

# Chapitre 3 : Analyse probabiliste des crues du XVI<sup>ème</sup> siècle à aujourd’hui

Mathieu

February 24, 2023

## Contents

<b>1</b>	<b>Introduction du chapitre</b>	<b>2</b>
<b>2</b>	<b>Méthodes d’analyse probabiliste d’un échantillon mixte de crues</b>	<b>2</b>
2.1	Concepts de base et hypothèses	2
2.2	Propagation des incertitudes hydrométriques de la période continue	3
2.3	Incertitudes autour du seuil de perception	3
2.4	Incertitudes autour de la durée de la période historique	4
2.5	Incertitudes autour du seuil de perception et de la durée de la période historique	4
2.6	Distribution empirique d’un échantillon mixte	5
<b>3</b>	<b>Données disponibles</b>	<b>5</b>
3.1	Échantillon mixte de crues du Rhône à Beaucaire	5
3.2	Homogénéité des données	5
3.2.1	Données continues	6
3.2.2	Données historiques	7
<b>4</b>	<b>Application aux crues du Rhône à Beaucaire</b>	<b>8</b>
4.1	Résultats pour la période récente dégradée (1816-2020)	9
4.1.1	Quel est l’apport de l’utilisation des crues historiques pour une longueur de chronique “courante” ?	10
4.1.2	Quel est l’impact de la méconnaissance du seuil de perception sur l’estimation des quantiles ?	11
4.1.3	Quel est l’impact de la méconnaissance de la durée de la période historique sur l’estimation des quantiles ?	11
4.1.4	Quel est l’impact de la méconnaissance du seuil de perception et de la durée de la période historique sur l’estimation des quantiles ?	12
4.1.5	Ajout d’un modèle pour lequel le débit des crues sup-seuil historiques sont connus ? A FAIRE ?	13
4.1.6	Discussion intermédiaire	13
4.2	Application à la période 1500-2020	14
4.2.1	BASELINE VS A pour l’échantillon 1 ?. Quel est l’apport des crues historiques pour l’analyse fréquentielle à Beaucaire ?	14

4.2.2	Modèle A pour les échantillons 1 et 2 : Quel est l'impact du choix de l'échantillon de crues historiques (S3 vs S4) . . . . .	14
4.3	Discussion . . . . .	14
5	Conclusion du chapitre	14

## 1 Introduction du chapitre

Quelques rappels d'intro aux méthodes d'analyse des crues histo: Stedinger and Cohn, 1986  
Solutions quand le débit est connu  
Solution quand le débit n'est pas connu  
Explication du concept de seuil de perception et de ses limites.  
durée de la période historique

**Méconnaissance du seuil** Le seuil de perception est un concept empirique qui ne prend une signification physique que dans certaines situations. Par exemple, imaginons une station hydrométrique dont la section n'a pas évolué au cours du temps et pour laquelle les débordements surviennent toujours au delà d'un même débit. Ces débordements (par exemple au-dessus d'une digue qui n'a subi aucune modification au cours du temps) laissent systématiquement une trace dans les écrits ou sur des infrastructures (marques de crue) suite aux dommages occasionnés. Cette situation parfaite existe rarement et le concept de seuil de perception peut être mis à mal par une variabilité temporelle de la perception des crues par les populations ripariennes. Néanmoins, l'utilisation du concept de seuil de perception est très souvent obligatoire pour exploiter les données de crues historiques.

**Méconnaissance de la durée histo** Cette durée est pourtant importante dans le calcul de la binomiale. Il est erroné de considérer que la période historique démarre à la première crue et cela peut entraîner une sur-estimation de la fréquence d'occurrence. Plusieurs méthodes existent dans la littérature pour calculer cette durée et se basent sur (A COMPLETER...).

## 2 Méthodes d'analyse probabiliste d'un échantillon mixte de crues

### 2.1 Concepts de base et hypothèses

L'utilisation d'occurrences de crues historiques pour l'analyse fréquentielle nécessite a minima l'hypothèse suivante : le débit de tous les événements de crue qui ont laissé une trace est supérieur à un seuil de perception. Il existe des méthodes d'analyse pour lesquelles il n'est pas nécessaire de connaître précisément le débit de ces événements (voir notamment Stedinger and Cohn (1986)). Le nombre de dépassements d'un seuil de perception pour une durée donnée est une information qui peut être exploitée en l'état. Ainsi, il faut considérer un échantillon mixte, composé d'une part de débits enregistrés en continu et d'autre part d'un nombre d'occurrences de crues supérieures à un seuil de perception.

De manière similaire au Chapitre 1 (REF), on suppose que le débit maximum annuel des périodes continues et historiques  $Q$  est une variable aléatoire *iid* qui suit une distribution GEV, de paramètres  $\theta = (\mu, \sigma, \xi)$  (respectivement : position, échelle, forme). Pour simplifier, on suppose ici que le paramètre de forme  $\xi$  est différent de zéro (forme Gumbel). Ainsi, on a la fonction de répartition de la GEV :  $F(x; \theta) = e^{-(1-\xi(\frac{x-\mu}{\sigma}))^{1/\xi}}$ . Les débits de l'échantillon de maximum annuels enregistrés en continu pendant  $j$  années

