de l'approche de régionalisation à appliquer : l'approche de référence d'Hosking et Wallis a été appliquée à partir des 9 sites disponibles, étant donné que l'hypothèse d'homogénéité de la courbe de croissance semble être raisonnable pour ces séries. Pour l'application de l'approche proposée, basée sur la relation d'index de crue, la région a été limitée à 7 sites (sans les deux stations à Bras et Chateauvert, voir figure 23b), en raison de l'homogénéité nécessaire de la relation d'index de crue dans ce dernier cas.

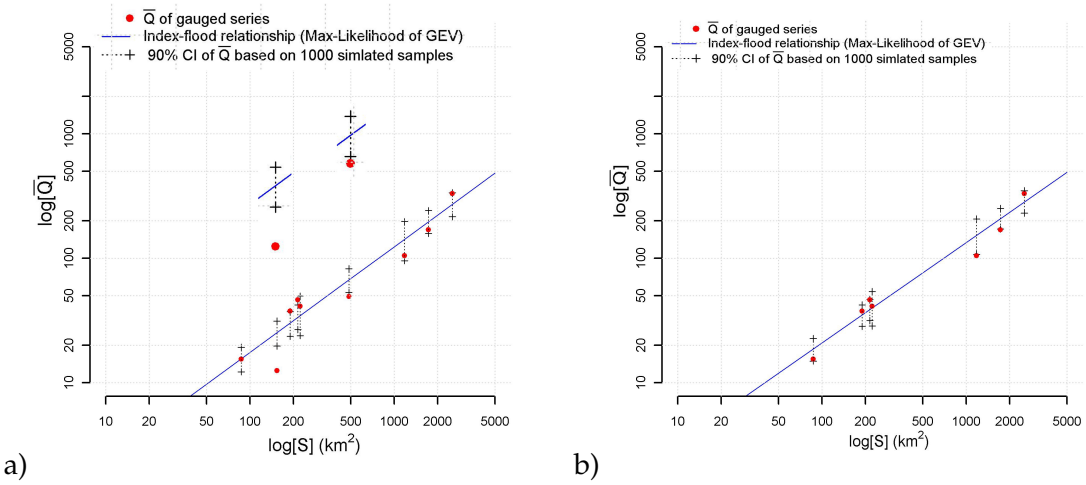


FIGURE 23 – *Comparaison des moyennes des séries jaugées et des moyennes associées de 1000 échantillons simulés (même surface et longueur d'observation) à partir du modèle régional optimisé (distribution GEV et la relation d'index de crue) : a) 9 sites jaugés (partie nord), et b) 7 sites jaugés (sous- partie nord).*

(iv) Données disponibles sur les crues extrêmes

L'information sur les crues extrêmes s'avère très limitée dans la région considérée. En effet, mise à part la crue en 2010, l'information disponible est limitée à deux crues historiques majeures qui se sont produites sur la rivière Nartuby en 1674 et 1827. En outre, la crue de 2010 a détruit une grande partie du réseau hydrométrique et par conséquent les informations disponibles sur cette crue sont limitées. Heureusement, le débit de pointe de la crue a pu être estimé sur plusieurs sites jaugés et non jaugés au sein de la région Payrastre et al. (2012). Ces informations sur les crues extrêmes ont finalement été introduites dans les jeux de données régionales de la façon suivante :

- les débits estimés des trois crues historiques connues sur la rivière Nartuby, en complément de la série jaugée à Trans-en-Provence. La crue de 2010 ayant détruit la station, elle a été considéré comme une crue historique. Chacune des trois crues historiques (1674, 1827 et 2010) prises en compte a été considérée comme étant la plus forte observée sur une période de 120 ans (période historique complète d'une durée de 360 ans avant la période d'enregistrements automatiques),
- la crue de 2010 sur la rivière Florieye a été considérée comme une crue extrême sur site non jaugé. Le débit de pointe de cette crue est l'un des plus élevés enregistrés en 2010. Cette crue a été considérée comme la plus forte observée sur la Florieye sur une période de longueur 300 ans.

TABLE 4 – *Détail des information sur les crues extrêmes prises en compte dans la région du Var.*

Location	Rivière	Date	Q (m^3/s)	S (km^2)	Période associée
Trans-en-Provence	Nartuby	1676	350	196	120
Trans-en-Provence	Nartuby	1827	400	196	120
Trans-en-Provence	Nartuby	06/2010	450	196	120
Draguignan	Florieye	06/2010	490	87	300

0.4.3 Résultats obtenus dans le cas de l'Ardèche

Simulations

Comme indiqué en introduction de ce section, de nouvelles simulations ont été conduites de façon à évaluer les performances théoriques de chacune des approches régionales étudiées, dans le contexte particulier de disponibilité des données spécifique à chaque étude de cas.

Les résultats présentés dans ce paragraphe se basent par conséquent sur des échantillons simulés dont les caractéristiques sont identiques au jeu de données réel de la région de l'Ardèche (168 maximum annuels jaugés sur 5 sites et 18 crues extrêmes recensées). Ces échantillons ont été générés à partir d'une distribution GEV et d'une relation d'index de crue ajustées (maximum de vraisemblance) à l'échantillon des données réelles (β_0= 0.76, ξ_0= 3.34, α_0= 2.24, et κ_0= -0.16). La procédure de simulation a été décrite au paragraphe 0.3.5. Les niveaux d'hétérogénéité introduits dans les séries simulées correspondent à ceux qui paraissent plausibles au regard des données réelles : l'hétérogénéité prise en compte sur la relation d'index de crue reste modérée (δ= 0.1), et celle sur les courbes de croissance locales est également limitée (fluctuation de valeurs de L_{CV} et $L_{skewness}$ correspondant à des échantillons d'une durée de 40 ans). Sur les 1000 échantillons simulés, l'application du test d'homogénéité d'Hosking et Wallis a conduit à retenir 799 échantillons pour lesquels la valeur H_1 est inférieure à 2.

Les résultats présentés par la suite sont focalisés sur deux des sites jaugés de la région : Saint Laurent (63 km^2), et Saint-Martin (2240 km^2). Les approches comparées sont les suivantes :

- **Cas 1** : approche proposée sur la base des 5 séries hydrométriques et des 18 crues extrêmes recensées,
- **Cas 2** : approche de référence d'Hosking et Wallis sur la base des 5 séries hydrométriques,

- **Cas 3** : approche locale avec la série hydrométrique et les crues historique disponibles localement : à Saint Martin l'information historique correspond à une crue extrême connue pour être la plus élevée sur une période de 50 ans ; à Saint Laurent, aucune information historique n'est disponible, et donc ce cas n'est pas pris en considération,
- **Cas 4** : approche locale intégrant uniquement la série hydrométrique disponible (Saint Laurent et Saint Martin).

(i) Qualité d'estimation théorique du quantile centennal

La figure 24 présente de façon synthétique la dispersion des 799 quantiles estimés $\hat{Q}_{i,ML}^{(100)}$ autour de la valeur réelle $Q_i^{(100)}$, selon le bassins versant considéré (Saint Martin ou Saint Laurent), et l'approche d'analyse adoptée.

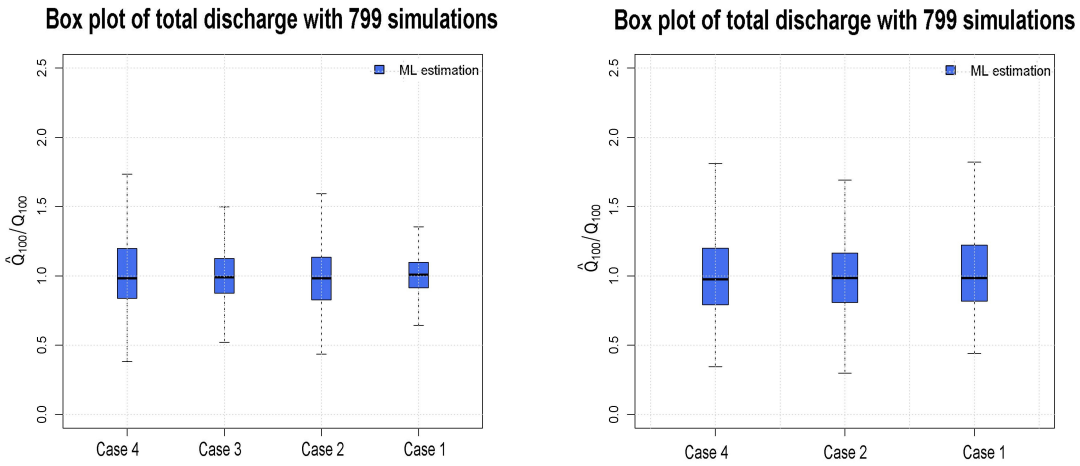
La figure 24.a montre que, dans le cas du bassin de Saint Martin (2240 km^2), les deux approches régionales doivent conduire à des améliorations significatives par rapport à l'approche locale (**cas 4**) sans information historique : la dispersion des estimations obtenues s'avère plus faible dans les deux cas, et ne fait pas apparaitre de biais. Néanmoins, l'approche de régionalisation proposée (**cas 1**), qui mobilise l'information disponible sur les crues extrêmes, semble clairement apporter une amélioration par rapport à l'approche de référence d'Hosking et Wallis (**cas 2**). Par ailleurs, et de façon surprenante, l'approche locale avec incorporation de l'information historique (**cas 3**), conduit à une dispersion des $Q_i^{(100)}$ encore plus faible qu'avec l'approche régionale de référence. Ces résultats peut être attribué aux hétérogénéités introduites dans les simulations qui limitent beaucoup la valeur des approches régionales. Toutefois, dans le cas de l'approche proposée, cet effet négatif lié à l'hétérogénéité de la région est partiellement compensé par la très grande durée d'observation associée aux crues extrêmes.

La figure 24.b, qui concerne le petit bassin versant de Saint Laurent (63 km^2), montre que pour ce bassin la hiérarchie entre les différentes approches testées s'avère significativement différente de ce qui était observé dans le cas du bassin de Saint Martin. La dispersion des $Q_i^{(100)}$ obtenue avec l'approche régionale proposée n'est pas significativement réduite par comparaison avec l'approche locale. L'approche régionale de référence (**cas 2**), pour sa part, aboutit à une légère amélioration, comparable à ce qui était obtenu à Saint Martin. Ce résultat est tout à fait cohérent avec les conclusions du section 0.3 : on retrouve ici les limites de l'approche proposée pour les bassins de petite surface, en liaison avec la plus grande incertitude sur la calibration de la relation d'index de crue pour ces surfaces.

Ces résultats révèlent globalement, de façon assez surprenante, que l'approche régio-

nale de référence n'apporte qu'une plus value très limitée, en raison de l'effet négatif lié aux hétérogénéités certainement présentes dans la région étudiée, hétérogénéités qui sont difficilement évitables dans la réalité. L'approche proposée est également affectée par ces hétérogénéités, mais dans un cas comme celui présenté ici où une quantité d'information importante sur les crues extrêmes non jaugées est disponible, cette approche semble susceptible d'apporter une réelle plus-value, sauf dans le cas des bassins de petite surface. Enfin, sur les sites où une information historique ou préhistorique locale est disponible, une approche locale mobilisant cette information peut s'avérer la meilleure alternative.

Il est important de rappeler ici certaines limites associées à ces résultats de simulation, qui reposent sur des hypothèses fortes. La principale d'entre elles est l'adéquation entre les échantillons de données considérés et le modèle statistique proposé pour les représenter (ici une distribution GEV). Cette hypothèse ne peut jamais être vérifiée à partir des données réelles. La conséquence est que les niveaux d'incertitude présentés ici sont probablement sous-estimés, et que la valeur ajoutée des informations historiques et régionales est de ce fait également probablement sous-estimée.



a) b)

FIGURE 24 – *Dispersion des 799 valeurs du débit centennal $\hat{Q}_{i,ML}^{(100)}$ estimées à partir d'échantillons simulés correspondant aux caractéristiques de la région de l'Ardèche : a) bassin de Saint Martin (2240 km^2), and b) bassin de Saint Laurent (63 km^2). Les valeurs présentées correspondent aux ratios $\hat{Q}_{i,ML}^{(100)}$ / $Q_i^{(100)}$, où $Q_i^{(100)}$ est la vraie valeur du quantile à estimer.*

(ii) Validité théorique des intervalles de crédibilité calculés

Les figure 25 et figure 26 présentent les distributions des valeurs de $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ (cf. paragraphe 0.3.4) aux deux stations de Saint Martin et Sain Laurent. Ces résultats indiquent que la dispersion des distributions a posteriori $\hat{F}_{Q_i^{(100)}}$, et la largeur associée des intervalles de crédibilité à 90%, s'avèrent correctement estimées pour l'ensemble des approches lo-

cales, et nettement sous-estimées avec les deux approches régionales considérées. En ce qui concerne les approches régionales, les résultats diffèrent une nouvelle fois selon la surface des bassins considérés. Pour le bassin de Saint Martin, les figure 25.a et b indiquent une sous estimation plus marquée dans le cas de l'approche de référence (**cas 2**). Le calcul des distributions $\hat{F}^{disp}_{Q_i^{(100)}}(Q_i^{(100)})$ permet d'évaluer la sous-estimation à 40% avec l'approche proposée (**cas 1**) et environ 80% avec l'approche de référence (**cas 2**). Cette hiérarchie est logique au regard des origines de cette sous-estimation, identifiées dans la section 0.3 : principalement l'hétérogénéité de la région dans le cas de l'approche proposée, à laquelle s'ajoute l'incertitude d'estimation des moyennes locales dans le cas de l'approche de référence. Dans le cas du bassin de Saint Laurent, la sous-estimation reste d'un niveau similaire avec l'approche de référence (évaluée à 80%), et dépasse les 80% avec l'approche proposée : nous retrouvons ici la dégradation des performances de l'approche proposée pour les bassins de petite surface.

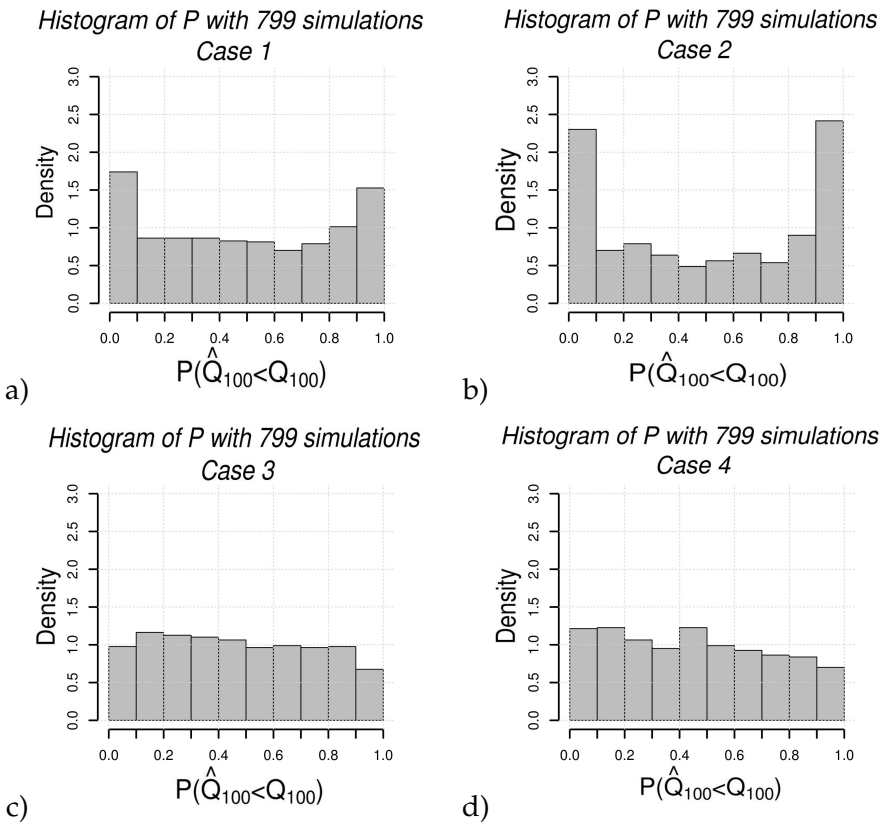


FIGURE 25 – *Distribution des 799 valeurs de $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ obtenues à la station de Saint Martin d'Ardèche : a) approche régionale proposée (**cas 1**), b) approche régionale de référence (**cas 2**), c) approche locale avec données historiques (**cas 3**), et d) approche locale sans données historiques (**cas 4**).*

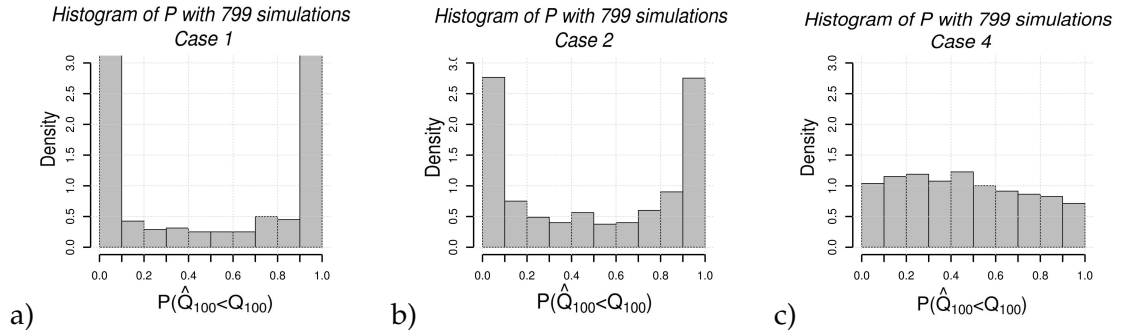


FIGURE 26 – *Distribution des 799 valeurs de* $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ *obtenues à la station de Saint Laurent : a) approche régionale proposée (**cas 1**), b) approche régionale de référence (**cas 2**), et d) approche locale sans données historiques (**cas 4**).*

Application aux données réelles :

Les figure 27, figure 28, ainsi que le tableau 5 présentent les résultats d'estimation obtenus à partir des jeux de données réels disponibles. Les différentes approches présentées sont identiques à celles considérées lors des simulations, à savoir :

- **Cas 1** : approche proposée sur la base des 5 séries hydrométriques et des 18 crues extrêmes recensées,
- **Cas 2** : approche de référence d'Hosking et Wallis sur la base des 5 séries hydrométriques,
- **Cas 3** : approche locale avec la série hydrométrique et les crues historiques disponibles localement : à Saint Martin l'information historique correspond à une crue extrême connue pour être la plus élevée sur une période de 50 ans, à Saint Laurent, aucune information historique n'est disponible, et donc ce cas n'est pas pris en considération,
- **Cas 4** : approche locale intégrant uniquement la série hydrométrique disponible (Saint Laurent et Saint Martin).

Ces résultats mettent en évidence des différences significatives entre les deux approches locales (**cas 3** et **4**) et les approches régionales (**cas 1** et **2**), et ce pour les deux bassins considérés. Dans les deux cas, les ajustements obtenus sur la base des approches régionales sont significativement réorientés par comparaison aux approches locales. Les intervalles de crédibilité obtenus paraissent par ailleurs plus resserrés avec les approches régionales. Toutefois, les simulations ont montré que ces intervalles sont sous-estimés : pour cette raison, les intervalles présentés dans le tableau 5 ont été corrigés à partir des degrés de

sous-estimation identifiés au paragraphe précédent (dernière colonne du tableau). Même corrigés, ces intervalles ne reflètent pas parfaitement la qualité d'estimation identifiée lors des simulations : la hiérarchie entre les approches s'avère différente de la hiérarchie théorique établie paragraphe précédent. Trois explications à ces différences peuvent être avancées :

- **(i)** le caractère particulier des échantillons étudiés ici : présence d'un événement rare dans la série de 30 ans enregistrée à Saint Laurent et absence d'événement fort dans la série de 93 ans disponible à Saint Martin d'Ardèche,
- **(ii)** l'hypothèse d'homogénéité de la région n'est pas valide, ou plutôt l'hétérogénéité prise en compte dans les simulations ne correspond pas à la réalité,
- **(iii)** la distribution GEV n'est pas parfaitement adaptée aux jeux de données manipulés ce qui peut amener à sous-estimer le nombre de degrés de liberté du problème d'inférence, et donc la largeur des intervalles de crédibilité.

Une analyse plus détaillée des résultats permet, dans chaque cas d'essayer de confirmer ou invalider ces hypothèses. Dans le cas de la station de Saint Laurent, la plus forte valeur de la série (crue de septembre 1980) apparait comme un horsain (Figure 27.a). Cette valeur reste la plus forte de la série régionale incluant les données des stations hydrométriques (Figure 27.b), ce qui explique que les intervalles de crédibilité soient réorientés dans la partie basse de l'intervalle initial. Enfin, l'incorporation des données sur les extrêmes confirme, cette fois sur des bassins de taille équivalente, que la valeur n'a été atteinte sur aucun des 5 à 10 bassins de taille similaire sur une durée de 50 ans. Ceci semble confirmer que la période de retour de la crue de 1980 est élevée, probablement supérieure à 100 ans, et donc le caractère particulier de l'échantillon mesuré à Saint Laurent. Ce cas montre finalement que même si l'intérêt des analyses régionales s'avère en théorie limité, il peut s'avérer significatif sur un cas particulier, les données régionales confirment le caractère exceptionnel du débit de la crue de 1980 à Saint Laurent.

Dans le cas de la station de Saint Martin d'Ardèche, les analyse basées sur les échantillons régionaux aboutissent à des estimations nettement supérieures à l'analyse locale, à un degré tel que les intervalles de crédibilité calculés ne se recoupent quasiment pas. La série de 93 ans (en incluant l'information historique) enregistrée à Saint Martin parait finalement improbable au regard des modèles régionaux identifiés. Ceci est de nature à remettre en question l'homogénéité de la région considérée. Toutefois, les recherches his-

toriques menées par Naulet (2002) sur ce site montrent que la plus forte valeur de la série locale (crue de 1958) a été dépassée 4 fois au cours du XIX^{me} siècle en 1827, 1846, 1890 et 1900. Les débits estimés des crues de 1890 et 1827 ont atteint d'après cette source 7000 m^3/s, valeur très cohérente avec l'estimation du quantile centennale issu des analyses régionales. Bien qu'il n'existe aucune certitude, les résultats issus de l'analyse locale pourraient donc également être attribués au caractère particulier de l'échantillon disponible. On peut même pousser plus loin l'analyse en constatant que la période de retour du débit maximum observé à Saint Martin durant les 93 années d'observations est inférieure à 20 ans selon les ajustements régionaux. Or la probabilité de ne pas observer de valeur de période de retour supérieure à 20 ans durant 93 années successives est de $(1 - 1/20)^{93}$, soit environ 1%. La série jaugée à Saint Martin est très peu probable compte tenu du modèle statistique calé. On peut se demander si le modèle statistique, et notamment la loi GEV, sont réellement adaptés au jeu de données étudié.

Finalement, l'application aux données réelles illustre essentiellement l'importance possible de la fluctuation d'échantillonnage et des écarts possibles entre le modèle statistique théorique et les données observées sur un cas d'étude réel. Etant donnée l'impossibilité de connaitre à l'avance le caractère particulier d'un échantillon, ces résultats plaident plutôt pour l'application systématique, à titre d'essai, des méthodes régionales permettant de mobiliser des informations complémentaires, et ceci même lorsque ces méthodes semblent n'apporter qu'une faible plus-value au regard des simulations théoriques.

TABLE 5 – *Comparaison des quantiles centennaux estimés et des bornes des intervalles de crédibilité à 90%* $\Delta CI = \hat{Q}_{i,95}^{(100)} - \hat{Q}_{i,5}^{(100)}$ *estimés à partir des données réelles pour les stations de Saint Laurent et Saint Martin.*

Surface (km^2)	N_{cont} (ans)	N_{hist} (ans)	Cas	$\hat{Q}_{i,5}^{(100)}$ (m^3/s)	$Q_{i,ML}^{(100)}$ (m^3/s)	$\hat{Q}_{i,95}^{(100)}$ (m^3/s)	$\Delta CI/Q_{i,ML}^{(100)}$ (%)	$\Delta CI/Q_{i,ML}^{(100)}$ (%)corr
63	30	0	**4**	411.5	**636.5**	1854.8	226.8	227
	168	0	**2**	423.1	**513.5**	706.5	55.2	100
	168	900	**1**	452.1	**511.5**	587.3	26.4	–
2240	43	0	**4**	3427.8	**4038.4**	6195.3	68.5	69
	43	50	**3**	3778.1	**4297.4**	5790.4	46.8	47
	168	0	**2**	5528.8	**6710.3**	9233.4	55.2	99
	168	900	**1**	5778.7	**6863.9**	8278.4	36.4	50

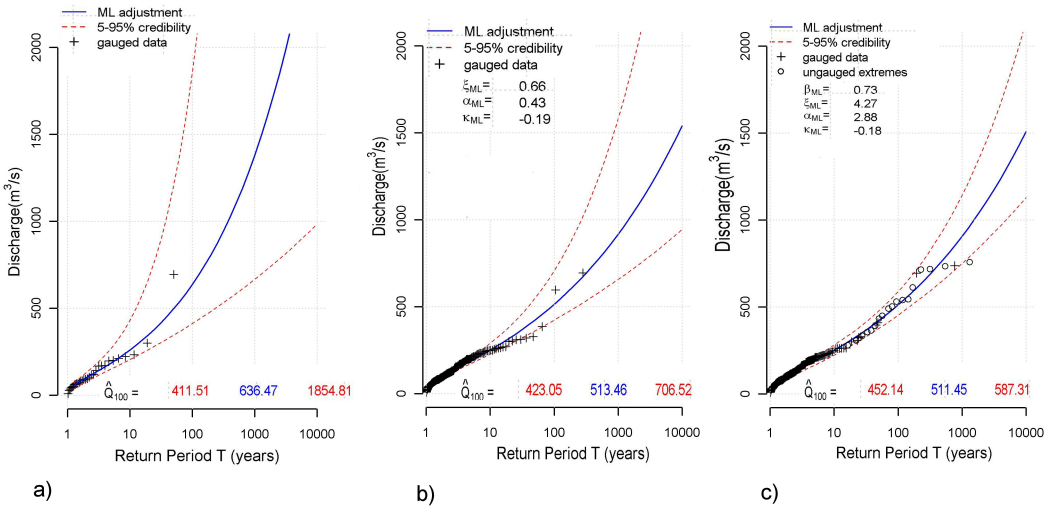


FIGURE 27 – *Forme des distribution GEV locales ajustées (maximum de vraisemblance) et des intervalles de crédibilité associés, à la station de Saint Laurent (63 km²) : a)* ***Cas 4*** *, b)* ***Cas 2****, et c)* ***Cas 1****.*

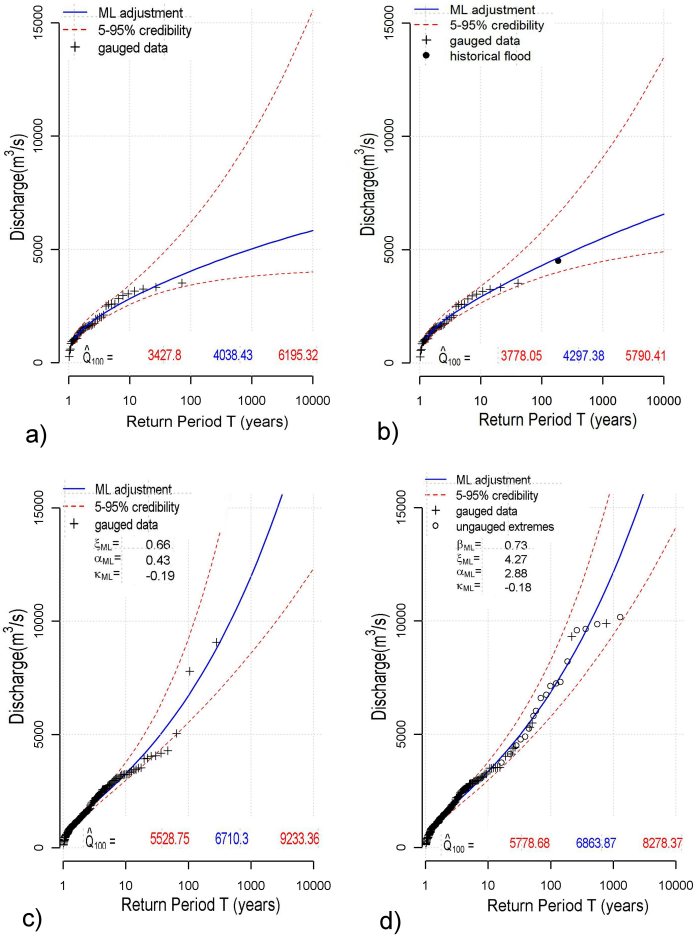


FIGURE 28 – *Forme des distribution GEV locales ajustées (maximum de vraisemblance) et des intervalles de crédibilité associés, à la station de Saint Martin (2240 km²) : a)* ***Cas 4****, b)* ***Cas 3****, c)* ***Cas 2****, et d)* ***Cas 1****.*

0.4.4 Résultats obtenus dans le cas du Var

Simulations :

Comme dans le cas de l'Ardèche présenté précédemment, les simulations relatives à l'étude de cas du Var ont été conduites à partir d'une distribution GEV (ξ_0= 3.34, α_0= 2.24, and κ_0= -0.16) et d'une relation d'index de crue (β_0= 0.77) ajustées à partir des données réelles disponibles. Les séries simulées sont de caractéristiques identiques au jeu de données réel, incluant 9 séries hydrométriques et 4 événements extrêmes, dont 3 correspondent à des crues historiques enregistrées sur la Nartuby à Trans-en-Provence (chacune étant la plus forte observée sur une période de 120 ans), et un au site non jaugé de la Florieye (87 km^2) sur lequel la plus forte crue sur une période de 300 ans est supposée connue.

Au regard des caractéristiques du jeu de données réel, les hétérogénéités introduites restent modérées à la fois sur la relation d'index de crue (δ= 0.1), et sur les formes de distributions locales (fluctuation de valeurs de L_{CV} et $L_{skewness}$ correspondant à des échantillons d'une durée de 40 ans). Sur les 1000 échantillons simulés, l'application du test d'homogénéité d'Hosking et Wallis a conduit à retenir 666 échantillons pour lesquels la valeur H_1 est inférieure à 2.

(i) Qualité d'estimation théorique du quantile centennal

La figure 29 présentée ci-après synthétise les résultats obtenus sur le bassin versant de la Nartuby à Trans-en-Provence (191 km^2), qui présente la particularité d'intégrer une longue chronique de données historiques. Les approches comparées sur cette figure sont similaires à celles présentées aux paragraphes précédents :

- **Cas 1** : l'approche proposée appliquée sur la base des séries hydrométriques régionales (limitées à 7 sites pour des raisons d'homogénéité de la relation d'index de crue), et des 4 crues extrêmes identifiées,
- **Cas 2** : l'approche de référence Hosking et Wallis sur la base des 9 sites jaugés,
- **Cas 3** : l'approche locale basée sur la série locale et l'information historique disponible (3 crues historiques, chacune étant la plus forte observée sur une période de 120 ans),
- **Cas 4** : l'approche locale avec la série hydrométrique disponible.

Les résultats obtenus confirment que les deux approches régionales améliorent considérablement l'estimation de $Q_i^{(100)}$, lorsqu'elles sont comparées à l'approche locale sans information historique. Encore une fois, l'approche proposée, grâce à la valorisation des informations sur les extrêmes présente des résultats sensiblement meilleurs que l'approche

de référence d'Hosking et Wallis. Mais un autre résultat s'avère surprenant : l'approche locale mobilisant l'information historique mène clairement à des estimations nettement moins dispersées que les deux approches régionales. Ce résultat, qui diffère de ce qui a été obtenu sur la rivière Ardèche, doit être relié à la particularité du jeu de données disponibles dans le cas de la région du Var : richesse de l'information historique locale à Trans-en-Provence, et limites de l'information sur les extrêmes disponibles sur les sites non-jaugés (1 seule crue).

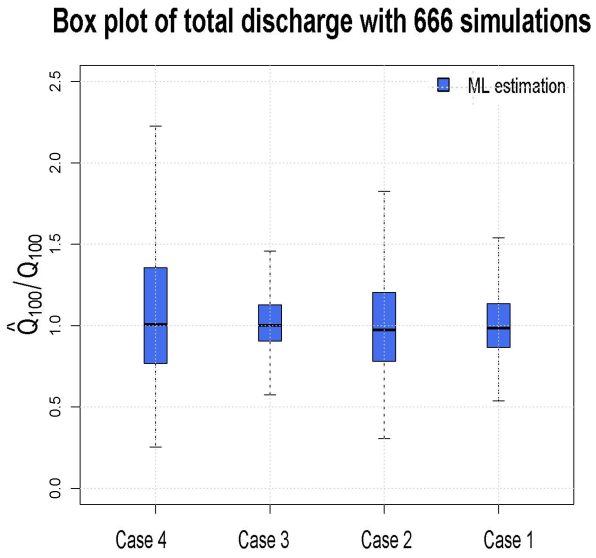


FIGURE 29 – *Dispersion des valeurs du débit centennal estimé $\hat{Q}_{i,ML}^{(100)}$, calculées à partir de 666 échantillons simulés, présentant des hétérogénéités dans la relation d'index de crue (δ=0.1) et les courbes de croissance locales (n= 40 ans), sur le bassin versant de la Nartuby à Trans-en-Provence.*

(ii) Validité théorique des intervalles de crédibilité calculés

La figure 30 présente l'évaluation distributions a posteriori $\hat{F}_{Q_i^{(100)}}$ estimées à Trans-en-Provence, à partir du calcul de $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ (cf. paragraphe 0.3.4). Ces résultats indiquent à nouveau une nette sous-estimation de la dispersion des distributions $\hat{F}_{Q_i^{(100)}}$ (et donc de la largeur des intervalles de crédibilité) avec les deux approches régionales, alors que l'estimation semble correcte avec les approches locales. Le calcul des distributions $\hat{F}^{disp}_{Q_i^{(100)}}(Q_i^{(100)})$ (non présenté ici) permet d'évaluer le niveau de sous-estimation à 70% dans le cas de l'approche régionale proposée. Une nouvelle fois, l'effet de sous-estimation s'avère encore plus marqué avec l'approche régionale de référence, avec la même explication que dans les cas précédents : l'incertitude supplémentaire introduite par l'estimation des moyennes locales, qui n'affecte que l'approche de référence.

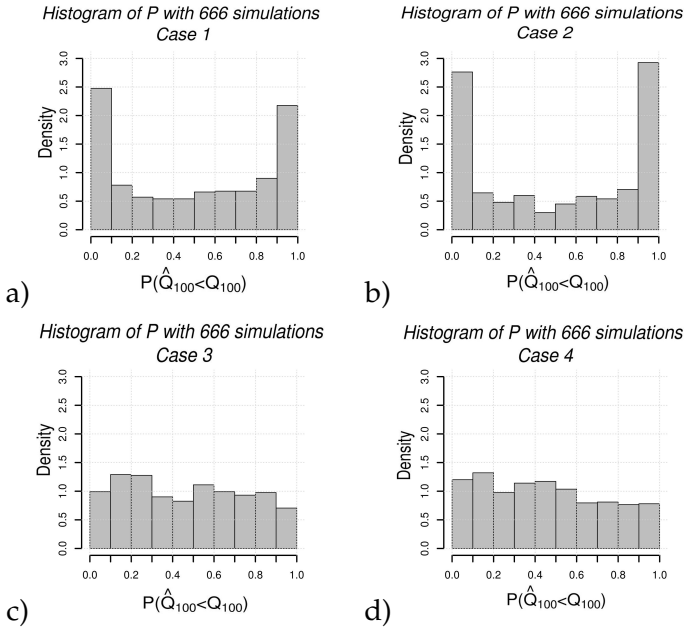


FIGURE 30 – *Distribution des 666 valeurs de* $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ *calculées sur le site de Trans-en-Provence (190* km^2*) : a)* ***cas 1****, b)* ***cas 2****, c)* ***cas 3****, et d)* ***cas 4****.*

Application aux données réelles :

TABLE 6 – *Estimations of the 100-year quantile at Trans-en-Provence gauging stations, including the Maximum Likelihood estimate* $Q_{i,ML}^{(100)}$*, the 90% credibility bounds* $[\hat{Q}_{i,5}^{(100)}\ ;\ \hat{Q}_{i,95}^{(100)}]$*, and associated width* $\Delta CI = \hat{Q}_{i,95}^{(100)}\text{-}\hat{Q}_{i,5}^{(100)}$*, and corrected* ΔCI *according to the theoretical simulation results.*

Surface (km^2)	N_{cont} (ans)	N_{hist} (ans)	Cas	$\hat{Q}_{i,5}^{(100)}$ (m^3/s)	$Q_{i,ML}^{(100)}$ (m^3/s)	$\hat{Q}_{i,95}^{(100)}$ (m^3/s)	$\Delta CI/Q_{i,ML}^{(100)}$ (%)	$\Delta CI/Q_{i,ML}^{(100)}$ (%)corr
190	34	0	**4a**	106.0	**147.0**	410.7	210	207.3
	35	0	**4b**	179.5	**313.7**	1085.6	290	288.8
	34	360	**3**	196.7	**263.5**	395.8	76	75.5
	249	0	**2a**	99.5	**112.2**	136.7	34	57
	250	0	**2b**	157.4	**179.6**	226.1	39	65
	188	660	**1**	174.1	**212.2**	287.0	53	90

Les résultats des estimations obtenues à partir des données réelles sont présentés dans le tableau 3 et 4 (estimation du quantile centennal), ainsi que sur les figure 31 et figure 32 (forme des distributions GEV ajustées et des intervalles de crédibilité). Les différentes approches d'estimation présentées reprennent celles mises en œuvre lors des simulations. Toutefois, pour les **cas 2** et **cas 4** (basées uniquement sur les séries hydrométriques), l'effet

de l'intégration de la crue de 2010 (initialement considérée comme une crue historique) dans la série de Trans-en-Provence a été examinée. Six situations sont donc distinguées :

- **Cas 1** : l'approche régionale proposée,
- **Cas 2a** : l'approche régionale de référence,
- **Cas 2b** : avec la crue de 2010,
- **Cas 3** : l'approche locale avec données historiques,
- **Cas 4a** : l'approche locale sans données historiques,
- **Cas 4b** : avec la crue de 2010.

Ces résultats montrent que l'analyse locale à Trans-en-Provence (**cas 4**) est fortement modifiée par l'incorporation de la crue de 2010, et à nouveau modifiée par l'incorporation des données historiques complètes (intervalle de crédibilité à 90% fortement réduit dans ce dernier cas). Les résultats basés sur les approches régionales, de leur côté, apparaissent encore plus sensibles à l'incorporation de l'information sur les crues extrêmes : dans le **cas 2**, l'estimation évolue de façon très importante lorsque la crue de 2010 est intégrée à la série jaugée de Trans-en-Provence, et évolue à nouveau significativement en cas d'utilisation de l'information complète sur les extrêmes avec l'approche proposée (**cas 1**).

Comme dans le cas de l'Ardèche, la hiérarchie entre les différentes approches appliquées, vue au travers des intervalles de crédibilité calculés, diffère des résultats de simulations. Ceci peut ici encore être attribué, au moins en partie, aux effets de la fluctuation d'échantillonnage, mais aussi éventuellement à la mauvaise adéquation entre le modèle théorique GEV et les jeux de données étudiés, ou à une mauvaise évaluation de l'hétérogénéité de la région.

Le résultat le plus surprenant obtenu ici est la relative incohérence (intervalles de crédibilité disjoints) entre l'estimation obtenue avec l'approche régionale de référence (**cas 2**) et les autres approches (**cas 1** et **cas 3** notamment). L'approche régionale de référence induit visiblement en erreur et conduit à une nette sous-estimation des quantiles de crues. La figure 32.c semble indiquer une certaine incompatibilité entre les séries hydrométriques régionales, le jeu de crues extrêmes, et la distribution GEV. Cette incompatibilité a pu être vérifiée par un test statistique, montrant que le jeu de données régional constitué à partir des séries hydrométriques est extrêmement improbable au regard de la distribution GEV ajustée à partir des données extrêmes (**cas 1**) : la période de retour du débit le plus élevé de l'échantillon régional jaugé ne dépasse pas 20 ans selon la distribution statistique ajustée dans le **cas 1**. Or la probabilité que le quantile centennal ne soit pas dépassé durant

249 tirages aléatoires de variables indépendantes est de $(1-1/20)^{249}$, soit $3x10^{-6}$! Cette incohérence peut avoir plusieurs origines possibles :

- des lacunes présentes dans les séries hydrométriques ont pu conduire à exclure des ces séries les débits de crues les plus importants. Un examen attentif des séries permet toutefois d'exclure cette hypothèse,
- les séries hydrométriques peuvent ne pas être indépendantes, notamment pour les événements extrêmes. La crue de 2010 par exemple a montré que ce type d'événement touche généralement plusieurs sites au sein de la région. De ce fait, le nombre équivalent d'années-stations de la série régionale peut être largement surestimé,
- le processus de génération des pluies extrêmes peut s'avérer non stationnaire, en liaison avec les évolutions du climat. Aucun cumul de pluie comparable aux 450 (mm) enregistrés en 2010 n'a par exemple été enregistré entre 1958 et 2009,
- la région peut être très hétérogène en terme de forme de distributions locales. A ce stade, cette hypothèse reste difficile à vérifier,
- enfin, la distribution parente des séries de données étudiées pourrait s'écarter significativement de la distribution GEV, avec une forte évolution des débits dans une gamme resserrée de périodes de retour, comme semble l'indiquer les périodes de retour empiriques présentées sur la figure 32.c. C'est l'hypothèse la plus probable. L'étude de cas du Var et dans une moindre mesure celle de l'Ardèche seraient alors les premières confirmations d'hypothèses avancées par certains auteurs (Gaume 2006, Rogger et al. 2011) sur les formes possibles des distributions de débits de crues qui pourraient s'écarter notablement des lois usuellement utilisées pour l'inférence statistique en hydrologie.

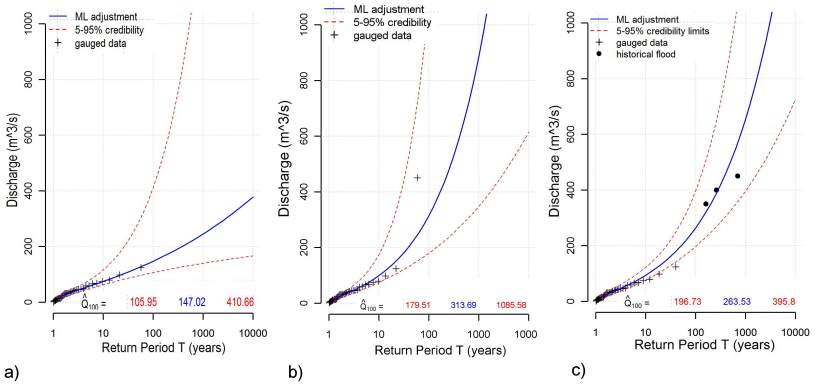


FIGURE 31 – *Forme des distributions GEV ajustées (maximum de vraisemblance), et des intervalles de crédibilité associés, à la station hydrométrique de Trans-en-Provence (190 km^2) à partir des approches locales : a)* ***Cas 4a****, b)* ***Cas 4b*** *avec crue de 2010, c)* ***Cas 3****.*

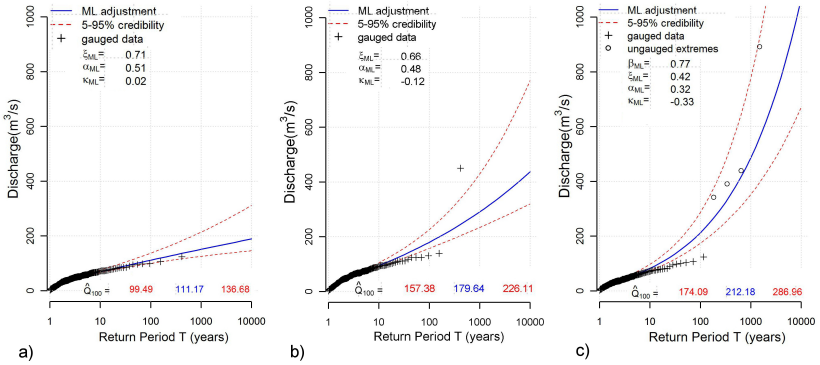


FIGURE 32 – *Forme des distributions GEV ajustées (maximum de vraisemblance), et des intervalles de crédibilité associés, à la station hydrométrique de Trans-en-Provence (190 km^2) à partir des approches régionales : a)* ***Cas 2a****, b)* ***Cas 2b*** *avec crue de 2010, c)* ***Cas 1****.*

0.4.5 Conclusions partielles

Cette section a permis d'illustrer les possibilités et les limites des deux approches d'analyse régionale étudiées, à partir de deux études de cas réelles. Des approches purement locales ont également été appliquées de façon à mieux mettre en évidence la plus value potentielle des approches régionales. L'analyse conduite s'est basée à la fois sur une application aux jeux de données réels, mais également sur des simulations permettant d'évaluer l'influence des hétérogénéités possibles des régions considérées, et d'obtenir une évaluation moins dépendante du caractère toujours particulier d'un jeu de données observées. Les résultats confirment globalement la pertinence de l'approche régionale proposée incluant les données sur les crues extrêmes, pour plusieurs raisons :

- Tout d'abord, dans les deux études de cas présentées, des régions géographiquement cohérentes ont pu être constituées en respectant la double contrainte d'une homogénéité suffisante de la relation d'index de crue, et des formes des distributions locales. Cette double contrainte d'homogénéité peut par ailleurs être vue comme un avantage : elle peut conduire à limiter l'étendue des régions prises en compte (cas du Var) qui sont, de ce fait, hydrologiquement plus cohérentes et moins contestables que dans l'approche régionale classique,
- Ensuite, l'application des méthodes d'analyse régionale a montré que l'approche proposée aboutit dans les deux cas à une qualité d'estimation qui dépasse celle de l'approche de référence, principalement en raison de la longue durée d'observation équivalente représentée par les données sur les crues extrêmes. Par ailleurs, l'approche régionale de référence semble globalement, sur ces études de cas, offrir

une plus-value très limitée par rapport aux approches locales, notamment celles mobilisant les données historiques. Toujours par rapport à une approche locale, l'approche régionale proposée semble susceptible d'apporter une réelle plus-value, en fonction des données sur les crues extrêmes mobilisables. Il faut toutefois se méfier de conclusions trop hâtives basées uniquement sur l'interprétation des intervalles de crédibilité issus de l'approche bayésienne : ces intervalles peuvent s'avérer fortement sous-estimés en cas d'hétérogénéité de la région considérée, et ils sont par ailleurs sensibles à la variabilité d'échantillonnage.

Ces deux études de cas illustrent enfin l'intérêt global des données sur les crues extrêmes, qui peuvent, au-delà de leur intégration dans l'analyse régionale des débits via la méthode proposée, apporter un éclairage intéressant sur les propriétés des régions considérées : présence éventuelle d'hétérogénéités régionales, forme possible de la distribution parente, caractère particulier d'un échantillon local, etc..

0.5 Conclusion et Perspectives

Le but de cette thèse était de poursuivre le développement et de tester plus avant une nouvelle méthode d'analyse fréquentielle régionale des crues permettant d'intégrer des valeurs de débits extrêmes estimées sur des bassins versants non jaugés. Les performances de la méthode proposées ont été évaluées au regard d'une méthode de référence largement utilisée, proposée initialement par Hosking and Wallis (1997). Dans les deux cas, une approche Bayésienne MCMC a été utilisée pour caler les distributions régionales afin d'estimer les paramètres des distributions optimales compte tenu des observations, mais aussi des intervalles de crédibilité. La nouvelle approche proposée permet potentiellement de mobiliser une grande quantité d'information supplémentaire, mais elle nécessite une hypothèse importante si on la compare à l'approche de référence : une relation d'index de crue est censée être respectée dans la région, considérée comme homogène de ce point de vue, comme de celui de la forme de distributions locales. Cette contrainte limite potentiellement l'extension de la région considérée, et représente également un paramètre supplémentaire à estimer qui doit être pris en compte dans le calcul des intervalles de crédibilité issus de l'approche bayésienne.

Dans la première partie de ce travail, un cadre spécifique de simulations Monte Carlo a été développé afin de comparer les deux méthodes et de tester la pertinence de la méthode d'estimation Bayésienne MCMC appliquée dans chaque cas. Auparavant, l'approche

proposée a du être adaptée (en particulier la formulation des vraisemblances) afin de permettre une estimation commune de l'ensemble des paramètres, qu'ils soient relatifs à la forme des distributions locales (ici une distribution GEV) ou à la relation d'index de crue. Les algorithmes MCMC appliqués portent finalement sur 4 paramètres avec l'approche proposée alors qu'ils se limitent à 3 paramètres avec l'approche de référence. L'application des deux approches à un grand nombre de jeux de données simulées, a ensuite permis de comparer les résultats d'estimation obtenus (distributions a posteriori des paramètres et estimation du maximum de vraisemblance), lorsque les deux approches sont appliquées à des jeux de données identiques et limités aux séries hydrométriques jaugées. Les résultats illustrent plusieurs limites des deux approches. L'approche de référence, en effet, apparait sensible à l'incertitude d'estimation des moyennes des séries locales. Cet effet disparaît en utilisant la relation d'index de crue (approche proposée), ce qui limite clairement l'incertitude finale pour les séries de longueur très limitée. L'introduction de la relation d'index de crue et d'un paramètre supplémentaire qui doit être étalonné, apparait finalement comme un facteur de stabilisation plutôt qu'une source d'incertitude supplémentaire. Cependant, il en résulte un certain biais d'estimation et une incertitude accrue de l'estimation pour les petites surfaces de bassins versants. Ces premiers résultats illustrent aussi les limites de chaque approche liées aux hypothèses d'homogénéité effectuées. Un non respect de ces hypothèses conduit à la fois à une plus grande incertitude des estimations, généralement associée à un biais, et également à une sous-estimation parfois forte de la dispersion des distributions a posteriori calculées (et donc de intervalles de crédibilité) avec l'approche bayésienne. Cet effet apparait moins marqué en cas d'hétérogénéités portant sur la relation d'index de crue, que pour celles portant sur les formes de distributions locales.

Dans la deuxième partie de ce travail, les deux approches régionales ont été appliquées à deux études de cas dans lesquelles des informations sur des crues extrêmes non jaugées sont disponibles : la région de l'Ardèche et la région du Var dans sud-est de la France. Des simulations ont été à nouveau menées dans cette partie afin de comparer la qualité d'estimation prévisible dans chaque cas à partir de différentes approches, mobilisant une quantité plus ou moins importante d'information : approche locale, approche locale avec information historique, et les deux approches régionales impliquant ou non des informations sur les crues extrêmes non jaugées. Ces simulations tiennent compte des hétérogénéités supposées des régions considérées. Les résultats montrent que l'incorporation d'informations sur les extrêmes avec l'approche régionale proposée apporte une plus-value assez nette par rapport à l'approche régionale de référence basée sur les séries jaugées.

Cette approche de référence voit ses performances limitées par les effets de l'hétérogénéité et conduit finalement à une qualité d'estimation assez proche des approches locales. C'est un premier résultat surprenant et nouveau qui remet en cause l'intérêt des approches régionales classiquement pratiquées. Ces simulations montrent également que la hiérarchie entre les différentes approches dépend fortement des spécificités de chaque étude de cas : surface du bassin versant considéré, longueurs des séries locale et régionale, quantité d'information historique locale et information régionale sur les extrêmes non jaugés. Enfin, ils confirment que les intervalles de crédibilité calculés par le biais de la procédure Bayésienne MCMC ne peuvent pas être directement considérés comme une indication de la précision de l'estimation des quantiles, puisque la largeur de ces intervalles peut s'avérer très largement sous-estimée, surtout en présence d'hétérogénéités très probables dans la région étudiée.