$\mathbf{q} = (q_t)_{t=1,\dots,j}$  sont ici supposés parfaitement connus et ne sont donc pas affectés d'une quelconque incertitude. L'échantillon historique est composé de  $k$  événements ayant dépassé le seuil de perception  $S$  sur une période de  $n$  années. Le seuil de perception n'a donc pas été dépassé pour les  $k - n$  années restantes. Ainsi, la probabilité de dépassement du seuil peut s'écrire :

$$\pi = \left(1 - F(S; \boldsymbol{\theta})\right) = 1 - e^{-\left(1 - \xi\left(\frac{S - \mu}{\sigma}\right)\right)^{1/\xi}} \quad (1)$$

On suppose que  $k$ , le nombre de dépassements du seuil de perception, peut être estimé par une loi binomiale de paramètres  $n$  et  $\pi$ , soit  $\mathcal{B}(n, \pi)$ . On peut alors écrire la fonction de vraisemblance (equation 2) qui est fonction d'un échantillon mixte de données composé des débits maximum annuels de la période continue  $(q_t)_{t=1,\dots,j}$  et du nombre de dépassements du seuil de perception  $k$  durant la période historique  $n$  :

$$L(\boldsymbol{\theta}; \mathbf{q}, k) = \underbrace{\prod_{t=1}^j f(q_t; \boldsymbol{\theta})}_{\text{a}} \underbrace{\left\{ \binom{n}{k} F(S; \boldsymbol{\theta})^{n-k} [1 - F(S; \boldsymbol{\theta})]^k \right\}}_{\text{b}} \quad (2)$$

Ici, le terme a représente la vraisemblance pour les données continues et le terme b pour les données historiques. L'application de la formule de Bayes permet de calculer la distribution a posteriori des paramètres  $\boldsymbol{\theta}$  sachant les données :

$$P(\boldsymbol{\theta} \mid \mathbf{q}, k) \propto L(\boldsymbol{\theta}; \mathbf{q}, k) P(\boldsymbol{\theta}) \quad (3)$$

Le terme  $P(\boldsymbol{\theta})$  représente ici la distribution a priori des paramètres qu'il faudra éliciter. La distribution a posteriori est explorée via une méthode Bayésienne MCMC (notamment décrites par Coles (2001)). Cette distribution a posteriori représente l'incertitude d'échantillonnage du modèle par  $r$  jeux de paramètres :  $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_r)$ . Le jeu de paramètres le plus probable est appelé maxpost et s'écrit  $\hat{\boldsymbol{\theta}}$ .

## 2.2 Propagation des incertitudes hydrométriques de la période continue

Dans la section précédente, l'incertitude autour des débits maximum annuels de la période continue est supposée négligeable. Cette incertitude pouvant atteindre 30 % à Beaucaire (REF Chap 1), il semble pragmatique de la considérer. Comme décrit dans la section REF du Chapitre 1, cette incertitude hydrométrique est représentée par  $s = 500$  réalisations :  $(q_t^{(i)})_{t=1,\dots,j; i=1,\dots,s}$ . Elle peut être propagée aux estimations des paramètres de l'équation 2 en estimant un jeu de paramètres pour chacune des  $s$  réalisations, soit  $(\boldsymbol{\theta}^{(i)})_{i=1,\dots,s}$ . Au total,  $r \times s$  jeux de paramètres sont estimés et représentent l'effet combiné de l'incertitude d'échantillonnage et de l'incertitude hydrométrique des données continues, on a donc  $(\boldsymbol{\theta}_p^{(i)})_{p=1,\dots,r; i=1,\dots,s}$ . Le jeu de paramètres le plus probable est calculé en utilisant l'échantillon maxpost de débits maximum annuels (CHAPITRE 1) sur lequel on estime le jeu de paramètres maxpost de l'équation 2. Le modèle décrit ci-dessus sera appelé "modèle A". La propagation des incertitudes hydrométriques de la période continue telle que décrite ici sera effectuée similairement pour les trois modèles définis dans les sections suivantes.

## 2.3 Incertitudes autour du seuil de perception

Afin de considérer dans le modèle proposé une prise en compte pragmatique de la méconnaissance du seuil de perception, il est possible de considérer ce seuil comme étant un paramètre à part entière du

modèle. Il ne s'agit pas ici de considérer un seuil de perception variable au cours du temps, quoi que de nombreux exemples dans la littérature font l'hypothèse de plusieurs seuils de perception successifs dont le débit est différent. Un seul seuil de perception est ici considéré pour l'ensemble de l'échantillon, mais sa valeur est incertaine et est déterminée par le modèle. L'impact de la méconnaissance du seuil est ainsi répercuté sur l'incertitude des résultats. Dans la section précédente, le seuil de perception faisait déjà partie du modèle (équation 2), mais sa valeur était supposée connue. Ici, la vraisemblance s'écrit différemment :

$$L(\boldsymbol{\theta}, S; \mathbf{q}, k) = \prod_{t=1}^j f(q_t; \boldsymbol{\theta}) \left\{ \binom{n}{k} F(S; \boldsymbol{\theta})^{n-k} [1 - F(S; \boldsymbol{\theta})]^k \right\} \quad (4)$$

On peut alors écrire une nouvelle distribution a posteriori :

$$p(\boldsymbol{\theta}, S | \mathbf{q}, k) \propto L(\boldsymbol{\theta}, S; \mathbf{q}, k) p(\boldsymbol{\theta}, S) \quad (5)$$

Les distributions de paramètres a posteriori de ce modèle reflètent donc l'incertitude hydrométrique de la période continue, l'incertitude d'échantillonnage, ainsi que l'incertitude autour du seuil de perception. Ce modèle sera nommé "modèle B" dans les sections suivantes.