L'application des différentes approches locales et régionales aux jeux de données réels illustre en dernier lieu les différences très importantes qui peuvent être observées, sur une étude de cas spécifique, par rapport aux résultats des simulations théoriques. Malgré le calcul des intervalles de crédibilité, qui peuvent être évalués et corrigés sur la base des simulations, quelques-unes des estimations obtenues s'écartent de façon assez surprenante des autres : c'est le cas par exemple de l'approche locale à Saint Martin d'Ardèche ou de l'approche de référence d'Hosking et Wallis dans l'étude de cas Var. Ces résultats illustrent finalement l'effet potentiel important de la variabilité d'échantillonnage, et également des différentes hypothèses sur lesquelles reposent les approches comparées, hypothèses qui ne sont pas (ou mal) prises en compte dans les simulations : homogénéité des régions considérées, indépendance des observations, mais aussi et peut-être surtout adéquation entre les jeux de données et la forme théorique des distributions calées (ici une distribution GEV). Dans ce contexte, les études de cas présentées illustrent finalement certains avantages des approches régionales mobilisant les informations sur les crues extrêmes : confirmation de l'occurrence de niveaux de crues qui pourraient apparaître comme aberrants par comparaison aux séries locales (étude de cas Ardèche), et identification d'anomalies dans les séries (par exemple dépendance anormale des séries), ou d'éventuelles limites de l'approche d'inférence sélectionnée (pertinence de la forme théorique des distributions).

Un certain nombre de résultats complémentaires ont été produits en réponse aux questions posées par les rapporteurs de la thèse. Ces résultats, présentés lors de la soute-

nance, apportent notamment des éclairages sur l'influence de certains choix effectués pour conduire les calculs dans le manuscrit :

- définition des valeurs seuils pour les périodes historiques dans le cas de la Nartuby à Trans-en-Provence (annexe 7) ;
- intégration des données historiques anciennes issues de la thèse de Naulet (2002) dans l'analyse statistique des débits de l'Ardèche à Saint-Martin (annexe 8) ;
- élimination des données fortement dépendantes dans l'échantillon régional de l'Ardèche : crues extrêmes ayant touché le même réseau hydrographique à la même date (annexe 9) ;
- pourcentage d'erreur pris en compte pour les données systématiques et extrêmes dans le calcul de la vraisemblance, formulation modifiée (annexe 10) ;
- Choix de la longueur de la série représentée par les crues extrêmes historiques ou non jaugées (annexe 11).

Ces résultats constituent un complément important des travaux menés et ont été ajoutés à la version finale de ce manuscrit (annexes 7 à 11) à la demande du jury.

Deux perspectives principales de recherche peuvent être envisagées à l'issue de ce travail. D'une part, les incertitudes affectant les débits de crues extrêmes reconstitués sur les bassins versants non jaugés n'ont pas été considérées et devraient l'être pour pouvoir porter un jugement parfaitement juste sur l'utilité de ce type de données dans les études statistiques. La prise en compte de ces incertitudes pourrait s'inspirer fortement de travaux antérieurs, menés sur l'utilité de l'information historique dans l'étude fréquentielle des crues, (Payrastre et al. 2011). Les résultats de ces travaux permettent d'être relativement optimistes sur l'intérêt des extrêmes non jaugés. Ces travaux indiquent en effet que le niveau de précision des débits extrêmes estimés et le nombre des extrêmes pour lesquelles une estimation de débit est disponible, ont peu d'influence sur les résultats d'inférence quand la période historique d'observation est fixée. Il reste à confirmer que la même conclusion peut être tirée pour les extrêmes non jaugés lorsqu'ils sont inclus dans les analyses régionales de fréquence des crues. Cela semble être le cas d'après les résultats présentés par Gaume et al. (2010).

D'autre part, les données relatives aux crues extrêmes non jaugées semblent, au travers des études de cas présentées, de nature à remettre en cause les formes théoriques des distributions statistiques classiquement utilisées pour l'inférence statistique en hydrologie, notamment dans le cas des bassins versants exposés aux crues éclair. Les distributions de

débits pourraient présenter de fortes évolutions des valeurs des quantiles pour des plages de périodes de retour limitées (courbes en S sur des représentations semi-logarithmique) du fait des distributions des pluies et de réactions à seuils des bassins versants (Gaume 2006, Rogger et al. 2011). Les premiers résultats présentés dans cette thèse laissent penser que les analyses statistiques régionales incluant des extrêmes non jaugés pourraient confirmer cette théorie si elle est juste. Les outils sont désormais disponibles et efficaces. Reste à les appliquer sur un nombre bien plus important d'études de cas pour confirmer ou infirmer ce que les premières applications de cette thèse ne permettent que de suggérer.

Abstract

This study presents additional developments about a flood frequency approach proposed by Gaume et al. (2010) that aims to incorporate available information on extreme floods at ungauged sites in a regional flood frequency analyses (RFFA). The extreme floods at ungauged sites, indeed, may represent a very large additional record length if compared with gauged series, and therefore may highly improve the accuracy of flood quantile estimates. However, this approach requires to calibrate an index flood relation, which represents a complement of parameterisation of the regional statistical model. The performances and robustness of this approach are tested and compared to the reference approach of Hosking and Wallis (1997). The comparisons are based both on simulations and case studies. The inference procedure is based on a GEV distribution associated with a specific likelihood formulation and a Bayesian MCMC algorithm for parameter estimation. In the first part, the inference results obtained without incorporation of extreme floods are compared based on simulations, with a focus on the effects of possible heterogeneities in the considered regions. In the second part, both approaches are applied to two case studies in the south-east of France : the Ardèche region (168 gauged records at 5 sites and 18 ungauged extremes), and the Var region (9 sites with 249 gauged records and 4 ungauged extremes). The regionalisation approaches are compared to a local analysis. This application finally confirms the very positive impact of the incorporation of information on extreme floods in regional analyses that often enables to outperform the results based on a conventional regional approach.

1 Introduction

1.1 Motivation

Today, many hydrological applications require flood frequency analyses (FFA) such as : (**i**) design of water control structures such as culverts, bridges, urban storm sewers, dams, spillways, and levees ; (**ii**) economic evaluation of flood protection projects ; (**iii**) land use planning and management ; (**iv**) flood insurance assessment, and other hydrologic purposes. If an inadequate understanding of the probabilistic behaviour of extreme floods may have significant economical impacts, underestimation of flood discharges will lead to increased flood risk, while overestimation of flood discharges will lead to increased construction costs. Therefore, a large number of studies have focused on the reduction of uncertainties in flood quantile estimates based on site-specific hydrologic information in the form of annual maximum flow series or peak-over-threshold series.

However, the local data series are often too short to provide reliable estimates of extreme design floods. Without additional information to refine the statistical estimation, frequency analyses result are generally highly uncertain. As an example figure 1.1 presents flood frequency analysis result obtained at the Chambonas gauging station, located on the Chassezac river of the Ardèche region in the south-east of France. The result corresponds to the fitting of a generalized extreme value distribution (GEV) to the local streamflow series (26 years), with the associated 90% credibility intervals. The Bayesian Monte Carlo Markov Chain (MCMC) procedure used to produce these statistical inference results is presented in detail in chapter 2. The width of these intervals confirms that a common problem in flood frequency analysis is the estimation of high return period quantiles : typically 50 to 1000 years. Indeed, based on the available dataset. In this sample, it is only possible to conclude that the 100-year peak discharge of the Chassezac river at Chambonas is comprised between 1700 and 8600 m^3/s, with a 10% chance to be wrong ! And the uncertainty range may be higher since some major hypotheses of the statistical inference procedure

may be wrong : i.e. the maximum annual peak discharges of this river follow a GEV distribution. This example illustrates that the continuous measured discharge series available at a single gauging site are rarely rich enough to provide reliable estimates of high return period quantiles.

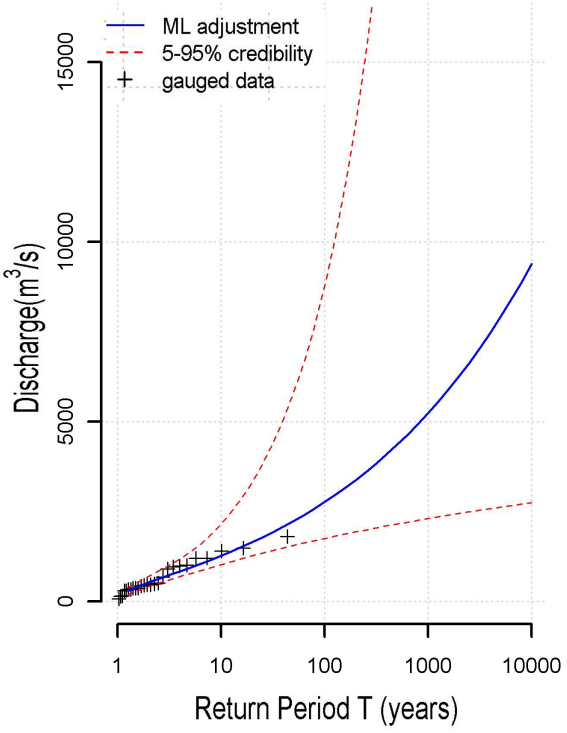


FIGURE 1.1 – *Fit of the GEV distributions to the systematic gauged data set (26 records) of the Chassezac watershed with associated 90% credibility intervals.*

For this reason, hydrologists have tried to make use of alternative sources of information. Several approaches developed consist in "temporal extension" of the data set, based on the incorporation of historical and paleoflood data (Hosking et al. 1985b, Hosking and Wallis 1986a;b, Stedinger and Tasker 1986, Cohn and Stadinger 1987, Gary and Stedinger 1987, Sutcliffe 1987, Minghui and Stedinger 1989, Sheffer et al. 2003, Reis et al. 2005, Werritty et al. 2006, Ribatet et al. 2007b; 2009, Neppel et al. 2010, Payrastre et al. 2011). Figure 1.2 also presents flood frequency analysis result at the Chambonas gauging station based on the local measured series as well as the available historical information. This historical information corresponds to one extreme flood known to be the highest one in a pre-gauging period of 50-year length, with estimated peak-discharge of about 3400 m^3/s. The algorithms developed by Payrastre et al. (2011) were used to account for this historic

data. The incorporation of this additional information leads to a significant reduction of the credibility interval for the 100-year discharge : 2200 to 6700 m^3/s (see Fig 1.2). But even if reduced, the uncertainty on the estimated high return period quantiles remains extremely large.

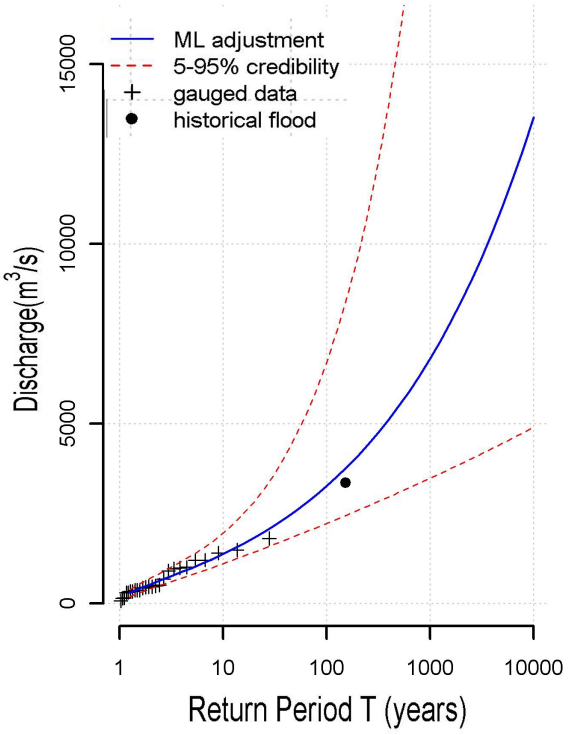


FIGURE 1.2 – *Fit of the GEV distributions to the systematic gauged data set (26 records) of the Chassezac watershed and historical available information with associated 90% credibility intervals.*

Other common approaches used to enrich the data sets available for the statistical inference correspond to a "spatial extension", and consist in merging statistically homogeneous data to build a large regional data sample (Hosking and Wallis 1997, Charles and Stedinger 1999, Ouarda et al. 2001, Kjeldsen et al. 2002, Merz and Bloschl 2003, Seidou et al. 2006, Ribatet et al. 2007a, Norbiato et al. 2007, Wallis et al. 2007, Kjeldsen and Jones 2009). Most of these approaches are based on the "simple scaling" hypothesis or "index flood" principle Dalrymple (1960) : within a statistically homogeneous region all local frequency distributions are identical apart from a site-specific scaling factor, the index flood, that usually corresponds to the mean of each-series. In this context, similar catchments are those that resemble each other in climatology and flood producing mechanisms. They are likely to be geographically contiguous. This approach is described in chapter 2. Figure 1.3 presents the

new inference results for the Chassezac watershed when the data from nearby statistically homogeneous sites are taken into account. Again, the increase of the information content of the analysed dataset, increase of records, leads to a significant reduction of the credibility intervals for the 100-year quantile : 2400 to 4000 m^3/s. This of course, provided that the hypotheses on which the regional flood frequency analysis is based are correct. But still, the computed intervals remain large for relatively small watersheds exposed to flash floods with high inter-annual variability of peak discharges.

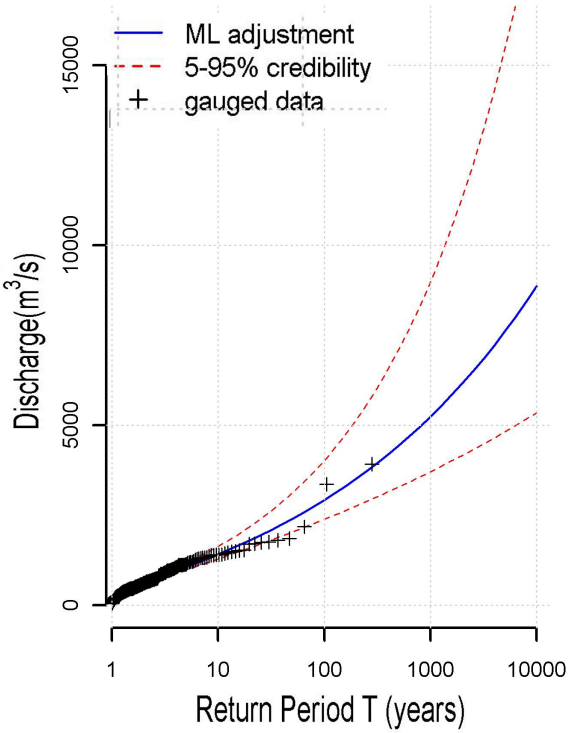


FIGURE 1.3 – *Fit of the GEV distributions to the Chassezac watershed with associated 90% credibility intervals using a regional dataset composed of 168 records at 5 gauging sites based on a reference of Hosking and Wallis approach (1997).*

Finally, the possibility to combine spatial and temporal extension can also be considered. However, Hosking et al. (1985a) discussed problems using historical information with the British flood studies index flood procedure, and illustrate the negligible impact of good historical information on the precision of the computed regional growth curve. Hosking and Wallis (1986a;b) showed that at-site historical and paleoflood data was of little benefit in improving regional growth curves in large regions (20 sites) : if the at-site systematic-record sample mean is used to scale the growth curve, historical and paleoflood data result

in little improvement in flood quantile estimators. Jin and Stedinger (1989) considered regional index flood analyses in which historical and paleoflood data was used to help estimate the site's scale parameter : in their examples historical information substantially improved quantile-estimator accuracy at sites where it was available. See also Stedinger and Tasker (1986), Cohn and Stadinger (1987), who also consider methods for improving flood estimates with historical and paleoflood information.

More recently, Gaume et al. (2010) observed that inventories of extreme floods observed at ungauged sites may also provide valuable additional information in the context of regional flood frequency studies, especially in flash flood prone areas where extreme flood events often affect watersheds of limited spatial extension that are ungauged. Observed extreme floods, indeed, are often known to be the largest ones over a relatively long period even at ungauged locations, and thus may represent a relatively large additional recording length if a corresponding peak discharge estimate can be provided. They could be introduced in regional flood frequency studies in a similar way as historical information available at gauged sites have been used in local flood frequency studies. Until now, such information was only seldom used. Based on the Bayesian approach and MCMC algorithms, Gaume (2010) proposed a methodology to incorporate this information, i.e. extreme discharge values observed at ungauged sites, in a regional flood frequency analyses. The present thesis aims at further developing and validating this new approach.

1.2 Objectives

This thesis presents additional developments of this approach aiming at incorporating available information on extreme floods at ungauged sites in a regional flood frequency analyses. The performances of this approach indeed may be limited by various factors :

- use of an index flood relation in order to estimate index flood values at ungauged sites, and corresponding assumption of homogeneity of this relationship in the considered region. This assumption represents an additional constraint for the delineation of homogeneous regions, and may lead to limit the extent of the considered regions,
- necessary calibration of the index flood relation, which may require additional information if compared to a conventional regionalisation approach, that can be a source of uncertainty,
- as for other regionalisation approaches, important assumption of homogeneity of the growth curves,

– homogeneity of the considered region, tested based on gauged series, and supposed to be to verified at ungauged sites. This requires a careful delineation of the region, based for instance on spatial proximity and/or catchment characteristics considerations. This may again limit the extent of the considered region.

The results presented hereafter aim to illustrate the possible advantages and limitations of incorporating information on extremes, based on the methodology defined by Gaume (2010). The study is conducted in two steps.

In the first step, the performances and robustness of the proposed approach are tested and compared to the conventional regionalisation approach of Hosking and Wallis (1997), based on equivalent datasets limited to gauged series. The comparisons are based on simulations for different dataset characteristics. The inference procedure used is based on a GEV distribution associated with a specific likelihood formulation and a Bayesian MCMC algorithm for parameter estimation. Herein, the Bayesian MCMC inference procedure has been modified in order to calibrate in the same time (**i**) the index flood relation and (**ii**) the regional growth curve, which enables a more reliable estimation of the credibility intervals associated to the estimated peak discharge distributions. The simulations enable to examine several concerns.

– First, the comparison of both approaches (the reference of Hosking & Wallis approach, and the proposed approach) is conducted with homogeneous datasets,
– Then, the sensitivities of both approaches to different sources of heterogeneities are examined : heterogeneities in the index flood relation, heterogeneities in the growth curves. The comparison of both approaches is conducted based on simulations under effect of both factors of heterogeneity.

In the second step, both approaches and the local approaches with or without historical information, are compared based on two cases studies : the Ardèche region (168 gauged records at 5 sites and 18 ungauged extremes) and the Var region (249 gauged records at 9 sites and 4 ungauged extremes) in the south-east of France. These applications aim at evaluating the positive impact of the incorporation of information on extreme floods at ungauged sites. Simulations are again conducted here in order to provide a global evaluation of impacts of supposed heterogeneities in the considered regions, and of the additional

information represented by extreme floods. Finally, the results of application of each approach on real datasets are presented and discussed.

1.3 Thesis Overview

This manuscript is organised as follows : Chapter 2 proposes a review of regional flood frequency analysis methods, and a detailed description of both approaches (the Hosking & Wallis reference approach and the approach proposed by Gaume), that will be compared. Chapter 3 presents the methodology used to compare both approaches based on Monte Carlo simulations, with progressive introduction of heterogeneities in the considered regions and illustrates the simulation results. Chapter 4 shows the results of application to two case studies : the Ardèche region and the Var region in the south-east of France. Conclusions and perspectives are presented in chapter 5.

Regional flood frequency analysis : state of the art 2

2.1 Introduction

Regional flood frequency analyses (RFFA) may be used either (**i**) to propose some estimates of quantiles at ungauged sites, or (**ii**) to improve the estimation of quantiles at locations and/or for return periods for which the available at-site record lengths appear to be too limited. In fact, RFFA is recommended for estimation of flood quantiles at sites with record length less than 5T, where T is the return period of interest (Reed et al. 1999). Design of water control structures such as highway culverts, bridges, airfields and small dams may require estimates of flood quantiles corresponding to 50 to 100-year return periods, whereas design of levees around cities and intermediate to large dams may require quantile estimates corresponding to 100 to 200-year recurrence interval (Chow et al. 1988). Floodway channels are designed for flood events corresponding to 500-year or even 1000-year return period. Therefore, the RFFA may be theoretically often required in design and operation of water resources systems, land use planning and management, flood insurance assessment, protection of populated areas, and others.

For both objectives, RFFA necessitate the delineation of statistically homogeneous regions. But the two types of approaches are based on different assumptions and methodologies :

– If the analysis aims at providing estimates of some characteristics of the local statistical distribution at ungauged sites - typically a quantile value - based on the available gauged series, the homogeneity assumption can be limited to the existence of a common relationship between catchment attributes and the considered characteristics of the growth curve. Most of the time, the relationship is adjusted using standard regres-

sion methods. Most of the recent developments were focussed on the determination of the best suited data pooling methods to delineate homogeneous regions and the test of robust regression techniques (see next section).

– If the objective is to merge data from various gauged sites to provide more accurate estimates of flood quantiles, the growth curves, possibly rescaled to account for the local specificities, should be considered as identical at all sites of the homogeneous region. It is generally supposed that, in a statistical homogeneous region, the local frequency distributions are identical apart from a site-specific scaling factor (simple scaling) called index flood Dalrymple (1960). The index flood, any value proportional to the average of the local distribution, may be estimated by the mean of each locally available series Hosking and Wallis (1997) or related to the characteristics of the catchment.

Both types of RFFA have been used extensively in hydrologic applications and sometimes combined. Stedinger (1989) suggested the use of regional regression to improve the estimation of the local index flood. Burn (1990a;b), Burn and Zrinji (1993), Zrinji and Burn (1994; 1996) propose to improve the index flood method by assigning to each site its own "region of influence" (see next session) defined by a measure of similarity. Tasker and Slade (1994) and Tasker et al. (1996) successfully extended the region of influence approach to regional regression methods.

Fill and Stedinger (1998) proposed to combine in other way the index flood and regional regression concepts using normalized quantile regression (NQR). With the NQR method, the regression model is not adjusted on the local quantile values but normalized quantile values (estimated local quantile, divided by the local estimated mean).

The next section will present the recent advances in RFFA methods used to evaluate characteristics of the growth curve (generally a quantile value) at ungauged sites. The two other sections will be devoted to the presentation of two RFFA methods aiming at merging data from statistically homogeneous sites to reduce the uncertainties in the estimated flood quantiles, namely : the relatively standard method proposed by Hosking and Wallis (1997) and a new approach proposed by Gaume et al. (2010) able to account for extremes estimated at ungauged sites. These two last methods will be used in this thesis, after some modifications described in chapter 3, and in the next two chapters of this thesis.

2.2 Recent advances in RFFA for the estimation of flood quantiles at ungauged sites

2.2.1 Improvement of the regression techniques

Regression can be used to derive simple models that predict the values of any characteristics Φ of the local growth curve : means, standard deviations, quantiles corresponding to various return periods, and normalized index flood as a function of physiographic river basin characteristics and other parameters. Consider the traditional log-linear model :

$$\Phi_i = a + b_1 log\,(Area_i) + b_2 log\,(Slope_i) + .. + e_i \tag{2.1}$$

Where drainage area and slope are two of many possible explanatory variables used to estimate Φ. A challenge in analyzing such models is that with available flood records, one only obtains sample estimates, denoted as y_i, of the streamflow statistics Φ_i . Thus the error ei combines : (**i**) the-sampling-error due to sampling variability and (**ii**) the regression model error. The e_i for different sites are cross-correlated if the y_i are based upon concurrent and cross-correlated flow records. Often these problems have been ignored and standard ordinary least squares (OLS) regression has been employed to obtain reasonable estimators of a model's parameters and its accuracy. Such procedures are satisfactory when the flood records are all sufficiently long.

Stedinger and Tasker (1985; 1986) suggested the use of Generalized Least Squares (GLS) regression as a regional hydrologic analysis methodology. GLS effectively assigns different weights to the observed streamflow statistics yi depending upon the local sampling uncertainties. Tasker et al. (1986), Tasker and Stedinger (1989) illustrated GLS use in regionalization studies for the skewness coefficient and of flow quantiles using physiographic basin characteristics. Advantages of the GLS procedure include more efficient parameter estimates and an unbiased model-error estimator.

An even more efficient approach is GLS regression accounting for correlated data, variable record length and hence sampling uncertainties and distinguishing sampling and regression model errors (Kuczera 1983, Stedinger and Tasker 1986, Tasker and Stedinger 1989, Reis et al. 2005, Griffis and Stedinger 2007). The use of regional GLS regression (Tasker et al. 1986, Tasker and Stedinger 1989, Pandey and Nguyen 1999, Madsen and Rosbjerg 1997, Madsen et al. 2002) and moreover Bayesian GLS estimation methods have been shown

to be more accurate in estimating flood quantiles and statistics than using at-site flood frequency analysis alone (Reis et al. 2005, Micevski and Kuczera 2009).

2.2.2 Delineation of homogeneous regions

In the past, homogeneous regions were often geographically contiguous regions with limits based on political, administrative, or physiographic boundaries. However, this practice is not correct, because the delineation of regions using these factors does not ensure hydrological homogeneity. Consequently, new regionalization methods have seeked for similarity between sites by examining catchment attributes, geographical location and at-site flood statistics. Examples of approaches to regionalization include (**i**) the method of residuals (MOR) (Thomas and Benson 1970, Wandle Jr. 1977, Glatfelter 1984, Choquette 1988) ; (**ii**) the canonical correlation analysis (CCA) (Cavadias et al. 2001, Ouarda et al. 2001) ; (**iii**) the region of - influence (ROI) approach and its extensions (Burn 1990a;b, Zrinji and Burn 1994) ; (**iv**) the hierarchical approach and its extension to ROI framework (Gabriele and Arnell 1991, Zrinji and Burn 1996, Haddad and Rahman 2011) ; and (**v**) cluster analysis (Hosking and Wallis 1997, Heinz and Stedinger 1998, Charles and Stedinger 1999, Ouarda et al. 2001, Kjeldsen et al. 2002, Javelle et al. 2002, Merz and Bloschl 2003, Seidou et al. 2006, Rao and Srinivas 2006b;a, Ribatet et al. 2007b;a, Norbiato et al. 2007, Wallis et al. 2007, Kjeldsen and Jones 2009).

- **the MOR :** In the MOR approach to RFFA, sites are grouped in regions according to the distance (residual) from a regression model relating flood quantile at each gauged site to the characteristics of the corresponding watershed, adjusted on the global set of sites. This approach has been widely used by the United States Geological Survey (USGS) for regionalization. It delineates regions in a rather arbitrary manner and the regions are often arranged to be coincident with recognized geographic and/or hydrologic boundaries, political or administrative areas. Therefore, the regions delineated using this approach, are likely to contain watersheds with a variety of geomorphologic characteristics for which flood-frequency characteristics may not be comparable (Wiltshire 1986, Bhaskar and O'Connor 1989).
- **the CCA :** In the canonical correlation based approach (Cavadias et al. 2001) to RFFA, basins are represented as points in the spaces of pairs of uncorrelated flood-related canonical variables and pairs of uncorrelated basin characteristics canonical variables to examine similarity in the corresponding point patterns. The problem with this

approach to regionalization is that the pattern recognition is based on a subjective visual judgement, and that patterns may not be found (Bobee and Rasmussen 1995).

- **The ROI :** The ROI approach Burn (1990a;b) allows each site to have its own region, rather than having fixed regions in the standard approach. The ROI of a target site consists of those sites in the study region whose distance to the target site in a weighted multidimensional attribute space does not exceed a chosen threshold value. The selection and weighting of variables (and sites) is one of the problems for which no strict mathematical solution is available Bobee and Rasmussen (1995). It becomes a point of concern as the number of attributes available for the analysis increase.
- **The hierarchical approach :** For fixed regions, Gabriele and Arnell (1991) proposed hierarchical approach to RFFA, which explicitly accounts for different spatial variability in different flood characteristics. As an example, skewness of annual maximum flood data is assumed to be constant over a larger area than the coefficient of variation (CV), which in turn is assumed to vary more progressively over space than the mean annual flood. Therefore, more sites are used to estimate the parameters of the statistical distribution controlling skewness than the number used to derive the parameters determining the CV. Zrinji and Burn (1996) incorporated the concept of hierarchical approach into the ROI framework by defining a set of ROIs for each site as opposed to a single ROI. The hierarchical ROI approach was found to perform well for the estimation of higher order moments. It was found that the hierarchical ROI approach improved flood estimates in the extreme range. Tasker et al. (1986) compared five different methods for developing regional regression models to estimate flood quantiles at ungauged sites in Arkansas. The methods looked at traditional flood estimation regression approach, multivariate techniques of cluster, discriminative analysis and the ROI approach based on geographical and catchment attribute space where the n gauged sites with the smallest distance make up the ROI for site i. The results concluded that the ROI approach outperformed the other methods, having the lowest root mean square error.

Recently, Eng et al. (2005) used different ROI approaches for estimating the 50-year average recurrence interval (ARI) flood quantile at ungauged sites in the south-eastern United States. Ordinary Least Squares regression (OLS) was used to regress flood statistics against catchment attributes for each ungauged site based on data from ROI containing the n closest gauging sites in both geographical (GROI) and catchment attribute space (CROI).

Model performance was based on the prediction errors from independent testing sites (cross-validation). This test showed for both the ROI approaches, that setting a threshold for the definition of the region based on a number n of sites rather than a distance in the phase-space of attributes led to better results. They also indicated that GROI procedure performed better than CROI.

Merz and Bloschl (2005) combined spatial proximity and catchment attributes to compare external drift kriging, georegression, and another variant of the ROI approach. They performed the assessment using a jackknife comparison of at-site estimated regionalized flood quantiles for 575 Austrian catchments. Consistently with the results obtained by Eng et al. (2005), the methods that only used catchment attributes appeared to perform relatively poorer to the methods that used geographical proximity. The ROI used in this study was then combined with multiple regressions. Merz and Bloschl (2005) were able to demonstrate that when spatial dependency was incorporated, the ROI showed less random errors.

Eng et al. (2007) proposed a hybrid ROI (HROI) which combined the GROI and CROI in a Generalized Least Squares (GLS) regression framework. They applied this method to 1091 gauged sites in the south-eastern part of the United States to estimate the 50-year flood quantile. Their study revealed that the HROI yielded smaller root mean square estimation errors and produced fewer extreme errors often found in either GROI or CROI. From this study it was concluded that for 50 years ARI flood quantile, the similarity with respect to catchment attributes was important, however it was incomplete and that the consideration of the geographical proximity of the sites provided a useful surrogate for characteristics that were not included in the analysis. Eng et al. (2007) went on to also present an enhanced GLS regression and ROI framework that is based on a "leverage-guided" ROI. They applied their method to 996 gauged sites in the south-eastern part of the United States. This new leverage-guided ROI regression provided improvements in terms of lower root mean square errors.

Overall, the uncertainty associated to estimate characteristics of growth curves at ungauged sites remain important, despite the real improvements in the methods in the recent years. These methods are generally evaluated through a cross-validation procedure : one gauge site is extracted in turn from the available data set for validation purposes. The regionally estimated characteristics (generally a quantile) are compared to the locally estimated ones. These local estimations are of course affected by sampling fluctuation which sets the maximum possible accuracy of the regional estimated quantiles to a generally low value,

especially if high return period quantiles or high order moments are considered. An alternative objective of RFFA is the reduction of the uncertainties associated to the estimated quantiles and moments, considering that merging various partly independent samples in a statistically heterogeneous region should inevitably reduce the uncertainty associated to sampling fluctuation in a statistical inference procedure. This is the objective of the two methods presented in the next sections that have been further evaluated in the work presented herein.

2.3 The RFFA proposed by Hosking and Wallis (1997)

2.3.1 General principle

The Hosking & Wallis procedure is based on the index-flood principle (Dalrymple 1960) : all local discharge series have identical statistical distributions in an homogeneous region apart from a site specific scaling factor : the index flood. This principle can be expressed by the equation :

$$Q_i^{(T)} = \mu_i q^{(T)} \tag{2.2}$$

Where i denotes a given site within the region (i= 1,..,s), s the total number of sites of the region, $Q_i^{(T)}$ is the quantile of return periord T at the i^{th} site, $q^{(T)}$ is the regional reduced quantile and μ_i is the index flood (or scaling factor). The index flood can be any value proportional to the average of each local distribution.

Usually, when only data from gauged sites are considered, the index flood can be estimated by the at-site sample mean (Hosking and Wallis 1997, Castellarin 2005, Castellarin et al. 2007). Let consider a regional sample **D** including s series $Q_{i,j}$ of annual maximal discharges recorded at gauged sites i=(1,..,s), and of respective length n_i (j=(1,..,n_i)). Let also consider μ_i as being the index flood value at site i, the index flood is naturally estimated by :

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Q_{i,j} \tag{2.3}$$

The quantile estimates at site i are obtained by combining the estimates of μ_i and $q^{(T)}$:

$$\hat{Q}_i^{(T)} = \hat{\mu}_i \hat{q}^{(T)} \tag{2.4}$$

This index flood procedure is based on the following assumptions : (**i**) observations at any given site are identically distributed ; (**ii**) observations at any given site are serially independent ; (**iii**) observations at different sites are independent ; (**iv**) frequency distributions at different sites are identical apart from the scaling factor ; and (**v**) the mathematical form of the regional growth curve is correctly specified.

The first two assumptions are plausible for many kinds of data, particularly for annual totals or extremes. It is a basic assumption of most flood frequency analysis methods that the events observed in the past are likely to be typical of what may be expected in the future (identically distributed). The last three assumptions are unlikely to be satisfied by environmental data. Correlation between nearby sites may be expected for many kinds of data. In addition, the last two assumptions will never be exactly valid in practice. At best they may be approximately attained, by careful choice of the limits of the considered region and of a frequency distribution that is consistent with the data.

The regional analysis procedure proposed by Hosking & Wallis aims to facilitate the two important steps of the RFFA based on the index flood principle : the delineation of statistically homogeneous regions and the identification of the best suited regional growth curve on the basis of the available samples. It includes both a measure of heterogeneity of the growth curves for delineation of the region, of and a goodness of fit measure based on L-Moments for the choice of an appropriate frequency distribution.

2.3.2 Inference procedure based on the L-Moments

The samples available in hydrology have generally a limited size, even in regional analyses. This poses some difficulties for the estimation of the parameters of statistical distributions based on the samples : the statistical inference. Some methods, and especially the widespread method of moments, appear to be very sensitive to sampling variability and particularly to the possible presence of outliers in the samples (large or high return period values). This is notably critical when 3-parameter distributions are calibrated and high order moments have to be computed. For this reason, Hosking and Wallis (1997) suggested to base the inference on the computation of the so-called L-Moments : which do not rely on power elevation of the available data, giving too much importance to the highest values, but on linear combinations of the ordered values (L standing for linear). To compute the L-Moments, the dataset has first to be sorted in increasing order Q_1, Q_2,.., Q_n. The expressions of the first four L-Moments are the following ones, with i the rank of the data in the sorted set.

$$\lambda_1 = \frac{1}{C_n^1} \sum_{i=1}^{n} Q_i \tag{2.5}$$

$$\lambda_2 = \frac{1}{2C_n^2} \sum_{i=1}^{n} \left(C_{i-1}^1 - C_{n-1}^1 \right) Q_i \tag{2.6}$$

$$\lambda_3 = \frac{1}{3C_n^3} \sum_{i=1}^{n} \left(C_{i-1}^2 - 2C_{i-1}^1 C_{n-1}^1 + C_{n-1}^2 \right) Q_i \tag{2.7}$$

$$\lambda_4 = \frac{1}{4C_n^4} \sum_{i=1}^{n} \left(C_{i-1}^3 - 3C_{i-1}^2 C_{n-1}^1 + 3C_{i-1}^1 C_{n-1}^2 - C_{n-1}^3 \right) Q_i \tag{2.8}$$

Usually, the computations are based on L-moment ratios rather than on the rough value of the moment for orders larger than 2 : $L_{CA} = \lambda_3/\lambda_2$ is called the $L_{skewness}$ and $L_K = \lambda_4/\lambda_2$ is called the $L_{kurtosis}$. L-Moments ratios take their values within the interval [-1, 1]. $L_{CV} = \lambda_2/\lambda_1$.

2.3.3 The Hosking and Wallis homogeneity test

Homogeneity test is an important component of the regional flood frequency analysis, particularly the methods based on the index flood assumption. Hosking and Wallis (1997) have proposed a statistical test based on L-Moments ratios for testing the heterogeneity of the regions. The test relies on the comparison of the between-site variation in sample L_{CV} (L coefficient of variation) with the variation that can be expected due to sampling fluctuation in an homogeneous region.

Again consider a regional sample **D** including s sites, with site i having record length n_i and sampling $L_{CV}^{(i)}$. Denote by L_{CV}^R the regional average L_{CV}, computed as follows :

$$L_{CV}^R = \frac{\sum_{i=1}^{s} n_i L_{CV}^{(i)}}{\sum_{i=1}^{s} n_i} \tag{2.9}$$

A weighted standard deviation V of the at site sample L_{CV} can then be computed :

$$V = \sqrt{\frac{\sum_{i=1}^{s} n_i \left(L_{CV}^{(i)} - L_{CV}^R \right)^2}{\sum_{i=1}^{s} n_i}} \tag{2.10}$$

To evaluate the likelihood of the value V computed on the available sample, a kappa distribution is fitted to the regional sample (L-Moments method). 500 regional samples comparable in size and structure to the available sample are generated using this distribution and the V value corresponding to each simulated sample is computed to build the

statistical distribution of V due to sampling fluctuation according to the size and characteristics of the regional sample. This distribution appears generally to be close to a normal distribution with mean φ_V and standard deviation σ_V. The critical values of the H_1 measure are determined as follows :

$$H_1 = \frac{(V - \varphi_V)}{\sigma_V} \tag{2.11}$$

The region is acceptably homogeneous if H_1<1 (H_1 has about 25% chance to be exceeded according to the normal distribution), possibly heterogeneous if 1$\leq$ H_1<2 (H_1 has 5% chance to be exceeded if equal to 2), and definitely heterogeneous if $H_1 \geq$2. However, measurements (rainfall and stream gauge measurements) often contain additional variability due to : gauges being moved during the many years of operation, frequent change of operators and level of diligence in timely measurement, missing data arising from inconsistent reporting, lack of attention measurement precision, instabilities and inaccuracies in the stage-discharge relation for stream gauges (Wallis 1997, Wallis et al. 2007, Castellarin et al. 2008). Wallis et al. (2007) suggested therefore releasing the constraints of the test : H_1<2 could be considered as acceptably homogeneous, while $H_1 \geq$3 (H_1 has 0.5% chance to be exceeded according to the normal distribution) is an indicator of heterogeneity. These threshold values have been retained hereafter. Although the L-Moments method is not the selected method used for statistical inference, the test proposed by Hosking & Wallis will be used to test the statistical homogeneity of the regional samples in this thesis.

2.3.4 The goodness of fit test for the regional distribution

Hosking and Wallis (1997) have also proposed a test based on the $L_{kurtosis}$ values of the samples to select the best suited type of statistical distribution. Alike the homogeneity test, the goodness of fit test compares a difference - between the $L_{kurtosis}$ (L_K^R) of the regional sample and the theoretical kurtosis of the calibrated model of the type DIST (L_K^{DIST}) - with variability that can be attributed to sampling fluctuation. A kappa distribution is therefore fitted to the regional sample and m samples with the same size as the actual sample are drawn from this distribution. Based on these simulated samples with theoretical kurtosis (L_K^R), an estimation bias B_K^4 and an estimation standard deviation can be computed :

$$\sigma_K = \sqrt{(N_{sim} - 1)^{-1} \left[\sum_{m=1}^{N_{sim}} \left(L_K^{(m)} - L_K^R \right)^2 - N_{sim} B_K^2 \right]} \tag{2.12}$$

$$B_K = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} \left(L_K^{(m)} - L_K^R \right) \tag{2.13}$$

A reduced difference Z^{DIST} between theoretical L_K^R and observed $L_{kurtosis}$ can be computed, which probability of exceedance is known if it is supposed that this difference is normally distributed.

$$Z^{DIST} = \frac{\left(L_K^{DIST} - L_K^R + B_K \right)}{\sigma_K} \tag{2.14}$$

Hosking & Wallis suggest not to reject a possible distribution type if $\mid Z^{DIST} \mid \leq 1.64$ (the value having 10% chance to be exceeded according to the normal distribution).

In the present study the probability distribution used to model the frequency distribution of the annual maximum of the peak discharge is the generalized extreme-value distribution (GEV) proposed by Jenkinson in 1955, which has been adopted in the UK for regional flood frequency analysis and also used by various authors (Hosking et al. 1985b, Heinz and Stedinger 1998, Lu and Stedinger 1992, Stedinger and Lu 1995, Coles and Powell 1996, Coles and Tawn 1996, Coles 2001, Cox et al. 2002, Seidou et al. 2006, Gaume et al. 2010). It has been selected because of its link to extreme value theory but above all for its flexibility. It covers a large range of shapes usually used to represent peak discharge distributions : from the bounded Weibull type distribution, to the Frechet type distribution, including the standard exponential distribution. The goodness of fit test has not been systematically realized, but did not lead to reject the GEV distribution in the few cases where it has been implemented.

2.4 Modification of the Hosking & Wallis inference procedure : Bayesian MCMC method

A common problem in statistics is the estimation, from a random sample of size n, of a probability distribution whose specification involves a finite number of unknown parameters. This procedure is called statistical inference. The commonly used inference methods are the method of moments, the method of L-Moments, and the maximum likelihood method.

Hosking et al. (1985b), Hosking and Wallis (1987) applied the L-Moments estimation method to extreme value distribution calibration. They found that it performs on average better than the method of moments and both methods do well in cases of small sample sizes if compared to the maximum likelihood estimation. However, these studies only refer to meteorological data. Delicado and Goria (2008) compared the three methods for the

asymmetric exponential power distribution and concluded that the L-Moments method was particularly well suited to very small sample sizes ($n \leq 10$).

On the other hand, the maximum likelihood method provides a consistent approach to parameter estimation problems. This method has desirable mathematical and optimality properties : it provides asymptotically minimum variance and unbiased estimators as the sample size increases. Moreover, in a Bayesian perspective and using the appropriate inference algorithms (Monte Carlo Markov Chains), likelihood based methods also enable the computation of posterior distributions and credibility limits for the calibrated model parameter values and the corresponding quantile values given the observed sample and the selected statistical distribution. Figure 2.1, taken form Payrastre (2011), illustrates the possible outcomes of the likelihood based Bayesian Monte Carlo Markov Chain inference procedures and how the computed posterior credibility intervals may indicate the information content of some added data in the inference procedure. In cases where the calibrated distribution is suited to the observed data set, the width of the posterior credibility intervals for the parameters or quantiles depends on the information content of the data set (i.e. is a measure of the information content of the data set).

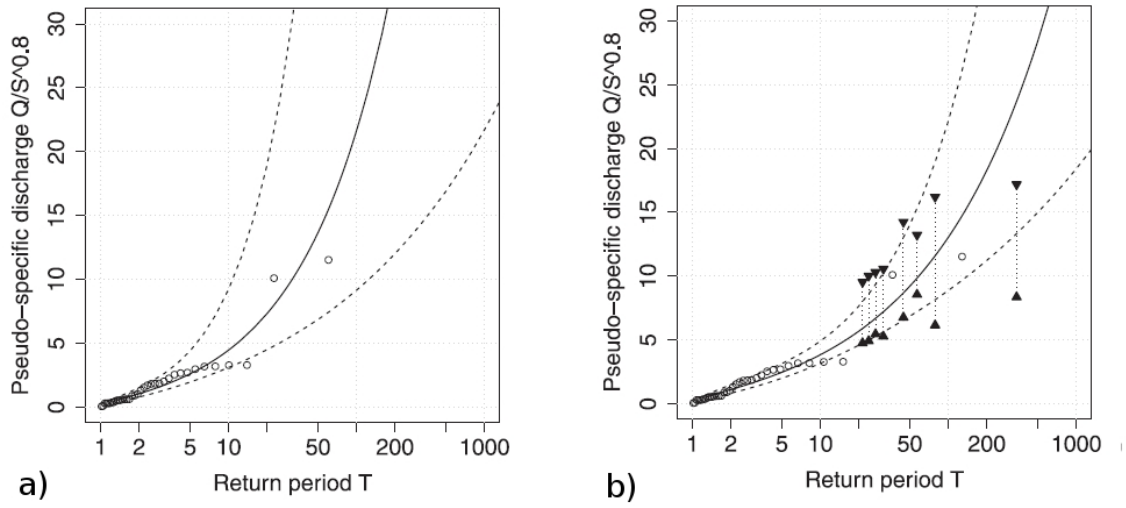


FIGURE 2.1 – *Adjusted General Extreme value distributions to annual maximum discharges and historic records available on the Lauqet river (Aude, France). Distribution corresponding to the maximum likelihood of the sample (black line) and 90% credibility intervals for the quantiles (dooted lines). a) the measured data only are considered (circles) and b) the historic records with their uncertainties (brackets) are accounted for.*

These likelihood based Bayesian MCMC inference approaches appeared therefore as particularly suited to compare various statistical inference procedures based on various sets of information. In the ideal case where the satistical model corresponds to the data and the inference procedure is unbiased - which will have to be verified (see chapter 3) -

any added information in the data set should lead to a more or less important reduction of the credibility intervals (see figure 2.1). This may nevertheless be conterbalanced by the approximations possibly done to account for this added information. The narrowing of the credibility intervals is therefore not systematically observed when the size of the dataset increases. The balance between the increasing size of the considered samples and the perturbing effects of the approximations necessitated for their inclusion - approximations on which the regional methods are based for instance - will be one of the major focus of this research.

A likelihood-based MCMC approach has been used for both, the standard Hosking & Wallis regionalization method and a modified version enabling the integration of ungauged extremes by Gaume (2010). This inference approach, already applied by Gaume (2010), is inspired by numerous previous works (Reis et al. 2005, Renard et al. 2006, Payrastre 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 maximizing the likelihood and of their posterior distribution according to the observed data set. It is based on three main ingredients which will be shortly described hereafter :

- the formulation of a likelihood of the data sample given the model,
- the Bayes theorem linking the likelihood of the sample given the model and particularly the value of its parameters and the posterior probability density function of these parameters given the data set,
- an efficient sampling method in the posterior distribution : Monte Carlo Markov Chain.

2.4.1 Likelihood of the observed sample

Considering the regional sample **D** described above, composed of s series of annual maximum peak discharges with local index flood values μ_i, the standard expression of the likelihood of the regional sample **D** would be the following :

$$\ell\left(\mathbf{D} \mid \theta\right) = \prod_{i=1}^{s}\left[\prod_{j=1}^{n_i} f_\theta\left(\frac{Q_{i,j}}{\mu_i}\right)\right] \tag{2.15}$$

Where f_θ denotes the regional probability density function (or density function of the reduced discharges), F_θ being 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 3-parameter GEV distribution, often used to describe peak discharge growth curves (Heinz and Stedinger 1998, Coles and Powell 1996, Seidou et al. 2006), has been selected (Eq.2.16 and Eq.2.17). It is a distribution type with a large spectrum and flexibility not constraining too much the shape of the posterior distribution. The huge sensitivity to sampling variability of the maximum likelihood adjustment is considered in this work as a positive aspect : the result of the inference must be firstly guided by the data and not too much by constraint imposed a priori on the distribution. To be really able to compare the information content of various data sets, the uncertainties related to the statistical inference should not be under-estimated due to unjustified constraint imposed on the shape of the peak discharge distributions. The vector θ comprises the position, scale and shape parameters (ξ, α, κ) of the GEV distribution.

$$F_\theta(Q) = \exp\left[-\left(1-\frac{\kappa(Q-\xi)}{\alpha}\right)^{1/\kappa}\right]_{\alpha>0} \tag{2.16}$$

$$f_\theta(Q) = \frac{1}{\alpha}\left(1-\frac{\kappa(Q-\xi)}{\alpha}\right)^{1/\kappa-1}\exp\left[-\left(1-\frac{\kappa(Q-\xi)}{\alpha}\right)^{1/\kappa}\right]_{\alpha>0} \tag{2.17}$$

2.4.2 Bayesian Monte Carlo Markov Chain algorithm

The Bayesian Monte Carlo Markov Chain procedure is now relatively common for hydrological applications (Reis et al. 2005, Renard et al. 2006, Ribatet et al. 2007b, Neppel et al. 2010). Recalling the Bayes' theorem, the likelihood of the sample given the parameters of the statistical model $\ell(\mathbf{D} \mid \theta)$ can be related to the likelihood or density of probability of the parameters given the sample $p(\theta \mid \mathbf{D})$ (posterior distribution) :

$$p(\theta \mid \mathbf{D}) = \frac{\ell(\mathbf{D} \mid \theta)\, p(\theta)}{p(\mathbf{D})} \tag{2.18}$$

Where $p(\theta)$ is the so called prior distribution of θ, which summarizes any prior or alternative knowledge on the distribution of θ, and $p(\mathbf{D})$ is the probability of the sample $\mathbf{D}$ which in unknown. When prior information on the distribution of θ does not exist, then $p(\theta)$ is generally taken as uniform. It is the case here, which implies that $p(\theta \mid \mathbf{D})$ is proportional to $\ell(\mathbf{D} \mid \theta)$. If the statistical model is selected, it is possible to compute the pobability density function of its parameters according to the observed data sample.

$p(\theta \mid \mathbf{D})$ being known, or more precisely $\ell(\mathbf{D} \mid \theta)$ which is proportional to it according to our hypotheses (uniform prior), parameter sets θ will be sampled according to $p(\theta \mid \mathbf{D})$ to build the posterior distributions and compute the corresponding credibility limits for

the discharge quantiles. The MCMC algorithms, combining random walk Monte Carlo methods with Markov chains, are a class of algorithms for the efficient sampling from multivariate random distributions (Tanner 1996, Robert and Casella.G 2004).

The MCMC algorithm will be further presented and some examples of outcomes will be given at the beginning of chapter 3.

2.5 The RFFA procedure proposed by Gaume (2010)

2.5.1 Calibration of a regional index flood relationship

Gaume et al. (2010) considered a situation in which a regional inventory of extreme floods at ungauged sites is available. This inventory includes h extreme peak discharges Q_k (k=(1,..,h)), each Q_k corresponding to the largest flood at site k during an inventory period of length n_k. In order to include this additional information in the regional dataset, Gaume proposed to use an index flood relationship related to the catchment area S :

$$\mu_i = c * S_i^{\beta} \; and \; \mu_k = c * S_k^{\beta} \tag{2.19}$$

Where S_i and S_k are the catchment areas at the corresponding sites, and c and β are coefficients to be determined. More complex relationships based on various climatic and physio-geographic watershed characteristics could have been used, but would have increased the corresponding number of parameters to be estimated.

Gaume et al. (2010) calibrated the c and β coefficients by regression according to the means of gauged discharge series, but did not consider the uncertainty associated with the c and β coefficients. It can be noticed that the presence of the c coefficient is just necessary to enable the regression based on observed samples means, and will not influence the inference results.

This approach represents two important additional assumptions in comparison with the conventional Hosking & Wallis approach :

- first, the assumption of homogeneity of the growth curves has to be extended to the considered ungauged sites. This represent an additional constraint for the delineation of the region, that may be solved based on considerations of spatial proximity or similar catchments attributes,
- then, the index flood relationship is supposed to be homogeneous within the region. Gaume (2010) verified the relevance of the relationship (homogeneity of the region)

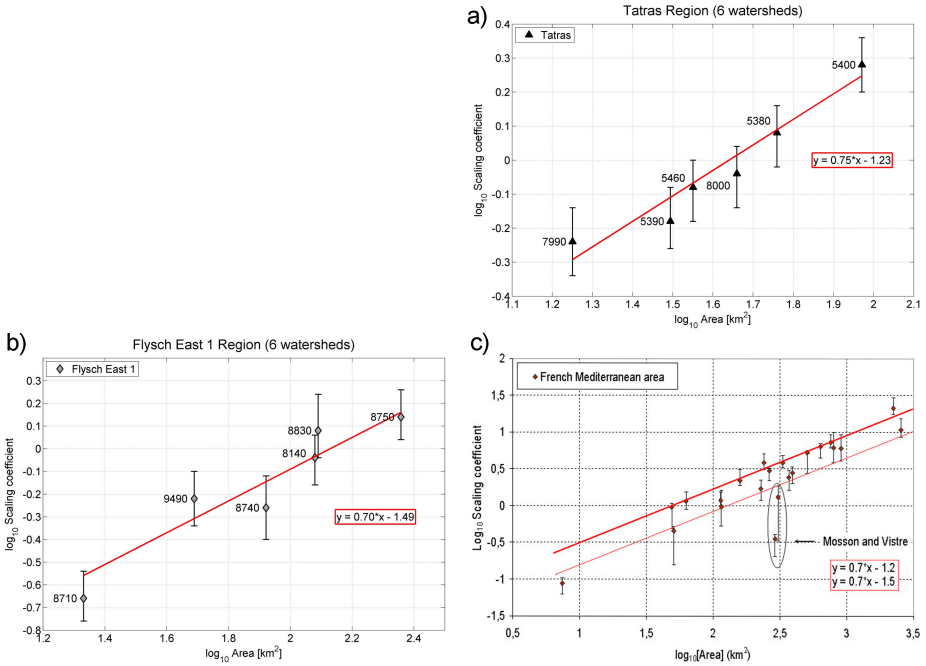


FIGURE 2.2 – *Best estimated scaling coefficients according to the Wilcoxon-Mann-Whitney test for the different regions selected by Gaume : (a) Tatras, (b) Flysch East, and (c) French Mediterranean area. 90% credibility bounds for the index flood value at each site according to the Wilcoxon, Mann and Withney test are presented.*

based on the results of a Wilcoxon-Mann-Withney test, used to associate a confidence interval to each local series mean [1]. Although this test shows that the observed sample means are not inconsistent with the proposed relationship, it cannot guarantee a perfect homogeneity of the region. For this reason, the possible presence of heterogeneity around the index flood relationship has been considered herein (see figure 2.2).

1. The WMW test, originally proposed by Wilcoxon (1945) and extended by Mann and Whitney (1947), is generally used to assess the null hypothesis that two independent samples are drawn from the same population or more precisely to assess if one of two random variables is stochastically larger than the other (Mann & Whitney, 1947). The WMW statistics is a rank sum test that is strongly controlled by the median values of the samples and is therefore often compared to the Student parametric test (Kendall and Stuart 1979). The idea of the implementation of the WMW test to support the evaluation of index flood relations consists in selecting a reference sample and in comparing it with all the other samples rescaled by a factor l. This factor is varied and the range of l values for which the WMW null hypothesis cannot be rejected at a significance level of p is an estimate of the $(1-p)\%$ confidence interval for the index flood of each sample given the choice of the reference sample and according to the WMW statistics (see figure 2.2)

2.5.2 Formulation of the likelihood

The index flood relation being calibrated, the regional h ungauged extremes may be taken into account in a regional flood frequency analysis in the same manner as historic records are introduced in a local flood frequency analysis (see Payrastre (2011) for instance).

$$\ell\left(\mathbf{D} \mid \theta\right) \quad = \quad \prod_{i=1}^{s}\left[\prod_{j=1}^{n_i} f_\theta\left(\frac{Q_{i,j}}{\mu_i}\right)\right]\prod_{k=1}^{h}\left[f_\theta\left(\frac{Q_k}{\mu_k}\right)\right]\prod_{k=1}^{h}\left[F_\theta\left(\frac{Q_k}{\mu_k}\right)\right]^{(n_k-1)} \tag{2.20}$$

In equation 2.20, the first term corresponds to probability of the gauged series. It is the only necessary term if continuous series of measured annual maximum discharges are used (modified Hosking & Wallis approach, equation 2.15). The second term is the probability of the ungauged extremes and the third complementary term is the probability of non-exceedance of the extreme value during the remaining $(n_k - 1)$ years at each ungauged site, n_k being the approximate period over which the extremes have been inventoried. Other, more complex, formulations may be used to account for the estimation uncertainties of the ungauged extremes Payrastre et al. (2011), which have been neglected by Gaume in its first tests and which will not be considered in this research (see conclusions and perspectives for some elements of discussion).

2.5.3 First implementation tests

The first implementation tests presented by Gaume (2010) demonstrated a real added value of the ungauged extremes. Nevertheless, the uncertainties associated with the calibration of the index flood relationship could not be considered and may modify significantly these conclusions. The integration of the index flood relation in the global calibration process and the evaluation of its contribution to the inference uncertainties was the first motivation of this research for a real consistent and fair comparison with the standard Hosking & Wallis regional approach. The proposed adaptation of the method initially proposed by Gaume to achieve this integration will be presented at the beginning of chapter 3.

2.6 Conclusions

Regional flood frequency analysis methods may have two different objectives : the estimation of generally low to moderate return period quantiles at ungauged sites or the building-up of larger data samples to reduce the uncertainties associated to the estimation of large return period quantiles.

Recently, Gaume (2010), proposed an original method aiming at including ungauged extremes in regional flood frequency analyses with the objective to further extend the data sample used and reduce the estimation uncertainties. The proposed method relies on both types of method since a regression technique is used to estimate a local index-flood value - basically the local expectancy (first moment) of the sample - and a regional growth curve is calibrated in a similar way to the standard Hosking & Wallis approach, once the index flood relationship is adjusted.

This relationship represents an additional homogeneity requirement if compared to the approach of Hosking & Wallis and may limit the possible extent of homogeneous regions.

Based on several case studies, Gaume (2010) showed the possible great value of such an approach, depending on the characteristics of available extreme values at ungauged sites. The simple 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 tested regions.

However, until now some debatable issues remain since the initial results presented were based on an index flood relationship adjusted a priori, and did not consider the uncertainties associated with the calibration of this relationship (one additional parameter to be estimated). Moreover, given that the index flood relationship cannot be perfect, the effects of intra-regional heterogeneity in this relationship should also be considered : a too large heterogeneity will probably lead to limit the extent of the region considered in the analysis and introduce additional uncertainty sources. Finally, the additional parametrisation introduced by the index flood relationship, combined with the necessary assumption of homogeneity of the index flood relationship within the region, may highly limit the performances of the proposed approach although additional information corresponding to ungauged extremes can be incorporated. Thus, the main objective of the research reported in this thesis is to update the method to directly account for the uncertainties induced by the calibration of the index flood relation and to conduct a complete analysis of the robustness of this approach.

Some other possible improvements of the method, benefiting from the recent developments achieved in regional regression methods or aiming at accounting for the uncertainties in the estimated values of the ungauged extremes were not considered herein but could be the focus of future developments of the method.

3 Proposed approach : improvements and comparisons based on simulations

3.1 Introduction

This chapter presents an overall comparison of both regionalisation approaches described in chapter 2, based on simulations. All the analyses conducted here consider datasets limited to gauged series. Thus, the two approaches will be compared here based on equivalent datasets (no extreme floods at ungauged sites will be considered). The differences observed in the inference results will therefore be explained only by methodological considerations.

In order to conduct a realistic comparison, some improvements have first been introduced in the proposed approach of Gaume (2010), in order to provide a better estimation of the uncertainty associated with the use of the index flood relationship. This required both an adaptation of the formulation of the index flood relationship, and of the expression of the likelihood of observations : these evolutions will be presented at the beginning of this chapter. These adaptations finally result in an increase of the number of parameters to be estimated with the MCMC algorithms (4 parameters instead of 3), and consequently the convergence of MCMC algorithms may become more problematic. This question of convergence was examined carefully using a common convergence diagnosis.

A Monte Carlo simulation procedure was then developed in order to enable a detailed comparison between the two regional approaches : i.e. the reference Hosking & Wallis approach and the proposed approach based on the index flood relationship, including the modifications described before. In each case, the analysis of simulations results was focused on the quality of estimation of the 100-year quantiles $Q_i^{(100)}$, and on the reliability

of 90% credibility intervals $[\hat{Q}_{i,5}^{(100)}, \hat{Q}_{i,95}^{(100)}]$ computed using the Bayesian MCMC procedure. A great advantage of such a comparison based on simulations is that it is not sensitive to sampling fluctuations. It also enables to conduct sensitivity analyses to various contexts of availability of data and of presence of heterogeneities within the considered regions. The sensitivity to the length of the local series was especially examined here. The impact of presence of heterogeneities in the index flood relationship and the local growth curves was also considered.

This chapter is organised as follows : the first section presents the evolutions introduced in the proposed approach. The next section describes in detail the methodology applied for simulations and the result evaluation. The results are then presented and commented, first in the case of homogeneous regions, and then with presence of heterogeneities in the considered regions. The main findings are finally summarised in the conclusion section.

3.2 Evolutions of the proposed approach

Gaume et al. (2010) calibrated the c and β coefficients of the index flood relationship (Eq. 2.19) by regression according to the means of gauged discharge series, and verified the homogeneity of this relationship in the considered regions based on the results of a Wilcoxon-Mann-Whitney test. However, they did not consider the uncertainty associated to the estimation of the c and β coefficients. A relevant comparison with the reference Hosking & Wallis approach should incorporate this additional source of uncertainty.

Therefore, in the present application the methodology has been adapted in order to enable a common estimation, based on the Bayesian MCMC algorithm, of both the index flood relationship and the growth curve parameters. This was achieved by a double adaptation of **(i)** the formulation of the index flood relationship (see next section), and **(ii)** the expression of the likelihood of observed samples (in order to incorporate the index flood relationship parameter in this expression, see section 3.2.2). Based on these new formulations, the number of parameters to be estimated with the Bayesian MCMC algorithm increases from 3 to 4, and logically the convergence of MCMC algorithms becomes more problematic. The methodology applied to ensure a good convergence of the algorithm, and the convergence diagnosis applied to verify this convergence will be detailed in section 3.2.3.

3.2.1 Index flood relationship and associated homogeneity analysis

It can be noticed, in the initial formulation of the index flood relationship (Eq. 2.19), that the presence of the c coefficient is just necessary to enable the regression based on observed samples means, and does not influence the inference results. Moreover, since this parameter just modifies the global scale of the reduced discharge series, it is redundant with the position (ξ) and scale (α) parameters of the regional growth curve. A common estimation of these three parameters in the Bayesian MCMC algorithm would not be possible : the convergence of the algorithm cannot be achieved in a situation for which infinity of vectors of parameters lead to the same likelihood value. Therefore, the formulation of the index flood relationship was adapted here to discard the c coefficient. The formulation used in the rest of this manuscript is the following :

$$\mu_i = S_i^{\beta} \tag{3.1}$$

Moreover, given that the β coefficient will not be fixed a priori but estimated using the MCMC algorithm, the analysis of the homogeneity of the considered regions has been conducted (case studies presented in chapter 4) in a different manner as described in chapter 2 (Wilcoxon-Mann-Whitney test). The test procedure applied here is based on the computation of the mean or median discharge at each site of the considered region, according to the adjusted maximum likelihood model, and of the 95% fluctuation range of observed means around this maximum likelihood value (computed by simulations), according to the length of each discharge series (effect of sampling variability). A simple comparison between the means of real observed samples and the fluctuation ranges associated with the maximum likelihood model may show that the observed sample means are (or not) consistent with the proposed relationship. However, as for the Wilcoxon-Mann-Whitney test, it cannot guarantee a perfect homogeneity of the region. For this reason, the possible presence of heterogeneities around the index flood relationship has been considered in the simulations conducted in this chapter and in chapter 4.

3.2.2 Formulation of the likelihood function

The inference approach applied herein is directly derived from Gaume (2010) (see section 2.5 in chapter 2), based on the likelihood of the available data sets and a Bayesian MCMC algorithm for parameter estimation. However, the joint estimation of the index flood relationship and the regional growth curve required here an adaptation of the likelihood formulations.

Considering again the regional sample **D** described in chapter 2, including both gauged flood series $Q_{i,j}$ and ungauged extremes Q_k, a logical expression of the likelihood of sample **D**, including the parameter β of the index flood relationship, may be the following :

$$\ell(\mathbf{D} \mid \theta) = \prod_{i=1}^{s} \left[\prod_{j=1}^{n_i} f_\theta \left(\frac{Q_{i,j}}{\mu_i(\theta)} \right) \right] \prod_{k=1}^{h} \left[f_\theta \left(\frac{Q_k}{\mu_k(\theta)} \right) \right] \prod_{k=1}^{h} \left[F_\theta \left(\frac{Q_k}{\mu_k(\theta)} \right) \right]^{(n_k-1)} \tag{3.2}$$

Where f_θ and F_θ are again the probability density function and the cumulative probability function of the statistical distribution selected to represent the regional growth curve, and θ corresponds to the parameters to be estimated. Here θ includes both the β parameter of the index flood relationship and the parameters of the GEV distribution, $\theta = (\beta, \xi, \alpha, \kappa)$.

However, a careful analysis of equation (3.2) shows that although this likelihood expression is similar to the one presented in chapter 2 (used in a quantity of previous studies), it is not suited for the present application. The index flood values $\mu(\theta)$, indeed, increase with parameter β, and the corresponding values of the density function $f_\theta(Q/\mu)$ also increase in the same proportion. Therefore, it is likely that the likelihood of the dataset will increase with β, leading to a biased estimate of β. To face this difficulty, an expression of the likelihood that is less sensitive to β has to be selected : this was achieved by systematically replacing the density function $f_\theta(Q/\mu)$ by the probability to observe a value in the range of $\pm 1\%$ around $\frac{Q}{\mu}$: $F_\theta\left(1.01 * \frac{Q}{\mu}\right) - F_\theta\left(0.09 * \frac{Q}{\mu}\right)$.

Note that this last expression does not increase with μ anymore. Using this expression, the likelihood of the dataset becomes :

$$\begin{aligned} \ell(\mathbf{D} \mid \theta) = & \prod_{i=1}^{s} \left[\prod_{j=1}^{n_i} \left[F_\theta \left(1.01 * \frac{Q_{i,j}}{\mu_i(\theta)} \right) - F_\theta \left(0.99 * \frac{Q_{i,j}}{\mu_i(\theta)} \right) \right] \right] \\ & \times \prod_{k=1}^{h} \left[F_\theta \left(1.01 * \frac{Q_k}{\mu_k(\theta)} \right) - F_\theta \left(0.99 * \frac{Q_k}{\mu_k(\theta)} \right) \right] \\ & \times \prod_{k=1}^{h} \left[F_\theta \left(\frac{Q_k}{\mu_k(\theta)} \right) \right]^{(n_k-1)} \end{aligned} \tag{3.3}$$

or

$$\begin{aligned} \ell(\mathbf{D} \mid \theta) = & \prod_{i=1}^{s} \left[\prod_{j=1}^{n_i} \left[F_\theta \left(1.01 * \frac{Q_{i,j}}{\mu_i(\theta)} \right) - F_\theta \left(0.99 * \frac{Q_{i,j}}{\mu_i(\theta)} \right) \right] \right] \\ & \times \prod_{k=1}^{h} \left[F_\theta \left(\frac{Q_k^U}{\mu_k} \right) - F_\theta \left(\frac{Q_k^L}{\mu_k} \right) \right] \\ & \times \prod_{k=1}^{h} \left[F_\theta \left(\frac{Q_k^U}{\mu_k} \right) \right]^{(n_k-1)} \end{aligned} \tag{3.4}$$

If the magnitudes of the ungauged extremes are known with some uncertainty and upper and lower estimates can be defined (Q_k^U ; Q_k^L), j= 1,..,h.

In the case of the reference approach of Hosking (1997) based on Eq.(2.3), only information from gauged sites can be used, and therefore the likelihood expressions (3.3) and (3.4) will be limited to their first term. The set of parameters θ to be estimated will also be limited to (ξ, α, κ).

In order to better illustrate the differences between likelihood functions 3.3 and 3.4., the MCMC algorithm has been applied to the same regional dataset using both formulations. The results are presented in figures 3.1, 3.2 and 3.3. The regional dataset used here was simulated and corresponds to the characteristics of the Ardèche case study (see chapter 4) : same catchment areas and record lengths (limited to gauged series), parameters of the growth curve (ξ= 3.34, α= 2.24 and κ= -0.16) and of the index flood relation (β= 0.76) determined based on the real Ardèche dataset. Figure 3.1 first compares the posterior distributions of the three parameters (ξ, α, κ) obtained using the reference Hosking & Wallis approach. This figure shows that with the reference approach of Hosking & Wallis, both formulations may be used with similar estimation results. However, figures 3.2 and 3.3 show that this is not the case anymore if the proposed approach is used : in this case, the use of formulation 3.2 clearly leads to overestimate parameter β and underestimate in the same time parameters α and ξ ; it also causes evident problems of convergence of the MCMC algorithm (see parameter κ on figure 3.2). On the other hand, the estimation obtained using formulation 3.3 (or Eq.3.4), even if affected by uncertainties, appears much more consistent with an apparently acceptable convergence of the MCMC algorithm.

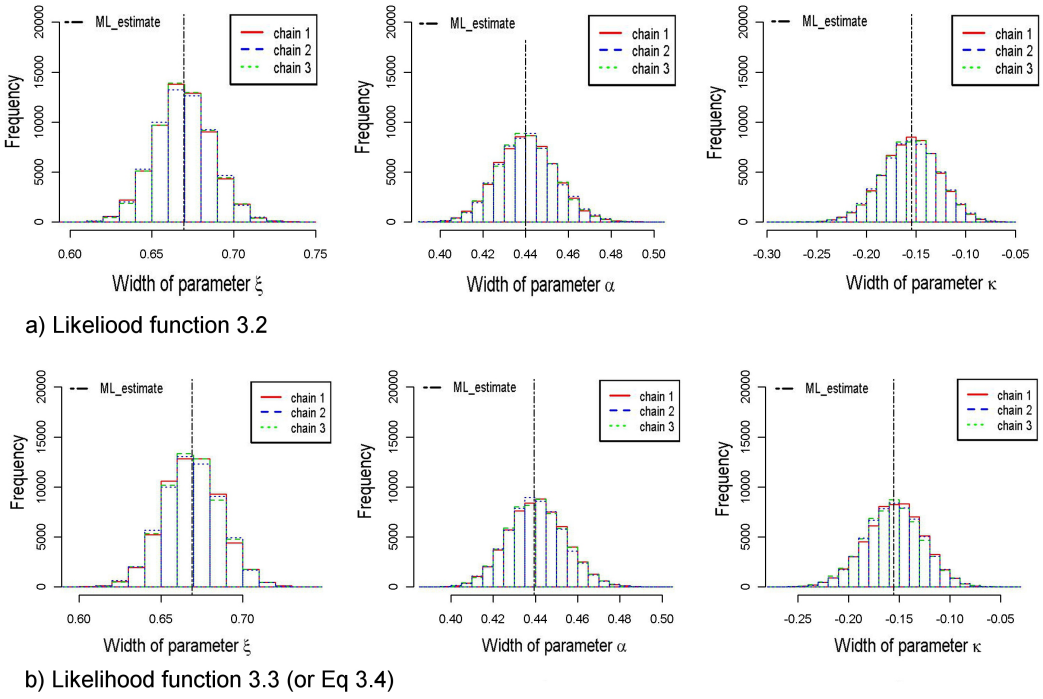


FIGURE 3.1 – *Posterior distributions of the growth curve parameters (ξ, α, κ) computed based on the reference Hosking and Wallis approach.*

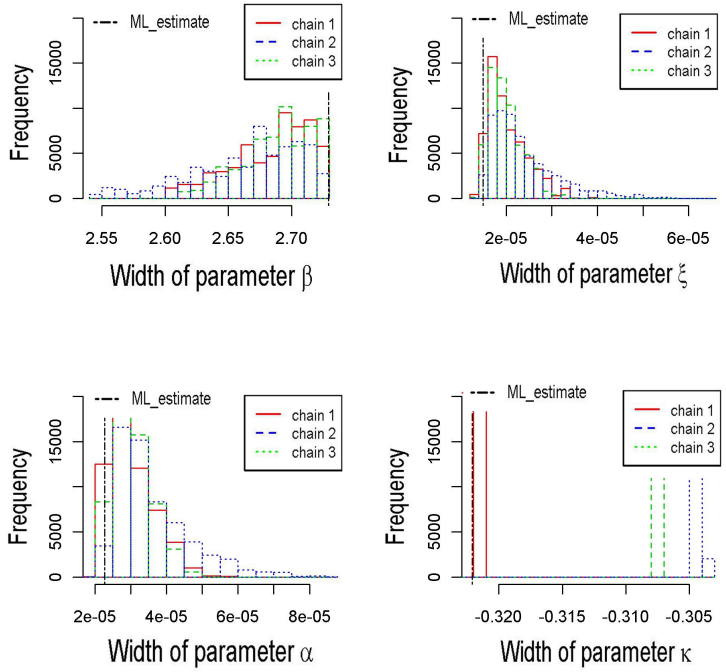


FIGURE 3.2 – *Posterior distributions of index flood relation and growth curve parameters (β, ξ, α, κ) computed based on the proposed approach and using likelihood expression 3.2.*

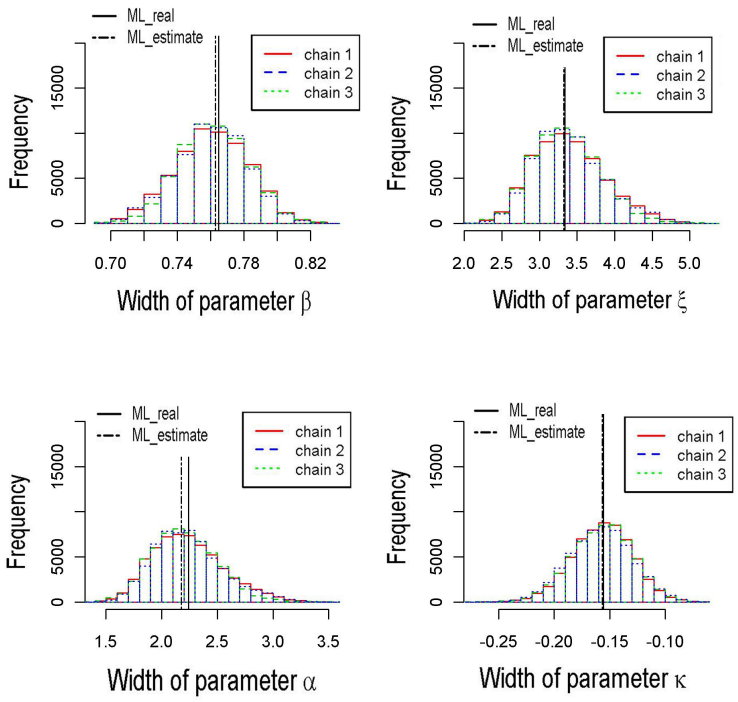


FIGURE 3.3 – *Posterior distributions of index flood relation and growth curve parameters (β, ξ, α, κ) computed based on the proposed approach and using likelihood expression 3.3 (or Eq.3.4).*

3.2.3 MCMC convergence diagnosis

An important issue when dealing with MCMC algorithms is to ensure a good convergence of the algorithms. Provided that the proposed approach, after the adaptations described in previous sections, now requires the estimation of 4 parameters, instead of three parameters for the reference approach, the convergence may be more problematic in this case and has to be carefully checked. Cowles and Carlin (1996), Salaheddine et al. (2006) provided an exhaustive review of the large number of convergence diagnostic methods proposed in previous publications (Gelman and Rubin 1992, Raftery and Lewis 1992, Geweke 1992, Roberts 1992; 1994, Ritter and Tanner 1992, Zellner and Min 1995, Liu et al. 1992, Garren and Smith 1993, Johnson 1994, Heidelberger and Welch 1983, Mykland et al. 1995, Yu 1994, Yu and Mykland 1994). We considered herein the diagnosis of Gelman and Rubin (1992) that is among the most popular in the statistical community. This test is based on running $M > 1$ MCMC chains in parallel with starting values that are over dispersed relative to the posterior distribution. The convergence is diagnosed when the chains have "forgotten" their initial values, and the output from all chains is undistinguishable. It is based on the comparison of within-chain and between-chain variances of the calibrated parameter

values, and is similar to a classical analysis of variance. The convergence diagnostic results in the computation of a coefficient **R** : values of **R** substantially above 1 indicate a lack of convergence, and the MCMC chains can be stopped as **R** is sufficiently close to 1. The detailed computation of **R** is presented in appendix **A.2**.

Figure 3.4 presents examples of application of the Gelman and Rubin test to MCMC algorithms applied on the both case studies that will be presented in chapter 4 : the Ardèche region including 168 records at 5 gauging stations and the Var region including 249 records at 9 gauging stations. The number of iterations of the MCMC algorithms applied varies from N= 40 000 to 300 000. To achieve fast convergence, a first MCMC chain was systematically run with a limited number of iterations ($N/2$) to adjust the variance of the "jump" distribution of a second chain (N iterations), used for the parameters estimation. The test was applied here with $M = 2$. The values of **R** presented on figure 3.4 show that the convergence is easily achieved with the reference approach (**R**= 1), meanwhile it is more uncertain with the proposed approach, especially as the number of iterations N is lower than 60 000. Hence, for the rest of the study, a number of iterations of N= **60 000** was systematically applied and the convergence systematically checked using the same test, and considering a threshold value **R** of 1.05.

3.3 Methodology of Monte Carlo Simulations

This section describes the simulation procedure developed to compare the two regional approaches described in chapter 2. The general principle is the following : in a first step, a large number of synthetic regional datasets are generated, including different possible degrees of heterogeneity ; the two estimation approaches to be compared are then applied to each of the synthetic series, including the estimation of the posterior distribution of parameters based on the Bayesian MCMC procedure described in previous sections, the estimation results are finally compared with the real parent distributions, with a focus on the quality of estimation of the 100-year quantiles $Q_i^{(100)}$, and on the reliability of credibility intervals $[\hat{Q}_{i,5}^{(100)}, \hat{Q}_{i,95}^{(100)}]$ computed using the Bayesian MCMC algorithm.

The following sections first present the procedure used for the generation of synthetic regional series, including different levels of heterogeneity. Section 3.3.5 presents the methodology applied for the exploitation and comparison of results.

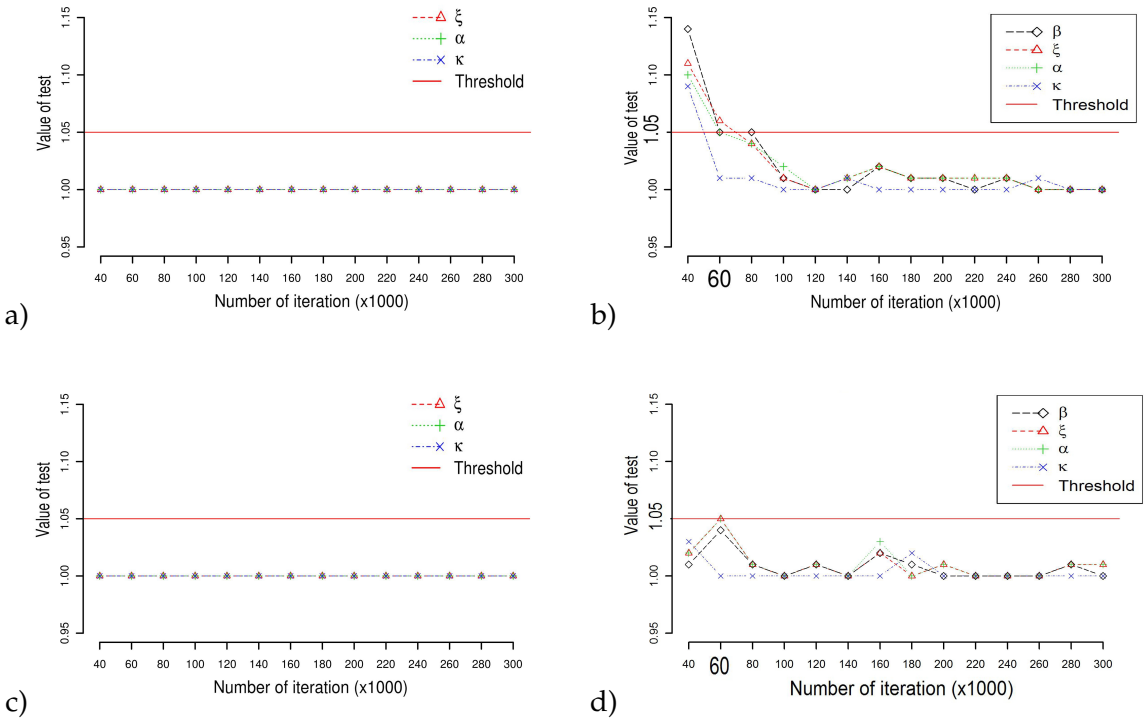


FIGURE 3.4 – *Gelman and Rubin's convergence diagnosis applied to MCMC algorithms : a) Ardèche region and reference approach, b) Ardèche region and proposed approach, c) Var region and reference approach, d) Var region and proposed approach.*

3.3.1 Generation of synthetic series - case of homogeneous regions

Each simulation conducted is based on 1000 synthetic regional samples, which characteristics (number of gauged sites, associated catchment surfaces, length of each gauged series, presence of information on extremes) are predefined and may correspond for instance to a case study for which the analysis has to be conducted. The samples are generated based on a regional growth curve and an index flood relationship which characteristics are fixed, for instance determined based on the real available datasets. The regional growth curve involved here will always correspond to a GEV distribution with three parameters (ξ_0, α_0, κ_0). The index flood relationship corresponds to the one presented at the beginning of this chapter, with a parameter β_0 (μ=S^{β_0}).

In the case of an homogeneous region, the generated samples respect perfectly the formalism of both the regional growth curve and the index flood relation (see Fig 3.5.a). In this specific case, the procedure of simulation of one synthetic regional sample is the following :

– a reduced discharge serie $q_{i,j}$ is generated at each site i of the considered region, with the corresponding length n_i (j=[1,..,n_i]). The at-site reduced discharges are di-

rectly derived from the expression of the cumulative probability function of the GEV distribution :

$$q_{i,j} = \xi_0 + \frac{\alpha_0}{\kappa_0}\left[1 - \left(-log\left(F_j\right)\right)^{\kappa_0}\right] \tag{3.5}$$

where, the non-exceedance probabilities F_j are generated from a random uniform distribution over the range [0,1],

– the rescaled discharge series at i^{th} gauged site is then computed using the index flood relationship (Eq. 3.1) :

$$Q_{i,j} = q_{i,j} * S_i^{\beta_0} \tag{3.6}$$

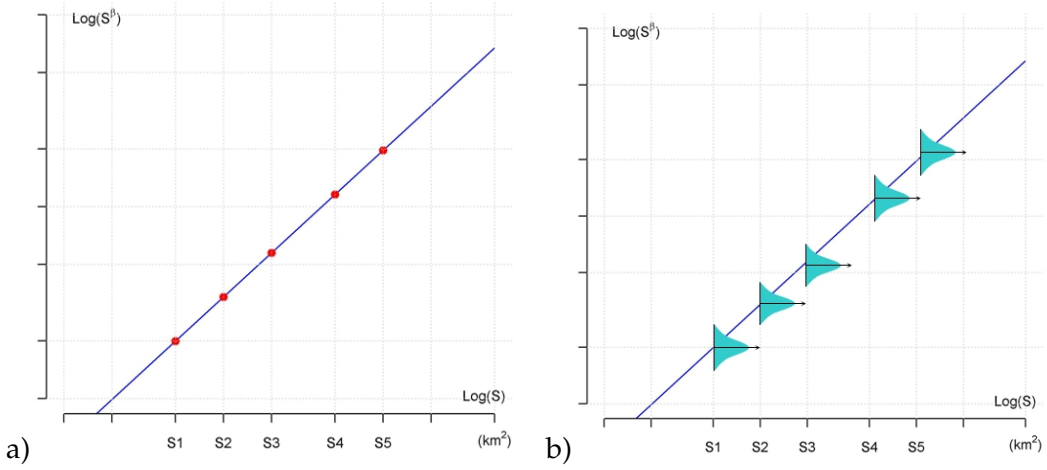
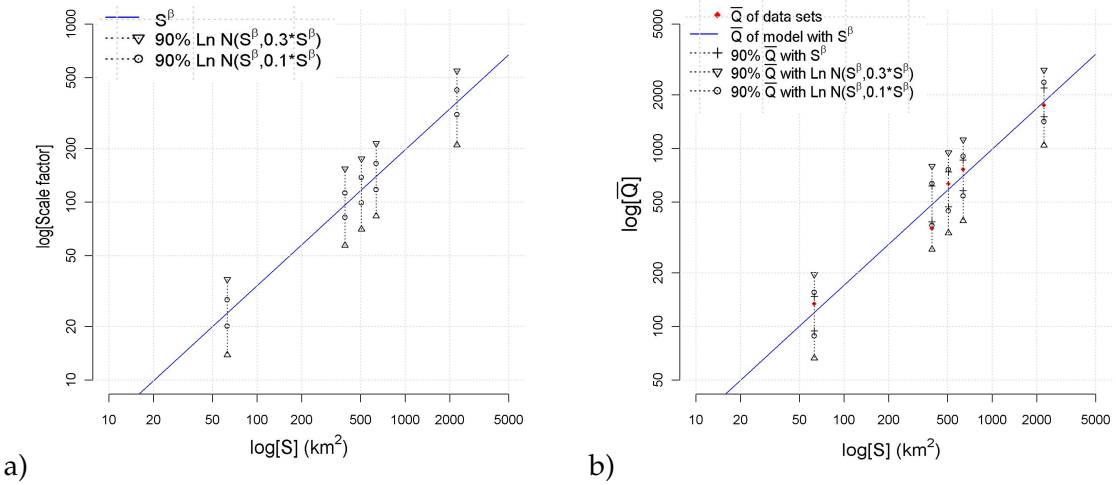


FIGURE 3.5 – *Relationship between the index flood values and the watershed surface used for synthetic series generation : a) case of homogeneous region, b) heterogeneity of the index flood relationship, based on a log-normal distribution.*

3.3.2 Introduction of heterogeneity in the index flood relation

The presence of heterogeneity in the index flood relation may be introduced in the rescaling phase of the simulated reduced discharge series (Eq. 3.6). The theoretical index flood value at the i^{th} site μ_i=S_i^{β} has been replaced here by a random value obtained from a log-normal distribution with mean S_i^{β} and standard deviation ($\delta * S_i^{\beta}$) (see Fig.3.5.b). In order to simulate different levels of heterogeneity δ has been successively set to 0.1 and 0.3. Figure 3.6 shows the corresponding dispersion of the computed index flood values and of means of generated samples, obtained for a simulation corresponding to the Ardèche region in France (see section 4.2.1 for a detailed presentation of this region). This figure suggests that for δ=0.1, the fluctuations of the sample means do not significantly increase in comparison

with the reference situation with no-heterogeneity. The means of the real Ardèche region gauged samples, represented on the graph, also suggest that the real heterogeneity of this region may correspond to δ=0 (no-heterogeneity) or to δ=1.



a) b)

FIGURE 3.6 – *Dispersion of a) the index flood values μ_i, and b) the means of simulated samples, associated with the random generation of μ_i using a log-normal distribution LN(S_i^β, $\delta * S_i^\beta$). The presented simulation (number of sites, catchment areas, samples sizes) corresponds to the Ardèche region.*

3.3.3 Introduction of heterogeneity in the local growth curves

Hosking and Wallis (1997) suggested to measure the heterogeneity of the local growth curves in a region based only on the differences between the samples L_{CV} and $L_{skewness}$. The introduction of heterogeneity in the local growth curves was conducted here in a similar manner, by selecting randomly L_{CV} and $L_{skewness}$ values that were supposed to correspond to the growth curves of each site in the region. This procedure requires three additional steps to be added to the initial simulation procedure described in section 3.4.1 (case of homogeneous datasets) : **(i)** evaluation of reasonable fluctuation ranges of L_{CV} and $L_{skewness}$, and choice of a random generation distribution **(ii)** incorporation of heterogeneities in the growth curves used to generate reduced discharge series at each site i, and **(iii)** application of the Hosking & Wallis heterogeneity test to each regional sample generated. These three steps are described in detail bellow :

Step 1 : Choice of a reasonable distribution for random generation of L_{CV} and $L_{skewness}$ values.

Figure 3.7 presents the fluctuations of L_{CV} and $L_{skewness}$ (or L_{CA}) values estimated based on 10.000 simulated samples including respectively 20 and 40 observations, and generated using a GEV distribution with parameters ξ_0, α_0, and κ_0. This figure clearly illustrates the correlation existing between the two L-Moments values. It also shows that the dispersion of estimated values is logically higher as the sample size used for the computation decreases. The dispersion observed on this figure is associated to sampling variability and not with heterogeneity of the growth curves. The assumption that was made here is that this dispersion may also be representative of reasonable heterogeneities of the local growth curves within a considered region, around the theoretical regional growth curve represented by parameters (ξ_0, α_0, κ_0).

In order to simulate L_{CV} and $L_{skewness}$ values with a dispersion corresponding to the figure 3.7, a bivariate normal distribution was adjusted to the corresponding L_{CV} and $L_{skewness}$ datasets. The parameters used to procedure the results presented in the figure 3.8 are detailed in table 3.1. In this table, ρ_0 corresponds to the coefficient of correlation,

$$\rho_0 = \frac{\sum_{i=1}^{m} \left(L_{CA}^i - \varphi_{L_{CA}}\right) * \left(L_{CV}^i - \varphi_{L_{CV}}\right)}{(n-1)\,\sigma_{L_{CA}}\sigma_{L_{CV}}} \tag{3.7}$$

and, $\varphi_{L_{CA}}$ and $\varphi_{L_{CV}}$ are the respective means of L_{CA} and L_{CV} ; $\sigma_{L_{CA}}$ and $\sigma_{L_{CV}}$ the standard deviations.

TABLE 3.1 – *Parameters of the bivariate normal distribution adjusted based on the 10 000 values of L_{CA} and L_{CV} presented in figure 3.8.*

n(year)	$\varphi_{L_{CA}}$	$\varphi_{L_{CV}}$	$\sigma_{L_{CA}}$	$\sigma_{L_{CV}}$	ρ_0
20-year	0.259	0.361	0.109	0.051	0.660
40-year	0.267	0.363	0.076	0.035	0.668

The parameters values presented in table 3.1 are just indicative : these coefficients have been systematically estimated based on a set of 10 000 simulated samples corresponding to the (ξ_0, α_0, κ_0) parameters of the reference regional growth curve used for each simulation.

Step 2 : Incorporation of heterogeneity in simulation of reduced discharge series.

For each synthetic regional sample to be generated, the bivariate normal distribution presented above has been used to generate random values L_{CV-i} and $L_{skewness-i}$ to be applied at each site i of the region. The procedure to be followed for random generation is the

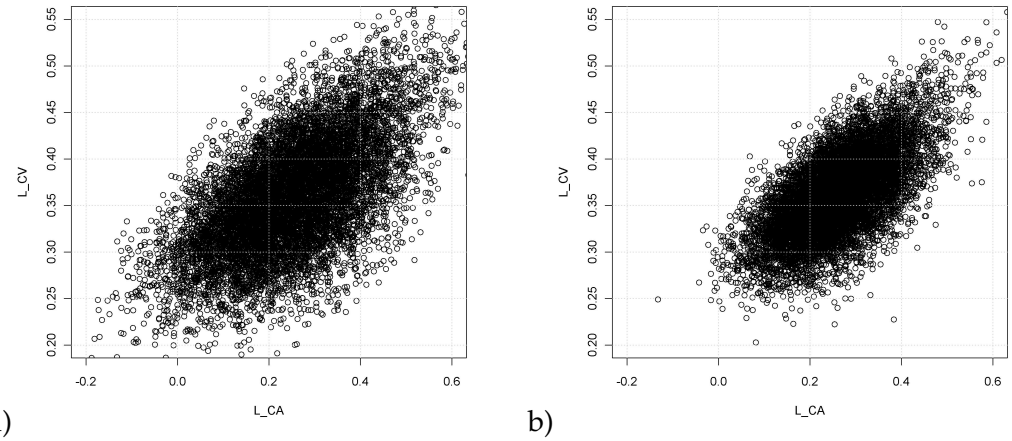


FIGURE 3.7 – *Dispersion of L_{CA} and L_{CV} values computed for 10.000 simulated samples, the number of observations in each sample being : a) n= 20-year, and b) n= 40-year. The considered samples were generated based on a GEV distribution with parameters ξ_0= 3.34, α_0= 2.24, and κ_0= -0.16, corresponding to the Ardèche region.*

following : first, the $L_{skewness-i}$ value is generated based on a normal distribution with mean $\varphi_{L_{CA}}$ and standard deviation $\sigma_{L_{CA}}$. The corresponding L_{CV-i} value is then obtained from a normal distribution with mean $\varphi_{L_{CV}} + \rho_0 \frac{\sigma_{L_{CV}}}{\sigma_{L_{CA}}}(L_{CA-i} - \varphi_{L_{CA}})$ and standard deviation $(1 - \rho_0^2)^* \sigma^2_{L_{CV}}$.

Two degrees of heterogeneity were systematically considered, the parameters of the bivariate distribution used corresponding successively to the L_{CV} and $L_{skewness}$ fluctuations observed for samples of 40 years length (limited heterogeneity), and 20 years length (high heterogeneity), An example of simulated L_{CV-i} and $L_{skewness-i}$ values is presented on figure 3.8.

The parameter values (ξ_i, α_i, κ_i) of the GEV distribution used to simulate the reduced discharge series at site i, were then directly estimated based on the L_{CV-i} and $L_{skewness-i}$ values, and considering an expected value φ_{q_i} being the same than for the reference growth curves :

$$\varphi_{q_i} = \xi_0 + \frac{\alpha_0 \left(\Gamma\left(1 + \kappa_0\right) - 1\right)}{\kappa_0} \tag{3.8}$$

No explicit solution exists to estimate κ_i, based on (φ_{q_i}, L_{CV-i}, L_{CA-i}) but the following approximation has an accuracy better than $9 * 10^{-4}$ for -0.5<L_{CA-i}<0.5

$$\kappa_i = \quad 7.8590 * c + 2.9554 * c^2 \quad with \quad c = \frac{2}{3 + L_{CA-i}} - \frac{log2}{log3} \tag{3.9}$$

The other parameters are given by :

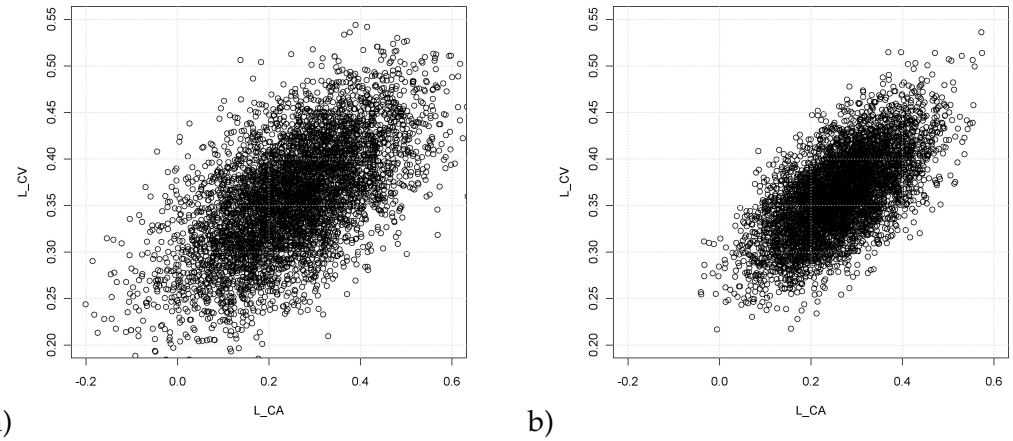
$$\alpha_i = \frac{\varphi_{q_i} * L_{CV-i} * \kappa_i}{(1 - 2^{\kappa_i}) * \Gamma(1 + \kappa_i)} \tag{3.10}$$

$$\xi_i = \varphi_{q_i} - \frac{\alpha_i * (1 - \Gamma(1 + \kappa_i))}{\kappa_i} \tag{3.11}$$

Finally, the generation of reduced discharge series at site i is made in the same manner as in the case of homogeneous regions, but is here based on the (ξ_i, α_i, κ_i) parameter values :

$$q_{i,j} = \xi_i + \frac{\alpha_i}{\kappa_i}\left[1 - \left(-log\left(F_{i,j}\right)\right)^{\kappa_i}\right] \tag{3.12}$$

The nonexceedance probabilities $F_{i,j}$ being again generated from a uniform random distribution over the range [0,1].



a) b)

FIGURE 3.8 – *Dispersion of L_{CV-i} and $L_{skewness-i}$ values generated based on a bivariate normal distribution, for 1000 regional samples including each 5 sites (5000 values) : a) case of important heterogeneity, bivariate distribution calibrated based on the dispersion observed for samples of n= 20-year length (Fig 3.7.a), and b) case of limited heterogeneity, bivariate distribution calibrated based on the dispersion observed for samples of n= 40-year length (Fig 3.7.b).*

Step 3 : Application of the Hosking & Wallis heterogeneity test.

The application of the Hosking & Wallis heterogeneity test Hosking and Wallis (1997) may help to filter detectable heterogeneous datasets. Therefore, a realistic simulation of impact of heterogeneities should be limited to samples that pass the test. For this reason, the homogeneity test was applied to each of the 1000 synthetic samples, and the samples identified as heterogeneous were systematically removed.

Hosking asserts that a region is acceptably homogeneous if the H_1 measure given by the test does not exceed 1. However, this threshold may also be placed at a higher value, given that Hosking (2007) considers H_1 values above 3 as representative of definitely heterogeneous region (see section 2.3.3). The H_1 measure, indeed, just gives an indication but cannot lead to a definitive conclusion. Figure 3.9 well illustrates the limits of this test, associated with sampling fluctuations : if the overall distributions of H_1 values significantly differ between homogeneous and heterogeneous samples, H_1 values above 1 may be observed for homogeneous samples, and conversely.

Based on the results of the Hosking & Wallis heterogeneity measure applied to each of the 1000 samples, three situations of heterogeneity were finally considered : (**case 1**) for homogeneous datasets, only the samples with $H_1 \leq 1$ value were selected (see Fig. 3.9.a), (**case 2**) for datasets corresponding to limited heterogeneity (L_{CV} and $L_{skewness}$ fluctuations observed for samples of 40 years length), the samples having $H_1 \leq 2$ value were selected (see Fig. 3.9.b), and (**case 3**) for highly heterogeneous datasets (L_{CV} and $L_{skewness}$ fluctuations observed for samples of 20 years length), the threshold on the H_1 value was set to $H_1 \leq 3$ (see Fig. 3.9.c).

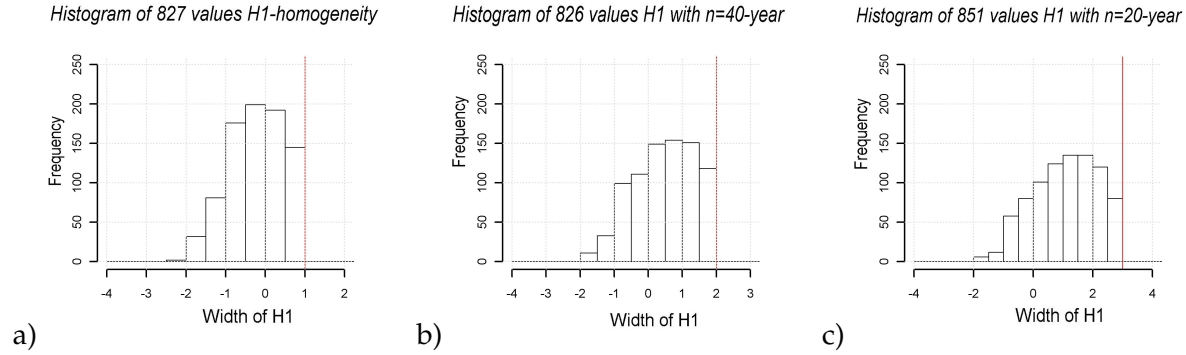


FIGURE 3.9 – *Histogram of H_1 values of selected samples, in three cases : a) **case 1** : homogeneous samples, 827/1000 samples selected, b) **case 2** : samples with limited heterogeneity, 826/1000 samples selected based on a H_1 threshold of 2, and c) **case 3** : samples with high heterogeneity, 851/1000 samples selected based on a H_1 threshold of 3.*

3.3.4 Introduction of heterogeneities in both index flood relationship and local growth curves

This last simulation procedure has been applied exclusively to the case studies that will be presented in chapter 4. Here the simulations conducted aim to provide an estimation of the performances of both regional approaches in a context of heterogeneity and of availability

of data that corresponds to each considered case study. Therefore, the simulations combine here the incorporation of heterogeneities in both the growth curves and in the index flood relationship. The simulation procedure remains similar to those presented in sections 3.4.2 and 3.4.3. It can be resumed as follows :

- calibration of the bivariate normal distributions to be used for introduction of heterogeneities in the local growth curves, generation of at site reduced discharge series incorporating heterogeneity on the growth curves used at each site,
- rescaling discharge series with introduction of heterogeneity in the index flood relationship,
- application to Hosking & Wallis homogeneity test and removal of all samples identified as heterogeneous,
- application to the inference approaches to be evaluated to each simulated regional sample.

3.3.5 Analysis of the quantile estimates and of the credibility intervals

The evaluation procedure relies on the application of the estimation approaches to be tested (local analysis, the reference Hosking & Wallis regional approach, proposed regional approach) to each of the synthetic samples obtained from Monte Carlo simulations. The entire posterior distribution of parameters is estimated in each case based on the Bayesian MCMC procedure. Of course, the cases for which convergence of MCMC algorithm is not achieved are removed.

The first exploitation of these results is based on the computation of discharge quantiles maximum likelihood estimates $\hat{Q}_{i,ML}^{(T)}$ corresponding to the modes of each posterior distribution $\hat{F}_{Q_i^{(T)}}$. The corresponding 90% credibility intervals $[\hat{Q}_{i,5}^{(T)}, \hat{Q}_{i,95}^{(T)}]$ are also computed. The final result is a set of 1000 (or less if some MCMC chains did not converge or if some samples were removed based on the Hosking & Wallis heterogeneity measure) of maximum likelihood quantiles $\hat{Q}_{i,ML}^{(T)}$ and corresponding 90% credibility intervals. The position of these estimated quantiles and the corresponding credibility intervals were compared at each site to the real quantile value to be estimated $Q_i^{(T)}$. The fluctuations observed around this value were reported on a boxplot graph as presented in the figure 3.11. This graph gives a synthetic overview of the fluctuations of the estimations obtained with each considered approach, by reference to the real value to be estimated.