## 2.4 Incertitudes autour de la durée de la période historique

L'équation 2 repose à la fois sur le fait que le seuil de perception  $S$  et la durée de la période historique  $n$  sont connus. De la même manière que décrit à la section précédente pour le seuil, la durée (et donc l'année qui marque le début) de la période historique peut être complexe à déterminer. Nous proposons ici de considérer la durée de la période historique comme étant un paramètre à part entière du modèle. Ici, le seuil de perception est en revanche supposé parfaitement connu. La vraisemblance s'écrit alors :

$$L(\boldsymbol{\theta}, n; \mathbf{q}, k) = \prod_{t=1}^j f(q_t; \boldsymbol{\theta}) \left\{ \binom{n}{k} F(S; \boldsymbol{\theta})^{n-k} [1 - F(S; \boldsymbol{\theta})]^k \right\} \quad (6)$$

On peut alors exprimer une nouvelle distribution a posteriori :

$$p(\boldsymbol{\theta}, n | \mathbf{q}, k) \propto L(\boldsymbol{\theta}, n; \mathbf{q}, k) p(\boldsymbol{\theta}, n) \quad (7)$$

Ainsi, la méconnaissance de la durée de la période historique est prise en compte par le modèle et est reportée dans l'incertitude des résultats. Ce modèle sera nommé "modèle C" dans les sections suivantes.

## 2.5 Incertitudes autour du seuil de perception et de la durée de la période historique

Enfin, on peut également considérer un modèle qui représente en même temps la méconnaissance du seuil de perception  $S$  et de la durée de la période historique  $n$ . Sa vraisemblance s'écrit :

$$L(\boldsymbol{\theta}, S, n; \mathbf{q}, k) = \prod_{t=1}^j f(q_t; \boldsymbol{\theta}) \left\{ \binom{n}{k} F(S; \boldsymbol{\theta})^{n-k} [1 - F(S; \boldsymbol{\theta})]^k \right\} \quad (8)$$

La distribution a posteriori de ce modèle s'écrit :

$$p(\boldsymbol{\theta}, S, n | \mathbf{q}, k) \propto L(\boldsymbol{\theta}, S, n; \mathbf{q}, k) p(\boldsymbol{\theta}, S, n) \quad (9)$$

Il sera nommé "modèle D" dans les sections suivantes.

## 2.6 Distribution empirique d'un échantillon mixte

Le classement en fréquence des crues dans le cas d'un échantillon mixte peut poser problème, notamment quand le débit des crues historiques n'est pas connu, n'ayant d'autre information que le dépassement d'un seuil de perception. Hirsch (1987) propose une méthode pour le classement en fréquence d'un échantillon mixte quand le débit des crues est connu. Pour un échantillon continu de crues classé par valeurs décroissantes :  $q(1) \geq \dots \geq q(j)$ , la fréquence empirique au dépassement s'écrit :  $f_i = P(Q > q(i)) = \frac{i-\alpha}{j+1-2\alpha}$ . Nous prenons ici  $\alpha = 0.5$  (REF HAZEN).

Pour un échantillon mixte composé d'un échantillon continu de  $j$  débits maximum annuels et de  $k$  crues historiques supérieures à un seuil  $S$  couvrant  $n$  années, il faut raisonner sous la forme de deux sous-échantillons. Le nombre de crues supérieures au seuil  $S$  sur la période complète est ici noté  $NS$ , et la durée de la période complète correspond à  $j + n$  années. On a d'après Hirsch (1987) :

$$P(Q > q(i)) = \begin{cases} \frac{i-\alpha}{NS+1-2\alpha} \frac{NS}{j+n}, & i = 1, \dots, NS \\ \frac{NS}{j+n} + \frac{j+n-NS}{j+n} \frac{(i-NS-\alpha)}{(j-NS+1-2\alpha)}, & i = NS + 1, \dots, j + k \end{cases} \quad (10)$$

Lorsque le débit des crues historiques est connu, on peut classer l'ensemble des crues sup-seuil (de la période continue et historique) par ordre décroissant en leur attribuant le rang  $i$ . Lorsque le débit des crues historiques n'est pas connu, il n'est pas possible de classer l'échantillon de crues sup-seuil. Une manière de contourner ce problème est de tirer aléatoirement le rang  $i$  de l'ensemble des crues sup-seuil. Ce classement n'est pas réaliste mais permet d'affecter une fréquence empirique aux crues. Ainsi, on peut comparer la fréquence empirique des observations de crues aux ajustements statistiques décrits dans les sections précédentes afin de vérifier leur cohérence.

## 3 Données disponibles

### 3.1 Échantillon mixte de crues du Rhône à Beaucaire

Les données de crues du Rhône à Beaucaire sont constituées de deux échantillons. Premièrement, un échantillon continu de débits maximum annuels mesurés de 1816 à 2020. Ces débits ont été estimés au Chapitre 1 (REF) et leur incertitude hydrométrique est représentée par 500 réalisations de l'échantillon. Deuxièmement, une collection de témoignages de crues historiques de 1500 à 1816, classées en deux catégories tel que décrit au chapitre 2 (REF). Les seuils de perception correspondant aux catégories ne sont pas précisément connus, mais on suppose que le seuil  $S3$  qui correspond aux crues des catégories C3 et C4 se situe aux alentours de  $7000 \text{ m}^3/\text{s}$ , et le seuil  $S4$  qui correspond aux crues de la catégorie C4 uniquement se situe aux alentours de  $9000 \text{ m}^3/\text{s}$ . La figure 1 présente l'ensemble des données disponibles.