In a second step, a global evaluation of the relevance of the posterior distributions

$\hat{F}_{Q_i^{(T)}}$ was conducted. This was achieved by the computation, according to the posterior distributions $\hat{F}_{Q_i^{(T)}}$, of the cumulative probability $\hat{F}_{Q_i^{(T)}}(Q_i^{(T)}) = \mathrm{P}(\hat{Q}_i^{(T)} < Q_i^{(T)})$ corresponding to the real quantile value $Q_i^{(T)}$. If the posterior distributions $\hat{F}_{Q_i^{(T)}}$ are well estimated, the values of $\hat{F}_{Q_i^{(T)}}(Q_i^{(T)})$ should be uniformly distributed. Figure 3.10 shows some examples of the distributions of $\hat{F}_{Q_i^{(T)}}(Q_i^{(T)})$ that may be obtained and the corresponding interpretation in terms of quality of estimation.

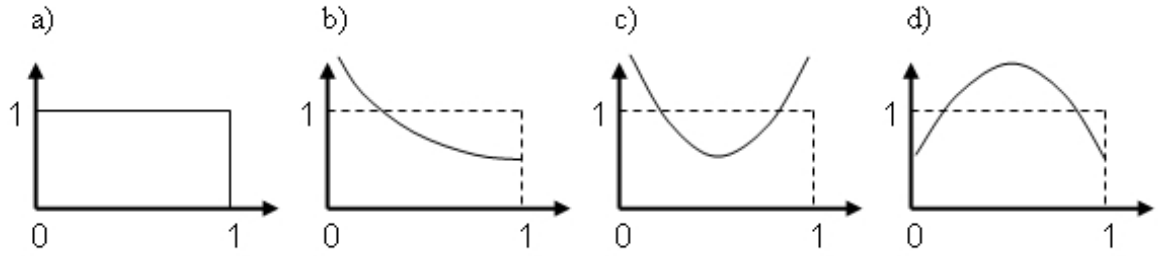


FIGURE 3.10 – *Examples of possible distributions of the $\hat{F}_{Q_i^{(T)}}(Q_i^{(T)})$ values, and corresponding interpretations : a) perfect estimation ; b) biased estimation of the flood quantile (over-estimation) ; c) too narrow estimated credibility limits ; and d) too large estimated credibility limits.*

This test procedure was systematically applied herein. Some of the results obtained show a tendency to underestimate the variance of the posterior distributions (too narrow estimated credibility limits, see in sections 3.5 and 3.6). In these cases, more dispersed distributions $\hat{F}^{disp}_{Q_i^{(T)}}$ have been generated, keeping the same median but with a standard deviation multiplicated by a factor m= (1.2 ; 1.4 ; 1.6 ; 1.8). This is obtained by applying the following formula to each estimated value $\hat{Q}_i^{(T)}$:

$$\hat{Q}^{(T)}_{i,disp} = m * \left(\hat{Q}_i^{(T)} - median\left(\hat{Q}_i^{(T)}\right)\right) + median\left(\hat{Q}_i^{(T)}\right) \tag{3.13}$$

The value of m necessary to retrieve a uniform distribution of $\hat{F}^{disp}_{Q_i^{(T)}}(Q_i^{(T)})$ gives an idea of the level of underestimation of the standard deviation of the posterior distribution $\hat{F}_{Q_i^{(T)}}$.

3.4 Simulation results - case of homogeneous datasets

3.4.1 100-year quantiles $\hat{Q}^{(100)}_{i,ML}$ estimation results

The results of simulations presented in this section were obtained considering a region which characteristics correspond to the Ardèche case study (see in section 4.2.1) : the region includes 5 gauged sites with respective catchment surfaces of 88, 392, 507, 636, and 2240 km^2. The total gauged available record length represents 168 years. The simulations were

conducted according to a regional GEV growth curve and an index flood relation adjusted on the real Ardèche dataset : ξ_0= 3.34, α_0= 2.24, κ_0= -0.16 for the growth curve, β_0= 0.76 for the index flood relation.

Figures 3.11 and 3.12 present the inference results obtained for 1000 homogeneous synthetic samples. Given that no heterogeneity was introduced in the considered region, the Hosking & Wallis heterogeneity test was not applied here. The fluctuations of the maximum likelihood estimates $\hat{Q}_{i,ML}^{(100)}$ (respectively $\hat{q}_{ML}^{(100)}$ for the regional growth curves), and associated 90% credibility bounds $[\hat{Q}_{i,5}^{(100)}, \hat{Q}_{i,95}^{(100)}]$, are presented for both at-site estimations and for the reduced regional growth curves. In order to facilitate comparisons, all estimations have been normalised by the real quantile value $Q_i^{(100)}$ (or $q^{(100)}$), and therefore the position of the real quantile is 1 in each of the presented cases. Both figures reveal the global similitude of at-site estimations obtained with the two regionalisation approaches : the maximum likelihood estimates and the width of computed credibility intervals fluctuate in similar ranges, particularly in the case of high catchment surfaces. However, clearly opposite behaviours are observed for the local growth curves : in the case of the reference approach, the maximum likelihood positions have a limited dispersion and the widths of credibility intervals appear much narrow if compared to at-sites estimates ; with the proposed approach, credibility intervals are clearly wider and the maximum likelihood positions appear more dispersed for the regional growth curves than for at site estimates.

Some explanations for these differences can be proposed. Let first consider the case of the proposed approach (see Fig 3.11.b). In this approach the index flood relationship (parameter β) has to be estimated based on each regional sample. According to equation (3.3), the sensitivity of the index flood value μ_i to β is increasing with the watershed area : for a catchment surface of 1 km^2, for instance, the index flood value does not vary with β. Therefore, the smallest surfaces have a lower weight in the determination of the index flood relationship, and as a consequence the calibrated index flood values are less accurate for small surfaces. The figure 3.11.b shows that this effect already appears for the lowest catchment surface of 63 km^2, for which the dispersion of estimated quantiles increases in comparison with larger watersheds. The regional growth curve corresponds here to an elementary catchment surface of 1 km^2 : this is largely in the extrapolation range of the calibration in the direction of small surfaces and explains the larger uncertainties observed in this case.

In the case of the reference approach, the regional growth curve has a different meaning : the reduced discharges correspond here to discharges values divided by the expec-

tancy of the distribution. Given that it does no depend on the calibration of an index flood relationship, the regional growth curve is better determined in this case (see Fig 3.11.a and 3.12.a). However, using this approach, the index flood values are estimated based on the mean of each at site series, and are therefore also affected by some uncertainty associated with sampling fluctuations. This uncertainty mainly affects at site estimates : it can be noticed on figures 3.11.a, and 3.12.a, c that the median positions do not change but the dispersion is higher for at sites estimates than for the regional growth curve. Given that this uncertainty is associated to the estimation of the mean of each at site series, it will decrease as the local record length available increases : this will be confirmed in the section 3.4.3 by running an additional simulation using different at site sample sizes.

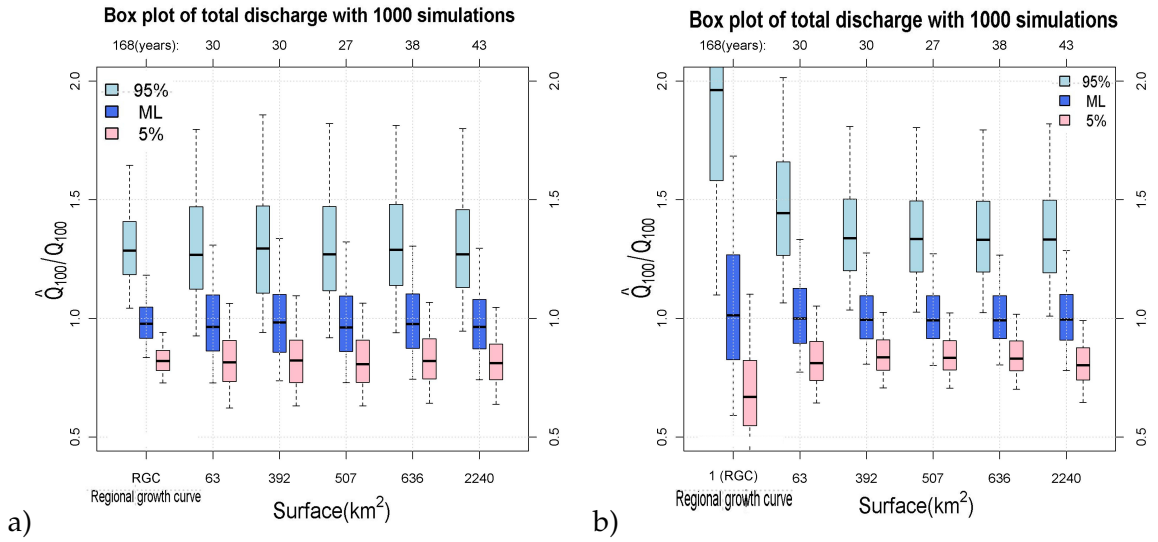


FIGURE 3.11 – *Dispersion of the estimated maximum likelihood values* $\hat{Q}_{i,ML}^{(100)}$ *(respectively* $\hat{q}_{ML}^{(100)}$ *for the regional growth curve), and of corresponding 90% credibility bounds* $[\hat{Q}_{i,5}^{(100)}, \hat{Q}_{i,95}^{(100)}]$*, computed based on 1000 simulated samples : a) reference approach, b) proposed approach. All the presented values have been divided by the real 100-year quantiles* $Q_i^{(100)}$ *(or* $q^{(100)}$*).*

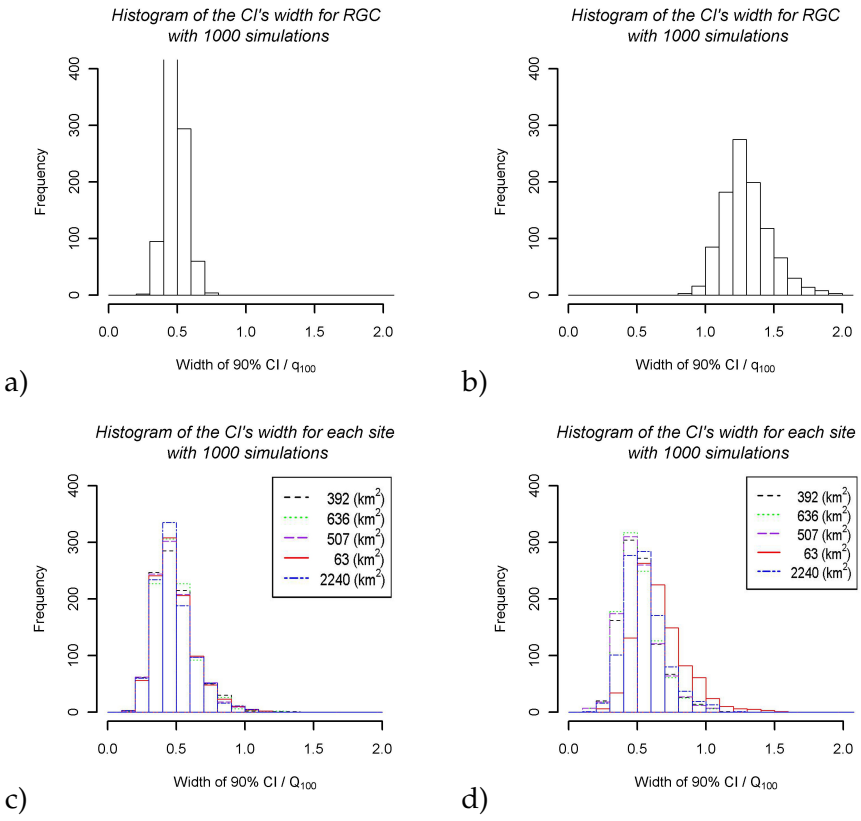


FIGURE 3.12 – *Width of the estimated 90% credibility intervals* $[\hat{Q}_{i,5}^{(100)}, \hat{Q}_{i,95}^{(100)}]$ *computed based on 1000 simulated samples : a) regional growth curve for reference approach, b) regional growth curve for proposed approach, c) at site estimations for reference approach, d) at site estimations for proposed approach. All the presented values have been divided by the real 100-year quantile* $Q_i^{(100)}$ *(or* $q^{(100)}$*).*

3.4.2 Reliability of estimated credibility intervals

The results of the computation of $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ are presented in figure 3.13 and figure 3.14. These results clearly illustrate some limitations of the posterior distributions $\hat{F}_{Q_i^{(100)}}$ estimated based on the Bayesian MCMC procedure.

In the case of the reference approach, the dispersion of the at site posterior distributions $\hat{F}_{Q_i^{(100)}}$, and hence the width of the 90% credibility intervals presented in the figures 3.11.a and 3.12.c, are clearly underestimated (see Fig 3.13.b, c). In the same time the dispersion of posterior distribution $\hat{F}_{q^{(100)}}$ appears to be overestimated for the regional growth curves (see Fig 3.13.a). What is observed here is again a consequence of the estimation of the index flood based on the at-site sample means : since the corresponding uncertainty is not taken into account in the Bayesian MCMC inference procedure, the computed at-site posterior distributions do not account for all sources uncertainty. If the dispersion of $\hat{F}_{Q_i^{(100)}}$ is increased by 20%, the distribution of $\hat{F}^{disp}_{Q_i^{(100)}}(Q_i^{(100)})$ becomes uniform (case of $\hat{F}^{disp}_{Q_i^{(100)}}(Q_i^{(100)})$

with m= 1.2, see Fig 3.13.d and Fig 3.13.e) : this suggests that the dispersion of the at-site posterior distributions $\hat{F}_{Q_i^{(100)}}$ may be underestimated by about 20% in this case.

In the case of the proposed approach, the effects observed for small catchment surfaces in figures 3.12, 3.13.b and 3.13.d are better illustrated (see Fig 3.15.a and Fig 3.15.b) : the large width of credibility intervals, indeed, is associated with a bias of estimation. This means that the posterior distributions tend to overestimate the real quantile $Q_i^{(100)}$. This effect of bias may be related to the distortions of the regional reduced sample associated with fluctuations of β : using a GEV model, such distortions easily lead to an overestimation of quantiles of high return periods. As revealed in the figure 3.15.c, for large catchment surfaces this effect of bias appears to be compensated by the rescaling phase that also depends on β : in this last case, the posterior distributions and associated 90% credibility intervals appear to be well estimated.

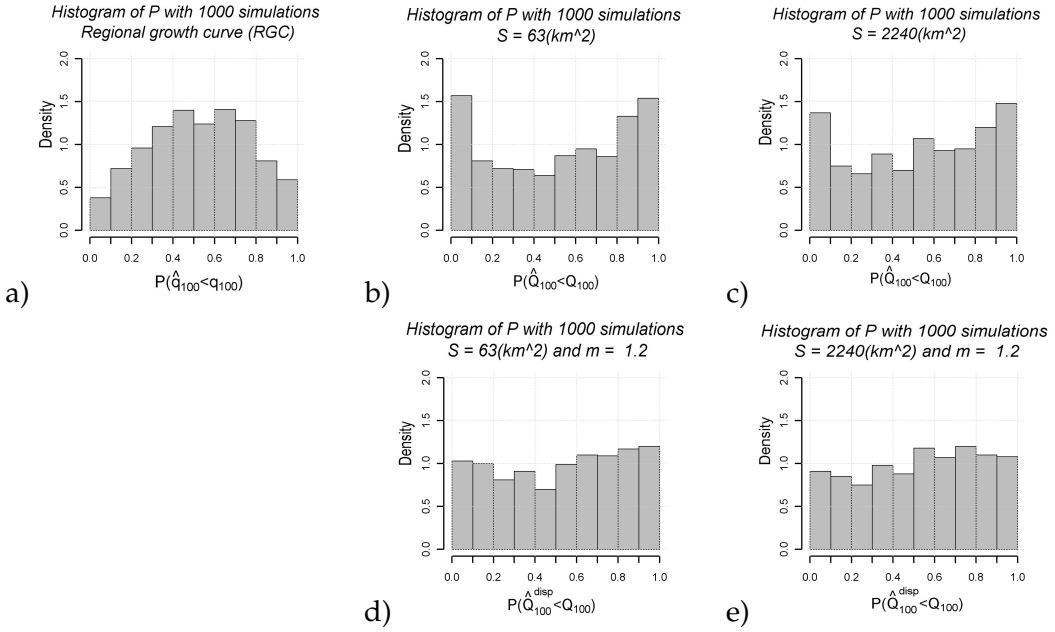


FIGURE 3.13 – *Values of $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ (or $\hat{F}_{q^{(100)}}(q^{(100)})$ in the case of the regional growth curve) computed based on 1000 simulated samples with case of the reference approach : a) regional growth curve, b) S= 63 km^2, c) S= 2240 km^2, and histograms of associated dispersed distributions $\hat{F}^{disp}_{Q_i^{(100)}}(Q_i^{(100)})$ for : d) S= 63 km^2 and m= 1.2, e) S= 2240 km^2 and m= 1.2*

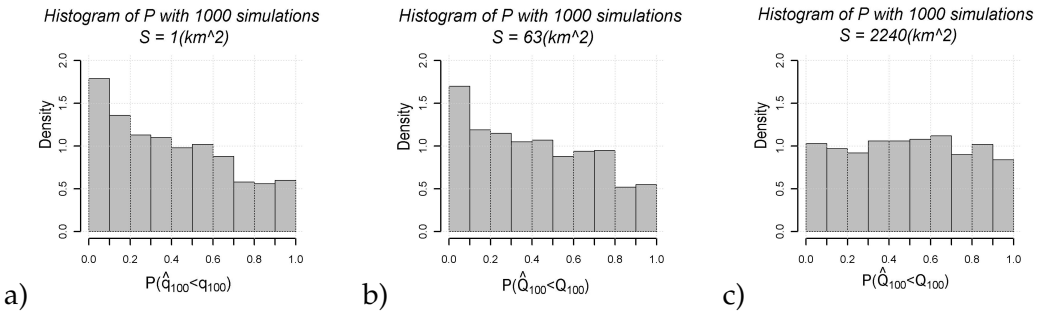


FIGURE 3.14 – *Values of* $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ *(or* $\hat{F}_{q^{(100)}}(q^{(100)})$ *in the case of the regional growth curve) computed based on 1000 simulated samples with case of the proposed approach : a) regional growth curve, b) S= 63 km*^{2}*, c) S= 2240 km*^{2}

3.4.3 Influence of the length of local series

This section presents an additional simulation conducted to illustrate the sensitivity of estimations to the lengths of at-site synthetic series. Here the regional growth curve used was the same as in previous section, but the number of at-site available observations was set to 88 for the 63 km^2 watershed, and respectively changed to 40, 25, 10, and 5 observations for the 392, 507, 636, and 2240 km^2 watersheds. Hence, the total number of observations included in the regional dataset (168) is the same as in the initial case. However, in this case the low number of observations available for the watersheds of high surfaces has a double effect : it increases the uncertainty associated with the estimation of the at-site mean, and it limits the weight of these watersheds in the estimation of β.

The results obtained are presented in the figures 3.15 and 3.16. They confirm the analysis conducted based on the initial simulation : with the reference approach, the dispersion of the maximum likelihood estimates $\hat{Q}_{i,ML}^{(100)}$ and of associated credibility bounds [$\hat{Q}_{i,5}^{(100)}$, $\hat{Q}_{i,95}^{(100)}$] increase with the decrease of the local series length, because of higher uncertainty on the at-site mean estimation. Meanwhile, the underestimation of the dispersion of the posterior distribution $\hat{F}_{Q_i^{(100)}}$ highly increases : for a local series limited to 5 years of observation, even with m= 1.2 an uniform distribution of $\hat{F}^{disp}_{Q_i^{(100)}}(Q_i^{(100)})$ cannot be obtained for the 2240 km^2 catchment surface (see Fig 3.16.e).

With the proposed approach, the lower weight given to the large catchment surfaces reduces the bias observed on the estimated regional growth curves. The dispersion of the maximum likelihood estimates $\hat{Q}_{i,ML}^{(100)}$ and of credibility bounds [$\hat{Q}_{i,5}^{(100)}$, $\hat{Q}_{i,95}^{(100)}$] also decrease for small surfaces, and increase for larger surfaces since the β parameter and thus the index flood less depends on the data available at these sites. The quality of estimation of the posterior distribution remains good for the large surfaces.

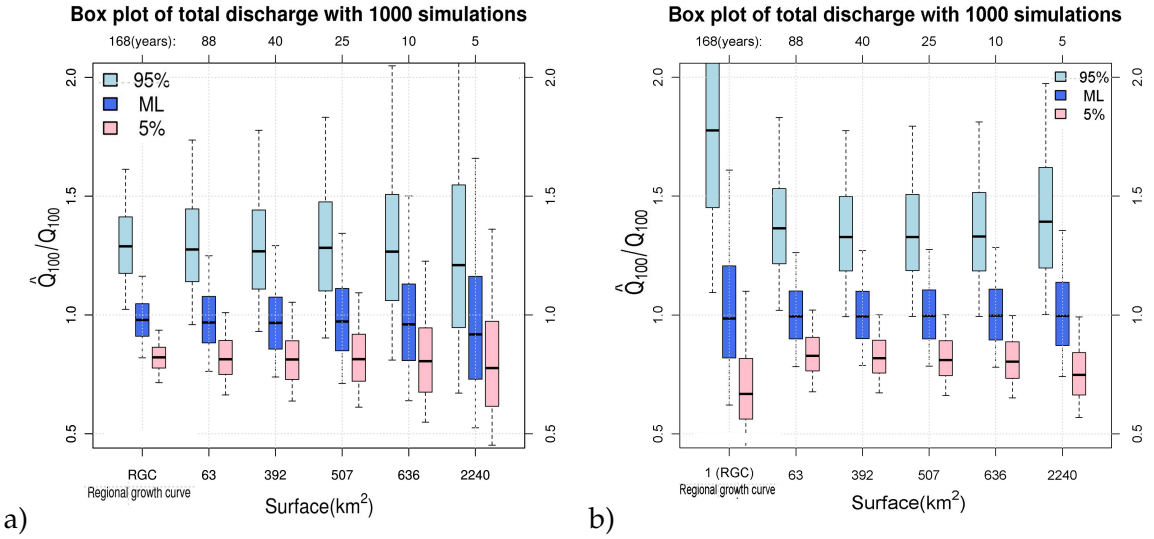


FIGURE 3.15 – *Dispersion of the estimated maximum likelihood values* $\hat{Q}_{i,ML}^{(100)}$ *(respectively* $\hat{q}_{ML}^{(100)}$ *for the regional growth curve), and of corresponding 90% credibility bounds* $[\hat{Q}_{i,5}^{(100)}, \hat{Q}_{i,95}^{(100)}]$*, computed using 1000 simulated samples with different at-site series length (88, 40, 20, 10, 5 years) : a) reference approach, b) proposed approach. All the presented values have been divided by the real 100-year quantiles* $Q_i^{(100)}$ *(or* $q^{(100)}$*).*

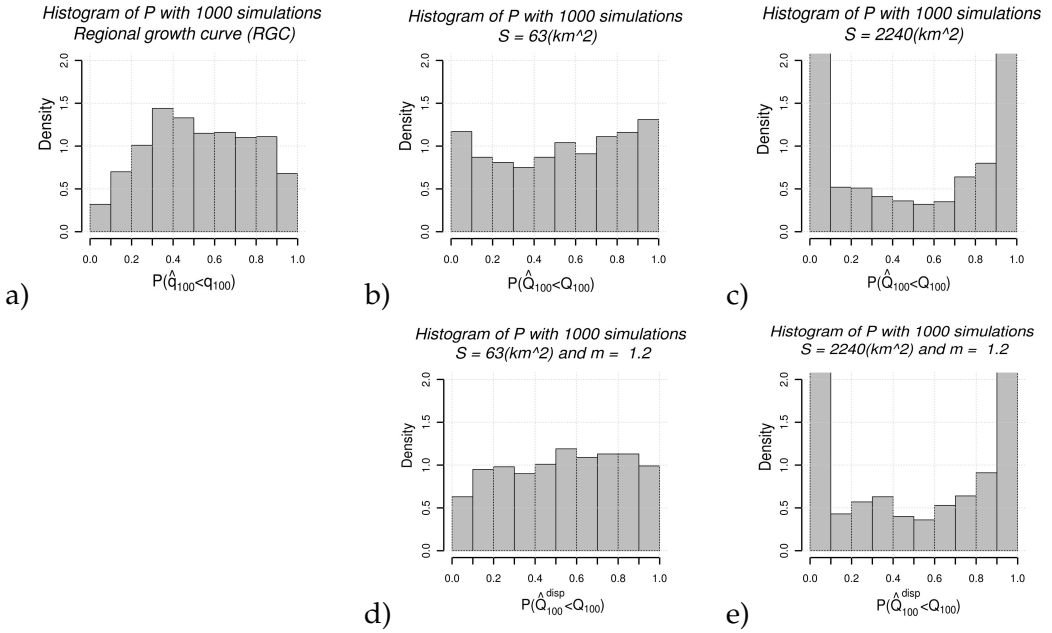


FIGURE 3.16 – *Values of* $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ *(or* $\hat{F}_{q^{(100)}}(q^{(100)})$ *in the case of the regional growth curves) computed based on 1000 simulated samples with different at-site series length (88, 40, 20, 10, 5 years) with reference approach : a), regional growth curves, b) S= 63 km*^{2}*, c) S= 2240 km*^{2}*, and histograms of associated dispersed distributions* $\hat{F}^{disp}_{Q_i^{(100)}}(Q_i^{(100)})$ *for : d) S= 63 km*^{2} *and m= 1.2, e) S= 2240 km*^{2} *and m= 1.2*

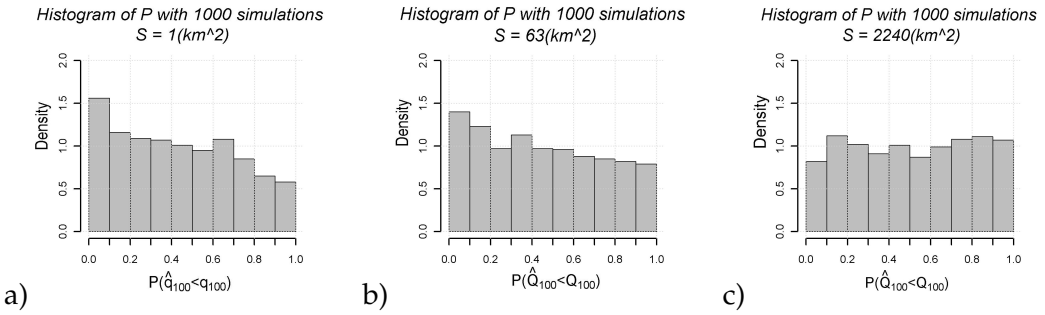


FIGURE 3.17 – *Values of $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ (or $\hat{F}_{q^{(100)}}(q^{(100)})$ in the case of the regional growth curves) computed based on 1000 simulated samples with different at-site series length (88, 40, 20, 10, 5 years) with proposed approach : a) regional growth curves, b) S= 63 km^2, c) S= 2240 km^2.*

3.5 Simulation results - case of heterogeneous datasets

3.5.1 Heterogeneity in the index flood relation

The consequences of the introduction of heterogeneities in the index flood relation are presented in figures 3.18, 3.19 and 3.20. The simulations presented here again correspond to the characteristics of the Ardèche region (see section 3.5.1). However, here the index flood values used to generate at-site series do not respect perfectly the theoretical index flood relationship : some heterogeneity was introduced based a log-normal random value centred on S^{β_0} and with standard deviation ($\delta * S^{\beta_0}$) with δ= 0.1 or 0.3 (see section 3.3.2).

The results confirm that the reference approach is not sensitive to such heterogeneities, which is completely logical since this approach does not rely on the index flood relationship (see Fig 3.18.a and c). With the proposed approach, figure 3.18.b, d and figure 3.19 show a general increase of the dispersion of maximum likelihood estimates $\hat{Q}_{i,ML}^{(100)}$ and credibility bounds [$\hat{Q}_{i,5}^{(100)}$, $\hat{Q}_{i,95}^{(100)}$], associated with an increase of the effect of bias on posterior distributions $\hat{F}_{Q_i^{(100)}}$. For important heterogeneities (coefficient of variation δ= 0.3), the bias is observed even for large catchments surfaces (see Fig 3.18.d and Fig 3.19). The dispersion of the posterior distribution computed on the regional growth curves is also largely underestimated (see Fig 3.19.a). However, as heterogeneity is more limited (coefficient of variation δ= 0.1, see Fig 3.18.a, b and Fig 3.20), the results observed remain close to those obtained in the case of homogeneous data sets : the bias remains limited except for the lowest surfaces ; the width of the credibility intervals remain close to those obtained with the reference approach (see Fig 3.20), and do not appear to be underestimated based on the $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ test.

These results finally suggest that the proposed approach may resist to a certain level of

heterogeneity of the proposed index flood relation within the considered regions. Important heterogeneities should nevertheless be avoided, which may lead to limit the extent of the considered region in comparison with the reference approach.

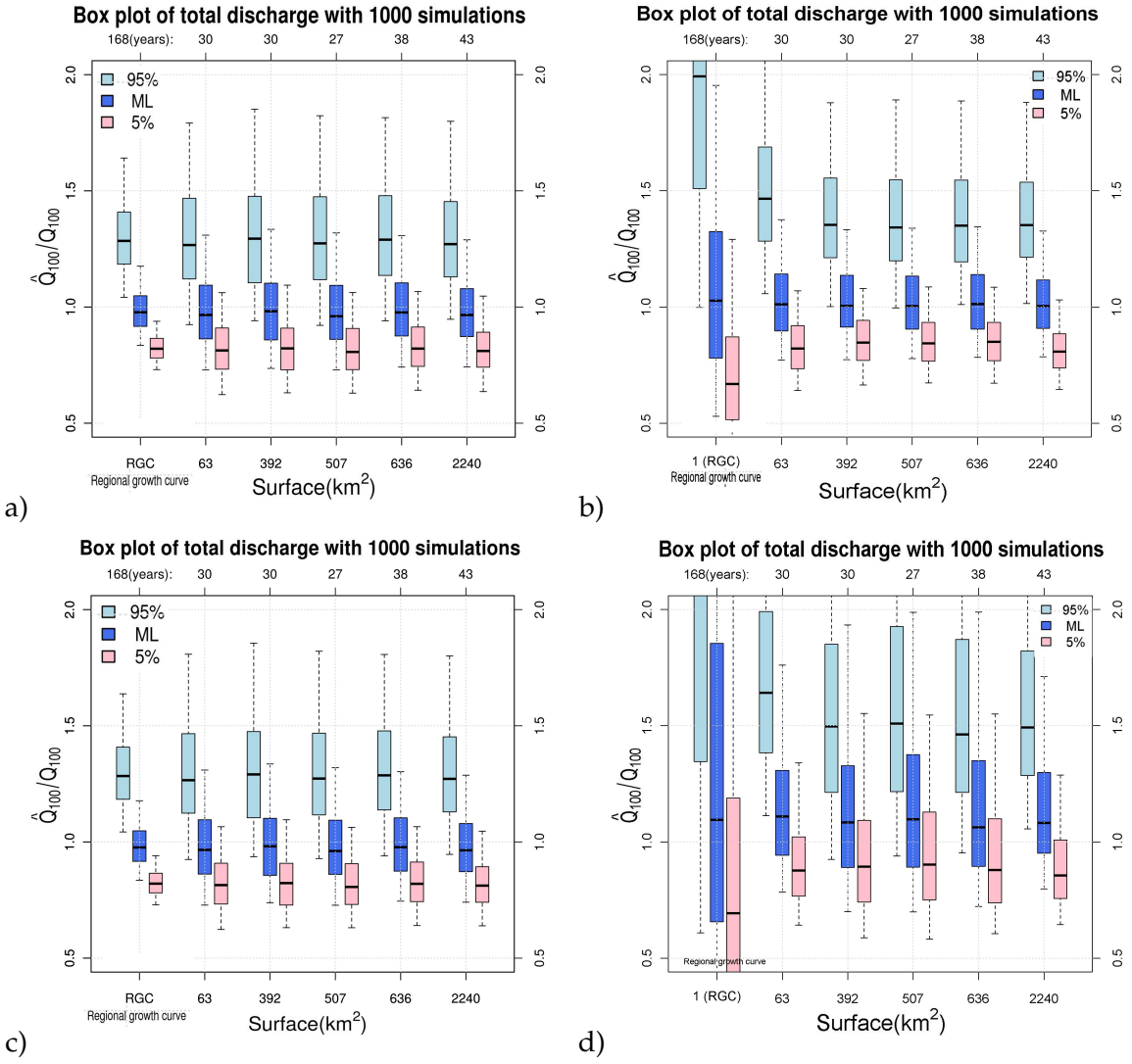


FIGURE 3.18 – *Dispersion of the estimated maximum likelihood values* $\hat{Q}_{i,ML}^{(100)}$ *(respectively* $\hat{q}_{ML}^{(100)}$ *for the regional growth curve), and of corresponding 90% credibility bounds* $[\hat{Q}_{i,5}^{(100)}, \hat{Q}_{i,95}^{(100)}]$*, computed using 1000 simulated samples, with introduction of heterogeneity in the index flood relationship : a) reference approach with* $\delta = 0.1$*, b) proposed approach with* $\delta = 0.1$*, c) reference approach with* $\delta = 0.3$*, d) proposed approach with* $\delta = 0.3$*. All the presented values are divided by the real 100-year quantile.*

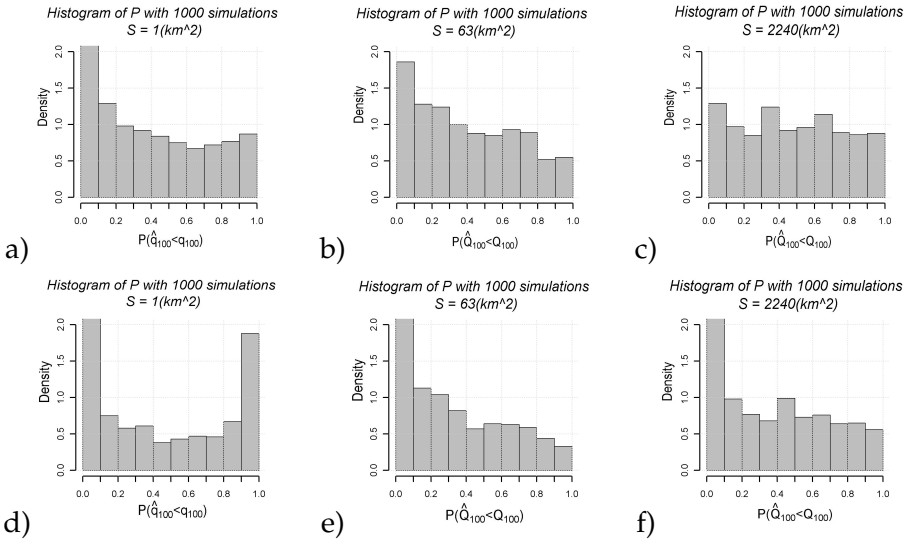


FIGURE 3.19 – *Distribution of* $\hat{F}_{Q_i^{(100)}}(Q_i^{(100)})$ *values (or* $\hat{F}_{q^{(100)}}(q^{(100)})$ *computed based on 1000 simulated regional samples with introduction of heterogeneity in the index flood relationship - case of the proposed approach : a) regional growth curves (*δ*= 0.1), b) S= 63 km*^{2} *(*δ*= 0.1), c) S= 2240 km*^{2} *(*δ*= 0.1), d) regional growth curves (*δ*= 0.3), e) S= 63 km*^{2} *(*δ*= 0.3), and f) S= 2240 km*^{2} *(*δ*= 0.3).*

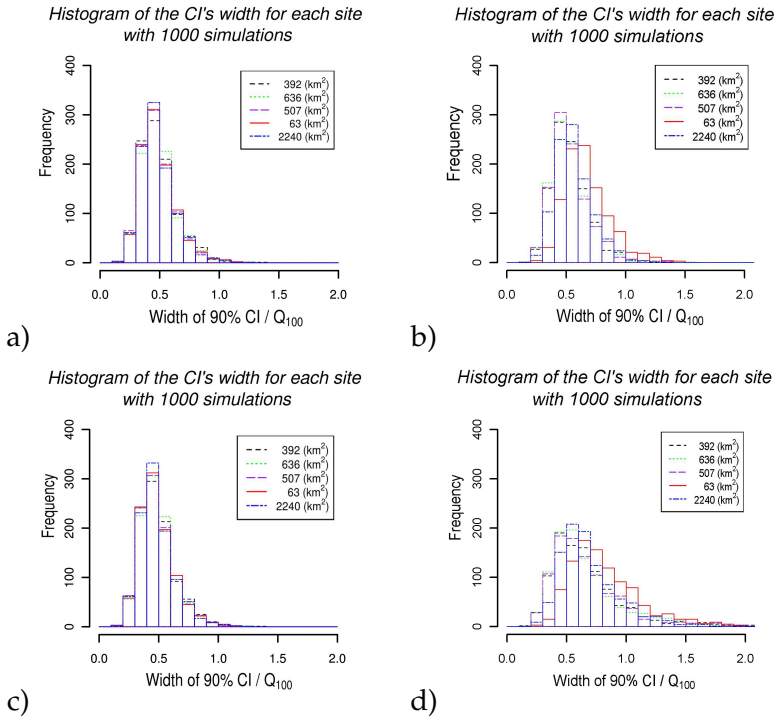


FIGURE 3.20 – *Distribution of the widths 90% credibility intervals on* $\hat{Q}_i^{(100)}$ *computed based on 1000 simulated samples with introduction of heterogeneity in the index flood relation : a) at site estimations for reference approach and* δ*= 0.1, b) at site estimations for proposed approach and* δ*= 0.1, c) at site estimations for reference approach and* δ*= 0.3, d) at site estimations for proposed approach and* δ*= 0.3. All the presented values are divided by the real 100-year quantiles* $Q_i^{(100)}$.

3.5.2 Heterogeneity in the local growth curves

The effects associated to the presence of heterogeneities in the local growth curves will be examined in this section based on specific simulation results : in order to avoid any effect associated with the estimation of the index flood relationship, the region considered here includes 5 gauged series corresponding to the same catchment surface of 1 km^2, and including the same record length (30 years). Therefore, in this specific case the proposed approach does not depend anymore on the index flood relationship since the considered series are already reduced to an elementary surface of 1 km^2. Since the β parameter cannot be estimated based on such data, the results presented here just correspond to the estimation of the local growth curves (parameters ξ, α, and κ). For the Hosking & Wallis approach, each generated series has just to be divided by its mean to be properly rescaled and enable the corresponding regional growth curve to be estimated.

The three levels of heterogeneity in the local growth curves as described in section 3.3.3 were successively introduced in 3 different sets of 1000 synthetic samples : after application of the Hosking & Wallis homogeneity test, the results presented hereafter are based on 827 regional samples in **case 1** (homogeneous regions and $H_1 \leq 1$), versus 826 samples in **case 2** (regions with limited heterogeneity and $H_1 \leq 2$), and 851 samples in **case 3** (higher heterogeneity and $H_1 \leq 3$).

The figure 3.21 presents the dispersion of estimated quantiles $\hat{q}_{i,ML}^{(100)}$ obtained in the three cases of heterogeneity and for each estimation approach (reference of Hosking & Wallis approach and proposed approach). The figure 3.22 presents the distributions of the $\hat{F}_{q_i^{(100)}}(q_i^{(100)})$ probabilities. These results are presented for only one site in the considered region since the watershed area (1 km^2) is the same for all the sites in this case.

These results clearly illustrate the effects of presence of heterogeneities in the local growth curves : (**i**) the estimated quantiles become more dispersed ; (**ii**) the dispersion also affects the credibility bounds [$\hat{q}_{i,5}^{(100)}$, $\hat{q}_{i,95}^{(100)}$] and the width of the credibility intervals increases (see figure 3.21) ; and **(iii)** the dispersion of estimated posterior distributions (and thus the width of credibility intervals) becomes highly underestimated (see figure 3.22.b, c, e, and f).

Figure 3.23.e and f shows that the dispersion of the estimated $\hat{q}_i^{(100)}$ distributions computed with the proposed approach should be increased by a factor of about 60% in **case 2** and 80% in **case 3** to obtain a uniform distribution of $\hat{F}^{disp}_{q_i^{(100)}}(q_i^{(100)})$. Meanwhile, in the case of the reference Hosking & Wallis approach (see Fig 3.23.b, c) the application of the same m coefficient do not completely remove the underestimation effects. This is logical since

the reference approach is still affected by the estimation of the mean of each series (see Fig 3.22.a) : this effect appears to be here of second order given that the selected record length (30 years) are relatively long.

Finally, it can be noticed that the application of the Hosking & Wallis homogeneity test does not remove the effects of heterogeneities introduced in the **case 2** and **case 3**. Even if the test is supposed to disqualify datasets on which a significant presumption of heterogeneity can be identified, the test is not very powerful, its interpretation remains subjective and its results are highly dependent on sampling variability : for instance, a significant part of simulated samples are disqualified in case 1 although an homogeneous region is considered in this case. In **case 2** and **case 3**, despite the presence of significant heterogeneities, the proportion of disqualified samples remains approximatively the same since the thresholds considered have been increased (subjectivity of the test).

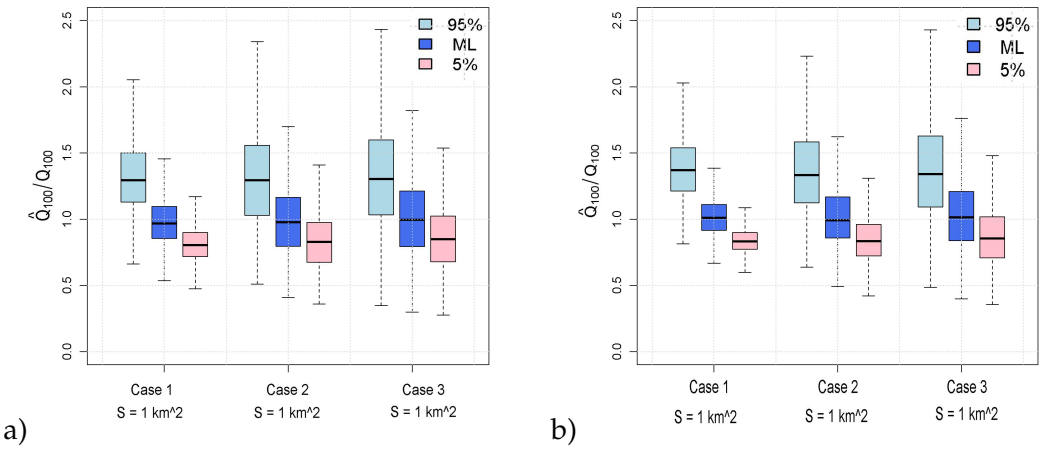


FIGURE 3.21 – *Dispersion of the estimated maximum likelihood values* $\hat{q}_{i,ML}^{(100)}$*, and 90% credibility bounds* $[\hat{q}_{i,5}^{(100)}, \hat{q}_{i,95}^{(100)}]$*, computed for three levels of heterogeneity in the growth curves defined in chapter 3 :* ***case 1*** *correspond the homogeneous regions with* $H_1 \leq 1$ *(827 samples),* ***case 2*** *to intermediate heterogeneous regions with* $H_1 \leq 2$ *(826 samples), and* ***case 3*** *to highly heterogeneous regions with* $H_1 \leq 3$ *(851 samples) : a) reference approach, b) proposed approach. All the presented values have been divided by the real 100-year quantiles* $q_i^{(100)}$*.*

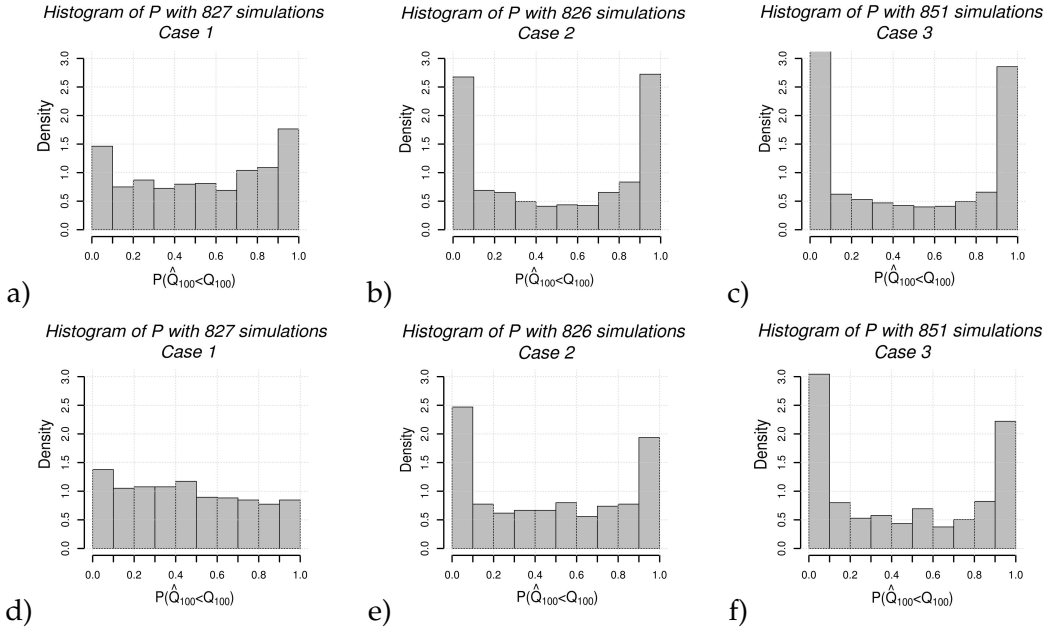


FIGURE 3.22 – *Values of $\hat{F}_{q_i^{(100)}}(q_i^{(100)})$ computed for the three cases of heterogeneity considered : a) reference approach - **case 1**, b) reference approach - **case 2**, c) reference approach - **case 3**, d) proposed approach - **case 1**, e) proposed approach - **case 2**, and f) proposed approach - **case 3**.*

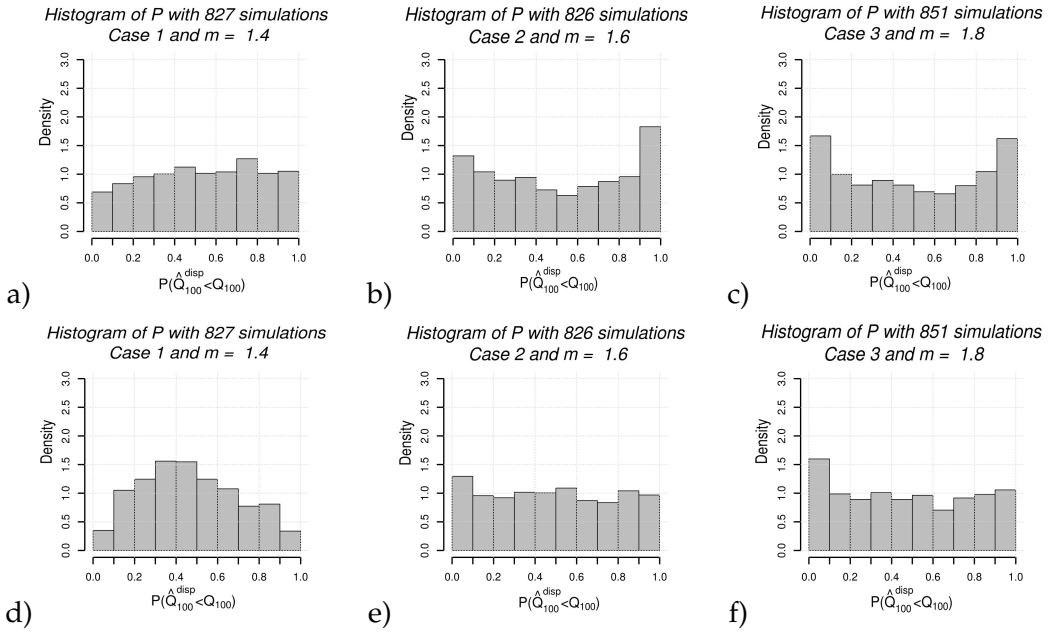


FIGURE 3.23 – *Values of $\hat{F}_{q_i^{(100)}}(q_i^{(100)})$ computed for the three cases of heterogeneity considered : a) reference approach - **case 1** with m= 1.4, b) reference approach - **case 2** with m= 1.6, c) reference approach - **case 3** with m= 1.8, d) proposed approach - **case 1** with m= 1.4, e) proposed approach - **case 2** with m= 1.6, and f) proposed approach - **case 3** with m= 1.8.*

3.6 Conclusions

This chapter enabled to compare the respective performances of the two regional flood frequency analysis approaches described in chapter 2 (the proposed approach and the reference of Hosking & Wallis approach), when applied to the same datasets (datasets limited to gauged series).

After some adaptation of the proposed approach of Gaume (2010) in order to provide a better estimation of the uncertainties associated with the calibration of the index flood relationship, a specific simulation framework has been developed in order to **(i)** compare the quantile estimation results obtained for the two approaches, **(ii)** estimate the relevance of the 90% credibility intervals obtained with the Bayesian MCMC procedure in each case, and **(iii)** evaluate the sensitivity of these results to the presence of heterogeneities in the datasets. Two sources of heterogeneity were examined : heterogeneity in the index flood relationship, and heterogeneity of the local growth curves.

The results obtained in the case of homogeneous datasets clearly identify some limitations of both of tested approaches. First, the reference Hosking & Wallis approach shows a real sensitivity to the lengths of the local series available, that directly impact the estimation of the at-site means. Therefore, a limited length of the gauged series results in an increase of uncertainty of the corresponding quantile estimates, and also in a clear underestimation of the widths of credibility intervals. This effect is particularly marked for series of less than 40 years. On the other hand, the proposed approach shows a certain sensitivity to the considered catchment surface : the uncertainty of estimation increases for very low surfaces and may become important for the regional growth curves for instance (surface of 1 km^2). This effect observed for small watersheds is associated to a bias of estimation of the posterior distributions and associated credibility intervals. However, for significant catchment surfaces (exceeding 100 km^2 in the presented case), this effect tends to disappear and in this case the computed credibility intervals appear to be well estimated. Finally, for sufficient catchment surfaces and a common length of local series (about 30 years), the simulations presented tend to indicate that the proposed approach, despite the presence of an additional parameter to be estimated (index flood relationship), is likely to perform better than the conventional Hosking & Wallis approach.

The introduction of heterogeneities in the simulated datasets generally tend to decrease the quality of the estimations. Heterogeneities in the index flood relationship were first considered. Logically, the reference approach does not show any sensitivity to such heterogeneities, meanwhile the proposed approach may be significantly affected : the estimated

quantiles become more dispersed, and a bias of estimation appears. These effects remain nevertheless limited if the heterogeneity is not too important. The introduction of heterogeneities in the local growth curves (simple scaling assumption) appeared much more problematic. Both of the tested methods are significantly affected, with similar consequences : global degradation of the quality of the estimations, associated with an important underestimation of the widths of credibility intervals (dispersion of estimated posterior distributions). Therefore, the question of heterogeneity of the growth curves in the considered regions appears here as an important concern as dealing with regional approaches, since a perfect homogeneity can never be guaranteed in real-world applications.

4 Application to case studies

4.1 Introduction

In the previous chapters, both regionalization approaches (the reference of Hosking & Wallis approach and the proposed approach) have been compared based on simulated data series corresponding to gauged sites. This chapter now presents their application to two case studies for which information on extreme floods at ungauged sites is available. Therefore, here the two approaches will not be compared in a situation for which the inference is based on the same data set : the added value of the information on extreme floods at ungauged sites will be evaluated. The considered case studies are both located in the south-east of France. The first one, the Ardèche region, has already been studied by Gaume (2010). The Ardèche data set includes 168 gauged records at 5 different sites, and 18 estimated extreme peak discharges among which 14 were observed at ungauged sites. This first case study represents a favourable case where a very large amount of information on extremes, representing an equivalent additional record period of about 900 years, could be retrieved. The second case concerns the Var region, affected by a catastrophic flash flood in June 2010. This second region includes 9 gauging stations (with a total of 249 years of records), and a much more limited set of information on extremes : 4 floods inventoried among which 3 were observed before the recording period at one of the gauged sites. The available information nevertheless represents a total record period of about 660 years in this case.