### 3.2 Homogénéité des données

L'homogénéité des données est un pré-requis essentiel à l'analyse fréquentielle des crues en contexte stationnaire, car cette dernière repose sur l'hypothèse que les variables étudiées sont dites *iid* (indépendantes et identiquement distribuées). C'est à dire que les processus statistiques pouvant être utilisés pour modéliser la distribution des crues ne changent pas dans le temps, et que pour une station donnée, une même distribution peut être utilisée pour modéliser les crues du XIX<sup>ème</sup> et du XXI<sup>ème</sup> siècle. Étant donné que deux types d'échantillons sont utilisés, deux types de tests statistiques sont utilisés pour étudier l'homogénéité de ces échantillons.

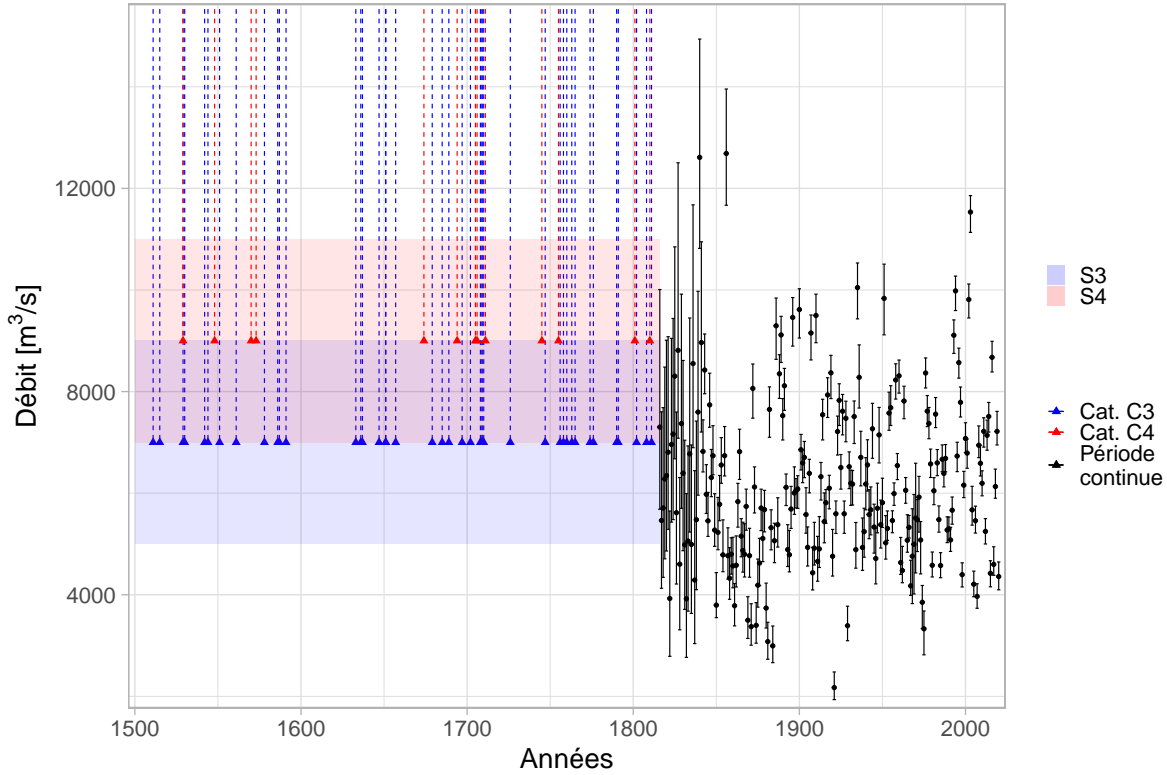


Figure 1: Échantillon de crues du Rhône à Beaucaire. L'incertitude autour des seuils de perception est représentée par les bandeaux bleus et rouge ("S3" et "S4")

### 3.2.1 Données continues

Deux tests seront utilisés pour qualifier l'homogénéité de l'échantillon de données continues : le test de Pettitt (REF) et le test de Mann-Kendall (REF). Ils permettent de détecter deux formes de non-homogénéités de la distribution séries temporelles : les ruptures et les tendances. Les ruptures sont des changements soudains (i.e. les données ont une distribution différente avant et après un instant  $t$ ), tandis les tendances représentent des changements progressifs dans la distribution des données au cours du temps. L'incertitude autour des débits maximum annuels à Beaucaire a été déterminée au Chapitre 1 (REF) et est représentée par 500 réalisations possibles de la série temporelle. Cette incertitude étant propagée aux estimations des quantiles de crue, il est nécessaire d'appliquer les tests à l'ensemble des réalisations.

Au risque d'erreur 5%, le test de Pettitt a conclu à l'existence d'une rupture pour 60% des 500 réalisations testés. ET DONC ?

L'application du test de Mann-Kendall a permis de conclure que seulement 20% des 500 réalisations de débits maximum annuels comportent une tendance temporelle. ET DONC ?

L'échantillon de données continues peut être considéré homogène suite aux tests statistiques réalisés.

### 3.2.2 Données historiques

Les données pré-enregistrements continus sont ici utilisés comme étant des occurrences de crues supposées supérieures à un seuil. Comme décrit par Lang et al. (1999), la fréquence des occurrences de crues sup-seuil est supposée suivre un processus de Poisson. Afin de vérifier l'homogénéité des occurrences sup-seuil, il est possible de calculer un intervalle de confiance autour du nombre cumulé de crues découlant du processus de Poisson. Si les occurrences de crues cumulées "sortent" de cet intervalle de confiance, alors leur fréquence d'occurrence est supposée non-stationnaire.

Ces intervalles de confiance ont été calculés pour l'échantillon de crues pré-enregistrements continus du Rhône à Beaucaire. Le début de la période historique est supposé débuter en 1500 et se termine à l'année des premiers enregistrements continus de hauteur d'eau, en 1816. Les deux échantillons testés ici reflètent deux seuils de perception,  $S3$  et  $S4$ , et on a  $S3 < S4$ .

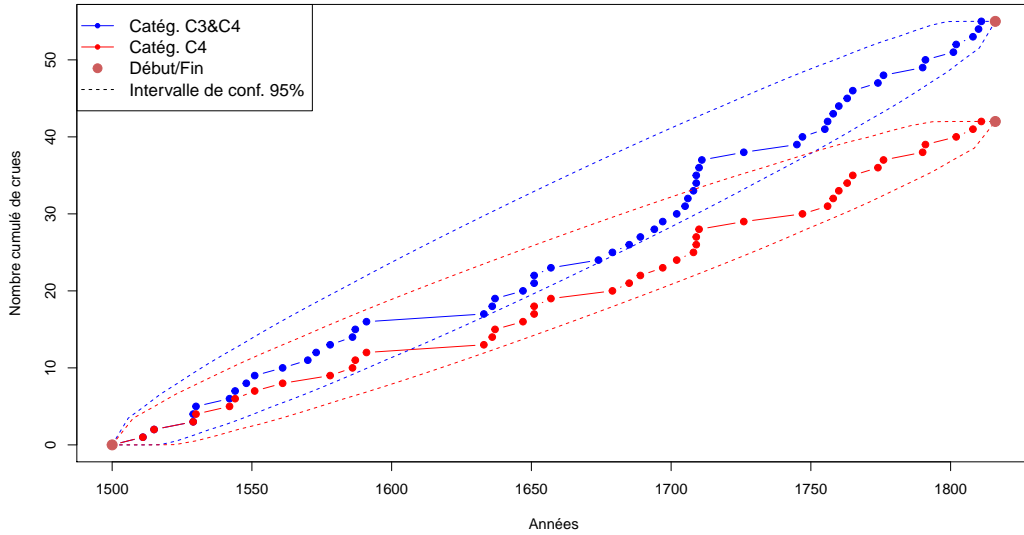


Figure 2: Nombre de crues cumulé et intervalles de confiance à 95% du processus de Poisson, pour deux échantillons d'occurrences de crues sup-seuil à Beaucaire (1500-1816)

Sur la figure 2, on remarque que les nombre cumulé de crues des deux échantillons ne sortent pas des intervalles de confiance à 95%, ils sont donc tous les deux homogènes. L'échantillon correspondant au seuil  $S3$  (en bleu) se rapproche de la borne inférieure de l'intervalle de confiance au XVII<sup>ème</sup> siècle, mais revient rapidement dans des valeurs moyennes à la faveur de nombreuses crues sup-seuil au début du XVIII<sup>ème</sup> siècle.

L'échantillon continu de débits maximum annuels (1816-2020) sera par la suite artificiellement "dégradé" pour reproduire des durées de chroniques plus usuelles. Ainsi, les crues dont le débit maxpost est supérieur au seuil considéré sont retenues dans l'échantillon. Cette période "dégradée" commence au début de la chronique, en 1816, et se termine en 1970, à la mise en fonctionnement de la station de Beaucaire Restitution. Deux seuils de perception similaires aux seuils  $S3$  et  $S4$  sont ici étudiés : 7000 et 9000  $m^3/s$ .

L'homogénéité de ces deux échantillons "artificiels" est testée dans la figure 3. Les deux échantillons de crues cumulés ne sortent pas des intervalles confiance à 95%, ils sont donc tous deux homogènes.