In the both considered case studies, one part of the information on extreme floods corresponds to historical information available at some of the gauged sites of the region. The analysis conducted has been focused in particular on these sites, which enabled to compare the results of a local analysis involving historical information with the results of the two considered regionalisation approaches.

For each case study, after a careful delineation of the regions that can be considered as

statistically homogeneous and the selection of the corresponding datasets, the analysis has been conducted in three steps :

- first, the methods have been applied on the various datasets to determine the general patterns of the best suited regional growth curve and index flood relation : i.e. to determine the possible range of the parameter values of the local statistical model,
- then, Monte Carlo simulations have been conducted drawing, from this statistical model, samples with the same characteristics as the available dataset : number of sites and records, level of heterogeneity in the index flood relation and in the local growth curves. The objective was to evaluate the theoretical performances of each considered approach, given the particular characteristics of the regional dataset and of the statistical distribution of the analysed data. The relevance of credibility intervals computed with the Bayesian MCMC procedure has been also evaluated a priori based on these Monte Carlo simulations,
- finally, the different approaches were tested on the real datasets and the results analysed at the light of the previously obtained theoretical results.

The chapter is organised as follows : the contents of available datasets, the homogeneity tests and the delineation of the regions are presented in the first section. In a second section, the results of both the preliminary simulations and the applications to the available datasets are presented and discussed.

4.2 The two application case studies

4.2.1 The Ardèche region

The Ardèche case study has already been selected by Gaume (2010) for the initial tests of the regionalisation method involving ungauged extremes. It has been analysed again herein to illustrate the changes introduced by (**i**) a redefinition of the content of the inventory of extremes, and (**ii**) the new inference procedure including the calibration of the index flood relation.

The region is located in the south-east of the Massif Central in the south of France (Cevennes region), one of the regions in Europe exposed to the most frequent and severe flash floods (Gaume et al. 2009). It is drained by three main streams : the Ardèche river (2380 km^2), the Eyrieux river (860 km^2), and the Doux river (635 km^2).

Available gauged series

Five gauging stations located in the south part of the region under the influence of the Mediterranean climate provide accurate, continuous and long series of measured discharges, representing a total recording period of 168 years. The situation of the 5 gauging stations is reported on figure 4.1 and the characteristics of corresponding samples are presented in table 4.1. The catchment areas at gauged sites vary from 63 to 2240 km^2, and the local record lengths are relatively homogeneous, varying from 27 to 43 years.

TABLE 4.1 – *The sample size and catchment surfaces of the selected gauging stations in the Ardèche region.*

Station	**Name river**	**Period of data**	**NP** (year)	**S** (km^2)
Beauvene	Eyrieux	1969-1998	30	392
Vogue	Ardèche	1966-2003	38	636
Chambonas	Chassezac	1971-1997	27	507
Saint Laurent	Borne	1968-1997	30	63
Saint Martin	Ardèche	1963-2005	43	2240

Homogeneity of the growth curves

The empirical growth curves of the available gauged series of maximum annual peak discharges are presented on figure 4.2. On this figure each series has been rescaled by the corresponding index flood, estimated using either the average of each series or the regionally calibrated index flood relationship (β = 0.76). In both cases, the shape of 5 reduced empirical growth curves appears similar in shapes and magnitudes, which suggests that the region may be considered as homogeneous as far as the shapes of the growth curves are considered. The tail of the Saint Laurent growth curve may appear discordant, but is determined by one single flood that may be considered as an outlier at this stage of the analysis.

However, the application of the Hosking & Wallis heterogeneity measure (see in section 2.3.3) does not clearly confirm the homogeneity of this region. The H_1 value based on the 5 gauged series is equal to 2.7, which suggests that the region is "possibly heterogeneous" according to the interpretation suggested by Hosking and Wallis (1997).

Despite this relatively high H_1 value, the region has been supposed homogeneous for the regional flood frequency analyses.

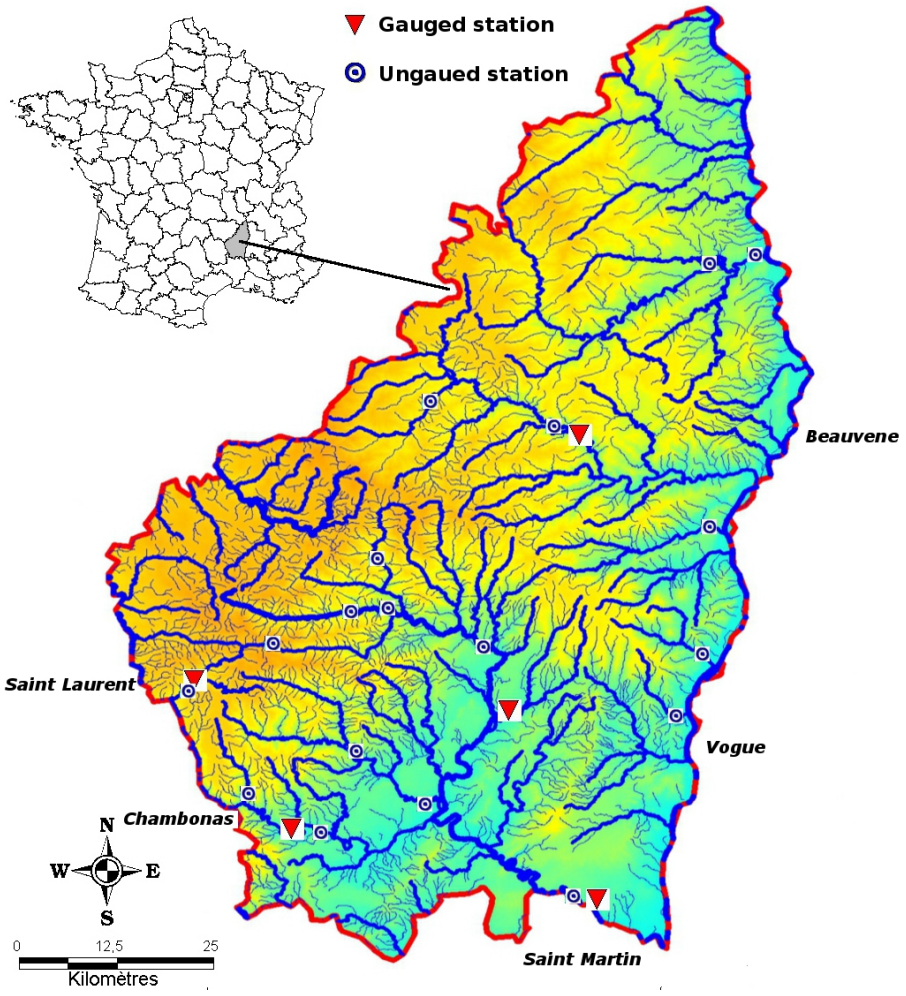


FIGURE 4.1 – *Location of 5 gauging stations and 18 ungauged sites within the Ardèche region.*

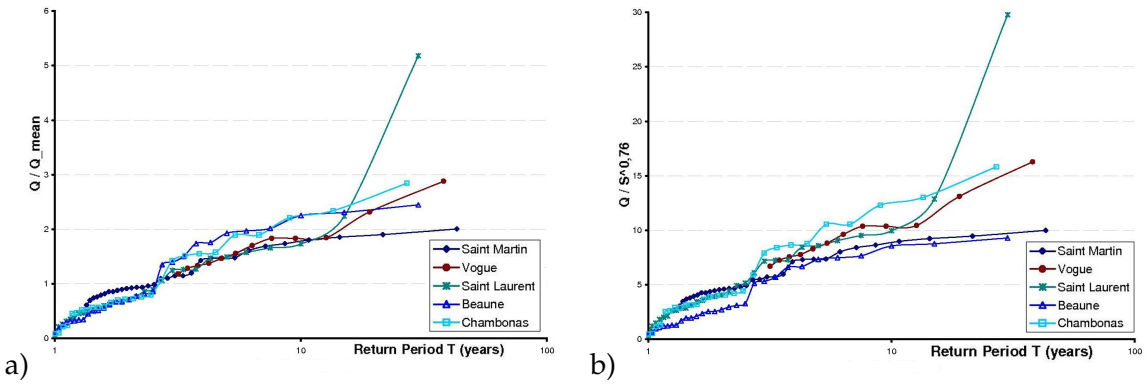


FIGURE 4.2 – *Empirical growth curves of 5 gauged series of the Ardèche set : a) discharge values divided by the average discharge of each series, and b) discharge values divided by* $S^{0.76}$.

Homogeneity of the index flood relation within the region

The region appears to be relatively homogeneous in terms of index flood relationship. This is first suggested by figure 4.2 showing that the index flood estimation procedure does

not affect too much the overall similarity between the reduced empirical growth curves. Figure 4.3 also confirms this impression. This figure compares the averages of the 5 gauge series of maximum annual peak discharges with average values computed from similar random samples generated with the best suited regional statistical model and index flood relationship. Overall, the distance between the averages of the samples and the calibrated index flood relationship (*continuous line*) may be attributed to sampling fluctuation.

The position of the average value of the Eyrieux river data set (392 km^2) may be considered as a little too low even if not completely improbable according to the proposed model [1]. This index flood model has been supposed to hold for the Ardèche south region according to these figures. In the next section, variability in the index flood relationship will be simulated with a value of δ= 0.1 (see chapter 3 in section 3.3.2), consistent with figure 4.3.

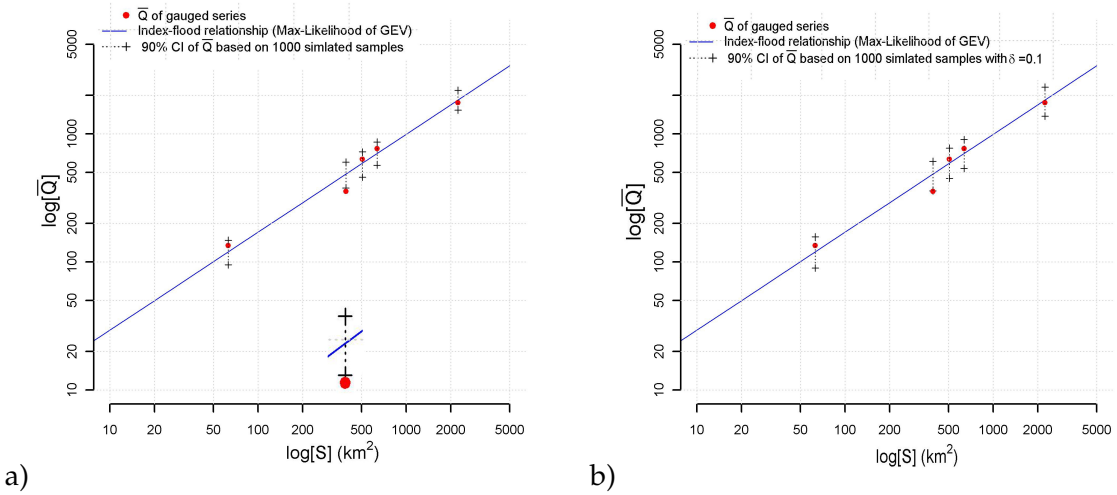


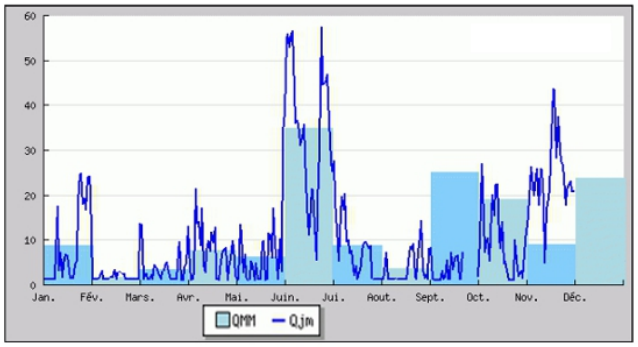
FIGURE 4.3 – *Comparison between the averages of the gauged series of maximum annual peak discharges and the distributions (only the 90% confidence bounds are shown) of the average values of random samples simulated using the adjusted regional model : a)intervals correspond to 90% fluctuations, b) intervals correspond to 90% normal sample fluctuations, corresponding to δ= 0.1.*

Available information on extreme flood events

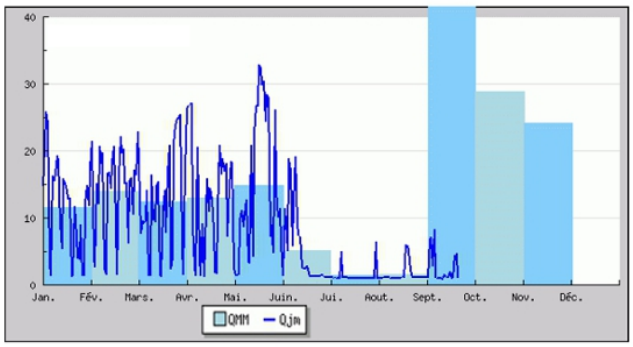
Using the large set of documentary sources available from the European flash flood data centre (Flood site and Hydrate European Projects), 14 additional estimated ungauged

1. Relatively lower peak discharge values on the Eyrieux river may be explained by the position further North of this watershed in the region with less influence of the Mediterranean climate inducing the large flash flood events.

extremes could be added to the regional dataset, corresponding to the highest observed peak discharge during the last 50 years at the considered locations. Similar information could be obtained at 4 gauged sites, providing estimates of the extreme peak discharge during the 50 years preceding the gauging period. The maximum estimated discharges and the catchments surfaces of the 18 sites for which information on extremes was finally selected are presented in table 4.2. Note that this set is reduced if compared to the set initially used Gaume (2010) that contained some redundancies (values reported at very close sites for the same event and river). The case of the Chambonas stream gauge on the Chassezac is also an interesting. The two largest floods that occurred during the gauging period in September 1980 and 1992 did apparently damage the stream gauge (see Fig 4.4), and the corresponding discharges are not reported in the available measured data series. This sorting out of the extreme discharges because of damaged measurement devices may lead to biased samples of maximums and statistical inference results. Even measured discharge series should be thoroughly criticized. It should, in particular, be verified that gaps do not precisely correspond to the largest flood events. An estimate of the largest peak discharge in Chambonas could be retrieved and introduced in the analysis.



a) Flood in 1992



b) Flood in 1980

FIGURE 4.4 – *Available series of discharges for the Chambonas gauging station on the Banque HYDRO database. The series is interrupted at the date of the two major floods occurred in the region.*

TABLE 4.2 – *Detail of the extreme floods reported to be the highest ones in the last 50 years, or last 50 years before the gauging period for gauged sites (bold).*

Location	**River name**	**Date**	**Q** (m^3/s)	**S** (km^2)	**Associated period**
Pont de Rolandy	Ardèche	22/09/1992	1150	150	50
Aubenas	Ardèche	26/09/1992	2200	480	50
Sauze St.Martin	**Ardèche**	**30/09/1958**	**4500**	**2240**	**50**
Rosieres	Beaune	04/10/1958	1820	210	50
Joyeuse	Beaune	30/09/1958	1000	100	50
Chambon	**Borne**	**30/09/1958**	**100**	**11**	**50**
Vans	Boudaric	03/11/1989	130	6	50
Burzet	Bourges	22/09/1992	350	47	50
Chambonas	**Chassezac**	**21/09/1980**	**3360**	**510**	**50**
Dorne	Dorne	1963	630	78	50
Lamastre	Doux	03/08/1963	970	242	50
Pont de Cesar	Doux	03/08/1963	1500	635	50
Barrage des Collanges	**Eyrieux**	**1963**	**1685**	**343**	**50**
Meyras	Fontoliere	22/09/1992	900	130	50
Meysee	Lavezon	30/09/1960	500	56	50
Rieut ord	Loire	01/09/1992	444	62	50
Pouzin	Ouveze	10/08/1967	700	140	50
Saliouse	Saliouse	1980	300	61	50

4.2.2 The Var region

The Var region is also located in the south-east of France. The region includes five main rivers : the Argens river, Gapeau river, Giscle river, Siagnole river and Grenouiller river. The Argens river has the largest catchment surface. This watershed includes several significant tributaries : Caramy river, Issole river, Cauron river, Reyran river, Aille river and Nartuby river (see Fig 4.5).

Available stream-gauge series

The gauged series available in the region include 472 records at 17 gauging stations (see table 4.3). Even if some of the series have a very limited length (Cogolin, Cabasse, Carces),

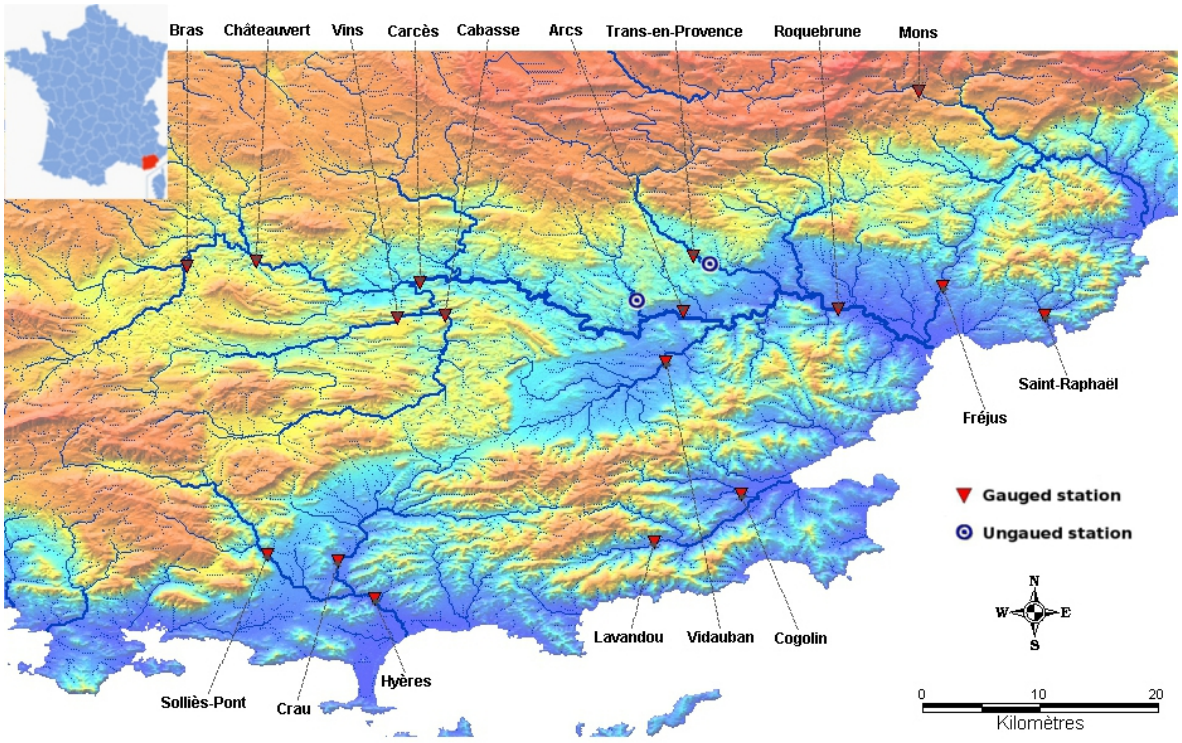
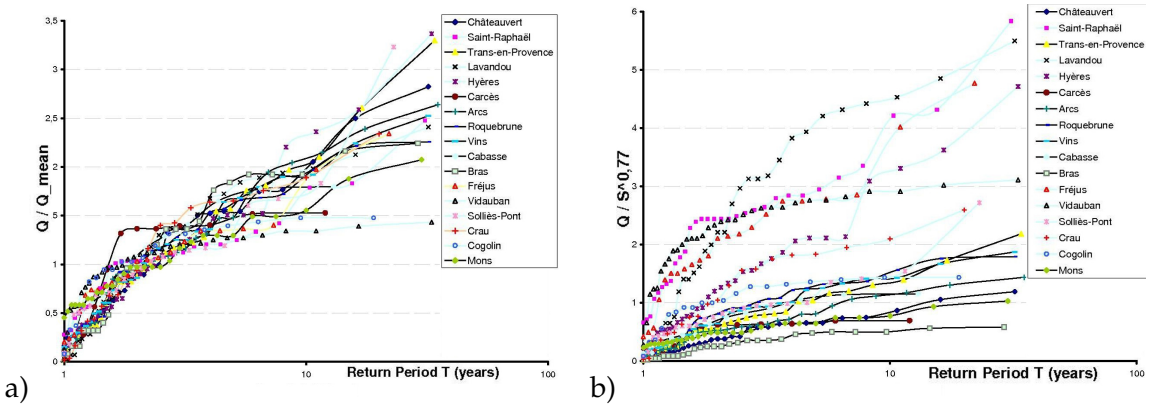


FIGURE 4.5 – *Location of 17 gauging stations and ungauged sites within the Var region.*



a) b)

FIGURE 4.6 – *Empirical growth curves of 17 gauged series in the Var region : a) values divided by the average discharge of each series, and b) values divided by* $S^{0.77}$.

most of them offer a record length of about 30 years. The catchment areas vary from 44 to 2530 km^2.

Homogeneity of the growth curves

Figure 4.5 presents the empirical growth curves based on each of the 17 gauged series. Each series has been reduced using an index flood value corresponding to the average discharge of the series (see Fig 4.6.a), or to a calibrated index flood relationship $S^{0.77}$(see Fig 4.6.b).

TABLE 4.3 – *The record periods and catchment areas of the gauging stations in the Var region.*

Region	**Station**	**Name river**	**Period of data**	**NP** (year)	**S** (km^2)
North part	Bars	Cauron	1975-2003	29	154
	Chateauvert	Argens	1972-2003	32	485
	Vins	Caramy	1972-2003	32	215
	Cabasse	Issole	1975-1988	13	223
	Carces	Argens	1989-2000	12	1181
	Arcs	Argens	1969-2003	35	1730
	Trans-en-Provence	Nartuby	1970-2003	34	190
	Roquebrune	Argens	1972-2003	32	2530
	Mons	Siagnole	1980-2009	30	87
Coastal part	Sollies-Pont	Gapeau	1968-2003	36	169
	Crau	Real Martin	1984-2003	20	277
	Hyeres	Gapeau	1971-2003	33	517
	Lavandou	Mole	1969-2000	32	44.4
	Vidauban	Aille	1971-2003	33	229
	Cogolin	Giscle	1985-2003	16	65.8
	Frejus	Reyran	1970-1991	22	71
	Saint Raphael	Grenouiller	1970-2000	31	48

This figure illustrates the important regional variability of the growth curves as well as the variability of the average annual peak discharge values revealed by the spread of the curves on the figure 4.6.b. The H_1 value of the Hosking & Wallis heterogeneity measure equal to 3.78 confirms this impression of heterogeneity of the growth curves.

Based on these first results, the region has been divided in two parts : (**i**) the north part including 9 gauging stations (see figure 4.6.b, sub-region presented by black lines) and (**ii**) the coastal part including the remaining 8 gauging stations (see points and light blue lines on the figure 4.6.b).

In the coastal part, the figures and the H_1 value again indicates significant heterogeneity (H_1 = 3, see table 4.4). The average annual peak discharges seem also spatially variable (see Fig 4.6.b). Only the north part of the region finally appears to be acceptably homogeneous (H_1= 1.5). Therefore, only this last region has been considered hereafter. It is by chance the region that has been affected by the extreme flash flood event of June 2010.

TABLE 4.4 – *The results of the Hosking & Wallis homogeneity tests (H_1 values).*

Region	Number of sites	Value of H_1	Implication of H_1 test
Var region	17	3.78	Definitely heterogeneous
Coastal part	8	3.0	Definitely heterogeneous
North part	9	1.5	**Acceptably homogeneous**

Homogeneity of the index flood relationship within the region

As shown in figure 4.7.a, sampling fluctuation hardly explains the distance between the estimated local sample average and the theoretical average according to the calibrated index flood relation for two sites in the North part. The locally estimated average annual peak discharge appears too low at Bras on the Cauron river (154 km^2) and also slightly too low at Chateauvert (485 km^2) on the Argens river. These two sites are located in the upstream part of the Argens river basin, which for various reasons could be exposed to less intense flood events on average. To take into account this possible heterogeneity, two different regions have been considered depending the tested regionalization method. The reference of Hosking & Wallis approach was applied using the 9 available series, given that the assumption of homogeneity of the growth curve appears to be reasonable for these series. For the application of the proposed approach based on the index flood relationship, the region was limited to 7 series (without two sites at Bras and Chateauvert, Fig 4.7.b), based on the necessary assumption of homogeneity of the index flood relation within the region.

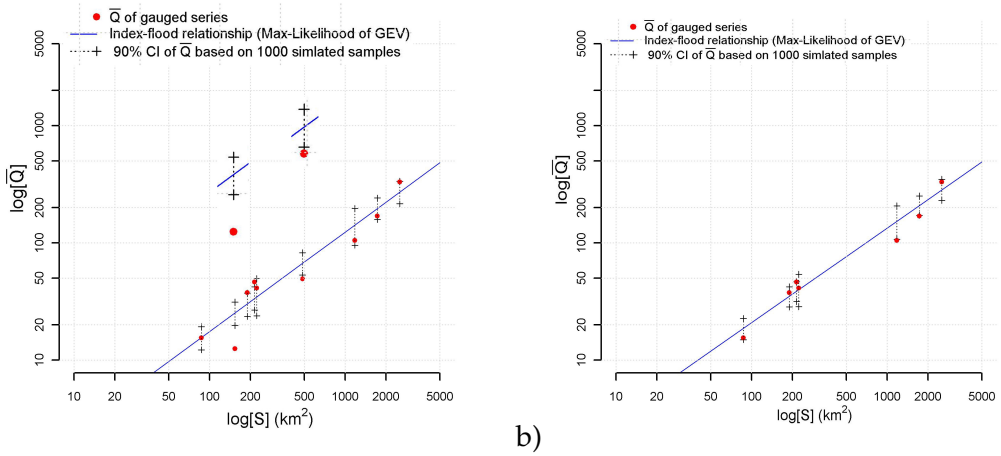


FIGURE 4.7 – *Comparison between the averages of the gauged series of maximum annual peak discharges and the distributions (only the 90% confidence bounds are shown) of the average values of random samples simulated using the adjusted regional model (maximum likelihood criterion) : a) 9 gauging stations (North part), and b) 7 gauging stations (Sub-north part).*

Available information on extreme flood events

The information on extremes is very limited in this region. No real information on past extremes could be for instance unearthed from the local archives except for the sites of Draguignan and Trans-en-Provence on the Nartuby river. Available information is limited to the peak discharges estimated after the June 2010 extreme floods that have affected the downstream part of the Argens basin (Payrastre et al. 2012) and to two major historical floods that occurred on the Nartuby River in 1674 and 1827. Finally, we decided to include information on extremes in the regional dataset as follows :

- the estimated peak discharges of historical floods documented on the Nartuby River in addition to the gauged series at Trans-en-Provence. Given that the gauging station was destroyed in 2010, the 2010 flood was considered as an historical flood. Each of the three historical floods finally considered (1674, 1827 and 2010) was supposed to be the highest one on a 120-year period (historical period of 360 years before the gauging period),
- the estimated peak discharge of the 2010 flood on the Florieye River was considered as an ungauged extreme flood. This flood is one of the highest recorded during the 2010 event. It was supposed to be the highest over a 300-year length period on the corresponding watershed (see Table 4.5).

Table 4.5 – *Detail of the information extreme floods available in the Var region.*

Location	River name	Date	Q (m^3/s)	S (km^2)	Associated period
Trans-en-Provence	Nartuby	1676	350	196	120
Trans-en-Provence	Nartuby	1827	400	196	120
Trans-en-Provence	Nartuby	06/2010	450	196	120
Draguignan	Florieye	06/2010	490	87	300

4.3 Test of the methods on the Ardèche datasets

4.3.1 Comparison of the inference methods based on simulations

Theoretical accuracy of the estimated quantiles

As indicated in introduction of this chapter, given that each case study represents a particular context for data availability, some simulations have been conducted in each case in or-

der to evaluate the theoretical performances of the different possible inference approaches. Thus, the simulations presented in this section are based on samples which characteristics correspond to the real data sample of the Ardèche region (5 gauged sites with the same record lengths, and 18 ungauged extremes representing each the maximal flood in a 50-year period). These samples were generated according to a regional GEV growth curve and an index flood relationship calibrated according to the maximum likelihood criterion on the Ardèche region dataset (β_0= 0.76, ξ_0= 3.34, α_0= 2.24, and κ_0= -0.16). The simulation procedure has been described in section 3.3.4. Based on the characteristics of the real dataset, the presence of heterogeneities in both the local growth curves and index flood relationship was considered in the simulations. The heterogeneity in the index flood relationship was introduced based on a standard deviation δ= 0.1 which seems to be consistent with the observations, and heterogeneity in the local growth curves based on L_{CV} and $L_{skewness}$ (or L_{CA}) fluctuations corresponding to samples of n= 40 years length. Only the simulated samples with a Hosking & Wallis heterogeneity measure H_1 lower than 2 were considered (see Fig .4.8). This led to limit the number of samples to 799 (for 1000 samples generated).

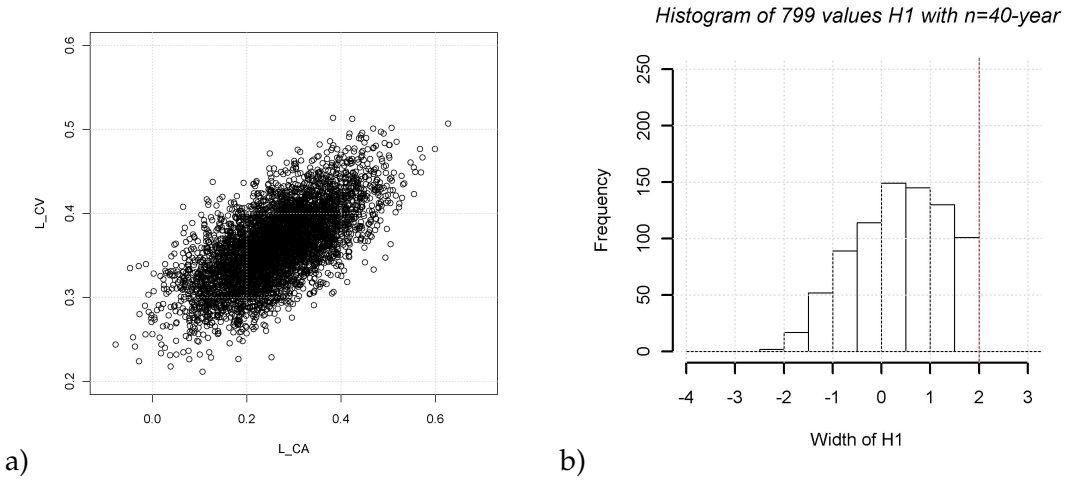


FIGURE 4.8 – *Simulation of samples corresponding to the Ardèche region case study : a) dispersion in the L_{CA} and L_{CV} used for the simulation of the random samples, and b) histogram of 799 H_1 values ($H_1 \leq 2$) corresponding to the 799 samples selected among the 1000 regional samples generated.*

The simulation results presented in figure 4.9 are focused on two gauged sites : St.Laurent (63 km^2), and St.Martin (2240 km^2) : the smallest and the largest watershed areas of the regional set. The presented inference results are all based on the application of the Bayes MCMC procedure on various samples. Four cases are tested for the sake of the comparison :

- **Case 1** : the proposed approach is used based on both gauged series (5 sites) and 18 ungauged extremes,
- **Case 2** : the reference of Hosking & Wallis approach is used based on the 5 gauged series only,
- **Case 3** : a local approach based on the local gauged series and at site available historical floods is used. At Saint Martin this historical information corresponds to one extreme flood known to be the highest one in a pre-gauging period of 50 years length (one threshold value with associated period of 50-years length). At Saint Laurent no historical information is available, and therefore this case is not considered,
- **Case 4** : the local approach based on the available gauged series only is applied at Saint Laurent and Saint Martin.

Figure 4.9 synthesises in a box plot the 799 values of the obtained ratios $\hat{Q}_{i,ML}^{(100)}/Q_i^{(100)}$, where $\hat{Q}_{i,ML}^{(100)}$ is the maximum likelihood estimate of the 100-year quantile and $Q_i^{(100)}$ is the real quantile value. This figure shows that at the Saint Martin gauging site (see Fig 4.9.a), both regional approaches should lead to significant improvements if compared to the local approach without historical information. The dispersion of the maximum likelihood estimates $\hat{Q}_{i,ML}^{(100)}$ is much lower in both cases, all estimations being unbiased (median ratio $\hat{Q}_{i,ML}^{(100)}/Q_i^{(100)}$ equal to 1). However, the proposed approach involving information on extremes clearly outperforms the reference regional flood frequency approach. Moreover, the local approach accounting for the available historical flood also leads to a lower dispersion of the estimated $\hat{Q}_{i,ML}^{(100)}$ than the reference regional approach. The regional heterogeneities that have been introduced in the simulated samples limit the usefulness of the regionalization approaches. However, in the case of the proposed approach, this negative effect of heterogeneities is more than compensated by the large amount of additional information provided by the ungauged extremes.

The ranking between the approaches is not the same at the Saint Laurent gauging site. Figure 4.9.b shows that the dispersion of $\hat{Q}_{i,ML}^{(100)}$ is not significantly reduced when the proposed regionalization approach (**case 1**) is used if compared to the approach based on the local data only (**case 4**). The reference regional approach (**case 2**) leads to slight improvements of the estimated 100-year quantile, comparable in magnitude to the theoretical improvements at Saint Martin. This result is consistent with the conclusions of chapter 3 : the efficiency of the proposed approach is limited for the smaller considered areas due to the growing uncertainties on the calibrated index flood relation, i.e. relation between the average of the

samples and the watershed areas. Overall, these first results are relatively satisfactory as far as the proposed regional flood frequency method is concerned. They surprisingly reveal that the standard regionalization method may be of little help for the estimation of flood quantiles, if there are some regional heterogeneities, which is certainly always the case in real-world applications. Local analyses valuating historical or paleoflood information may provide more accurate flood quantile estimates. In cases for which a significant amount of information on ungauged extremes may be available, the proposed approach may lead to a clear improvement of the accuracy of the estimated quantiles in pesence of some regional variability. This is true except for the smallest considered watershed areas.

It is important to mention here that these conclusions are based on theoretical samples and that they are valid provided that the hypotheses on which the Monte Carlo simulations are based correspond to the real-world case study. On the one hand, the exact statistical distribution of the observed data is never known and can differ from the calibrated distribution. The degrees of freedom of the inference problem exceed the 3 parameters of the calibrated GEV. Considering this, the conducted tests underestimate probably the added values of the historical or regional information. Conversely, the relevance of the homogeneity hypothesis can never be totally verified. It can never be verified for instance that the shapes of the distributions remain similar for larger quantiles and that for instance extreme peak discharge values observed on small watersheds really inform about high return period quantiles of large watersheds or vice versa. This leads probably to under-estimate the uncertainties associated to the quantiles estimated through regional flood frequency methods.

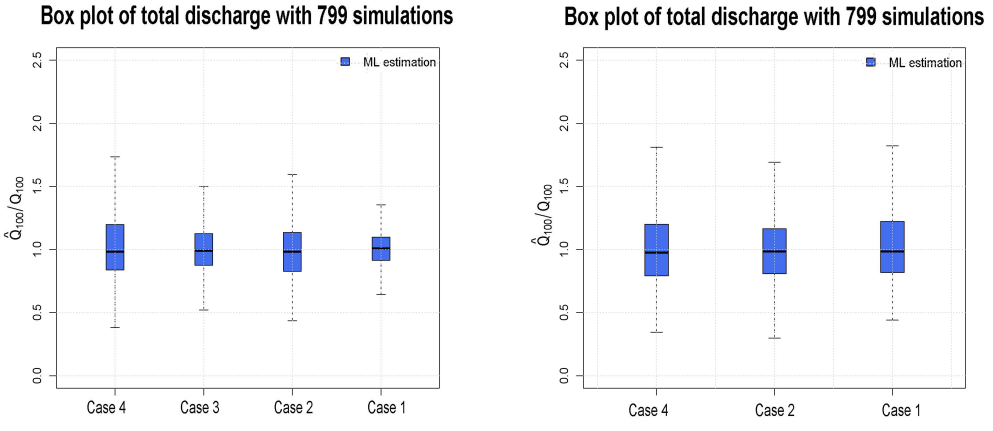


FIGURE 4.9 – *Dispersion of the estimated maximum likelihood values* $\hat{Q}_{i,ML}^{(100)}$*, computed using 799 simulated samples, with regional variability in the index flood relation (*δ*= 0.1) and growth curves (n=40-year) : a) Saint Martin, and b) Saint Laurent. All values have been divided by the real 100-year quantile* $Q_i^{(100)}$*.*

Theoretical reliability of the computed credibility intervals

Figures 4.11 and 4.13 present the distributions of $\hat{F}_{Q_i^{(100)}}\left(Q_i^{(100)}\right)$ obtained respectively at the Saint Martin and Saint Laurent gauging sites. These figures show that the dispersion of the posterior distributions $\hat{F}_{Q_i^{(100)}}$, and associated width of 90% credibility intervals, is correctly estimated in the case of the local approaches, but clearly underestimated in the case of both regional approaches. In the case of the largest catchment surface (Saint Martin), figure 4.12 shows that the dispersion of the estimated $\hat{F}_{Q_i^{(100)}}$ distribution should be increased by a factor of about 40% in **case 1** and of 80% in **case 2**, to obtain reliable credibility limits (i.e. simulated uniform distribution of $\hat{F}^{disp}_{Q_i^{(100)}}\left(Q_i^{(100)}\right)$, see chapter 3 for an explanation). According to the simulation results presented in chapter 3, this underestimation effect can be associated with the presence of heterogeneities in the regional growth curves (**case 1** and **case 2**), and also to the uncertainty introduced by the estimation of a local average for the reference regional approach (**case 2**). This last factor explains that the underestimation is larger in **case 2**. For the lowest catchment surface (Saint Laurent), the figures 4.13 and 4.14 show that the underestimation effect highly increases with the proposed approach (**case 1**) and remains the same for the reference approach. In this case, a correction factor of 80% the dispersion of the estimated $\hat{F}_{Q_i^{(100)}}$ distribution is not sufficient to obtain a uniform distribution of $\hat{F}^{disp}_{Q_i^{(100)}}\left(Q_i^{(100)}\right)$ in **case 1**. This large underestimation for small catchment areas is explained by the larger uncertainties in the calibrated index flood relation (i.e. in the relation between local sample average and catchment area (see Fig 4.10)).

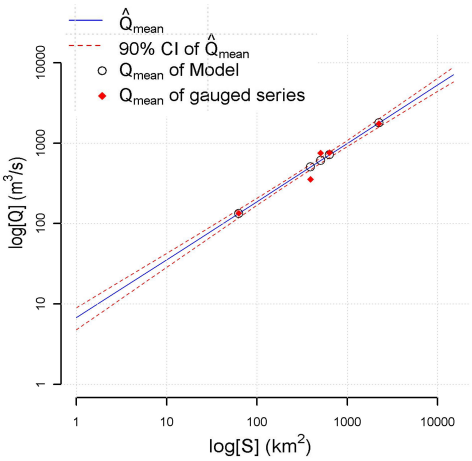


FIGURE 4.10 – *Means of estimated posterior distributions* $\hat{F}_{Q_i^{(100)}}$ *obtained in the* ***case 1*** *and associated 90% credibility intervals computed on the index flood relationship (and compare with the means of gauged series (red points).*

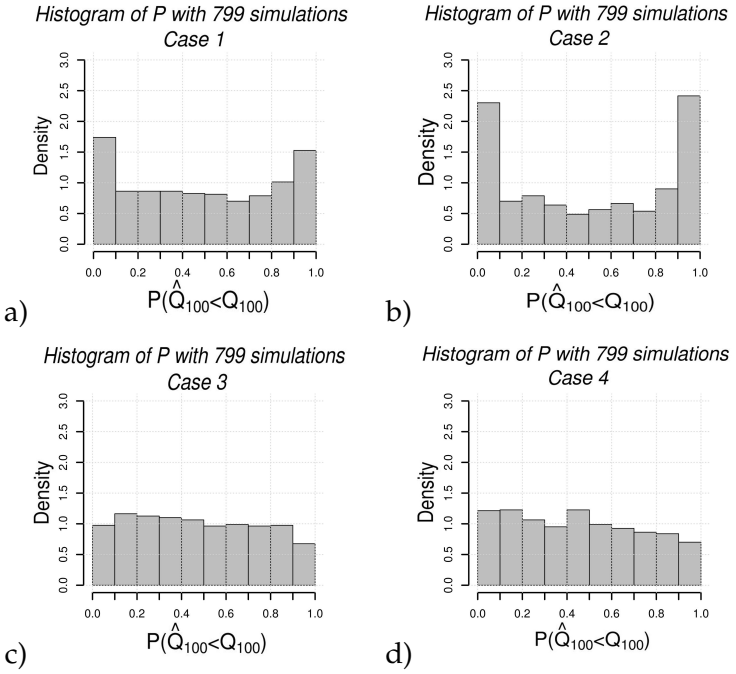


FIGURE 4.11 – *Distribution of the values of* $\hat{F}_{Q_i^{(100)}}\left(Q_i^{(100)}\right)$ *at Saint Martin gauging site for the 799 simulated samples : a)* ***case 1****, b)* ***case 2****, c)* ***case 3****, and d)* ***case 4****.*

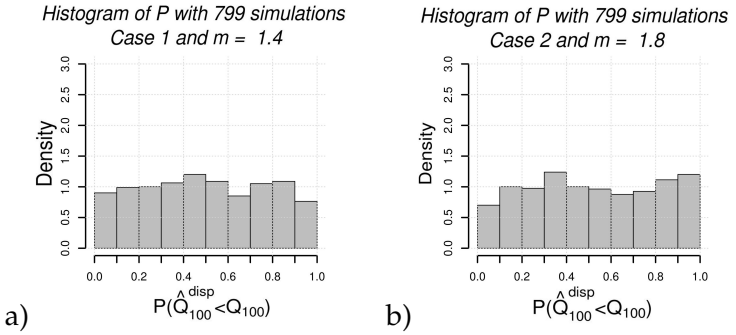


FIGURE 4.12 – *Distributions of the values of* $\hat{F}^{disp}_{Q_i^{(100)}}\left(Q_i^{(100)}\right)$ *at Saint Martin gauging site for the 799 simulated samples : a)* ***case 1*** *and correction factor m= 1.4, b)* ***case 2*** *and m= 1.8.*

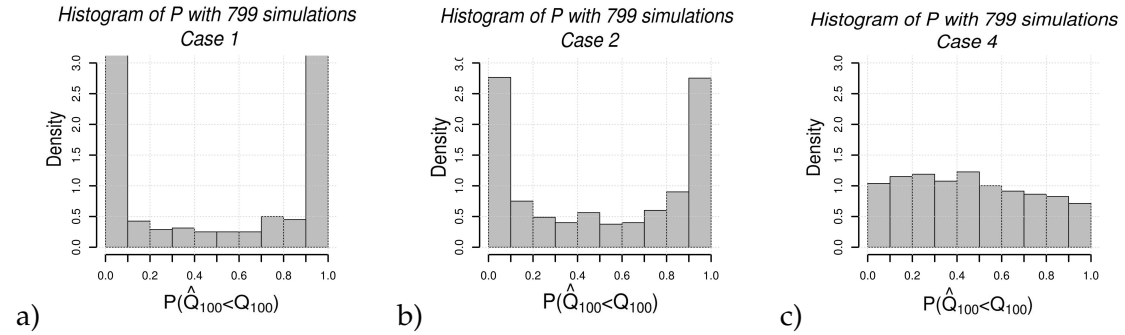


FIGURE 4.13 – *Distributions of the values of* $\hat{F}_{Q_i^{(100)}}\left(Q_i^{(100)}\right)$ *at Saint Laurent gauging site computed for the 799 simulated samples : a)* ***case 1****, b)* ***case 2****, and c)* ***case 4****.*

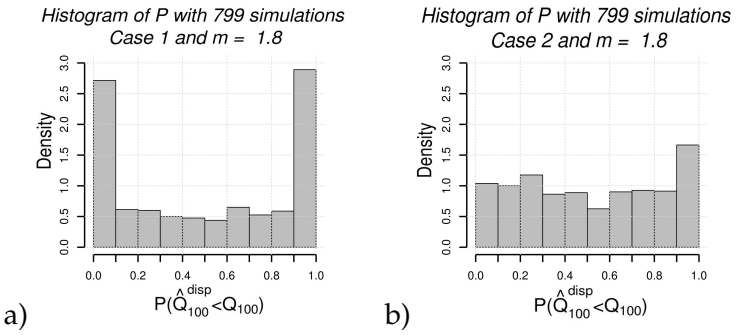


FIGURE 4.14 – *Distributions of the values of* $\hat{F}^{disp}_{Q_i^{(100)}}\left(Q_i^{(100)}\right)$ *at Saint Laurent gauging site computed for the 799 simulated samples : a)* ***case 1*** *and m= 1.8, b)* ***case 2*** *and m= 1.8.*

4.3.2 Application of the methods to the Ardèche datasets

Figures 4.15 and 4.16 present the shapes of the estimated GEV distributions and associated 90% credibility intervals at the two gauging stations of Saint Laurent (63 km^2 corresponding to the Borne watershed), and Saint Martin (2240 km^2 corresponding to the outlet of the entire Ardèche watershed). Table 4.6 compares the 100-year quantiles estimates and corresponding initial and corrected credibility intervals, obtained with the four inference approaches already used for simulations : (**case 1**) regional analysis with availability of information on extreme floods, (**case 2**) regional analysis limited to gauged series, (**case 3**) local gauged series and local historical information, and (**case 4**) local gauged series.

TABLE 4.6 – *Estimations of the 100-year quantile at Saint Laurent and Saint Martin gauging stations, including the Maximum Likelihood estimate* ($\hat{Q}^{(100)}_{i,ML}$)*, the 90% credibility bounds* $[\hat{Q}^{(100)}_{i,0.05}$; $\hat{Q}^{(100)}_{i,0.95}]$*, and associated width* $\Delta CI = \hat{Q}^{(100)}_{i,0.95}$ - $\hat{Q}^{(100)}_{i,0.05}$*, and corrected* ΔCI *according to the simulation results.*

Surface (km^2)	N_{cont} (years)	N_{hist} (years)	Case	$\hat{Q}^{(100)}_{i,0.05}$ (m^3/s)	$\hat{Q}^{(100)}_{i,ML}$ (m^3/s)	$\hat{Q}^{(100)}_{i,0.95}$ (m^3/s)	$\Delta CI/\hat{Q}^{(100)}_{i,ML}$ (%)	$\Delta CI/\hat{Q}^{(100)}_{i,ML}$ (%)corr
63	30	0	**4**	411.5	**636.5**	1854.8	226.8	227
	168	0	**2**	423.1	**513.5**	706.5	55.2	100
	168	900	**1**	452.1	**511.5**	587.3	26.4	–
2240	43	0	**4**	3427.8	**4038.4**	6195.3	68.5	69
	43	50	**3**	3778.1	**4297.4**	5790.4	46.8	47
	168	0	**2**	5528.8	**6710.3**	9233.4	55.2	99
	168	900	**1**	5778.7	**6863.9**	8278.4	36.4	50

These results show significant differences between quantiles obtained using local analyses (**cases 3** and **4**) and regional analyses (**cases 1** and **2**) for both considered sites. Moreover, the computed credibility intervals are narrower in **cases 1** and **2**, the lower values

of widths being obtained in **case 1**. However, according to simulations results, the widths of these credibility intervals may be underestimated. To get an idea, the factors of correction adjusted in the previous section have been applied to the credibility bounds (see Table 4.6). The results appear closer to the theoretical simulations, if not completely consistent. This illustrates the effect of the specificities of the given samples on the variability of the inference results. In both cases, the regional analyses lead to a significant reorientation of the statistical adjustments if compared to the local approaches even if both results remain consistent according to the credibility limits that are not totally disconnected. This may suggest either (**i**) that the locally available samples are particular - presence of a relatively rare event in the 30-year long Saint Laurent series and absence of large flood events in the 93 year-long Saint Martin series - or (**ii**) that the regional homogeneity hypothesis is not valid, or even (**iii**) that the GEV distribution is not perfectly suited to the data and that its use in the inference procedure leads to under-estimate the degrees of freedom of the inference problem and hence the credibility intervals.

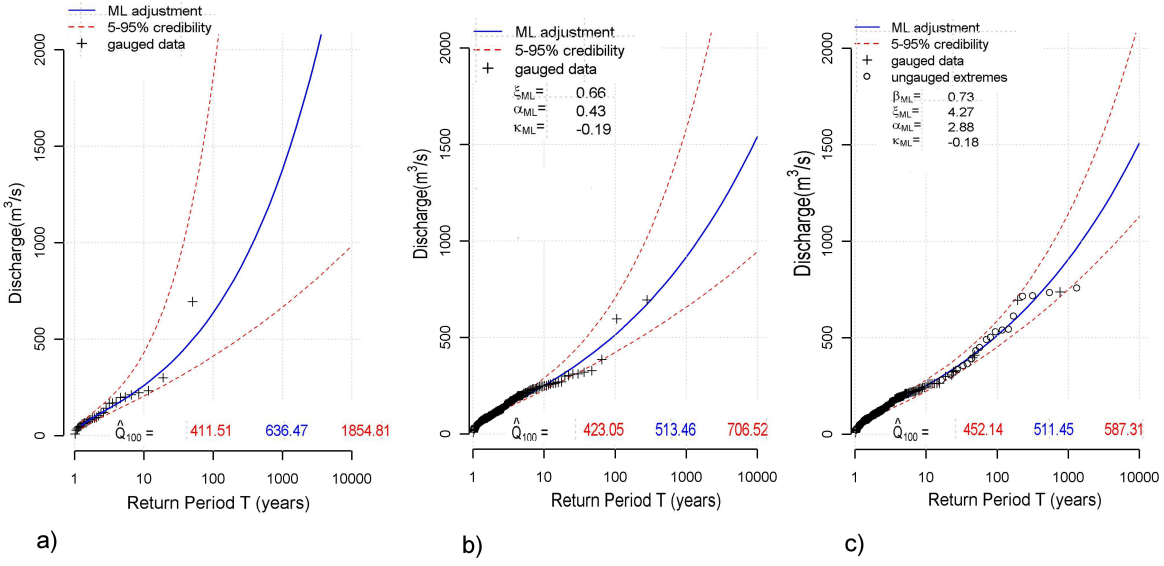


FIGURE 4.15 – *Fitted GEV distributions and associated credibility intervals at Saint Laurent gauging station (63 km²) : a)* ***Case 4*** *- local approach, b)* ***Case 2*** *- regional gauged datasets for reference approach, and c)* ***Case 1*** *- proposed approach including the whole set of ungauged extremes (50 years * 18 = 900 years).*

A detailed analysis of the inference results and data sets provide some clues to confirm or invalidate these hypotheses. Concerning the Saint Laurent site, the locally based statistical adjustment appears to be oriented by the presence in the data set of the high discharge value occurred in September 1980 (about 700 m^3/s), that appears as an outlier if compared to the rest of the series (see Fig 4.15.a). This value remains the highest of the rescaled regio-

nal gauged dataset (see Fig 4.15.b), which explains the narrowing of the adjusted credibility intervals around lower quantile values. But can this new adjustment be really considered as more accurate than the previous one ? Difficult to conclude since it is based on a regional homogeneity hypothesis that can hardly be totally verified, especially if it is considered that the peak discharge distributions are not necessarily of the GEV type. The shape of the calibrated regional growth curve (see Fig 4.15.b) is determined by discharges measured on watersheds which areas are much larger than the area of the Borne river at Saint Laurent (ses Table 4.1). The Hosking & Wallis homogeneity tests shows that the shapes of the peak discharge distributions of the various rivers are similar, regardless of the watershed areas, for low to medium return periods. It does not say anything about the tails of the distributions that may different for different watershed areas. The inclusion of the ungauged extremes confirms the adjustment based on the regional gauged data set (see Fig 4.15.c), even if the computed credibility interval is probably largely under-estimated according to the simulation results (see Fig 4.14). Apart from this confirmation, the strength of the ungauged data set is that it includes information on watersheds with areas comparable to the area of the Borne watershed (see Table 4.2). The maximum rescaled discharge of the Borne river series has been exceeded on none of the 5 to 10 watersheds with similar areas during the last 50 years, confirming empirically the relatively high return period (over 100 years) for this event suggested by both regional analyses, independently on most of the hypotheses on which the regional flood frequency model is based. Finally, this example illustrates that even if the added value of the regional flood frequency analyses may be theoretically limited on average (see Fig 4.9), it can nevertheless be significant on specific case studies.