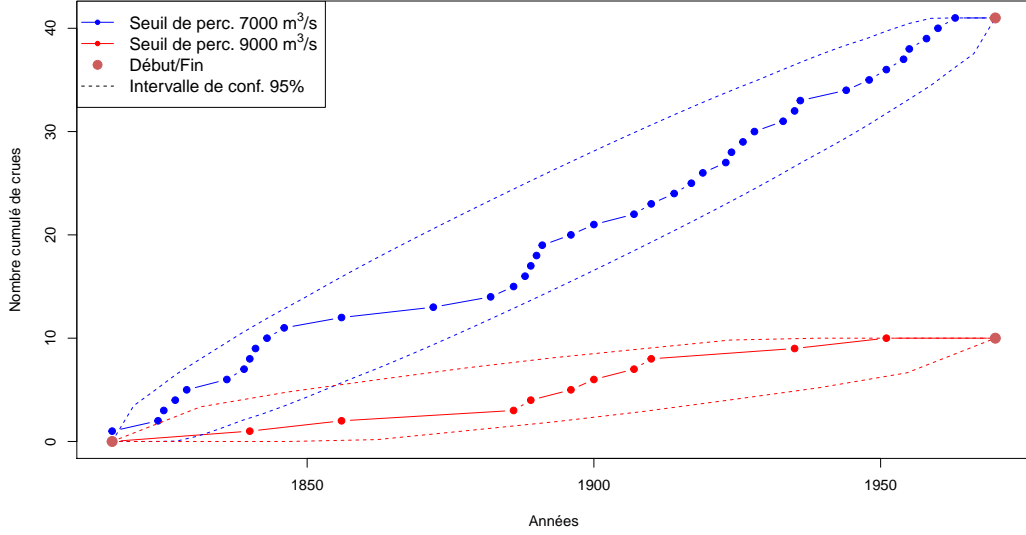


Figure 3: Nombre de crues cumulé et intervalles de confiance à 95% du processus de Poisson, pour deux échantillons d'occurrences de crues sup-seuil à Pont de Beaucaire (1816-1969)

## 4 Application aux crues du Rhône à Beaucaire

Les 4 modèles décrits dans les sections précédentes (A, B, C et D) sont appliqués aux 4 échantillons de crues du Rhône à Beaucaire présentés dans le tableau 1. Les échantillons 1 et 2 représentent la combinaison des données pré-1816 décrites dans le chapitre 2 REF avec les données continues 1816-2020 estimées au chapitre 1 (REF). Les échantillons 3 et 4 sont basés sur les débits estimés au chapitre 1(REF) qui ont été dégradés pour créer artificiellement un échantillon mixte de données continues (1970-2020) et ponctuelles (1816-1969). Il s'agit ici de tailles d'échantillon plus usuelles et dont le seuil de perception et la durée de la période historique sont parfaitement connus, contrairement aux échantillons 1 et 2. Ainsi, pour les modèles faisant l'hypothèse que le seuil de perception et/ou la durée de la période historique sont inconnus, on jugera notamment la capacité du modèle à converger vers des valeurs acceptables.

n°	Échantillon	Période historique	Nb. sup-seuil histo.	Période continue	Nb. sup-seuil cont.	Seuil de perc. [ $m^3/s$ ]	A priori seuil [ $m^3/s$ ]	Date de début histo.	A priori date de début histo.
3	1816-2020 dégradé S3	1816-1969	41	1970-2020	16	7000	$\mathcal{N}(7000, 2000)$	1816	$\mathcal{U}(1816, 1316)$
4	1816-2020 dégradé S4	1816-1969	10	1970-2020	4	9000	$\mathcal{N}(9000, 2000)$	1816	$\mathcal{U}(1840, 1340)$
1	1500-2020 S3	1500-1815	55	1816-2020	57	7000	$\mathcal{N}(7000, 2000)$	1500	$\mathcal{U}(1511, 1111)$
2	1500-2020 S4	1500-1815	13	1816-2020	14	9000	$\mathcal{N}(9000, 2000)$	1500	$\mathcal{U}(1529, 1129)$

Table 1: Caractéristiques des échantillons de crues du Rhône à Beaucaire. REVOIR L'ORDRE DES LIGNES

Pour les modèles qui font l'hypothèse d'un seuil de perception connu (A et C), le seuil retenu est



présenté dans le tableau 1. Il en est de même pour les modèles qui font l’hypothèse d’une durée de période historique connue (B et D). Ici, par souci de clarté, ce n’est pas la durée de la période historique  $n$  qui sera examinée mais la date de début de la période historique. Lorsque cette date est supposée méconnue, l’a priori débute à la date de la première crue de l’échantillon.

#### 4.1 Résultats pour la période récente dégradée (1816-2020)

Les 4 modèles décrits dans la section (REF) ont été appliqués à l’échantillon 4 du tableau 1. Les estimations pour les crues centennales et millénales sont présentés dans la figure 4, dans laquelle les 4 modèles GEV-Binomiale sont comparés au modèles GEV (chapitre 1 (REF)) pour la chronique continue totale (1816-2020) et la chronique continue dégradée (1970-2020).