The case of the Saint Martin gauging station provides another insight into the possibilities and limits of the regional approaches for flood frequency analysis. Both regional approaches provide consistent estimates, which is by the way satisfactory as far as the ungauged extremes are concerned, and lead to significantly larger flood quantiles if compared to the two estimates based on the locally available data (see Table 4.6 and Fig 4.16). This reorientation of the adjustment is significant since the intersection between the 90% credibility intervals obtained in **case 1** (regional analysis including ungauged extremes) and **case 3** (local analysis with extremes) is very limited. In other words, according to the regionally calibrated GEV distribution, the locally observed 93-year series is unlikely. But, the regional flood frequency results are questionable since the Saint Martin watershed is by far the largest of the region. A detailed historical study had been previously conducted

at this specific site (Naulet 2002). According to this study, the largest peak discharge of the considered period of time, occurred in 1958, has been exceeded four times during the 19^{th} century in 1827, 1846, 1890 and 1900. Two estimated values in 1890 and 1827 have reached 7000 m^3/s giving credit to the regionally estimated 100-year quantiles (see Table 4.6). This consistent result is not a demonstration of course of the relevance of both regional flood frequency approaches but is very encouraging and an incentive to further test these methods on new case studies. Finally, the low probability of the data set observed during the 20^{th} century according to the finally calibrated GEV distribution could be sign, but with no certainty :

- of either non-stationary processes with a reduced occurrence of large flood events at Saint Martin during the 20^{th} century. But this is not confirmed by the regional analysis since floods of magnitudes comparable to the floods occurred at Saint Martin during the 19^{th} century according to the calibrated regional models, have been observed on other watersheds in the same area during the same period of time (see Appendix **A.6**),
- or of the inadequacy of the GEV distribution. Threshold - type or S-shaped distributions showing sudden increases of the magnitudes of the discharge quantiles, suggested by some authors (Gaume 2006, Rogger et al. 2011) could explain the apparently erratic occurrence of large discharge values in the observed series as seems to be the case at Saint Martin and Saint Laurent.

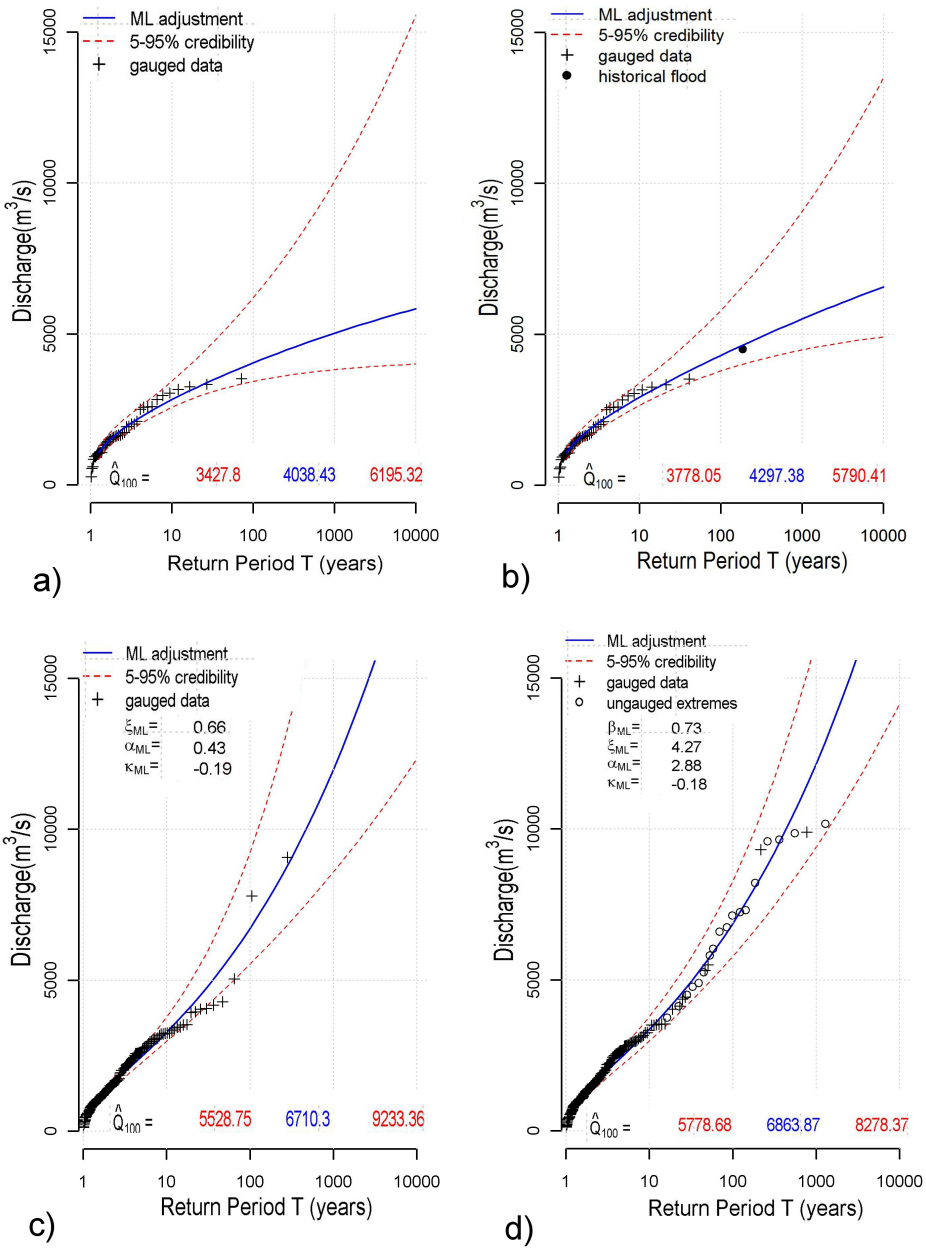


FIGURE 4.16 – *Fitted GEV distributions and associated credibility intervals at Saint Martin : a)* ***Case 4*** *- local approach, b)* ***Case 3*** *- local approach using historical information (one historical flood of 4500 m^3/s and associated period of 50-year), c)* ***Case 2*** *- reference regional approach based on gauged datasets, and d)* ***Case 1*** *- proposed approach including the whole set of ungauged extremes (50 years * 18 = 900 years).*

4.3.3 Partial conclusions

This first application case study provides already rich results, illustrating the potential usefulness of the proposed alternative approach to local or standard regional flood frequency analyses.

– On the two tested watersheds, the regional flood frequency analyses, although based

on simple hypotheses provided results that appeared to be consistent with other sources of information, especially information on historical floods. The case of Saint Martin is particularly striking : the shape of the regionally calibrated distribution which is very different form the locally calibrated distributions is mostly controlled by extreme discharges observed on much smaller watersheds in the region, but still consistent with long historical records that had not been used for the inferences. The simple hypotheses on which both regionalization methods are based were not infirmed by the available dataset. Of course, the suitability of these hypotheses will have to be further tested on a larger set of case studies before any general conclusion can be drawn.

- Some theoretical simulations accounting for possible regional variability, that can exist even in samples considered as homogeneous according to the tests, show that regional flood frequency analyses are not necessarily efficient for reducing the uncertainties in the estimated quantiles. In particular, the standard approach of Hosking & Wallis, based on gauged samples only, appears to theoretically hardly improve the estimates of the quantiles if compared to the local method and to be outperformed by the local approach including historical data. This is a new result.
- The same theoretical simulations indicate that the credibility limits computed with the Bayes MCMC algorithm on regional samples may be largely underestimated since the possible regional variability is not accounted for. This limits the usefulness of the MCMC algorithm, even if the range of the correction factor to be applied to the computed credibility intervals can be anticipated on the basis of the simulation tests as shown here.
- The results on the real-world application differ from the theoretical forecast : the added value of the two regional methods and especially of the proposed one has shown to be larger than the foreseen added value. This illustrates the importance of sampling fluctuation : performances of the methods on one single sample may differ significantly from the average performances estimated on a large number of samples. Since the specificities of the given samples are never known a priori, this pleads for the systematic trial of inclusion of additional information in flood frequency studies. It is not possible to say in advance if the standard or proposed flood frequency studies will be useful or not. The discrepancy between theory and practice, may also indicate that the background hypotheses of the method are not valid and especially that the GEV distribution - selected because it is the most flexible among the avai-

lable theoretical distributions with simple mathematical formulation - is not suited to the observed data sets. This can only be verified if the same tendency is observed repeatedly on a large sample of case studies.

4.4 Test of the methods on the Var datasets

4.4.1 Comparison of the inference methods based on simulations

Theoretical accuracy of the estimated quantiles

This section again presents simulations results that aim to illustrate the hierarchy between the different inference approaches in the particular context of the data sets available in the Var case study. The simulations conducted, indeed, exactly correspond to the characteristics of the Var region : 9 gauged sites with the same surfaces and same record lengths as the real datasets, and 4 extreme floods among which 3 correspond to the 360-year historical period documented on the Nartuby watershed (190 km^2), and the last one corresponds to an ungauged watershed of 87 km^2 for which the maximal flood over a 300-year period has been reported. The corresponding samples were generated according to a GEV distribution (ξ_0= 0.42, α_0= 0.32, and κ_0= -0.33) and an index flood relation (β_0= 0.77), both calibrated on the Var regional dataset. The simulation procedure was again the same, as already described in previous sections. The presence of heterogeneities in both the regional growth curves and index flood relationship was again simulated here with the same parameters as the Ardèche case study. Only the samples with $H_1 \leq 2$ heterogeneity measures based on the Hosking & Wallis homogeneity test were taken into account. This led to limit the number of samples considered to 666 (for 1000 generated samples).

The results presented hereafter will be focused on the Nartuby watershed at Trans-en-Provence (191 km^2) for which a large amount of historical information on extreme floods is available, including the 2010 flood that is one of the highest observed on this river. The inference approaches compared are the same as for the Ardèche case study. However, according to the problems of homogeneity of the index flood relationship identified on the real dataset, the number of gauged sites included in the regional analyses is different here between **case 1** and **case 2** :

- **Case 1** : the proposed approach based on both regional gauged series (limited to 7 sites) and the 4 extreme floods identified (including 3 historical floods on the Nartuby

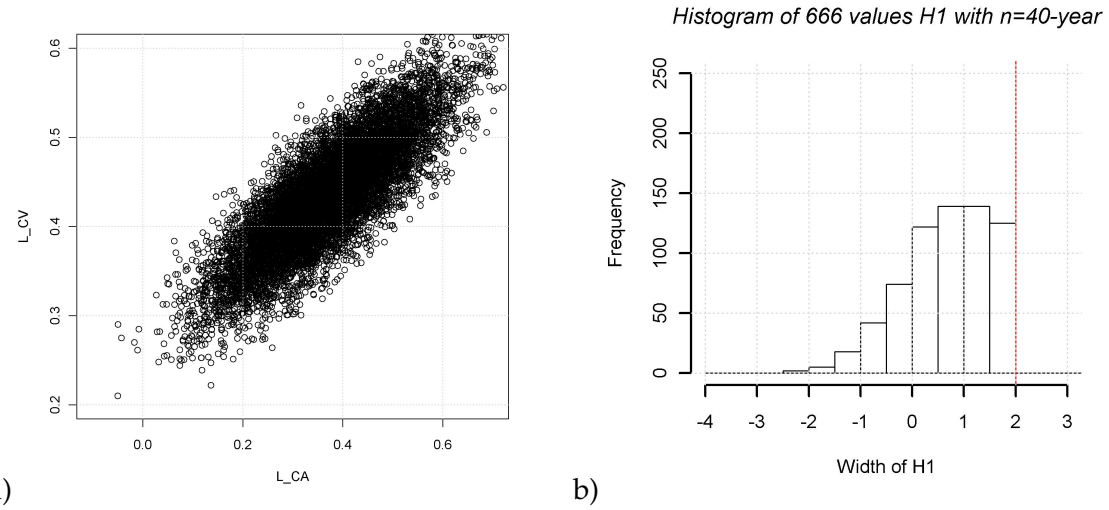


FIGURE 4.17 – *Simulation of samples corresponding to the Var region case study : a) dispersion in the L_{CA} and L_{CV} considered for introduction of heterogeneity in the local growth curves (based on L_{CA} and L_{CV} values computed for random samples of n= 40-year length), and b) histogram of 666 H_1 values ($H_1 \leq 2$) corresponding to the 666 samples selected among the 1000 regional samples generated.*

river, each one being the highest one on a 120-year period, and one flood on the ungauged Florieye river, 87 km^2, being the highest one on a 300-year period),

- **Case 2** : the reference Hosking & Wallis approach based on full regional gauged series (9 sites),
- **Case 3** : the local approach based on gauged series and historical information (3 floods, each one being associated with a 120-year historical period),
- **Case 4** : the local approach based only on the local gauged series.

Figure 4.18 presents the fluctuations of the 666 maximum likelihood estimates of the 100-year quantile $\hat{Q}_{i,ML}^{(100)}$ at Trans-en-Provence, obtained based on simulated samples. As in the previous section, each value has been divided by the real quantile value $Q_i^{(100)}$. This figure again confirms that both regional approaches improve significantly the estimation of $Q_i^{(100)}$ (lower dispersion of $\hat{Q}_{i,ML}^{(100)}$) if compared to local approach without historical information. Again, the proposed approach involving information on extremes performs significantly better than the reference of Hosking & Wallis approach in this case. But a most surprising result is obtained here : this time the local approach involving historical information clearly leads to the most accurate estimation on average. This is explained by the fact that, most of the information on ungauged extreme floods concerns the Nartuby river and is already included in the historical data set. The regional set does only provide one

additional ungauged extreme, which is not sufficient to compensate for the uncertainties induced by the simulated regional variability.

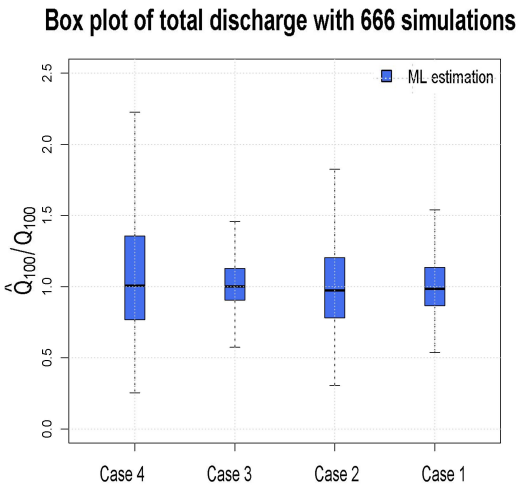


FIGURE 4.18 – *Dispersion of the estimated maximum likelihood values* $\hat{Q}_{i,ML}^{(100)}$ *at Trans-en-Provence gauging site (190 km*^{2}*), computed using 666 simulated samples, with introduction of heterogeneities in the index flood relation (*δ*= 0.1) and the local growth curves (n= 40-year). All values have been divided by the real 100-year quantile* $Q_i^{(100)}$

Theoretical reliability of the computed credibility intervals

The accuracy of the estimated posterior distributions $\hat{F}_{Q_i^{(100)}}$ is presented on figure 4.19. As for the Ardèche case study, this figure indicates an important underestimation of the dispersion of $\hat{F}_{Q_i^{(100)}}$, and therefore of the width of the 90% credibility intervals computed with both regional approaches when regional variability in the growth curves and index flood relation exist (see Fig 4.19.a and Fig 4.19.b). Figure 4.20 suggests that the level of underestimation observed in **case 1** (proposed approach involving extremes), could be of about 70%, as the distribution of $\hat{F}^{disp}_{Q_i^{(100)}}\left(Q_i^{(100)}\right)$ becomes close to uniform for a correction factor of the credibility intervals of m= 1.7 (see Fig 4.20.a). For the reference Hosking & Wallis approach, with m= 1.7 the dispersion of $\hat{F}^{disp}_{Q_i^{(100)}}\left(Q_i^{(100)}\right)$ still appears to be underestimated (see Fig 4.20.b), suggesting that the level of underestimation exceeds 70% in this case. The same explanation as for the Ardèche case study can be put forward to explain this result : the underestimation observed is mainly associated with the presence of regional heterogeneities and with the local average estimation uncertainties (for the Hosking & Wallis approach) that are not considered in the inference procedure.

In comparison, the dispersion of the posterior distributions computed in the case of

local approaches (**cases 3** and **case 4**) appear to be close to perfect (see Fig 4.19.c, d), validating the Bayesian MCMC procedure for the computation of credibility limits when all possible sources of uncertainties are taken into account. In these cases, the 90% credibility intervals computed on $Q_i^{(100)}$ can be considered as reliable.

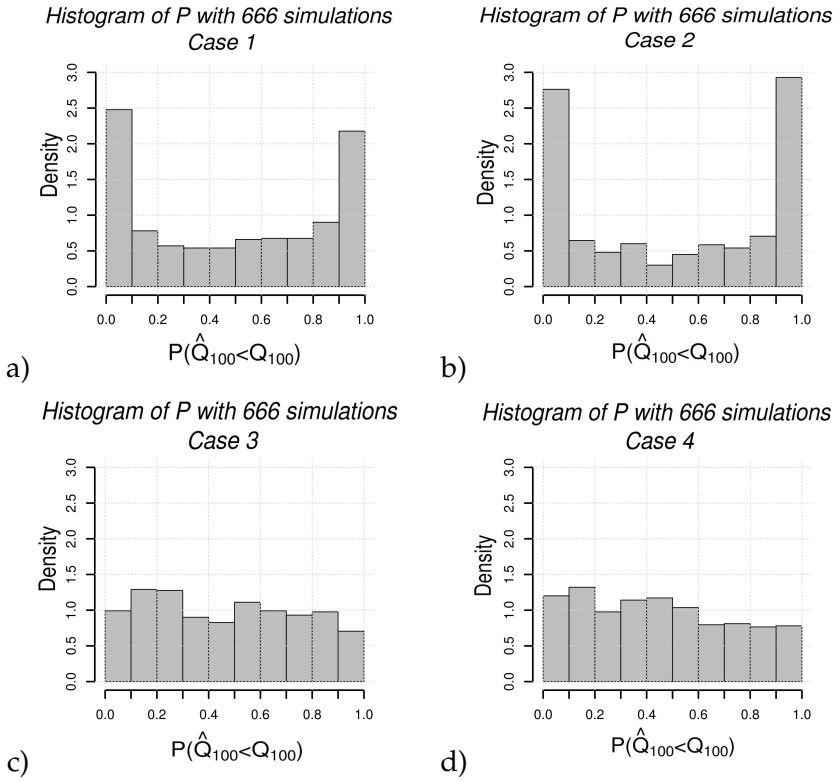


FIGURE 4.19 – *Distributions of* $\hat{F}_{Q_i^{(100)}}\left(Q_i^{(100)}\right)$ *computed at Trans-en-Provence gauging site (190 km²) based on 666 simulated samples : a)* ***case 1****, b)* ***case 2****, c)* ***case 3****, and d)* ***case 4****.*

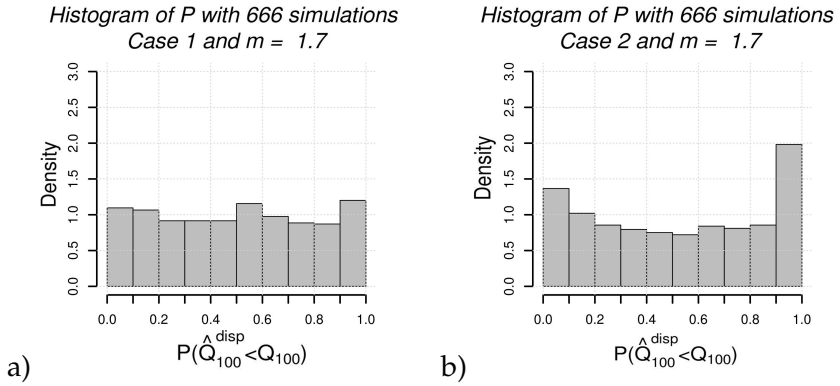


FIGURE 4.20 – *Distributions of* $\hat{F}^{disp}_{Q_i^{(100)}}\left(Q_i^{(100)}\right)$ *computed at Trans-en-Provence gauging site (190 km²), based on 666 simulated samples : a)* ***case 1*** *and m= 1.7, b)* ***case 2*** *and m= 1.7.*

4.4.2 Application of the methods to the Var datasets

Figures 4.21 and 4.22 present the shapes of the GEV distributions and associated 90% credibility intervals computed at Trans-en-Provence (190 km^2), based on the datasets available in the Var region. The inference approaches considered (**case 1** to **4**) are the same as for simulations. In **case 4** and **case 2**, the effect of the incorporation of the 2010 flood in the Trans-en-Provence gauged series is tested (see Fig 4.21.b and Fig 4.22.b). The estimated maximum likelihood 100-year peak discharge and the associated credibility bounds before and after the corrections accounting for regional variability are summarized in table 4.7.

TABLE 4.7 – *Estimations of the 100-year quantile at Trans-en-Provence gauging stations, including the Maximum Likelihood estimate ($\hat{Q}_{i,ML}^{(100)}$), the 90% credibility bounds [$\hat{Q}_{i,0.05}^{(100)}$; $\hat{Q}_{i,0.95}^{(100)}$], and associated width ΔCI = $\hat{Q}_{i,0.95}^{(100)}$ - $\hat{Q}_{i,0.05}^{(100)}$, and corrected ΔCI according to the theoretical simulation results.*

Surface (km^2)	N_{cont} (years)	N_{hist} (years)	Case	$\hat{Q}_{i,0.05}^{(100)}$ (m^3/s)	$\hat{Q}_{i,ML}^{(100)}$ (m^3/s)	$\hat{Q}_{i,0.95}^{(100)}$ (m^3/s)	$\Delta CI/\hat{Q}_{i,ML}^{(100)}$ (%)	$\Delta CI/\hat{Q}_{i,ML}^{(100)}$ (%)corr
190	34	0	**4a**	106.0	147.0	410.7	210	207.3
	35	0	**4b**	179.5	313.7	1085.6	290	288.8
	34	360	**3**	196.7	263.5	395.8	76	75.5
	249	0	**2a**	**99.5**	**112.2**	**136.7**	34	57
	250	0	**2b**	157.4	179.6	226.1	39	65
	188	660	**1**	174.1	212.2	287.0	53	90

As in the case of the Ardèche region, the estimated ranking of the various approaches according to the estimated credibility intervals, does not exactly correspond to the theoretical results due probably to both : the effect of sampling fluctuation and the distance between the statistical model on which the theoretical results are based and the effective statistical distributions of the data.

The adjusted distribution in the case of the regional approach based on the gauged data (see Fig 4.22.a) is the most striking result for the Var. It appears clearly inconsistent with the other calibrated statistical distributions and especially the ones based on the local gauged and historical information (**case 3** and Fig 4.21.c) and the regional information including ungauged extremes (**case 1**, Fig 4.22.c). This is particularly clear for high return period quantiles. This indicates that the regional gauged set, the ungauged extreme set and the GEV distribution are incompatible.

This can be verified with various statistical tests. The first is based on the likelihood value of the observed samples already used by (Payrastre et al. 2005). A statistical distribu-

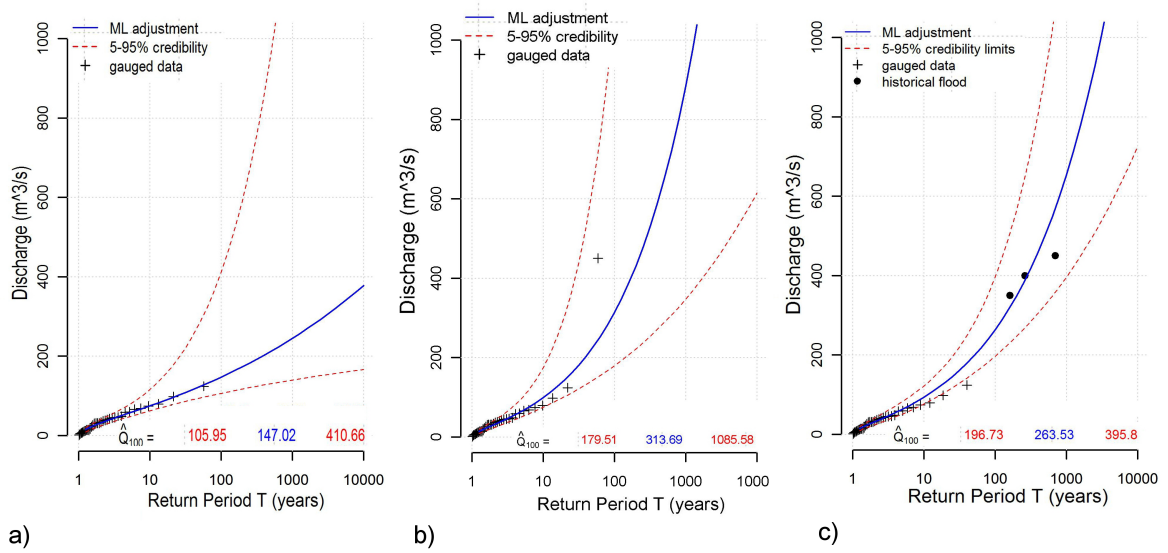


FIGURE 4.21 – *Fitted GEV distributions and associated credibility intervals at Trans-en-Provence gauging site (190 km^2) : a)* ***Case 4a*** *- local approach with a gauged series (34-year) ending in 2009 ; b)* ***Case 4b*** *with 2010 flood - local approach with a series (35-year) including the peak discharge (450 m^3/s) of the 2010 flood ; and c)* ***Case 3*** *- local approach combining the 34-year gauged series and the three historical floods.*

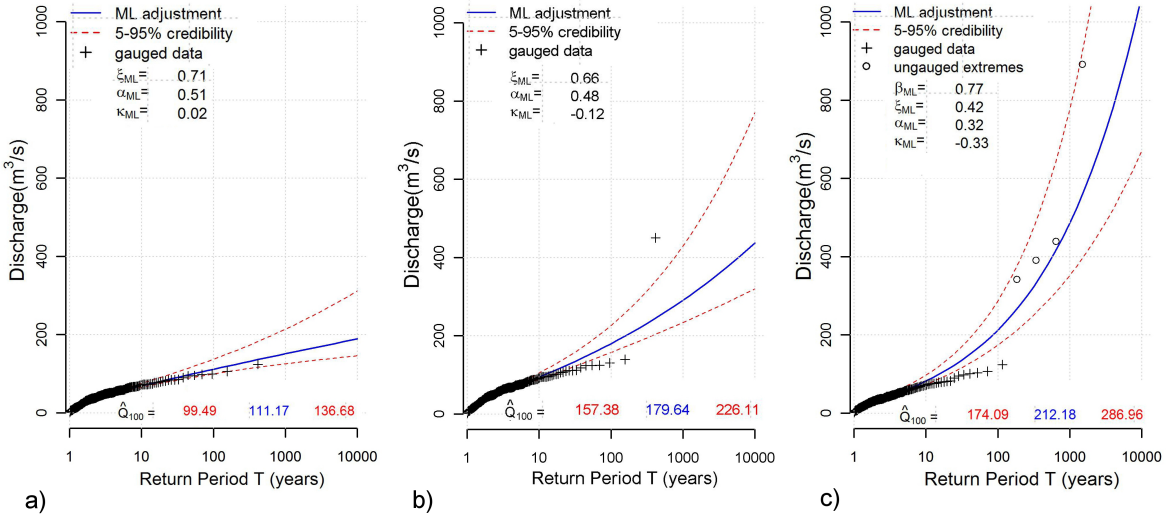


FIGURE 4.22 – *Fitted GEV distributions and associated credibility intervals at Trans-en-Provence gauging site (190 km^2) : a)* ***Case 2a*** *- the reference Hosking & Wallis regional analysis based on gauged series (9 sites) without the 2010 flood at Trans-en-Provence, b)* ***Case 2b*** *- the reference Hosking & Wallis regional analysis (9 sites) including the 2010 flood at Trans-en-Provence, and c)* ***Case 1*** *- proposed regional analysis including 7 gauged series and 4 extreme floods*

tion being selected, random samples with the same characteristics as the observed sample (number and type of observations) can be drawn and the probability density function of their likelihood can be empirically build. The position of the likelihood of the observed

sample given the selected distribution can then be compared to this density, to evaluate if the observed sample is likely to result from this selected statistical distribution.

Figure 4.23.a presents the density function of log-likelihood values built based on 1000 simulated samples (7 sites and 4 ungauged extremes) using the calibrated maximum likelihood distribution obtained in **case 2** (regional sample of gauged data, Fig 4.22.a). The log-likelihood value of the available regional set including the 4 ungauged extremes is about -6000. Without any doubt, the occurrence of the 4 ungauged extremes is incompatible with the calibrated distribution obtained in **case 2**. Conversely, Figure 4.23.b presents the density function of log-likelihood values built based on 1000 simulated samples (9 gauged sites) using the calibrated maximum likelihood distribution obtained in **case 1** (regional sample of gauged data and 4 ungauged extremes, Fig 4.22.c). In this case, the log-likelihood of the observed sample of data at 9 gauged sites is equal to -1298. Surprisingly, the gauged sample, even if it does not contain any peak discharge values corresponding to return periods exceeding 20 years according to the regional statistical distribution calibrated in **case 1**, does not appear as unlikely. This puzzling result reveals the limited power of tests based on the likelihood of the sample, already mentioned by (Payrastre et al. 2005).

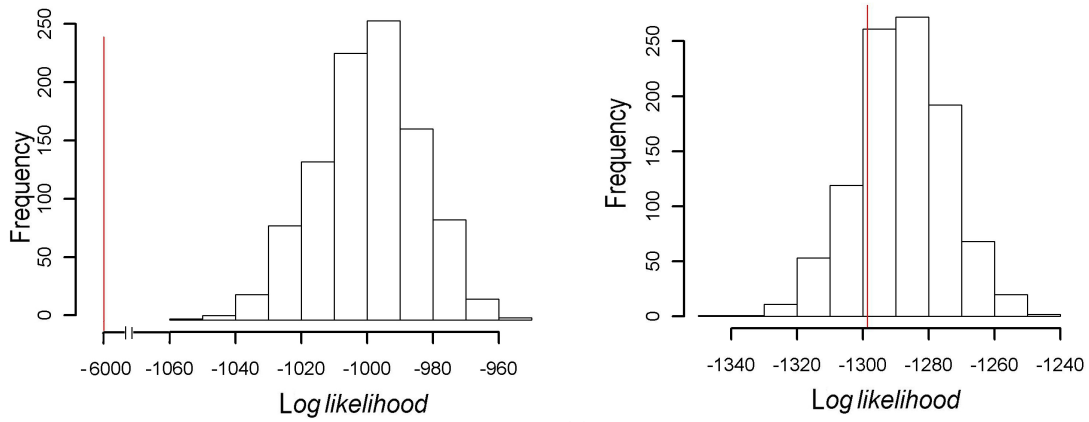


FIGURE 4.23 – *Histogram of likelihood values of 1000 simulation samples with proposed approach : a) 7 sites and 4 ungauged extremes, and b) 9 sites. A real likelihood value is red line.*

Simple probability computations offer another possible way for testing the compatibility between the samples and the adjusted distributions. According to the distribution calibrated in **case 1** (complete regional sample), the return period of the maximum rescaled discharge value of the gauged regional sample did not exceed 20 years. What is the probability of drawing 249 successive independent and identically distributed values that

do not exceed the 20-year quantile ? The answer is $(1 - 1/20)^{249} = 3.10^{-6}$. It is extremely unlikely ! Clearly, the regional gauge sample is not consistent with the finally calibrated GEV distribution in **case 1**.

Five main explanations can be put forward to explain the discrepancy between the regional gauged set, the ungauged extreme set and the GEV distribution :

– The regional gauged series may be censored : i.e. the largest values may not be reported in the records and the corresponding period appears as gaps with missing data. The Ardèche set has shown some examples of data censoring. The Var data set has been therefore checked and it appears that the gaps are very limited in the discharge series, and do apparently not correspond to heavy rainfall events (see Table 4.8), at least over the period 1959-1995 (see Appendix **A.5**).

TABLE 4.8 – *The largest recorded local daily rainfall accumulations in the Var over the period 1958-1995 (source : Meteo France, Inventaire des situations a precipitations diluviennes surle Languedoc-Roussillon la Provence-Alpes cote d'azur et la corse, periode 1958-1994)*

Year	**Date**	**Position**	**X**(mm)
1958	18-19 December	Callas	260
1961	11-12 November	Adrets de l'estered	232
1972	27-28 December	Collobrieres	193
1973	02 October	Saint Mandrier	209
1973	13 October	Saint Raphael	223
1976	18 July	Grimaud	224
1980	27 August	Grimaud and Le Levant	200 and 236

– The gauged series may not be independent especially as far as large floods are concerned. The example of the 2010 shows, that large rainfall events inducing extreme floods have a certain spatial extent and generally affect more than one watershed in the same time. The equivalent number of independent station-years may have been overestimated for the regional gauged series and hence the credibility intervals largely underestimated. But nevertheless, the total extent of the considered region is large enough to limit the possible influence of dependencies on the conclusions : the 2010 extreme flood did only significantly affect the downstream part of the Argens watershed and only one of the considered gauged watersheds.
– The region may be heterogeneous as far as extreme floods are considered even if this is not revealed by the test of Hosking & Wallis. It is hard to conclude at this

stage. Only an inventory of extreme floods in the upper part of the Argens watershed (Western part of the region) could help verify this hypothesis.

– The process of generation of extreme floods may be non-stationary with a succession of relatively calm periods and meteorologically and hydrologically more active periods, for climatic reasons. No rainfall event comparable as the 2010 event more than 450 (*mm*) maximum local daily rainfall accumulation - has for instance been recorded over the period 1958-2010 over the Var region (see Appendix **A.4**).

– As already mentioned on the Ardèche case study, the statistical distribution of the maximum annual peak discharges, and may be of the maximum local point rainfall accumulations - may differ significantly from the GEV distribution with fast evolution of the quantile values over short ranges of return periods. The shape of the distribution could be closer to the shapes suggested by the plotting positions on figures 4.21.c and 4.22.c. This could also explain the observed apparent higher added value of the proposed regional flood frequency method if compared to the theoretical simulation results in the two case studies. To illustrate this idea, if we consider that the actual return period of the largest rescaled discharge of the regional gauged set is much closer to the length of this set, for instance 150 to 200 years, than the observed regional gauged set becomes credible : the probability of drawing 249 successive independent and identically distributed values that do not exceed the 150-year quantile is around 20%. While, if we consider that the return period of the ungauged extreme discharges in Trans-en-Provence is at least 250 years, the probability of having exceeded 3 times or more the 250 year quantile within 400 years at Trans-en-Provence is also close to 20%. Again, one isolated example does not demonstrate anything but the repetition of similar examples could help to give a new insight into the possible shape of the statistical distributions of peak discharges. The results show the importance of taking ungauged extremes into account in statistical inference in this perspective.

4.5 Conclusions

This chapter aimed at illustrating the possibilities and limits of the proposed approach on the basis of two real-world applications. The statistical analyses of the available data sets has been combined with some theoretical simulations, conducted to evaluate the possible effects of intra-regional variability on the results of the two tested regional approaches. Overall, and despite some limits, especially concerning the computed credibility limits that may generally be too optimistic, these two application case studies confirm the relevance of the proposed method.

In the two areas, geographically consistent regions could be delineated based on both constraints : on the growth-curves and on the index flood relationship. Moreover, the hypotheses on which the method is based, in particular the simple index flood relationship, did not lead to results inconsistent with local flood frequency analyses including or not historical extremes. The Saint Martin case on the Ardèche region, where historical data could be unearthed after the analyses and used for validation purposes, is particularly striking in that perspective (Naulet 2002). But above all, the proposed approach seems to provide more accurate estimates than the reference regional flood frequency method based on gauged data only, this theoretically but also in practice. The explanation for this result is simple : the added value of the regional methods depends on the size (measured in record station-years) of the regional samples. Moreover, if gauged data exist, the additional information (regional or historical) may be censured with limited impact on the inference results (Payrastre et al. 2011) : what really matters is the information content on the extremes. The record length represented by ungauged extremes in a given region rapidly exceeds by far, if some inventories are conducted, the total recording length at the gauged sites as illustrated by the two case studies (see tables in this chapter). In the two cases, according to the size of the regional sets, the added value of the reference regional flood frequency approach seem to be very limited if compared to the standard local analyses, if the regional variability is considered. This is to our knowledge a relatively surprising and new result.

Some may argue that the regional data set could have been increased in the Hosking & Wallis approach looking at gauged sites in wider areas around the considered region of interest. It is true, but is it reasonable to consider data from very various climatologic and hydrologic settings as statistically homogeneous, even if their heterogeneity is not demonstrated by a test ? Regional flood frequency analyses may also be misleading as illustrated by the case study of the Var. The proposed methods should preferably not be based on too many hypotheses that can not be verified.

Finally, these case studies have illustrated that ungauged extremes can be considered as a valuable and rich additional information for flood frequency studies and that they can be included, provided some hypotheses, in statistical inference procedures. In the given examples, their inclusion in the analyses even led to question the adequacy of the GEV distribution which shows how informative this data can be. It is nevertheless important to note, that regional flood frequency analyses are not always as straightforward as shown here. Some tests were also conducted on datasets from the Adige mountainous region in North-East Italy. But it has been there impossible to define homogeneous sub-regions, neither for the growth curves nor for the index flood relationship, due to the combination of a huge hydrological variability between karstic and non-karstic areas and of a pluviographic gradient from the sea to the inner Alps. Regional flood frequency analyses may no always be possible. To be really complete, uncertainties in estimated extreme peak discharges (ungauged or historic extremes but also extremes on gauged sites exceeding by far the range of values for which the stream gauge rating curve is accurate) should also be taken into account for the evaluation of the various approaches. According to recent studies (Payrastre et al. 2011), these uncertainties will nevertheless probably not modify the conclusions of this work unless the estimates of extreme values are systematically biased.

Conclusions and perspectives

5

The aim of this thesis was to develop and test thoroughly a new method of regional flood frequency analysis aiming at incorporating information on extreme floods documented at ungauged sites. The performances of this method have been evaluated in comparison with a reference regional flood frequency approach initially proposed by Hosking and Wallis (1997). In both cases, a Bayesian MCMC approach has been used to estimate the model parameters given the observations, and to compute associated credibility intervals.

Although the new proposed approach enables to account for a large amount of additional information, it necessitates an important assumption if compared to the reference regionalisation approach of Hosking & Wallis : an index flood relationship is supposed to hold in the considered region. This homogeneity requirement adds to the expected homogeneity of the local growth curves, and therefore many limit the possible extent of the region considered. It also represents an additional parameter to be estimated that should be included in the computation of credibility intervals.

In the first part of this work, a specific Monte Carlo simulation framework has been developed in order to compare both methods and to test the relevance of the Bayesian MCMC estimation method applied in each case. Previously, the proposed approach was adapted (in particular the formulation of the likelihoods) in order to enable a common estimation of both, the growth curve (here a GEV distribution) and the index flood relationship. Finally, the MCMC algorithms applied aim to estimate 4 parameters in the case of the proposed approach, and 3 parameters in the case of the reference method. The two approaches were applied to a high number of synthetic datasets, and the estimations results (based on posterior distributions and maximum likelihood estimation of parameters) were compared

to the real quantile values at each site of the considered regions.

The simulations conducted were first focused on the comparison of both approaches when applied to identical datasets, limited to gauged series. The results illustrate several limitations of both approaches. The reference approach, indeed, appear to be sensitive to the uncertainty of estimation of local means : these results in an increase of estimation uncertainty and also in an underestimation of the width of credibility intervals. This effect disappears by using the index flood relationship (proposed approach), which clearly limits the final uncertainty for local series of limited length (less than 40 years). The introduction of the index flood relationship with an additional parameter to be calibrated, finally appears to be a stabilization factor rather than a factor of complexity and a source of uncertainty. However, this relationship results in some bias and increased uncertainty of estimation for low catchment surfaces.

These first simulations also illustrated some limitations of both approaches associated with the homogeneity assumptions made in each case. The effects of two types of heterogeneities were examined : heterogeneity of the index flood values around the theoretical relationship, and heterogeneity of the shapes of the local growth curves. Such heterogeneities generally result in larger uncertainties of estimation, sometimes associated with a bias, and in an underestimation of the width of credibility intervals. Logically, heterogeneities in the index flood relationship only affect the proposed approach : this approach fortunately appeared to resist to limited heterogeneity levels. The effects of heterogeneities in the local growth curves appeared much more problematic : they highly affect both tested approaches, even for limited heterogeneity levels.

In a second part of this work, the two regional approaches were applied to two case studies for which information on ungauged extreme floods is available : The Ardèche region and the Var region in the south east of France. Some simulations were again conducted here in order to compare the quality of estimation to be expected in different situations : local approach, local approach with historical information, and the two regional approaches involving or not information on ungauged extremes. These simulations account for the supposed heterogeneities on both the local growth curves and the index flood relationship in the considered regions. The simulation results first illustrate that the incorporation of information on extremes with the proposed regional approach may be of

great value and may clearly outperform the reference regional approach based on gauged series. This reference approach is affected by the effects of heterogeneity and finally leads to a quality of estimation that is not far from a local approach. This surprising result questions the real added value of conventional regional flood frequency approaches. The simulation results also show that the hierarchy between the different approaches highly depends on the particularities of each case study : catchment surface considered, length of local and of regional gauged series, quantity of local historical information and of regional information on extremes. Finally, these results also confirm that the credibility intervals computed with the Bayesian MCMC procedure should not be directly considered as an indication of the accuracy of quantile estimation, since the width of these intervals may be highly underestimated in presence of heterogeneity.

The application of the different local and regional approaches to the real datasets finally illustrates the very important differences that may be observed, for a specific case study, if compared to theoretical simulation results. Both applications show that despite the computation of credibility intervals that can be evaluated and corrected based on simulations, some of the inference results may still be particularly misleading : this is the case for instance for the local approach at Saint Martin in the Ardèche region or for the reference regional approach in the Var region. These results well illustrate the great potential effects of sampling variability, and also the limits associated to the assumptions made in the two regional approaches, that are not accounted for in theoretical simulations : independence of observations and suitability of the proposed theoretical statistical distribution.

In this context, the case studies presented also illustrate some advantages of approaches involving information on extreme floods : confirmation of the occurrence of intense floods that may appear as outliers if compared to local series (Ardèche case study), identification of anomalies in the considered series (for instance presence of dependences) or possible limits of the inference procedure (relevance of the shape of the theoretical growth curve).

Some additional results have been obtained to answer some questions raised by the reviewers of this PhD. These results, presented during the defense of the thesis, illustrate in particular the possible influence of some of the hypotheses on which the computations presented in this manuscript are based :

- determination of the threshold values for the various historical periods in the case of the Nartuby at Trans-en-Provence (appendix 7) ;
- integration of the ancient historical information on floods included in the PhD of Naulet (2002) in the statistical analysis of the discharges of the Ardèche at Saint-Martin (appendix 8) ;
- elimination of the obviously dependent data in the Ardèche regional sample : extreme floods affecting the same river network at the same date (appendix 9) ;
- percentage of error for systematic and extreme discharges considered for the computation of the likelihood, modified formulation (appendix 10) ;
- determination of the duration of the period over which the historic or ungauged extremes are considered to be the largest observed value (appendix 11).

These results are a useful complement to the results presented in this text and have therefore been added in the appendices (7-11) of the final version of this PhD according to the suggestion of the jury.

Two perspectives can be imagined at the end of this work. First, the question of the uncertainties affecting the discharge estimates of ungauged extreme floods was not considered here, and should be examined in order to definitely confirm the usefulness of such data for statistical studies. This could be achieved in a similar manner as in previous works conducted on the usefulness of historical records for flood frequency analyses (Payrastre et al. 2011). These results indicated that the accuracy level of the estimated historical extreme discharges and the number of extremes for which an estimated discharge is available, do not affect significantly the value of historical data when the historical period of record is fixed. It remains to be confirmed that the same conclusion can be drawn when ungauged extremes are included in regional flood frequency analyses. It seems to be the case according to the results presented by Gaume (2010).

Lastly, information on ungauged extremes floods seems, based on the presented case studies, susceptible to put into question the theoretical shape of the statistical distributions commonly used for inference purposes in hydrology, particularly in the case of watersheds prone to flash floods. Discharge distributions may present important evolutions of quantile values concentrated on limited range of return periods, due to rainfall distributions and threshold-based reactions of watersheds (Gaume 2006, Rogger et al. 2011). The results presented here suggest that regional analyses including information on ungauged extremes

may be an interesting way to confirm (or infirm) this theory. This might be achieved by analysing a much larger number of similar case studies, which is possible since the methods are now available.

Appendix

A

A.1 MCMC Metropolis algorithm

The Metropolis Hastings algorithms are efficient methods to draw samples from a distribution with knowing a probability density function P(θ). The algorithm generates a Markov chain in which each state θt+1 depends only on the previous state xt. The algorithm uses a Gaussian density N(θt, Sθ) to generate candidate value $\theta^{'}$ at the step $t+1$, which depends on the current state θt, to generate a new proposed sample $\theta^{'}$. This candidate is accepted as the next value $\theta t + 1 = \theta^{'}$ if a drawn from U(0,1) satisfies

$$a < \frac{P\left(\theta^{'}\right)}{P\left(\theta t\right)} \tag{A.1}$$

If the candidate is not accepted, then the current value of x is retained (θt+1=θt). The Markov chain is started from a random initial value θ0 and the algorithm is run for many iterations until this initial state is forgotten and stabilization of it outcomes is reached. The samples corresponding to the courage of the algorithm, which are discarded, are known as burn in. The remaining set of accepted values of θ represent a sample with probability density function P(θ). As a Gaussian proposal density (or a lognormal one for definite positive parameters) is used, the variance parameter Sθ^2 has to be tuned during the burn in period. This is done by calculating the acceptance rate, which is the fraction of proposed samples that is accepted in a window of the last N samples. The desired acceptance rate depends on the target distribution, however it has been shown theoretically that the ideal acceptance rate for a one dimensional Gaussian distribution is approx 50%, decreasing to approx 23% for an N dimensional Gaussian target distribution. If Sθ^2 is too small the chain will courage slowly (i.e., the acceptance rate will be too high, so the sampling will move around the space slowly and converge slowly to P(θ)). If Sθ^2 is too large the acceptance rate will be very low because the proposals are likely to land in regions of much lower probability density. The desired acceptance rate is fixed here to 34% in our case study.

A.2 The diagnostic of Gelman and Rubin

The diagnostic of Gelman and Rubin (1992) assumes that M Markov Chains have been simulated independently. If convergence has been obtained, inference is made by computing the sample mean and variance of each chain.

In the case of convergence the obtained M inference results should be similar. The idea is to estimate σ^2, the variance of θ and compute the ratio of this estimator, denoted by $\hat{V}$ and σ^2. The estimator $\hat{V}$ is constructed by mixing $M * N$ draws from all chains and considers both the variance between and within the chains. With $\theta_i^t = \theta\left(x_i^t\right)$ is the function of interest evaluated at the t^{th} observation from chain i.

The variance $\frac{B}{N}$ between the M chain means can be computed as :

$$\frac{B}{N} = \frac{1}{M-1}\sum_{i=1}^{M}\left(\bar{\theta}_i - \bar{\theta}\right)^2 \tag{A.2}$$

Where

$$\bar{\theta}_i = \frac{1}{N}\sum_{t=N_0+1}^{2N}\theta_i^t \tag{A.3}$$

$$\bar{\theta} = \frac{1}{N}\sum_{i=1}^{M}\bar{\theta}_i \tag{A.4}$$

The mean of the M within-sequence variances s_i^2 is given by :

$$W = \frac{1}{M}\sum_{i=1}^{M}s_i^2 \tag{A.5}$$

Where

$$s_i^2 = \frac{1}{N-N_0-1}\sum_{t=N_0+1}^{2N}\left(\theta_i^t - \bar{\theta}\right)^2 \tag{A.6}$$

The estimator $\hat{V}$ is then defined by :

$$\hat{V} = \frac{N-1}{N}W + \left(1+\frac{1}{M}\right)\frac{B}{N} \tag{A.7}$$

The ratio is computed :

$$\hat{R}_c = \left(\frac{d+3}{d+1}\right)\frac{\hat{V}}{W} \tag{A.8}$$

Where d is the estimated degrees of freedom of a Student-t distribution and can be estimated by the method of moment. In the case, if $\hat{R}_c$ is close to 1, we can conclude that each of the M chains has converged.

A.3 Generalized extreme-value distribution (GEV)

Definition :

The generalized extreme value distribution (GEV) is a generalized three-parameter extreme value distribution : ξ (location), α (scale), κ (shape).

Range of x : -$\infty < x \leq \xi + \alpha/\kappa$ if $\kappa > 0$; -$\infty < x < \infty$ if κ=0 ; $\xi + \alpha/\kappa \leq x < \infty$ if $\kappa < 0$.

Probability density function :

$$f(x) = \alpha^{-1} e^{-(1-\kappa)y - e^{-y}} \tag{A.9}$$

Where,

$$y = -\kappa^{-1} log\left(\frac{1-\kappa(x-\xi)}{\alpha}\right) \quad with\ \kappa \neq 0 \tag{A.10}$$

The cumulative probability density function :

$$F(x) = exp\left[-\left(1 - \frac{\kappa(x-\xi)}{\alpha}\right)^{1/\kappa}\right]_{\alpha>0} \quad with\ \kappa \neq 0 \tag{A.11}$$

Quantile function :

$$x(F) = \xi + \frac{\alpha\left(1 - (-logF)^{\kappa}\right)}{\kappa} \quad with\ \kappa \neq 0 \tag{A.12}$$

Special cases : κ=0 is the Gumbel distribution ; κ=1 is a reverse exponential distribution.

A.4 The records of the 9 gauging stations in the Var region

Table A.1 – *Station : Trans-en-Provence on the Nartruby River (190 km^2) ; Code hydro : Y5235010 (X :935648, Y :1842521) ; β=0.77 ;$\bar{Q}$=37.6*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$
1	1970	42	0.74	1.12	18	1987	98	1.72	2.61
2	1971	33	0.58	0.88	19	1988	17	0.30	0.45
3	1972	31	0.55	0.82	20	1989	1	0.02	0.03
4	1973	124	2.18	3.30	21	1990	66	1.16	1.76
5	1974	26	0.46	0.69	22	1991	14	0.25	0.37
6	1975	4	0.07	0.11	23	1992	45	0.79	1.20
7	1976	31	0.55	0.82	24	1993	68	1.20	1.81
8	1977	40	0.70	1.06	25	1994	58	1.02	1.54
9	1978	14	0.25	0.37	26	1995	74	1.30	1.97
10	1979	38	0.67	1.01	27	1996	36	0.63	0.96
11	1980	12	0.21	0.32	28	1997	59	1.04	1.57
12	1981	6	0.11	0.16	29	1998	11	0.19	0.29
13	1982	9	0.16	0.24	30	1999	31	0.55	0.82
14	1983	46	0.81	1.22	31	2000	79	1.39	2.10
15	1984	48	0.84	1.28	32	2001	22	0.39	0.59
16	1985	14	0.25	0.37	33	2002	30	0.53	0.80
17	1986	44	0.77	1.17	34	2003	7	0.12	0.19

TABLE A.2 – *Station : Mons on the Siagnole River (87 km^2) ; Code hydro : Y5515410 (X :953534, Y :1862675) ; β=0.77 ; $\bar{Q}$=15.4*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^{β}	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^{β}	$Q/\bar{Q}$
1	1980	9	0.29	0.58	16	1995	29	0.93	1.88
2	1981	12	0.39	0.78	17	1996	15	0.48	0.97
3	1982	11	0.35	0.71	18	1997	23	0.74	1.49
4	1983	15	0.48	0.97	19	1998	10	0.32	0.65
5	1984	15	0.48	0.97	20	1999	15	0.48	0.97
6	1985	12	0.39	0.78	21	2000	23	0.74	1.49
7	1986	11	0.35	0.71	22	2001	20	0.64	1.30
8	1987	15	0.48	0.97	23	2002	18	0.58	1.17
9	1988	8	0.26	0.52	24	2003	13	0.42	0.84
10	1989	7	0.22	0.45	25	2004	10	0.32	0.65
11	1990	9	0.29	0.58	26	2005	9	0.29	0.58
22	1991	9	0.29	0.58	27	2006	20	0.64	1.30
13	1992	24	0.77	1.56	28	2007	19	0.61	1.23
14	1993	20	0.64	1.30	29	2008	14	0.45	0.91
15	1994	32	1.04	2.07	30	2009	16	0.51	1.04

TABLE A.3 – *Station : Roquebrune on the Argens River (2530 km^2) ; Code hydro : Y5312010 (X :948137, Y :1838024) ; β=0.77 ; $\bar{Q}$= 330.8*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^{β}	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^{β}	$Q/\bar{Q}$
1	1972	450	1.08	1.36	17	1988	54	0.13	0.16
2	1973	549	1.32	1.66	18	1989	97	0.23	0.29
3	1974	337	0.81	1.02	19	1990	55	0.13	0.17
4	1975	305	0.73	0.92	20	1991	51	0.12	0.15
5	1976	520	1.25	1.57	21	1992	506	1.21	1.53
6	1977	733	1.76	2.22	22	1993	747	1.79	2.26
7	1978	138	0.33	0.42	23	1994	423	1.01	1.28
8	1979	397	0.95	1.20	24	1995	651	1.56	1.97
9	1980	58	0.14	0.18	25	1996	444	1.06	1.34
10	1981	123	0.29	0.37	26	1997	377	0.90	1.14
11	1982	153	0.37	0.46	27	1998	219	0.52	0.66
12	1983	132	0.32	0.40	28	1999	323	0.77	0.98
13	1984	397	0.91	1.15	29	2000	557	1.33	1.68
14	1985	307	0.74	0.93	30	2001	142	0.34	0.43
15	1986	154	0.37	0.47	31	2002	361	0.87	1.09
16	1987	570	1.37	1.72	32	2003	272	0.65	0.82

TABLE A.4 – *Station : Arcs on the Argens River (1730 km²); Code hydro : Y5202010 (X :935105, Y :1835965); β=0.77 ; $\bar{Q}$=169.4*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$
1	1969	182	0.58	1.07	19	1987	120	0.39	0.71
2	1970	175	0.56	1.03	20	1988	30	0.10	0.18
3	1971	330	1.06	1.95	21	1989	22	0.07	0.13
4	1972	360	1.17	2.15	22	1990	40	0.13	0.24
5	1973	405	1.30	2.39	23	1991	28	0.09	0.17
6	1974	144	0.46	0.85	24	1992	190	0.61	1.12
7	1975	155	0.50	0.91	25	1993	346	1.11	2.04
8	1976	294	0.94	1.74	26	1994	215	0.69	1.27
9	1977	447	1.44	2.64	27	1995	251	0.81	1.48
10	1978	62	0.20	0.37	28	1996	222	0.71	1.31
11	1979	249	0.80	1.47	29	1997	67	0.22	0.40
12	1980	38	0.12	0.22	30	1998	162	0.52	0.96
13	1981	85	0.27	0.50	31	1999	156	0.50	0.92
14	1982	56	0.18	0.33	32	2000	197	0.63	1.16
15	1983	100	0.32	0.59	33	2001	68	0.22	0.40
16	1984	168	0.54	0.99	34	2002	152	0.49	0.90
17	1985	163	0.52	0.96	35	2003	184	0.59	1.09
18	1986	62	0.20	0.37					

TABLE A.5 – *Station : Vins on the Caramy River (215 km²) ; Code hydro : Y5105010 (X :910981, Y :1834130) ; β=0.77 ; $\bar{Q}$= 46.3*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$
1	1972	104	1.66	2.24	17	1988	28	0.45	0.60
2	1973	83	1.33	1.79	18	1989	15	0.24	0.32
3	1977	39	0.62	0.84	19	1990	17	0.27	0.37
4	1975	63	1.01	1.36	20	1991	9	0.14	0.19
5	1976	62	0.99	1.34	21	1992	40	0.64	0.86
6	1977	117	1.87	2.52	22	1993	76	1.22	1.64
7	1978	19	0.30	0.41	23	1994	54	0.86	1.17
8	1979	58	0.93	1.25	24	1995	58	0.93	1.25
9	1980	13	0.21	0.28	25	1996	56	0.90	1.21
10	1981	7	0.11	0.15	26	1997	24	0.38	0.52
11	1982	12	0.19	0.26	27	1998	88	1.41	1.90
12	1983	28	0.45	0.60	28	1999	78	1.25	1.68
13	1984	38	0.61	0.82	29	2000	89	1.42	1.92
14	1985	38	0.61	0.82	30	2001	14	0.22	0.30
15	1986	23	0.37	0.50	31	2002	34	0.54	0.73
16	1987	47	0.75	1.01	32	2003	52	0.83	1.12

TABLE A.6 – *Station : Chateauvert on the Argens River (485 km^2) ; Code hydro : Y5032010 (X :898748, Y :1840110) ; β=0.77 ; $\bar{Q}$= 49.2*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$
1	1972	123	1.05	2.50	17	1988	19	0.16	0.39
2	1973	86	0.74	1.75	18	1989	8	0.07	0.16
3	1974	44	0.38	0.89	19	1990	16	0.14	033
4	1975	49	0.42	1.00	20	1991	18	0.15	0.37
5	1976	74	0.63	1.50	21	1992	76	0.65	1.54
6	1977	139	1.19	2.82	22	1993	101	0.86	2.05
7	1978	16	0.14	0.33	23	1994	67	0.57	1.36
8	1979	65	0.56	1.32	24	1995	87	0.74	1.77
9	1980	16	0.14	0.33	25	1996	76	0.65	1.54
10	1981	30	0.26	0.61	26	1997	13	0.11	0.26
11	1982	12	0.10	0.24	27	1998	42	0.36	0.85
12	1983	26	0.22	0.53	28	1999	35	0.30	0.71
13	1984	44	0.38	0.89	29	2000	49	0.42	1.00
14	1985	75	0.64	1.52	30	2001	68	0.58	1.38
15	1986	12	0.10	0.24	31	2002	31	0.27	0.63
16	1987	22	0.19	0.45	32	2003	36	0.31	0.73

TABLE A.7 – *Station : Bras on the Cauron River (154 km^2) ; Code hydro : Y5005210 (X :892835, Y :1839193) ; β=0.77 ; $\bar{Q}$= 12.5*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^{β}	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^{β}	$Q/\bar{Q}$
1	1975	12	0.25	0.96	16	1990	4	0.08	0.32
2	1976	15	0.31	1.20	17	1991	2	0.04	0.16
3	1977	28	0.58	2.24	18	1992	6	0.12	0.48
4	1978	4	0.08	0.32	19	1993	5	0.10	0.40
5	1979	14	0.29	1.12	20	1994	24	0.50	1.92
6	1980	2	0.04	0.16	21	1995	24	0.50	1.92
7	1981	1	0.02	0.08	22	1996	22	0.46	1.76
8	1982	4	0.08	0.32	23	1997	11	0.23	0.88
9	1983	10	0.21	0.80	24	1998	2	0.04	0.16
10	1984	18	0.37	1.44	25	1999	23	0.48	1.84
11	1985	17	0.35	1.36	26	2000	27	0.56	2.16
12	1986	12	0.25	0.96	27	2001	4	0.08	0.32
13	1987	17	0.35	1.36	28	2002	17	0.35	1.36
14	1988	12	0.25	0.96	29	2003	24	0.50	1.92
15	1989	1	0.02	0.08					

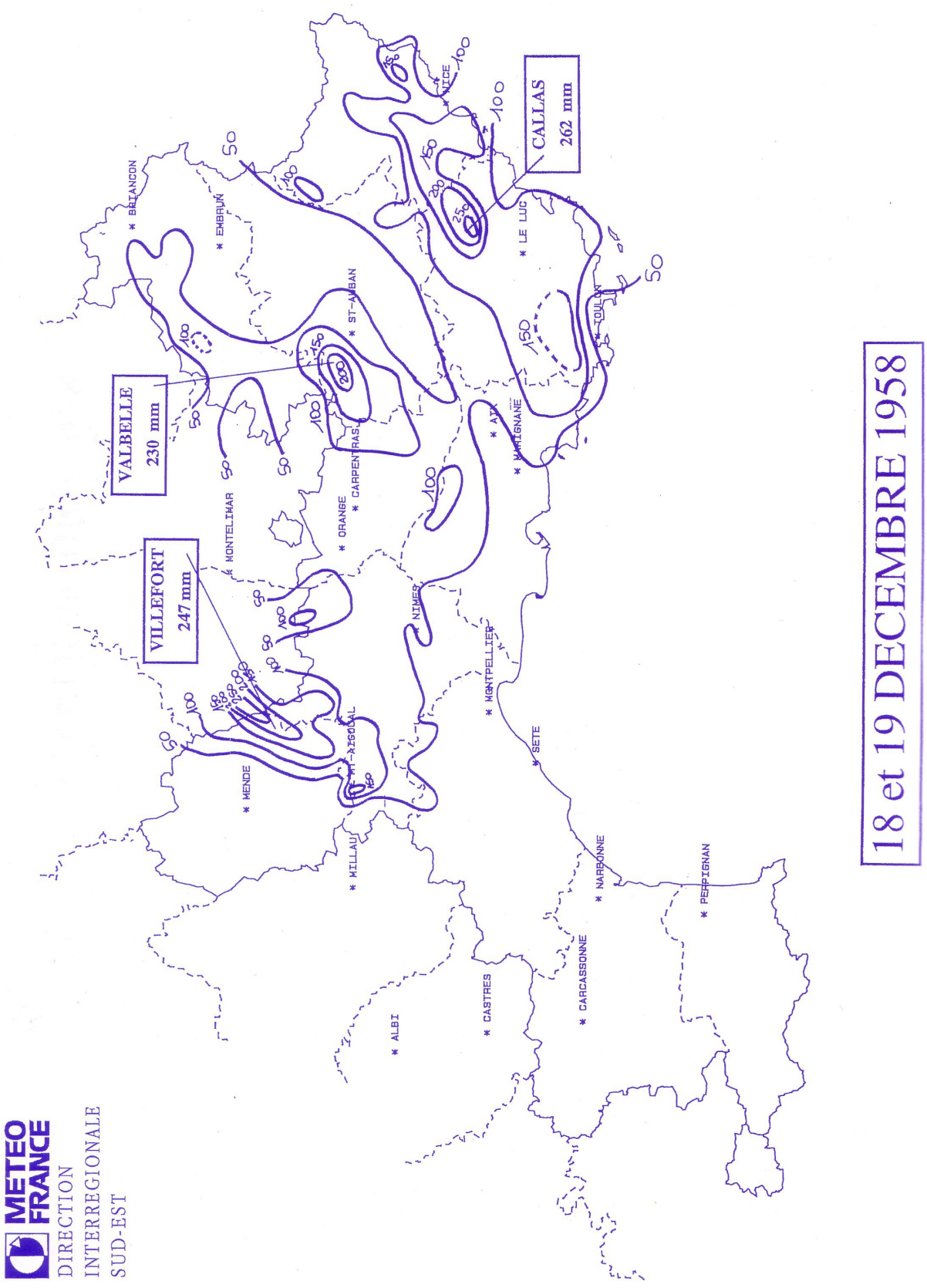
TABLE A.8 – *Station : Cabasse on the Issole River (223 km²) ; Code hydro : Y5106610 (X :914946, Y :1834591) ; β=0.77 ; $\bar{Q}$= 41.2*

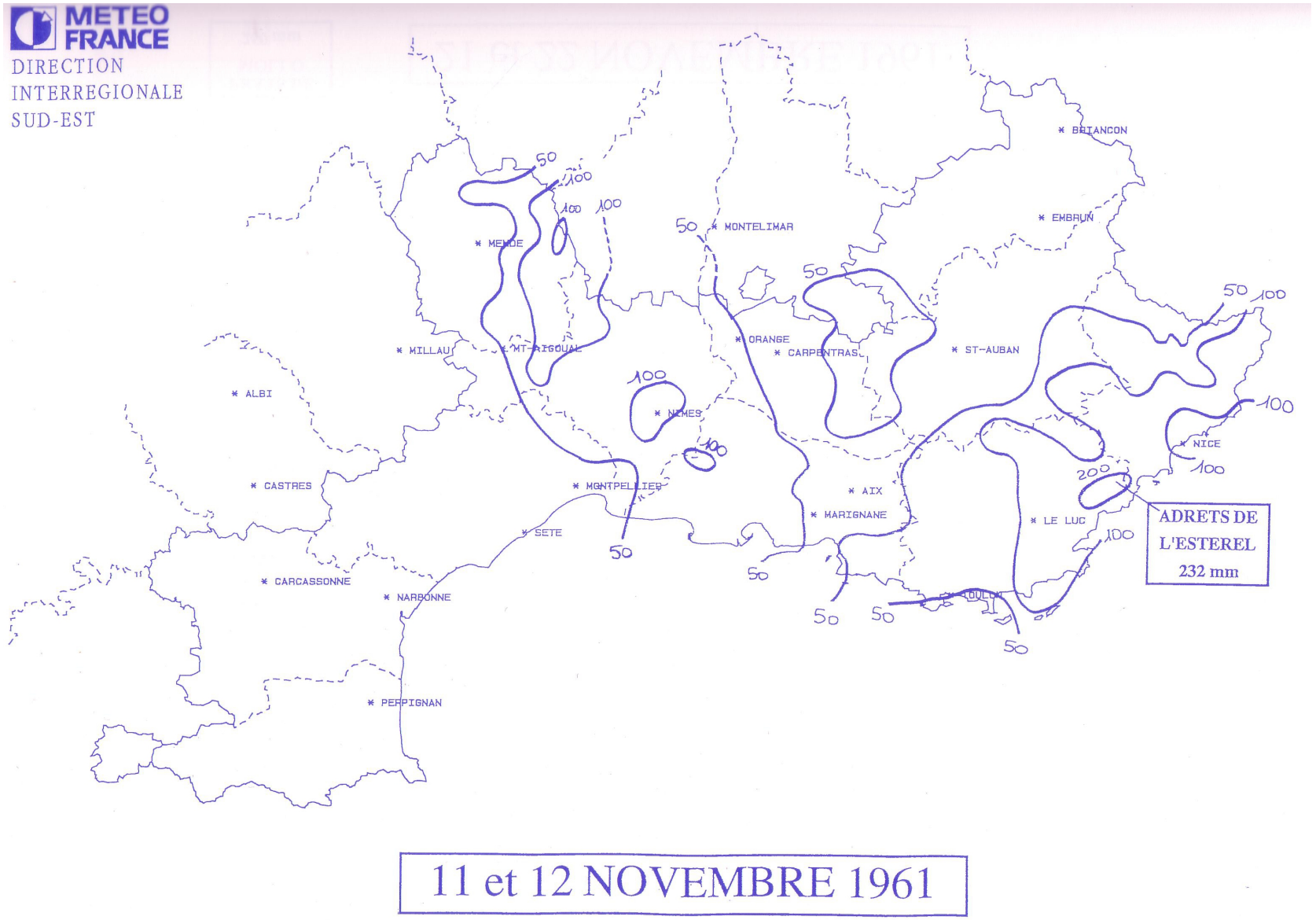
N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$
1	1975	9	0.14	0.22
2	1976	29	0.45	0.70
3	1977	58	0.90	1.41
4	1978	41	0.64	1.00
5	1979	48	0.75	1.17
6	1980	53	0.82	1.29
7	1981	21	0.33	0.51
8	1982	73	1.14	1.77
9	1983	68	1.06	1.65
10	1984	74	1.15	1.80
11	1985	8	0.12	0.19
12	1986	28	0.44	0.68
13	1987	25	0.39	0.61

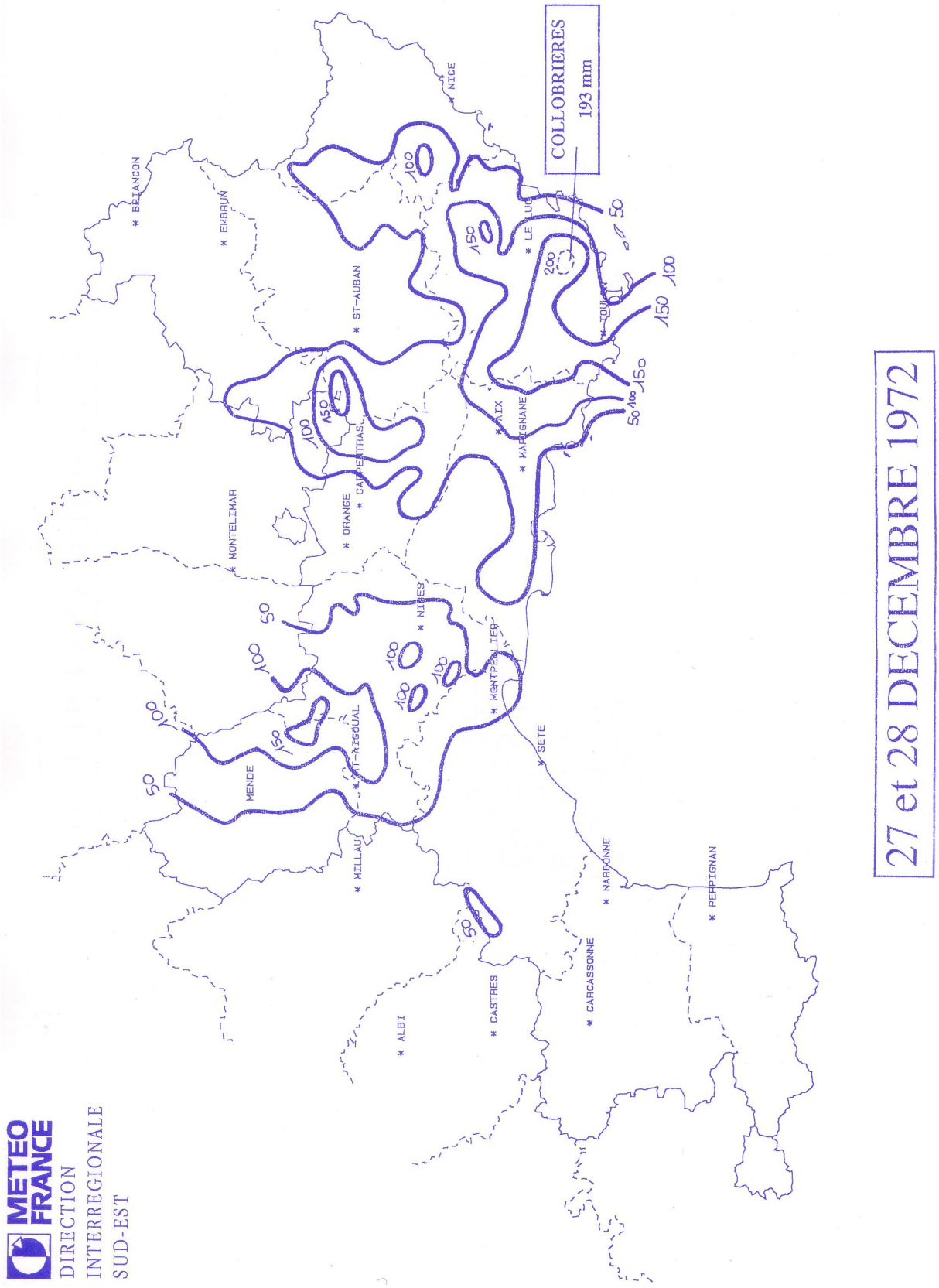
TABLE A.9 – *Station : Carces on the Argens River (1181 km²) ; Code hydro : Y5112010 (X :912616, Y :1838315) ; β=0.77 ; $\bar{Q}$= 104.8*

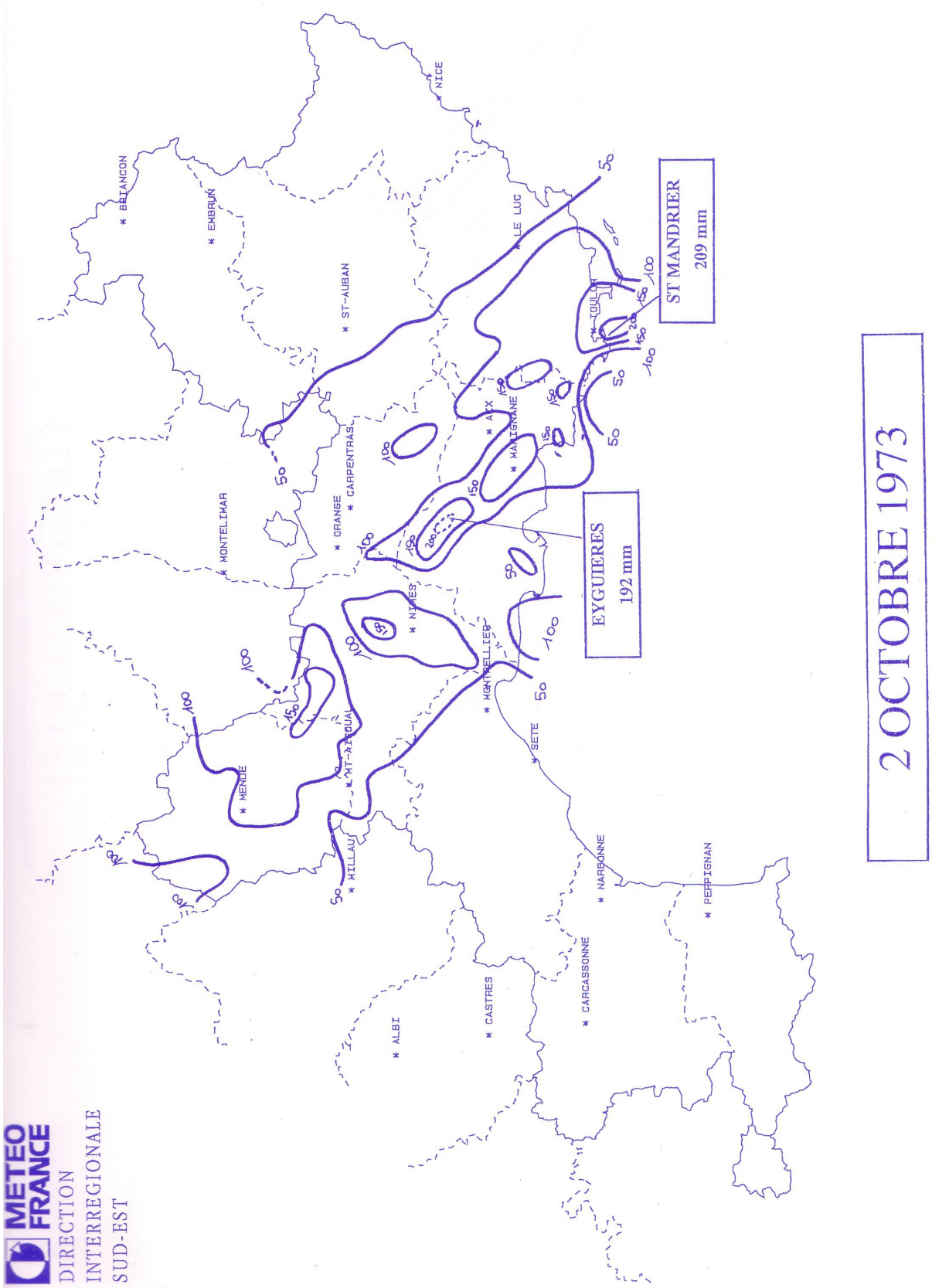
N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$
1	1989	12	0.19	0.29
2	1990	29	0.45	0.70
3	1991	22	0.34	0.53
4	1992	99	1.54	2.41
5	1993	160	2.49	3.89
6	1994	143	2.22	3.47
7	1995	147	2.29	3.57
8	1996	146	2.27	3.55
9	1997	59	0.92	1.43
10	1998	138	2.15	3.35
11	1999	143	2.22	3.47
12	2000	159	2.47	3.86

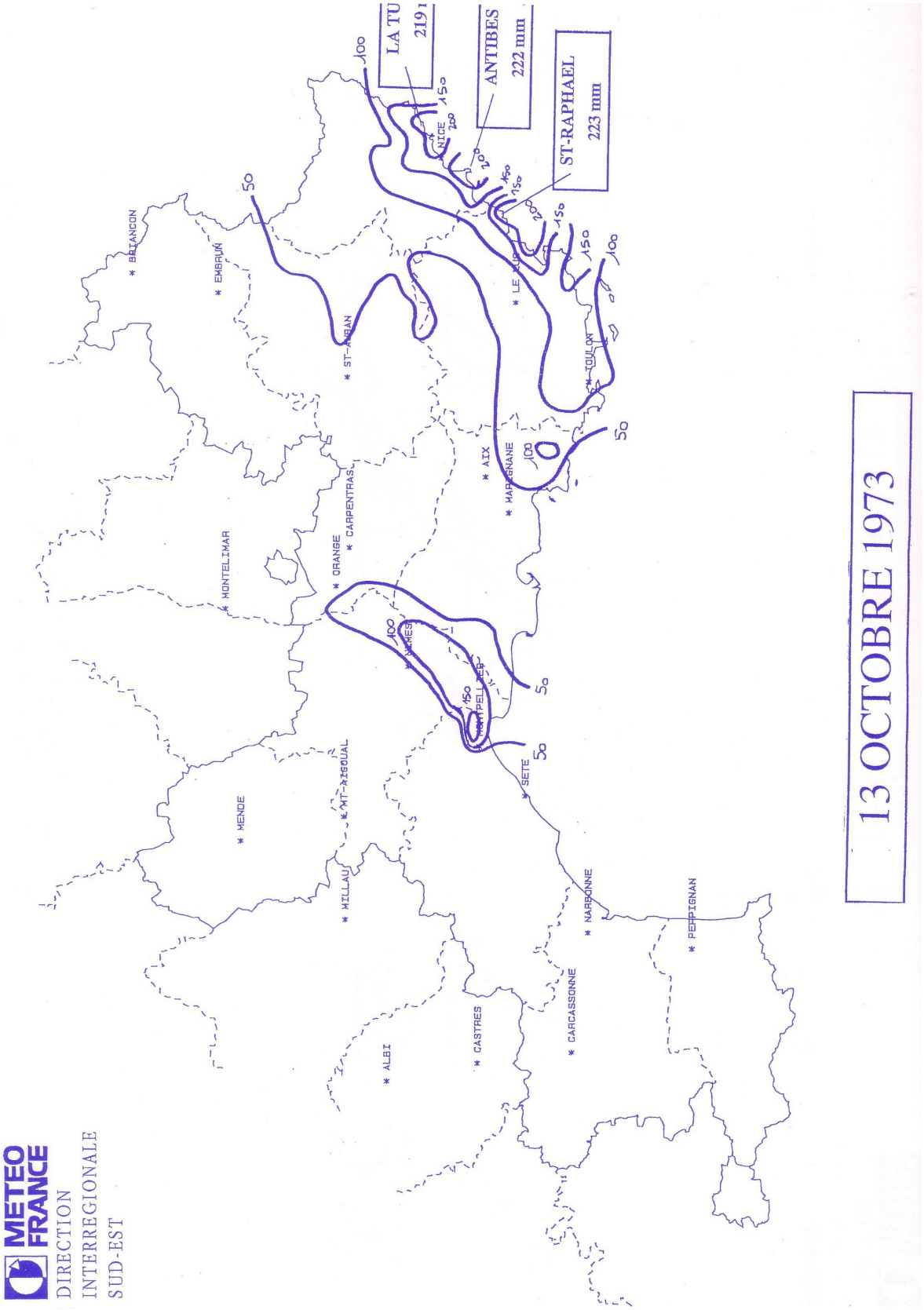
A.5 The highest rainfall accumulation in 24 hours the Var region

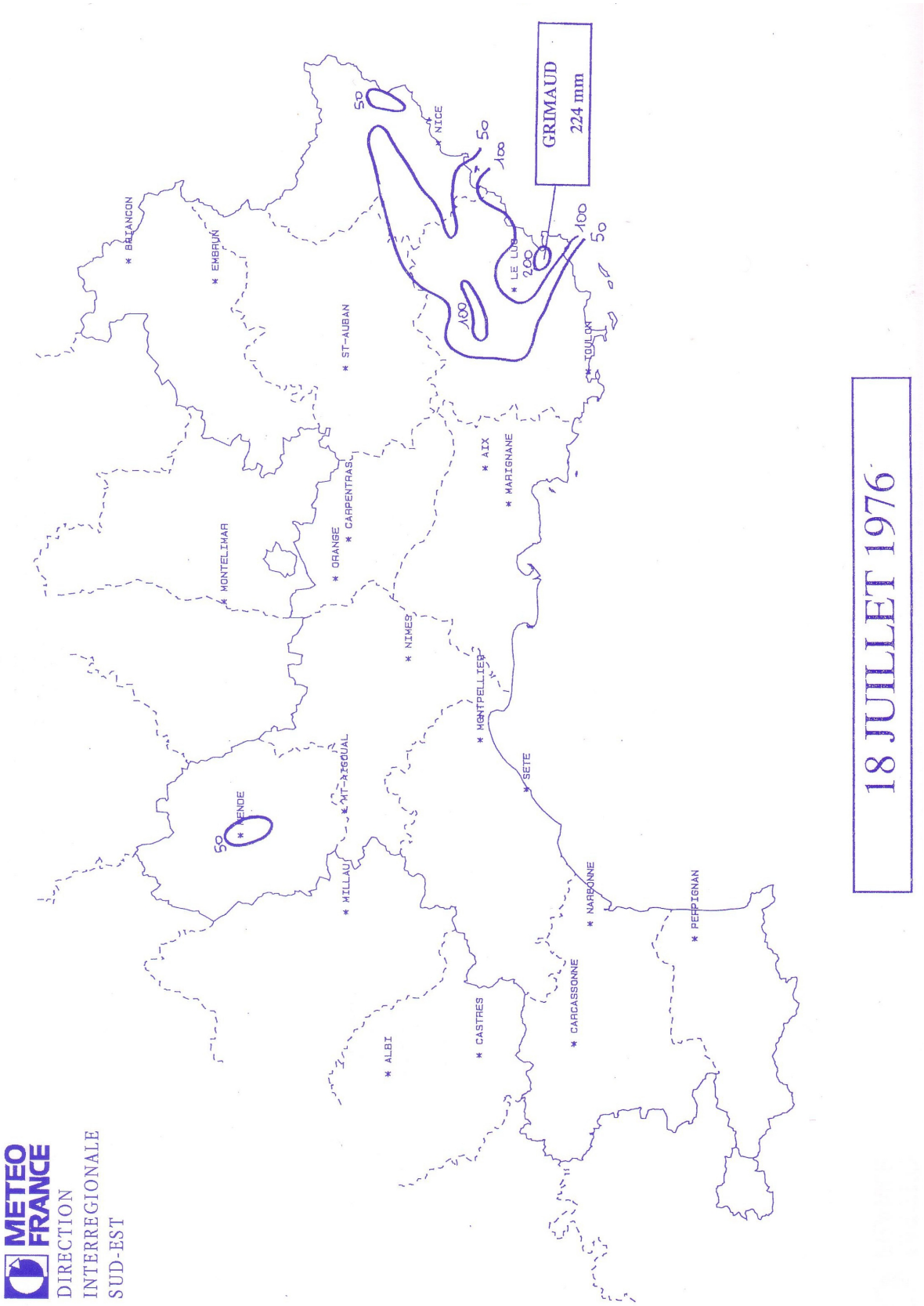

METEO FRANCE
DIRECTION
INTERREGIONALE
SUD-EST
18 et 19 DECEMBRE 1958
VALBELLE
230 mm
VILLEFORT
247 mm
CALLAS
262 mm
BRIANCON
EMBRUN
MONTELIMAR
ORANGE
CARPENTRAS
ST-AUBAN
NICE
LE LUC
TOULON
AIX
MARIGNANE
NIMES
MONTPELLIER
SETE
MENDE
MT-AIGOUAL
MILLAU
ALBI
CASTRES
CARCASSONNE
NARBONNE
PERPIGNAN

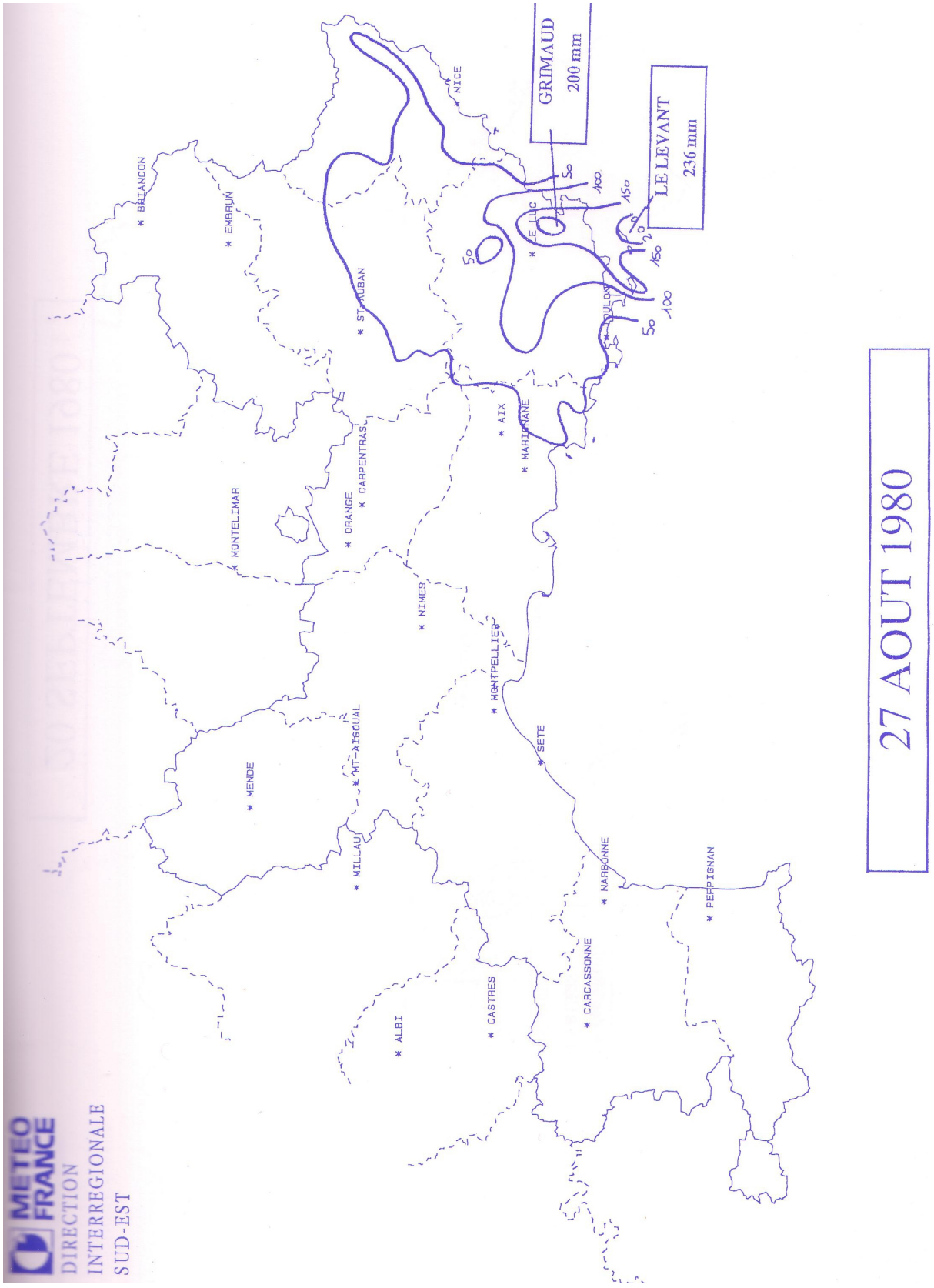

METEO FRANCE
DIRECTION
INTERREGIONALE
SUD-EST
MENDE
MILLAU
MT-AIGOUAL
ALBI
CASTRES
CARCASSONNE
NARBONNE
PERPIGNAN
SETE
MONTPELLIER
NIMES
MONTELIMAR
ORANGE
CARPENTRAS
BRIANCON
EMBRUN
ST-AUBAN
AIX
MARIGNANE
LE LUC
NICE
ADRETS DE
L'ESTEREL
232 mm
11 et 12 NOVEMBRE 1961


COLLOBRIERES
193 mm
27 et 28 DECEMBRE 1972
METEO FRANCE
DIRECTION
INTERREGIONALE
SUD-EST
NICE
BRIANCON
EMBRUN
ST-AUBAN
LE LUC
TOULON
AIX
MARIGNANE
CARPENTRAS
ORANGE
MONTELIMAR
NIMES
MONTPELLIER
SETE
MENDE
MILLAU
ALBI
CASTRES
CARCASSONNE
NARBONNE
PERPIGNAN


METEO FRANCE
DIRECTION
INTERREGIONALE
SUD-EST
2 OCTOBRE 1973
ST MANDRIER
209 mm
EYGUIERES
192 mm
NICE
BRIANCON
EMBRUN
ST-AUBAN
LE LUC
TOULON
MONTELIMAR
ORANGE
CARPENTRAS
AIX
MARIGNANE
NIMES
MONTPELLIER
SETE
MENDE
MT-AIGOUAL
MILLAU
NARBONNE
PERPIGNAN
ALBI
CASTRES
CARCASSONNE


METEO FRANCE
DIRECTION
INTERREGIONALE
SUD-EST
13 OCTOBRE 1973
LA TU
219
ANTIBES
222 mm
ST-RAPHAEL
223 mm
BRIANCON
EMBRUN
ST-AUBAN
NICE
TOULON
AIX
MARIGNANE
ORANGE
CARPENTRAS
MONTELIMAR
MENDE
MILLAU
SETE
NARBONNE
PERPIGNAN
CARCASSONNE
CASTRES
ALBI


GRIMAUD
224 mm
18 JUILLET 1976
METEO FRANCE
DIRECTION
INTERREGIONALE
SUD-EST
BRIANCON
EMBRUN
NICE
ST-AUBAN
LE LUC
TOULON
AIX
MARIGNANE
CARPENTRAS
ORANGE
MONTELIMAR
NIMES
MONTPELLIER
SETE
MILLAU
MT-AIGOUAL
MENDE
ALBI
CASTRES
CARCASSONNE
NARBONNE
PERPIGNAN
50
100
200


METEO FRANCE
DIRECTION
INTERREGIONALE
SUD-EST
GRIMAUD
200 mm
LE LEVANT
236 mm
27 AOUT 1980
NICE
BRIANCON
EMBRUN
ST-AUBAN
LE LUC
TOULON
AIX
MARIGNANE
CARPENTRAS
ORANGE
MONTELIMAR
NIMES
MONTPELLIER
SETE
MENDE
MT-AIGOUAL
MILLAU
ALBI
CASTRES
CARCASSONNE
NARBONNE
PERPIGNAN
50
100
150

A.6 The records of the 5 gauging stations in the Ardèche region

Table A.10 – *Station : Saint Martin on the Ardèche River (2240 km^2) ; Code hydro : V506401 (X :776696, Y :1926247) ; β=0.76 ; $\bar{Q}$= 1751.2*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$
1	1963	3160	8.98	1.80	23	1985	267	0.76	0.15
2	1964	1930	5.49	1.10	24	1986	1620	4.61	0.93
3	1965	1300	3.70	0.74	25	1987	1690	4.81	0.97
4	1966	540	1.54	0.31	26	1988	926	2.63	0.53
5	1967	1050	2.99	0.60	27	1989	594	1.69	0.34
6	1968	1220	3.47	0.70	28	1990	1070	3.04	0.61
7	1969	1330	3.78	0.76	29	1991	2580	7.34	1.47
8	1970	3250	9.24	1.86	30	1992	1640	4.66	0.94
9	1971	1500	4.26	0.86	31	1993	2010	5.71	1.15
10	1972	1540	4.38	0.88	32	1994	3040	8.64	1.74
11	1973	2570	7.31	1.47	33	1995	2820	8.02	1.61
12	1974	840	2.39	0.48	34	1996	2590	7.36	1.48
13	1975	1020	2.90	0.58	35	1997	1390	3.95	0.79
14	1976	2500	7.11	1.43	36	1998	1570	4.46	0.90
15	1977	3330	9.47	1.90	37	1999	1730	4.92	0.99
16	1978	930	2.64	0.53	38	2000	1430	4.07	0.82
17	1979	1600	4.55	0.91	39	2001	1500	4.26	0.86
18	1980	2100	5.97	1.20	40	2002	2960	8.42	1.69
19	1981	1070	3.04	0.61	41	2003	2010	5.71	1.15
20	1982	3510	9.98	2.00	42	2004	1040	2.96	0.59
21	1983	974	2.77	0.56	43	2005	1640	4.66	0.94
22	1984	1920	5.46	1.10					

TABLE A.11 – *Station : Vogue on the Ardèche River (636 km^2) ; Code hydro : V5014010 (X :764859, Y :1951259) ; β=0.76 ; $\bar{Q}$= 1763.8*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$
1	1966	315	2.33	0.41	20	1985	754	5.58	0.99
2	1967	461	3.41	0.60	21	1986	555	4.11	0.73
3	1968	269	1.99	0.35	22	1987	210	1.55	0.27
4	1969	396	2.93	0.52	23	1988	405	3.00	0.53
5	1970	1300	9.62	1.70	24	1989	2200	16.29	2.88
6	1971	260	1.92	0.34	25	1990	616	4.56	0.81
7	1972	550	4.07	0.72	26	1991	619	4.58	0.81
8	1973	980	7.25	1.28	27	1992	1190	8.81	1.56
9	1974	301	2.23	0.39	28	1993	1400	10.36	1.83
10	1975	1770	13.10	2.32	29	1994	1410	10.44	1.85
11	1976	570	4.22	0.75	30	1995	520	3.85	0.68
12	1977	880	6.51	1.15	31	1996	730	5.40	0.96
13	1978	571	4.23	0.75	32	1997	1050	7.77	1.37
14	1979	447	3.31	0.59	33	1998	1020	7.55	1.34
15	1980	1400	10.36	1.83	34	1999	671	4.97	0.88
16	1981	493	3.65	0.65	35	2000	1120	8.29	1.47
17	1982	152	1.13	0.20	36	2001	904	6.69	1.18
18	1983	738	5.46	0.97	37	2002	212	1.57	0.28
19	1984	875	6.48	1.15	38	2003	709	5.25	0.93

TABLE A.12 – *Station : Saint Laurent on the Borne River (63 km²) ; Code hydro : V5045810 (X :728167, Y :1954010) ; β=0.76 ; $\bar{Q}$= 134*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$
1	1968	83.4	3.58	0.62	16	1983	200	8.58	1.49
2	1969	170	7.29	1.27	17	1984	8.7	0.37	0.06
3	1970	47.2	2.03	0.35	18	1985	115	4.93	0.86
4	1971	80.7	3.46	0.60	19	1986	169	7.25	1.26
5	1972	167	7.17	1.25	20	1987	65.9	2.83	0.49
6	1973	28.3	1.21	0.21	21	1988	60.7	2.60	0.45
7	1974	61.7	2.65	0.46	22	1989	49.8	2.14	0.37
8	1975	300	12.87	2.24	23	1990	94.5	4.05	0.71
9	1976	142	6.09	1.06	24	1991	222	9.52	1.66
10	1977	72.8	3.12	0.54	25	1992	89.4	3.84	0.67
11	1978	66.5	2.85	0.50	26	1993	120	5.15	0.90
12	1979	694	29.78	5.18	27	1994	103	4.42	0.77
13	1980	34.9	1.50	0.26	28	1995	232	9.95	1.73
14	1981	211	9.05	1.57	29	1996	197	8.45	1.47
15	1982	91.2	3.91	0.68	30	1997	42.3	1.81	0.32

TABLE A.13 – *Station : Beauvene on the Eyrieux River (392 km^2) ; Code hydro : V4144010 (X :774080, Y :1988475) ; β=0.76 ; $\bar{Q}$= 355.2*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$
1	1969	103	1.10	0.29	16	1984	291	3.11	0.82
2	1970	238	2.54	0.67	17	1985	43.1	0.46	0.12
3	1971	90	0.96	0.25	18	1986	183	1.96	0.52
4	1972	198	2.12	0.56	19	1987	479	5.12	1.35
5	1973	685	7.32	1.93	20	1988	306	3.27	0.86
6	1974	62.5	0.67	0.18	21	1989	115	1.23	0.32
7	1975	113	1.21	0.32	22	1990	121	1.29	0.34
8	1976	716	7.66	2.02	23	1991	159	1.70	0.45
9	1977	624	6.67	1.76	24	1992	800	8.55	2.25
10	1978	182	1.95	0.51	25	1993	533	5.70	1.50
11	1979	276	2.95	0.78	26	1994	254	2.72	0.72
12	1980	700	7.49	1.97	27	1995	496	5.30	1.40
13	1981	239	2.56	0.67	28	1996	820	8.77	2.31
14	1982	869	9.29	2.45	29	1997	618	6.61	1.74
15	1983	220	2.35	0.62	30	1998	122	1.30	0.34

Table A.14 – *Station : Chambonas on the Chassezac River (507 km^2) ; Code hydro : V5045020 (X :742057, Y :1937046) ; β=0.76 ; $\bar{Q}$= 632.8*

N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$	N^0	Year	$Q_{max}(m^3/s)$	Q/S^β	$Q/\bar{Q}$
1	1971	330	2.90	0.52	15	1985	62.5	0.55	0.10
2	1972	461	4.05	0.73	16	1986	479	4.21	0.76
3	1973	900	7.91	1.42	17	1987	673	5.92	1.06
4	1974	148	1.30	0.23	18	1988	457	4.02	0.72
5	1975	363	3.19	0.57	19	1989	290	2.55	0.46
6	1976	1800	15.83	2.84	20	1990	134	1.18	0.21
7	1977	1200	10.55	1.90	21	1991	444	3.90	0.70
8	1978	359	3.16	0.57	22	1992	419	3.68	0.66
9	1979	505	4.44	0.80	23	1993	957	8.42	1.51
10	1980	53.4	0.47	0.08	24	1994	1400	12.31	2.21
11	1981	359	3.16	0.57	25	1995	999	8.79	1.58
12	1982	1200	10.55	1.90	26	1996	1480	13.02	2.34
13	1983	330	2.90	0.52	27	1997	299	2.63	0.47
14	1984	984	8.65	1.55					

A.7 Influence of the definition of the thresholds for the historical period in the case of the Nartuby at Trans-en-Provence (Var)

In the manuscript, three different historical periods of 120 years with one single maximum reported for each period have been defined (figure A.1.a and likelihood a). The threshold values for the computation of the non-exceedance probabilities in the likelihood are different for each period and equal to the maximum estimated discharge. One reviewer suggested another possibility : the three reported extreme discharge values could as well be considered as the 3 largest flood events over the 360 year historical period and one single threshold defined for the whole period (figure A.1.b and likelihood b). This representation is less consistent with the data : many flood events are reported before the 19^{th} century but without the possibility to evaluate their discharge. The 1674 flood seems to be the largest event, but there is no certainty that the 1827 or 2010 peak discharges were not exceeded in the past. Hypothesis (a) with higher threshold values in the past is more consistent with the available information and less constraining than hypothesis (b). It should therefore provide larger credibility intervals. But to what extent ? The influence of the model proposed for the historical data has been tested here and models (a) and (b) compared.

- The likelihood corresponding hypothesis (a) (used in chapter 4) :

$$\ell = P(X_1) * P(X_2) * P(X_3) * [1 - P(X_1)]^{120-1} * [1 - P(X_2)]^{120-1} * [1 - P(X_3)]^{120-1} \quad \text{(A.13)}$$

- The likelihood corresponding hypothesis (b) :

$$\ell = P(X_1) * P(X_2) * P(X_3) * [1 - P(X_3)]^{360-3} \quad \text{(A.14)}$$

The obtained results with both hypotheses are very similar as illustrated in Figure A.2, with slightly larger credibility intervals in case a (hypothesis 1) as foreseen.