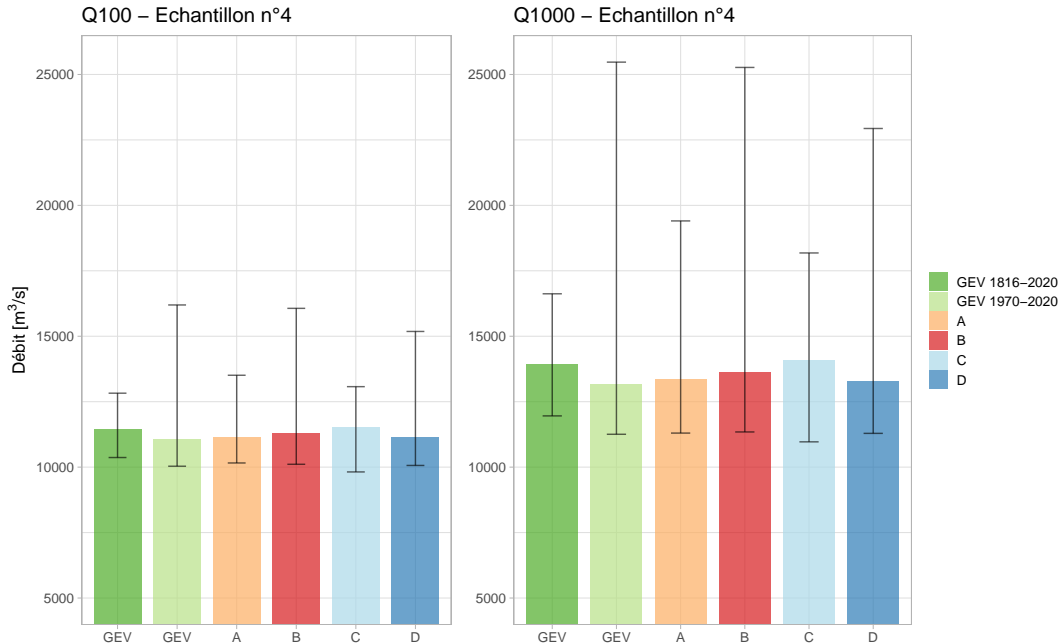


Figure 4: Estimations maxpost et incertitudes à 95% pour Q100 et Q1000 par 6 modèles différents pour l’échantillon 4 (1816-2020 dégradé, S4)

On observe tout d’abord dans la figure 4 que l’incertitude du modèle GEV pour la période 1970-2020 est la plus importante de tous les modèles (VALEUR EN % ?). Cinquante années de données continues ne permettent pas d’obtenir une précision acceptable. Cela souligne l’intérêt de valoriser les données historiques dans une situation de ce type. Globalement, les estimations maxpost de l’ensemble des modèles sont proches. En revanche, les enveloppes d’incertitude sont très variables d’un modèle à l’autre, tout particulièrement pour les estimations millénales.

La figure 5 présente l’ensemble des quantiles jusqu’à la crue décamillénale pour les 6 modèles. Ici, le rang des crues sup-seuil des deux périodes a été tiré aléatoirement comme décrit dans la section (REF). Les estimations sont cohérentes avec les observations pour les 6 modèles. On remarque que l’incertitude croît rapidement au delà la crue centennale pour les modèles B, D et GEV 1970-2020.