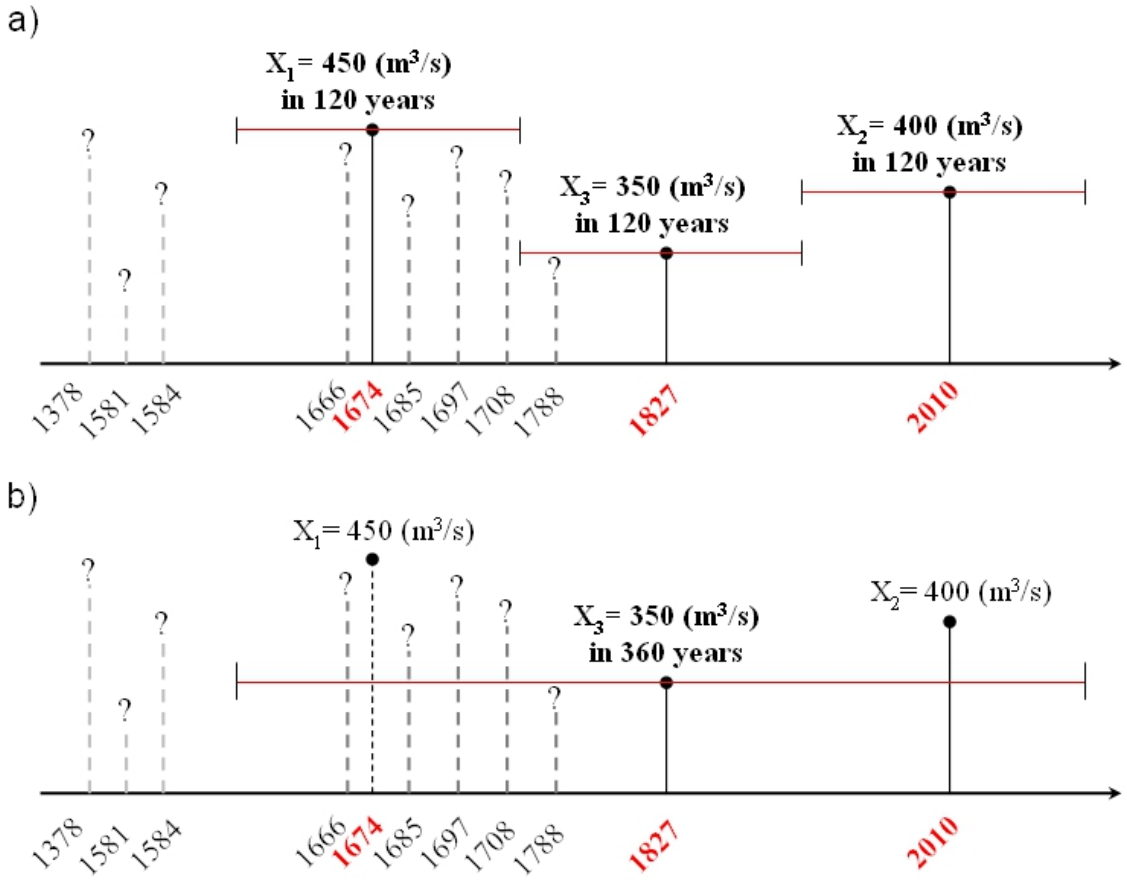


FIGURE A.1 – *Hypotheses of threshold for three extreme floods at Nartuby : a) hypothesis 1 (proposed), and b) hypothesis 2.*

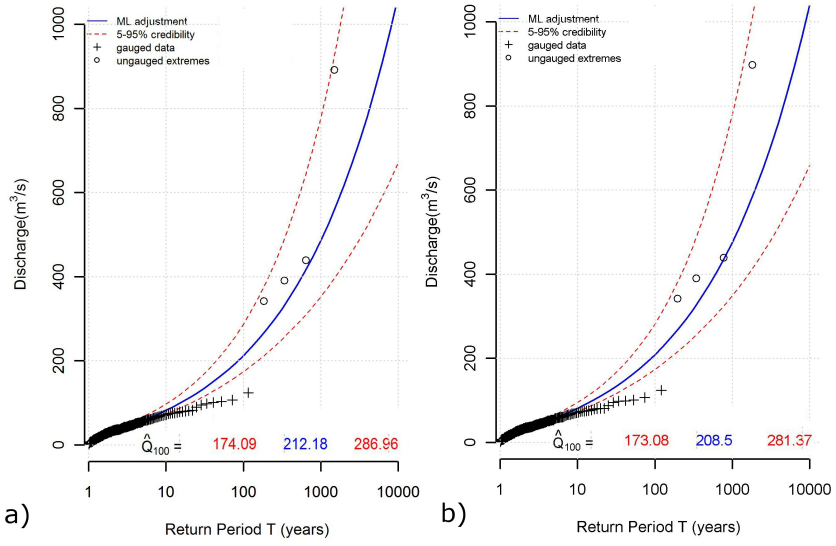


FIGURE A.2 – *Fitted GEV distributions and associated credibility intervals at Trans en Provence gauging station (193 km^2) with regional datasets of 7 sites incorporation with three values of historical flood at Nartuby and a value of extreme flood at Florieye : a) hypothesis 1, and b) hypothesis 2.*

A.8 Valuation of the ancient historical information available before the 20^{th} century at Saint Martin in the Ardèche region : data extracted from the PhD of Robin Naulet (cite)

Information on historical floods at Saint Martin (over the period 1772 to 1960) where retrieved by Robin Naulet (2002) (figure A.3), but have not been valuated in the manuscript. How would this additional information modify the conclusions about the estimation of the flood quantiles at this gauging station ? The computations were therefore conducted with this additional peace of information.

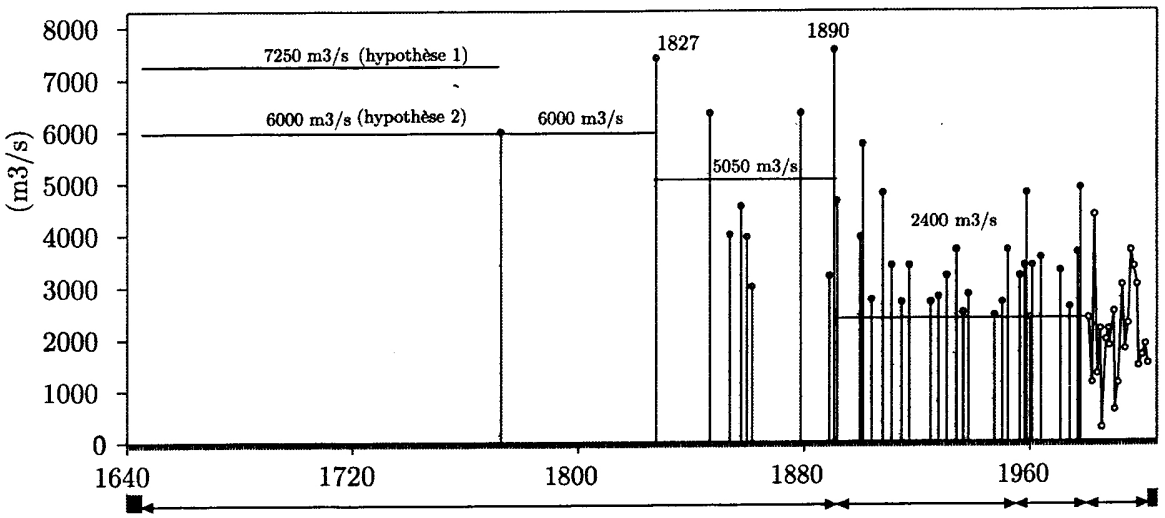


Figure A.3 – *Chronic of annual maximum flood at Saint Martin reconstructed by Robin Naulet (2002).*

Figure A.4 presents the shape of GEV distributions and associated 90% credibility intervals computed at Saint Martin (2240 km^2), based on the datasets available in the Ardèche region. Figure A.4.a and b consider local approach and combining information of historical floods corresponding hypothesis 1 (value of threshold = 7250 m^3/s) and hypothesis 2 (value of threshold = 6000 m^3/s) in the period (1772-1827). Figure A.4.c presents result of the proposed approach is used based on both gauged series and 18 ungauged extremes.

The local statistical analysis results including the ancient historical floods appear very consistent with the proposed regional analysis results including ungauged extremes (figure A.4). It is important here to recall that the local analysis results including information on the 20^{th} century only led to much lower quantile estimates. The new valuated local historical data confirm the regional estimates. This is a very reassuring result for the proposed approach : it validates a posteriori the estimated range of quantile values through the proposed regional approach.

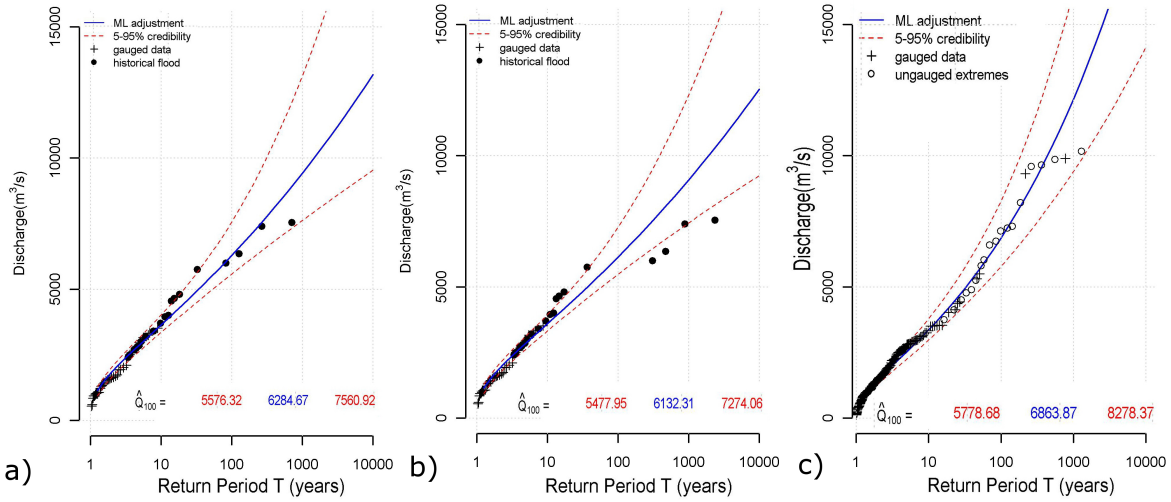


FIGURE A.4 – *Fitted GEV distributions and associated credibility intervals at Saint Martin gauging station (2240 km^2) with regional datasets of 5 sites incorporation with : a) information of historical floods corresponding hypothesis 1 ; b) hypothesis 2 ; and c) information of extreme floods.*

The same simulations, as those presented in the manuscript, were also conducted including the ancient historical extreme events to compare theoretically the accuracy of each considered flood frequency method in the case of the Saint Martin gauging station on the Ardèche river (figure A.5).

Without surprise, the important increase of the historical dataset at Saint Martin favours the local approach including historical data. If the regional variability is considered, the regional approaches may hardly help to reduce the uncertainties on the estimated quantiles in cases were the local data set is rich enough. Nevertheless, these simulation results are only valid if the real statistical distribution does not differ too much from the GEV distribution, which of course is never sure a priori. The theoretical simulations give only a tendency but can never demonstrate that an additional (regional) information will not be useful in a specific case study.

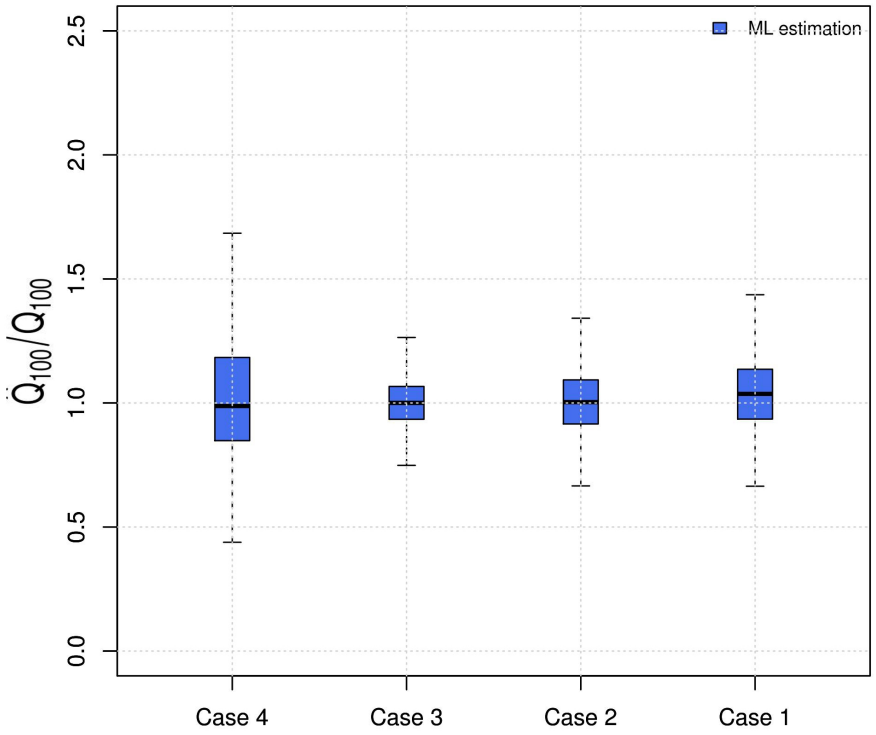


FIGURE A.5 – *Dispersion of the estimated maximum likelihood values* $\hat{Q}_{i,ML}^{(100)}$*, computed using 780 simulated samples, with regional variability in the index flood relation (*δ*= 0.1) and growth curves (n=40-year)at Saint Martin. All values have been divided by the real 100-year quantile* $Q_i^{(100)}$.

A.9 Elimination of the obviously dependent data in the Ardèche regional sample : extreme floods affecting the same river network at the same date

Some of the regional extremes reported for the Ardèche have occurred at the same date and have been produced by the same rainfall event. Some are event located on the same river network with an upstream-downstream relation (table A.15). The ungauged extremes of the original series can obviously not be considered as independent which is one of the basic hypotheses of the statistical model fitted to the data. What can be the consequence of this discrepancy between the model and the data set ? The type of dependence between the data is complex, it may in particular vary with the magnitude or return period of the floods. The influence on the statistical calibration results is therefore difficult to assess or study. The dependence may induce some biases, but the dominant effect will certainly be an underestimation of the credibility intervals due to the over-rating of the information content of the data set if the dependence between data is neglected. Independence can never be ensured in regional studies due to the spatial extent of the meteorlogical flood

generating processes. Gauged series can also hardly be considered as strictly independent. To have an idea of the possible influence of dependencies on the statistical calibration results, the series of ungauged extremes available in the Ardèche region has been further refined to remove obviously related extremes. One single extreme value has been randomly selected when more than one extreme value was reported for the same river system and the same date. The discarded values appear in grey color in table A.15. The refined series of regional ungauged extremes counts 13 values.

TABLE A.15 – *List of the ungauged extremes reported in the Ardèche region. Values discarded to reduce dependencies appear in light grey.*

Location	**River name**	**Date**	**Q** (m^3/s)	**S** (km^2)	**Associated period**
Pont de Rolandy	Ardeche	22/09/1992	1150	150	50
Aubenas	Ardeche	26/09/1992	2200	480	50
Sauze St.Martin	Ardeche	30/09/1958	4500	2240	50
Rosieres	Beaune	04/10/1958	1820	210	50
Joyeuse	Beaune	30/09/1958	1000	100	50
Chambon	Borne	30/09/1958	100	11	50
Vans	Boudaric	03/11/1989	130	6	50
Burzet	Bourges	22/09/1992	350	47	50
Chambonas	Chassezac	21/09/1980	3360	510	50
Dorne	Dorne	1963	630	78	50
Lamastre	Doux	03/08/1963	970	242	50
Pont de Cesar	Doux	03/08/1963	1500	635	50
Barrage des Collanges	Eyrieux	1963	1685	343	50
Meyras	Fontoliere	22/09/1992	900	130	50
Meysee	Lavezon	30/09/1960	500	56	50
Rieutord	Loire	01/09/1992	444	62	50
Pouzin	Ouveze	10/08/1967	700	140	50
Saliouse	Saliouse	1980	300	61	50

As illustrated in figure A.6, the refinement of the regional extreme dataset has only a slight influence on the credibility intervals. Since the 5 excluded events contain nevertheless some additional information if compared to the dataset of 13 extremes, the correct estimates of the credibility intervals when the samples of 18 extremes is used lies somewhere in between the results presented in figures A.6.a (underestimation) and A.6.b (overestimation).

For practical use, the dilemma can be summarized in the following way : either the whole information on extremes is valuated accepting the risk of underestimating the width of the credibility intervals, or some information is discarded to obtain more accurate credibility intervals but with less accurate quantile estimates.

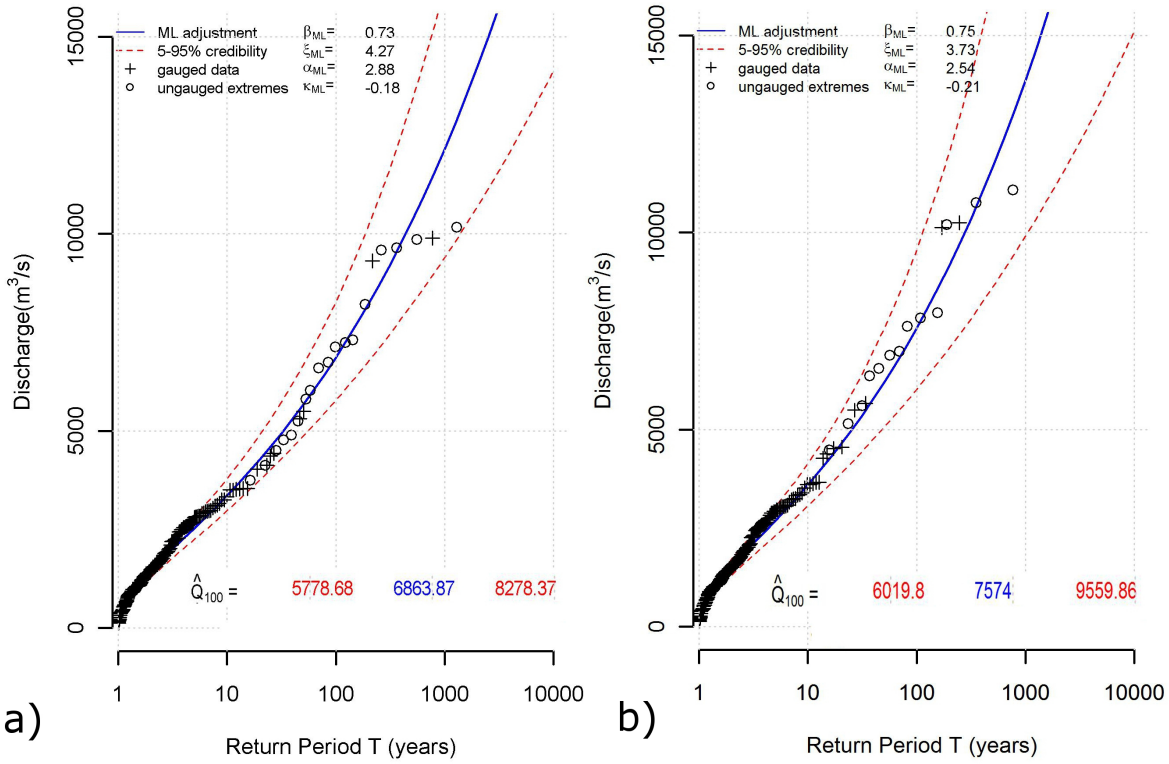


FIGURE A.6 – *Fitted GEV distributions and associated credibility intervals at Saint Martin gauging station (2240 km²) with regional datasets of 5 sites incorporation with : a) 18 ungauged sites (dependence datasets), and b) 13 ungauged sites (independence datasets).*

A.10 Influence of the percentage of error for systematic and extreme discharges considered for the computation of the likelihood, modified formulation

In the manuscript, a modified version of the likelihood function has been proposed including cumulative density functions only. The density functions usually used are replaced by the difference between the cumulative density functions corresponding to an upper and lower possible value for the considered discharges. In the manuscript, a difference of 1% of the average estimated discharge value has been selected arbitrary to define these upper and lower possible values. What is the influence of this arbitrary choice ? To test this hypothesis, three alternative choices have been tested : (a) 1% replaced by 0.1%, (b) initial version, (c) 1% replaced by 30% for the ungauged extremes that can reasonably be considered as much less reliable as the gauged discharges, (d) 1% replaced by 5%.

Figure A.7 illustrates the relatively moderate influence of this choice on the maximum likelihood estimate and on the credibility intervals. The intervals only very slightly increase when the considered percentage of uncertainty in the estimated discharge values increases. That remains true even if the uncertainty on the ungauged extreme values is relatively large (case c). This result is consistent with the results obtained by Payrastre (2005), who studied the influence of the uncertainties affecting reconstructed historical peak discharges on the results of flood frequency analyses and concluded that these uncertainties had, to a certain extent, little influence on the inference results. This is a strong argument for the valuation of ungauged extremes in flood frequency studies.

$$\begin{aligned}\ell\left(\mathbf{D} \mid \theta\right) = & \prod_{i=1}^{s}\left[\prod_{j=1}^{n_i}\left[F_\theta\left((1+\alpha) * \frac{Q_{i,j}}{\mu_i\left(\theta\right)}\right) - F_\theta\left((1-\alpha) * \frac{Q_{i,j}}{\mu_i\left(\theta\right)}\right)\right]\right] \\ & \times \prod_{k=1}^{h}\left[F_\theta\left((1+\alpha) * \frac{Q_k}{\mu_k\left(\theta\right)}\right) - F_\theta\left((1-\alpha) * \frac{Q_k}{\mu_k\left(\theta\right)}\right)\right] \\ & \times \prod_{k=1}^{h}\left[F_\theta\left((1+\alpha) * \frac{Q_k}{\mu_k\left(\theta\right)}\right)\right]^{(n_k-1)} \end{aligned} \tag{A.15}$$

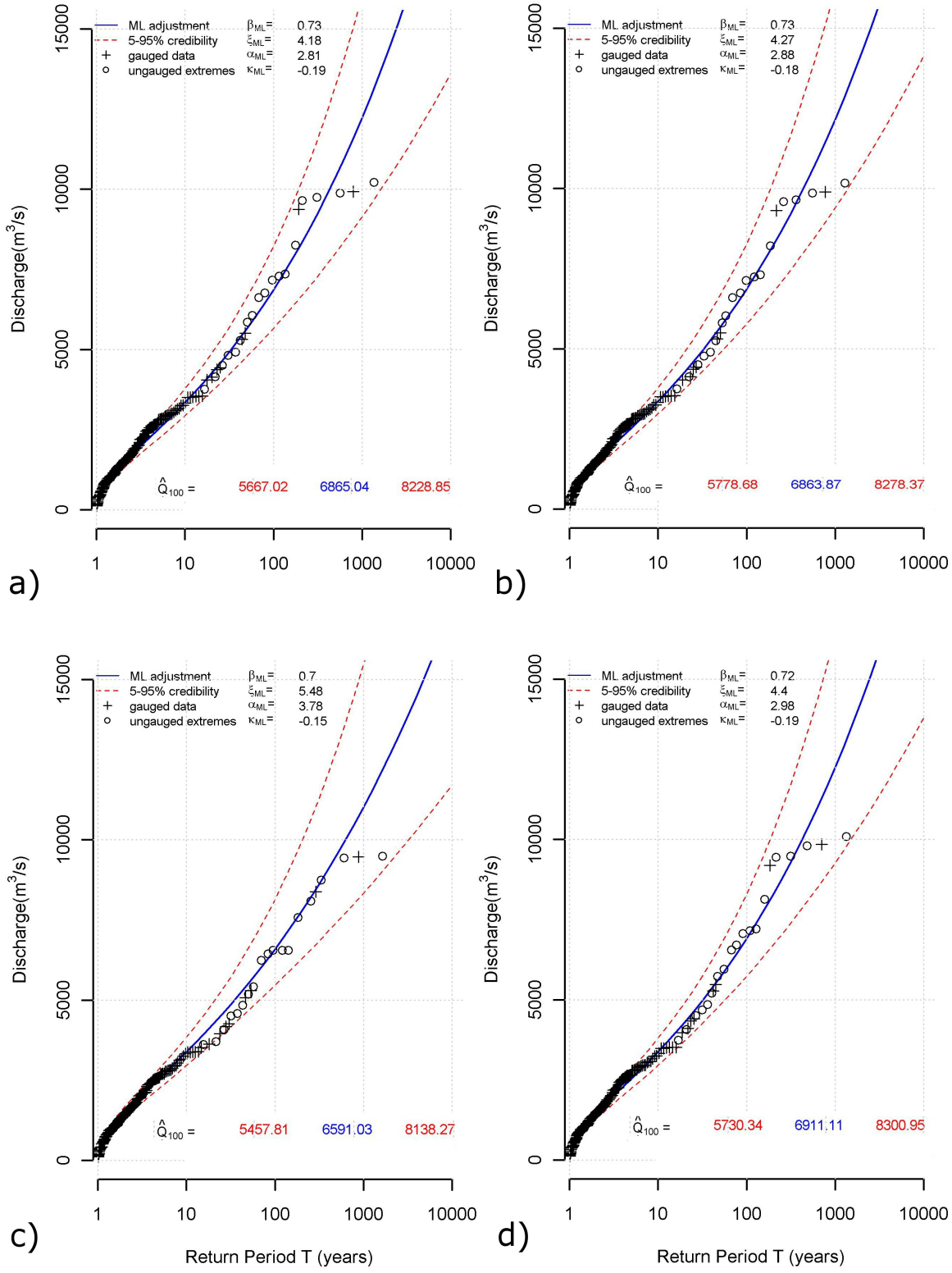


FIGURE A.7 – *Fitted GEV distributions and associated credibility intervals at Saint Martin gauging station (2240 km^2) with regional datasets composed of 5 gauged sites and 18 ungauged extremes : a) α = 0.001 for all data ; b) α = 0.01 for all data ; c) α = 0.01 for gauged dicharges and α = 0.3 for ungauged extremes ; and d) α = 0.05 for all data.*

A.11 Influence of the selected duration of the period over which the historical or ungauged extremes are considered to be the largest observed value : test for the Var case study

One last and important question concerns the lengths of the periods over which the historical or ungauged extremes can be considered as the largest observed values. This choice is based on the available information, but remains partly arbitrary and can easily vary by a factor of two depending on the point of view and the starting date of the considered period. For instance, if we know that an extreme flood invent occured in the 1710 on the Florieye but with no other information. What is the length of the period attributed to the 2010 flood event. Is it 300 years (period of time separating the two known extreme floods) ? Or is it half of this length (150 years) ? Or should we at least evaluate the hypothesis that the 2010 discharge is larger than the 1710 and the corresponding period covered by our inventory of extremes on the Florieye may exceed significantly 300 years ? The influence of this length on the inference results has been tested in the case of the regional analysis in the Var (figure A.8).

This length has a clear influence on the plotting position of the 2010 Florieye flood on the figures, but a surprising moderate influence on the inference results. To explain this results which may appear surprising, let us note that the 2010 event is a rare event. Due to sampling variability, it may occur once in a period of 150 years as well as in a period of 600 years. The likelihood approach, implicitly accounts for the possible influence of sampling variability and is therefore not very sensitive to the lengths of the periods attributed to the extremes as long as they remain consistent with the statistical extrapolations, which is the case here. The conclusions would have been different if a period of 1 million years would have been attributed to the 2010 flood event. But this would not have been consistent with observations. What the observations indicate is that the 2010 discharge on the Florieye has an empirical return period of a few centuries. And this is sufficient information for the inference procedure. Again, this moderate sensitivity is a strong argument for the incorporation of ungauged extremes in flood frequency studies.

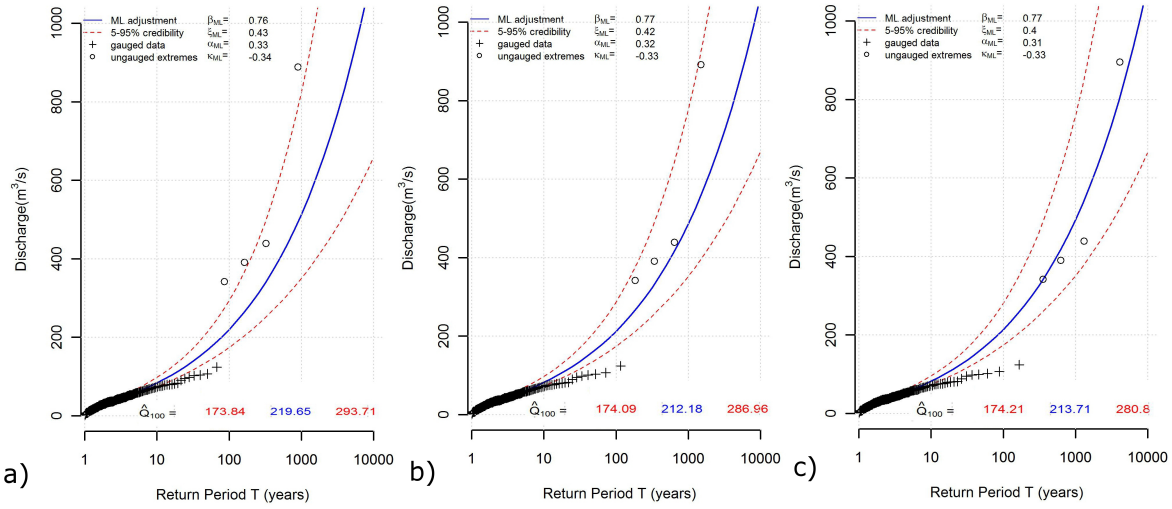


FIGURE A.8 – *Fitted GEV distributions and associated credibility intervals at Trans en Provence gauging station (193 km^2) with regional datasets of 7 sites incorporation with three values of historical flood at Nartuby (3x 120-year) and a value of extreme flood at Florieye with the return period : a) 150-year ; b) 300-year ; and c) 600-year.*

References

Bhaskar, N. and C. O'Connor (1989). Comparison of method of residuals and cluster analysis for flood regionalization. *Journal of Water Resources Planning and Management 115(6)*, 793–808. (Cité page 74.)

Bobee, B. and P. Rasmussen (1995). Recent advances in flood of frequency analysis. *US National Report to International Union of Geodesy and Geophysics 33*, 1111–1116. (Cité page 75.)

Burn, D. (1990a). An appraisal of the Region-of-Influence approach to flood frequency analysis. *Hydrological Sciences Journal 35(2)*, 149–165. (Cité pages 72, 74 et 75.)

Burn, D. (1990b). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research 26(10)*, 2257–2265. (Cité pages 72, 74 et 75.)

Burn, D. and Z. Zrinji (1993). Regional flood frequency analysis with a hierarchical region of influence approach. In *Proceedings of the international conference on Stochastic and Statistical methods in hydrology and environmental engineeging, University of Waterloo, Waterloo, Ontario*. (Cité page 72.)

Castellarin, A. (2005). Probabilistic behavior of a regional envelope curve. *Water resources research 41*, W06018. (Cité page 77.)

Castellarin, A., D. Burn, and A. Brath (2008). Homogeneity testing : How homogeneous do heterogeneous cross-correlated regions seem ? *Journal of Hydrology 360*, 67–76. (Cité page 80.)

Castellarin, A., R. Vogel, and N. Matalas (2007). Probabilistic envelope curves for design flood estimation at ungauged sites. *Water Resources Research 43*, W04406. (Cité page 77.)

Cavadias, G., T. Ouarda, B. Bobée, and C. Girard (2001). A canonical correlation approach to the determination of the homogeneous regions for regional flood estimation of ungauged basins. *Hydrological Sciences Journal 46(4)*, 499–512. (Cité page 74.)

Charles, N. and J. Stedinger (1999). Development of regional regression relationship with censored data. *Water resources research 35(4)*, 775–784. (Cité pages 5, 65 et 74.)

Choquette, A. (1988). Regionalization of peak discharges for streams in Kentucky. *U.S.Geological Survey Water-Resources Investigations Report*, 87–4209. (Cité page 74.)

Chow, V., D. Maidment, and L. Mays (1988). *Applied hydrology*. McGraw-Hill, New York. (Cité page 71.)

Cohn, T. and J. Stadinger (1987). Use of historical information in a maximum likelihood framwork. *Journal of Hydology 96 (1-4)*, 215–223. (Cité pages 5, 64 et 67.)

Coles, S. (2001). *An introduction to statistic modeling of extreme values*. Springer, London. (Cité page 81.)

Coles, S. and E. Powell (1996). Bayesian methods in extreme value modelling. A review and new developments. *Int. Stat. Rev 43*, 1–48. (Cité pages 81 et 84.)

Coles, S. and J. Tawn (1996). A bayesian analysis of extreme rainfall data. *Appl. Stat 45*, 463–478. (Cité page 81.)

Cowles, M. and B. Carlin (1996). Markov Chain Monte Carlo convergence diagnostics : a comparative review. *Statist-Assoc 91*, 883–904. (Cité page 95.)

Cox, D., V. Isham, and P. Northrop (2002). Flood : some probabilistic and statistical approaches. *Phil. Trans. R. Soc. Lond A360(1796)*, 1389–1408. (Cité page 81.)

Dalrymple, T. (1960). Flood frequency analyses. *Water Supply : Geological Survey, Reston,Virginia, USA*, 1543–A. (Cité pages 5, 6, 8, 65, 72 et 77.)

Delicado, D. and M. Goria (2008). A small comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power. *Computational statistics and data analysis 52(3)*, 1661–1673. (Cité page 81.)

Eng, K., P. Milly, and G. Tasker (2007). Flood regionalization : A Hybrid Geographic and Predictor-Variable Region-of-Influence Regression Method. *Journal of Hydrology 12(6)*, 585–591. (Cité page 76.)

Eng, K., J. Stedinger, and A. Gruber (2007). Regionalization of streamflow characteristics for the Gulf-Atlantic Rolling Plains using Leverage-Guided Region-of-Influence Regression. *EWRI World Water and Environmental Resources Congress, American Society of Civil Engineers*. (Cité page 76.)

Eng, K., G. Tasker, and P. Milly (2005). An analysis of Region-of-Influence methods for Flood Frequency Regionalization in the Gulf-Atlantic rolling plains. *Water Resour. Assoc 41(1)*, 135–143. (Cité pages 75 et 76.)

Fill, H. and J. Stedinger (1998). Using regional regression within index flood procedures and an empirical Bayesian estimator. *Journal of Hydrology 210*, 128–145. (Cité page 72.)

Gabriele, S. and N. Arnell (1991). A hierarchical approach to regional flood frequency analysis. *Water Resources Research 27(6)*, 1281–1289. (Cité pages 74 et 75.)

Garren, S. and R. Smith (1993). Convergence diagnostics for Markov Chain samplers. In *Technical Report, Department of Statistics, University of North Carolina*. (Cité page 95.)

Gary, D. and J. Stedinger (1987). Regional regression of flood characteristics employing historical information. *Journal of Hydology 96*, 255–264. (Cité pages 5 et 64.)

Gaume, E. (2006). On the asymptotic behavior of flood peak distributions. *Hydrology and Earth System Sciences 10(2)*, 233–243. (Cité pages 54, 60, 140 et 158.)

Gaume, E., V. Bain, P. Bernardara, O. Newinger, M. Barbuc, A. Bateman, L. Blaskovicova, G. Bloschl, M. Borga, A. Dumitrescu, I. Daliakopoulos, J. Garcia, A. Irimescu, S. Kohnova, A. Koutroulis, L. Marchi, S. Matreata, V. Medina, E. Preciso, D. Sempere-Torres, G. Stancalie, J. Szolgay, J. Tsanis, D. Velasco, and A. Viglione (2009). A compilation of data on European flash floods. *Journal of Hydrology 367*, 70–80. (Cité pages 35 et 122.)

Gaume, E., L. Gaal, A. Viglione, J. Szolgay, S. Kohnová, and G. Bloschl (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*. (Cité pages 6, 8, 10, 12, 34, 35, 59, 61, 67, 72, 81, 85 et 90.)

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science 7*, 457–511. (Cité pages 14, 15, 16 et 95.)

Geweke, J. (1992). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Oxford University Press, 169-193. (Cité page 95.)

Glatfelter, D. (1984). Techniques for estimating magnitude and frequency of floods on streams in Indiana. *U.S. Geological Survey Water-Resources Investigations Report*, 84–4134. (Cité page 74.)

Griffis, V. and J. Stedinger (2007). The use of GLS regression in regional hydrologic analyses. *Journal of Hydrology 204*, 82–95. (Cité page 73.)

Haddad, K. and A. Rahman (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. (Cité page 74.)

Heidelberger, P. and P. Welch (1983). Simulation run length control in the presence of an initial transient. *Operations Research 31*, 1109–1144. (Cité page 95.)

Heinz, D. and J. Stedinger (1998). Using regional regression within index flood procedures and an empirical Bayesian estimator. *Journal of Hydology 210*, 128–145. (Cité pages 74, 81 et 84.)

Hosking, J. and J. Wallis (1986a). Paleoflood hydrology and flood frequency analysis. *Water resources research 22*, 543–550. (Cité pages 5, 64 et 66.)

Hosking, J. and J. Wallis (1986b). The value of historical data in flood frequency analysis. *Water resources research 22*, 1606–1612. (Cité pages 5, 64 et 66.)

Hosking, J. and J. Wallis (1987). Parameter and quantile estimation for the Generalized Pareto distribution. *Technometrics 29*, 339–349. (Cité page 81.)

Hosking, J. and J. Wallis (1997). *Regional frequency analysis : An approach Based on L-Moments*. Cambridge University Press, London, UK. (Cité pages 5, 6, 8, 36, 39, 56, 61, 65, 68, 72, 74, 77, 78, 79, 99, 102, 123 et 155.)

Hosking, J., J. Wallis, and E. Wood (1985a). An appraisal of the regional flood frequency procedure in the UK Flood Studies Report. *Journal Hydol. Sci 30(1)*, 85–109. (Cité page 66.)

Hosking, J., J. Wallis, and E. Wood (1985b). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics 27*, 251–61. (Cité pages 5, 64 et 81.)

Javelle, P., T. Ouarda, M. Lang, B. Bobée, G. Galea, and J. Gresillon (2002). Development of regional flow-duration frequency curves based on the index-flood method. *Journal of Hydrology 258*, 249–259. (Cité page 74.)

Jin, M. and J. Stedinger (1989). Flood frequency analysis with regional and historical information. *Water Resour. Res 25(5)*, 925–36. (Cité page 67.)

Johnson, V. (1994). Studying convergence of Markov Chain Monte Carlo algorithms using coupled sampling paths. In *Technical Report, Institute for Statistics and Decision Sciences, Duke University*. (Cité page 95.)

Kendall, M. and A. Stuart (1979). *The advanced theory of statistic*. Charles Griffin and Company Limited, London. (Cité page 86.)

Kjeldsen, T. and D. Jones (2009). A formal statistical model for pooled analysis of extreme. *Hydrology Research 40(5)*, 465–480. (Cité pages 5, 65 et 74.)

Kjeldsen, T., J. Smithers, and R. Schulze (2002). Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method. *Journal of Hydrology 255*, 194–211. (Cité pages 5, 65 et 74.)

Kuczera, G. (1983). Effect of sampling uncertainty and spatial correlation on an empirical Bayes procedure for combining site and regional information. *Journal of Hydrology 65*, 373–389. (Cité page 73.)

Liu, C., J. Liu, and D. Rubin (1992). A variational control variable for assessing the convergence of the Gibbs sampler. *Proc. Statist. Comp. Sec. of ASA 21*, 74–78. (Cité page 95.)

Lu, L. and J. Stedinger (1992). Sampling variance of normalized GEV/PWM quantile estimators and regional homogeneity test. *Journal of Hydology 138*, 223–245. (Cité page 81.)

Madsen, H., P. Mikkelsen, D. Rosbjerg, and P. Harremoes (2002). Regional estimation of rainfall intensity duration curve using generalized least squares regression of partial duration series statistics. *Water Resources Research 38(11)*, 1–11. (Cité page 73.)

Madsen, H. and D. Rosbjerg (1997). Generalized least squares and empirical Bayes estimation in regional partial duration series index-flood modelling. *Water Resources Research 33(4)*, 771–782. (Cité page 73.)

Mann, H. and D. Whitney (1947). *On a test of whether one of two random variables is stochastically larger than the other.* Annals of Mathematical Statistics 18, 50-60. (Cité page 86.)

Merz, R. and G. Bloschl (2003). A process typology of regional floods. *Water Resources Research 39(12)*, 1340. (Cité pages 5, 65 et 74.)

Merz, R. and G. Bloschl (2005). Flood frequency regionalisation-spatial proximity vs.catchment attributes. *Journal of Hydology 302*, 283–306. (Cité page 76.)

Micevski, T. and G. Kuczera (2009). Combining site and regional flood information using a Bayesian Monte Carlo approach. *Water Resources Research 45*, W04405, doi : 10.1029/2008WR007173. (Cité page 74.)

Minghui, J. and J. Stedinger (1989). Flood frequency analysis with regional and historical information. *Water Resources Research 25*, 925–936. (Cité pages 5 et 64.)

Mykland, P., L. Tierney, and B. Yu (1995). Regeneration in Markov Chain samplers. *Journal of the American Statistical Association 90*, 233–241. (Cité page 95.)

Naulet, R. (2002). *Utilisation de l'information des crues historiques pour meilleure prédétermination du risque d'inondation. Application au bassin de l'Ardèche a Vallon Pont-d'Arc et St-Martin d'Ardèche*. Ph. D. thesis, Universite Joseph Fourier- Grenoble 1,Grenoble, France. (Cité pages 48, 140 et 152.)

Neppel, L., B. Renard, M. Lang, P. Ayral, D. Coeur, E. Gaume, N. Jacob, O. Payrastre, K. Pobanz, and F. Vinet (2010). Flood frequency analysis using historical data : accounting for random and systematic errors. *Hydrological Siences Journal 55*, 192–208. (Cité pages 5, 10, 64 et 84.)

Norbiato, D., M. Borga, M. Sangati, and F. Zanon (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. (Cité pages 5, 65 et 74.)

Ouarda, T., C. Girard, G. Cavadias, and B. Bobee (2001). Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology 254*, 157–173. (Cité pages 5, 65 et 74.)

Pandey, G. and V. Nguyen (1999). A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology 225*, 92–101. (Cité page 73.)

Payrastre, O., E. Gaume, and H. Andrieu (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. (Cité pages 147 et 149.)

Payrastre, O., E. Gaume, and H. Andrieu (2011). Usefulness of historical information for flood frequency analyses : Developments based on a case study. *Water resources research 47*, W08511. (Cité pages 5, 10, 59, 64, 83, 87, 152, 153 et 158.)

Payrastre, O., E. Gaume, P. Javelle, B. Janet, P. Fourmigue, P. Lefort, A. Martin, B. Boudevillain, P. Burunet, G. Delrieu, Y. Aubert, E. Dautrey, L. Durand, m. Lang, L. Boissuer, J. Douvinet, C. Martin, and I. Ruin (2012). Hydrological analysis of the catastrophic flash flood of 15th june 2010 in the area of Draguignan (Var, France). *Societe hydrotechnique de France, Actes du Congrès " Événements extrêmes fluviaux et maritimes"*, 1–8. (Cité pages 41 et 131.)

Raftery, A. and S. Lewis (1992). *How many iterations in the Gibbs sampler ?* Oxford University Press, 763-773. (Cité page 95.)

Rao, A. and V. Srinivas (2006a). Regionalization of watersheds by Fuzzy cluster analysis. *Journal of Hydrology 318(1-4)*, 57–79. (Cité page 74.)

Rao, A. and V. Srinivas (2006b). Regionalization of watersheds by Hybrid cluster analysis. *Journal of Hydrology 318(1-4)*, 37–56. (Cité page 74.)

Reed, D., D. Jakob, and A. Robson (1999). *Statistical procedures for flood frequency estimation. Flood Estimation Handbook*. Institute of Hydrology, Wallingford, UK. (Cité page 71.)

Reis, D., J. Stedinger, and E. Martins (2005). Bayesian GLS regression with application to LP3 regional skew estimation. *Water resources research 41*, W10419, doi :10.1029/2004WR00344. (Cité pages 5, 10, 64, 73, 74, 83 et 84.)

Renard, B., V. Garreta, and M. Lang (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. (Cité pages 10, 83 et 84.)

Ribatet, M., T. Ouarda, E. Sauquet, and J. Gresillon (2009). Modeling all exceedances above a threshold using an extremal dependence structure : Inferences on several flood characteristics. *Water resources research 45*, W03407. (Cité pages 5 et 64.)

Ribatet, M., E. Sauquet, J. Gresillon, and T. Ouarda (2007a). A regional Bayesian POT model for flood frequency analysis. *Stoch Environ Res Risk Assess 21*, 327–339. (Cité pages 5, 65 et 74.)

Ribatet, M., E. Sauquet, J. Gresillon, and T. Ouarda (2007b). Usefulness of the reversible jump Markov Chain Monte Carlo in regional flood frequency analysis. *Water resources research 43*, W08403. (Cité pages 5, 64, 74 et 84.)

Ritter, C. and M. Tanner (1992). Facilitating the gibbs sampler : The Gibbs stopper and the Griddy Gibbs sampler. *Journal of the American Statistical Association 87*, 861–868. (Cité page 95.)

Robert, C. and Casella.G (2004). *Monte Carlo statistical methods*. Springer, New York. 2nd ed. (Cité page 85.)

Roberts, G. (1992). *Convergence diagnostics of the Gibbs sampler.* Oxford University Press, 775-782. (Cité page 95.)

Roberts, G. (1994). Methods for estimating L2 convergence of Markov Chain Monte Carlo. In *Bayesian analysis in statistics and econometrics : essays in honor or Arnold Zellner, Wiley, New York.* (Cité page 95.)

Rogger, M., A. Viglione, R. Merz, R. Kirnbauer, H. Pirkl, and G. Bloschl (2011). Towards understanding the differences between deterministic and probabilistic flood hazard estimation methods. *Risk in Water Resources Management 347*, 16–21. (Cité pages 54, 60, 140 et 158.)

Salaheddine, A., F. Anne-Catherine, and B. Bernard (2006). Comparison of methodologies to assess the convergence of Markov Chain Monte Carlo methods. *Computational statistics and data analysis 50*, 2685–2701. (Cité page 95.)

Seidou, O., T. Ouarda, M. Barbet, P. Bruneau, and B. Bobée (2006). A parametric Bayesian combination of local and regional information in flood frequency analysis. *Water resources research 42*, W11408. (Cité pages 5, 65, 74, 81 et 84.)

Sheffer, N. A., Y. Enzel, G. Benito, T. Grodek, N. Poart, M. Lang, R. Naulet, and D. Coeur (2003). Paleofloods and historical floods of the Ardèche river, France. *Water resources research 39(12)*, 1379. (Cité pages 5 et 64.)

Stedinger, J. (1989). *Using historical and regional information in flood frequency analysis*. Proceeding of the pacific international seminar on water resources systems, Tomanu, Japan. (Cité page 72.)

Stedinger, J. and L. Lu (1995). Appraisal of regional and index flood quantile estimators. *Stochastic Hydrology and Hydraulics 9*, 49–75. (Cité page 81.)

Stedinger, J. and G. Tasker (1985). Regional hydrologic analysis 1. ordinary, weighted and generalized least squares compared. *Water Resource Research 21*, 1421–32. (Cité page 73.)

Stedinger, J. and G. Tasker (1986). Correction to regional hydrologic analysis, 1.ordinary, weighted, and generalised least squares compared. *Water Resour.Res 22(5)*, 844. (Cité pages 5, 64, 67 et 73.)

Sutcliffe, J. (1987). The use of historical records in flood frequency analysis. *Journal of Hydrology 96*, 159–171. (Cité pages 5 et 64.)

Tanner, M. (1996). *Tool for statistical inference : Methods for the Exploration of Posterior, Distributions and Likelihood Functions*. Springer, Evanston, USA. (Cité page 85.)

Tasker, G., J. Eychaner, and J. Stedinger (1986). Application of Generalised Least Squares in regional hydrologic regression analysis. *U.S. Geological Survey Water Supply, Washington, DC Water Supply Paper 2310*, 107–115. (Cité pages 73 et 75.)

Tasker, G., S. Hodge, and C. Barks (1996). Region of influence regression for estimating the 50 year flood at ungauged sites. *Water Resour. Bulletin 32(1)*, 163–170. (Cité page 72.)

Tasker, G. and R. Slade (1994). An interactive regional regression approach to estimating flood quantiles, in Fontane DG and Tuvel HN eds. *Water policy and management, solving the problems, Proceedings of the Twenty-First Annual Conference, May 23-26, 1994 : Denver, Colo., Denver Society of American Society of Civil Engineers*, 782–785. (Cité page 72.)

Tasker, G. and J. Stedinger (1989). An operational GLS model for hydrologic regression. *Journal of Hydology 111*, 361–375. (Cité page 73.)

Thomas, D. and M. Benson (1970). Generalization of streamflow characteristics from drainage-basin characteristics. *US Geological Survey Water-Supply, Paper 1975, pp. 55*. (Cité page 74.)

Wallis, J. (1997). Personal communication. (Cité page 80.)

Wallis, J. R., M. Schaefer, B. Barker, and G. Taylor (2007). Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. *Hydrology and Earth System Siences (11)*, 415–442. (Cité pages 5, 65, 74 et 80.)

Wandle Jr., S. (1977). Estimating the magnitude and frequency of floods on natural-flow streams in Massachusetts. *Water Resources Investigations Report*, 39–77. (Cité page 74.)

Werritty, A., J. Paine, N. Macdonald, J. Rowan, and L. McEwen (2006). Use of multi-proxy flood records to improve estimates of flood risk : Lower river Tay, Scotland. *Catena 66(1-2)*, 107–119. (Cité pages 5 et 64.)

Wilcoxon, F. (1945). Individual comparisons by ranking methods. In *Biometrics Bulletin, 1,80-83*. (Cité page 86.)

Wiltshire, S. (1986). Regional flood frequency analysis. Multivariate classification of drainage basins in Britain. *Hydrological Sciences Journal 31(3)*, 335–346. (Cité page 74.)

Yu, B. (1994). Monitoring the convergence of Markov samplers based on estimated L1 error. In *Technical Report, Department of Statistics, University of California at Berkeley*. (Cité page 95.)

Yu, B. and P. Mykland (1994). Looking at Markov samplers through cusum path plots : A simple diagnostics idea. In *Technical Report, Department of Statistic, University of California at Berkeley*. (Cité page 95.)

Zellner, A. and C. Min (1995). Gibbs sampler convergence criteria. *J. Am. Statist. Assoc 90*, 921–927. (Cité page 95.)

Zrinji, Z. and D. Burn (1994). Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology 153*, 1–21. (Cité pages 72 et 74.)

Zrinji, Z. and D. Burn (1996). Regional flood frequency with hierarchical region of influence. *Journal of Water Resources Planning and Management 122(4)*, 245–252. (Cité pages 72, 74 et 75.)

Titre Amélioration des approches Bayésiennes MCMC pour l'analyse régionale des crues.

Résumé Cette thèse présente des développements et une évaluation de l'approche d'analyse régionale des débits proposée par Gaume (2010). Celle-ci consiste à incorporer dans l'analyse régionale les informations relatives aux crues extrêmes observées sur des bassins non jaugés. Une évaluation des performances et de la robustesse de la méthode est proposée ici par comparaison avec une approche régionale plus classique proposée par Hosking et Wallis (1997). Les comparaisons sont basées à la fois sur des simulations et sur l'étude de cas. La procédure d'inférence utilisée s'appuie sur une distribution GEV, associée à une formulation spécifique de la vraisemblance et une approche Bayésienne MCMC pour l'estimation des paramètres. Dans un premier temps, les résultats d'inférence obtenus sans tenir compte des crues extrêmes sont comparés sur la base de simulations, en s'intéressant aux effets possibles d'hétérogénéités au sein des régions considérées. Ensuite, les deux approches sont appliquées à deux régions. Cette application confirme l'impact très positif de l'incorporation de l'information sur les crues extrêmes dans l'analyse régionale qui permet souvent de dépasser les performances obtenues avec une analyse régionale classique.

Mots-clés Crues, Risque Hydrologique, Statistique, Bayes, Valeurs extrêmes, MCMC.

Title Improvement of Bayesian MCMC approaches for regional flood frequency analyses

Abstract This thesis presents additional developments of an approach initially proposed by Gaume (2010), that aims to incorporate available information on extreme floods at ungauged sites in a regional flood frequency analyses (RFFA). The performances and robustness of this approach are tested and compared to a reference approach proposed by Hosking & Wallis (1997). The comparisons are based both, on simulations and case studies. The inference procedure is based on a GEV distribution associated with a specific likelihood formulation and a Bayesian MCMC algorithm for the estimation of the parameters. First, the inference results obtained without incorporating extreme floods are compared based on simulations, with a focus on the effects of possible heterogeneities in the considered regions. Next, both approaches are applied to two regions. This application finally confirms the very positive impact of the incorporation of information on extreme floods in RFFA, that enables to outperform the results based on a conventional regional approach.

Keywords Flood, Risk assessment, Statistics, Bayes, Extreme values, MCMC.