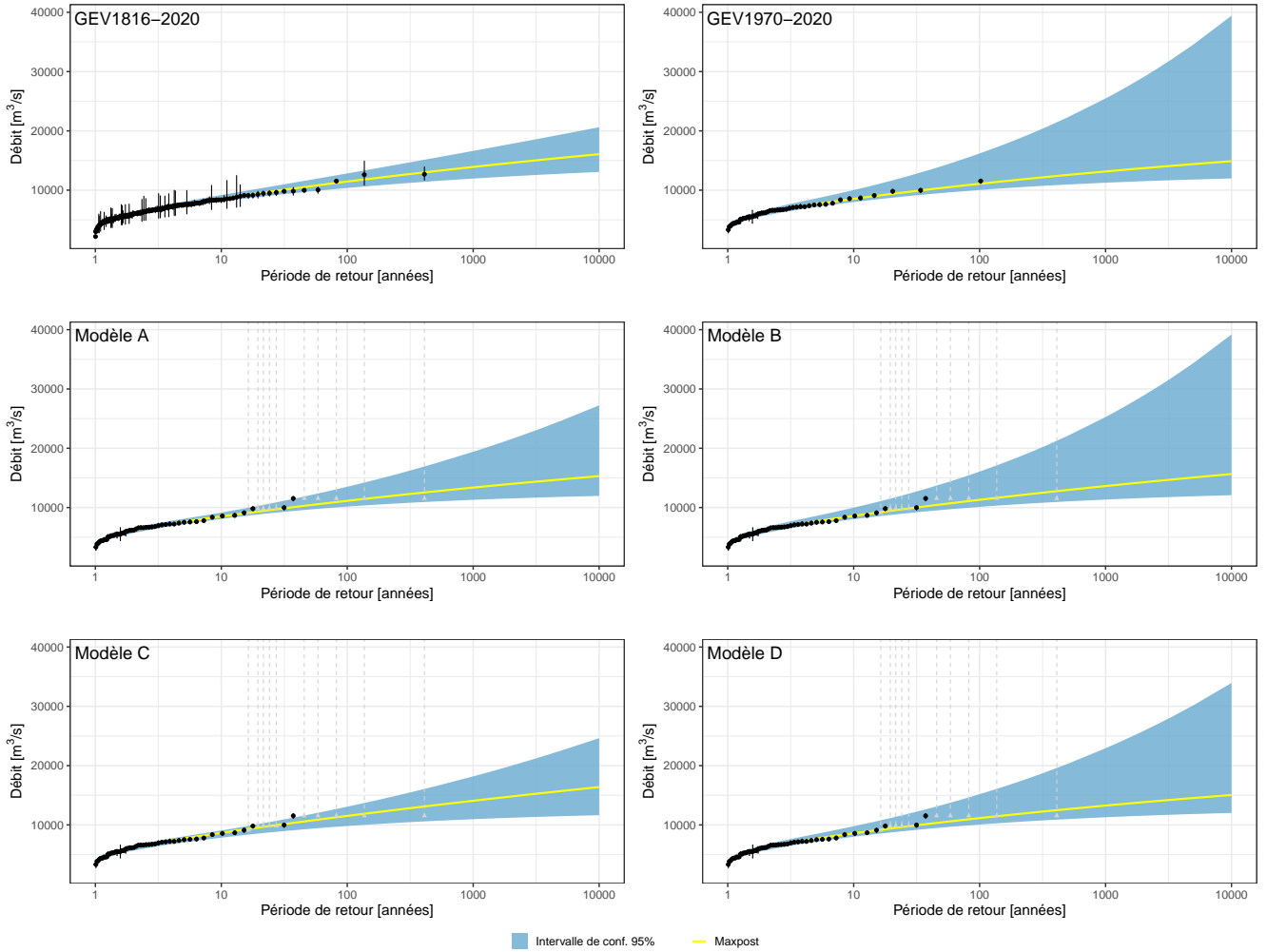


Figure 5: Quantiles estimés par 6 modèles pour l'échantillon 1 (1816-2020 dégradé,  $S_4$ ). Les observations sont en noir pour la période continue et en gris pour la période historique.

#### 4.1.1 Quel est l'apport de l'utilisation des crues historiques pour une longueur de chronique "courante" ?

Une durée de chronique continue trop courte devant le quantile visé entraîne des résultats très incertains (GEV 1970-2020 en vert clair sur la figure 4). Ici, la durée de la chronique continue de 50 ans est très petite devant la période de retour visée de 100 ou 1000 ans. Si on se trouve dans l'impossibilité de reconstituer des débits en continu au delà de 50 ans, on remarque que l'utilisation d'occurrences de crues historiques permet de réduire l'incertitude (modèle A en orange). Évidemment, l'utilisation de témoignages de crues ne permet pas d'atteindre la précision obtenue avec 200 ans de chronique continue (GEV 1816-2020 en vert foncé), mais l'incertitude obtenue s'en rapproche. Une part de cette incertitude pour l'ensemble des modèles provient notamment de l'estimation du paramètre de forme, qui gouverne le comportement de la queue de distribution. On retrouve les valeurs a posteriori du paramètre de forme dans la figure 6. On notera que l'ensemble des estimations sont proches de zéro et légèrement positives, on se trouve donc dans le cas "queue légère" de la distribution GEV. On remarque que l'estimation de ce paramètre est plus précise dans le cas GEV 1816-2020 que dans tous les autres cas. Dans le cas du modèle A, l'estimation

du paramètre de forme est similaire à celle du modèle GEV pour l'échantillon 1970-2020. En effet, seuls les paramètres de la distribution GEV sont estimés par le modèle A, le seuil de perception et la durée de la période historique étant supposés précisément connus, ce qui n'est pas toujours le cas.

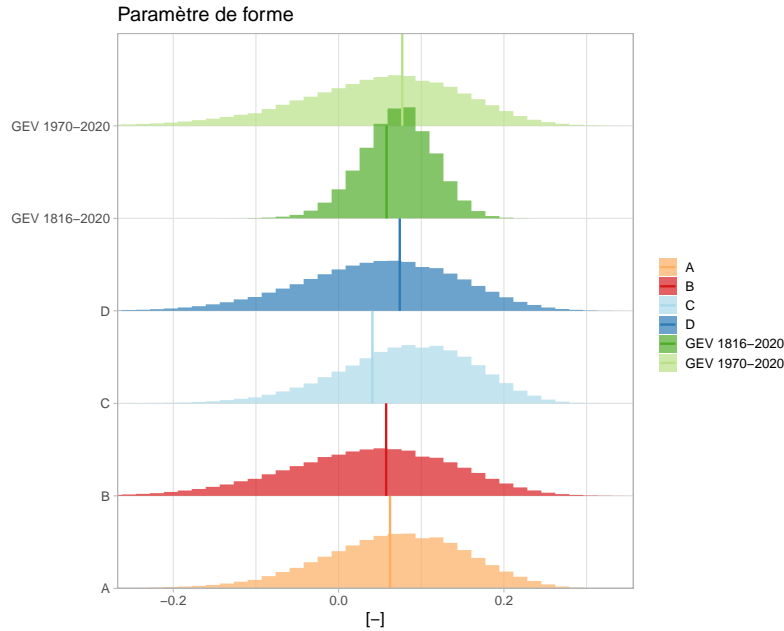


Figure 6: Distributions a posteriori du paramètre de forme pour les 4 modèles estimés sur l'échantillon 4. Les estimations des modèles GEV sont également indiquées. Les droites verticales représentent les valeurs maxpost.

#### 4.1.2 Quel est l'impact de la méconnaissance du seuil de perception sur l'estimation des quantiles ?

L'utilisation du modèle B reflète la méconnaissance du seuil de perception, celui-ci devenant alors un paramètre à part entière du modèle. Sur la figure 4, on constate que l'incertitude autour des quantiles estimés par le modèle B est bien plus importante que pour le modèle A. La méconnaissance du seuil a donc des conséquences importantes sur les estimations. La vraie valeur du seuil de perception pour l'échantillon 4 est  $S_4 = 9000 m^3/s$ . On retrouve les distributions a priori et a posteriori du seuil dans la figure 7. On remarque que l'a posteriori pour le modèle B est proche de la vraie valeur et que le modèle a effectivement permis d'améliorer la connaissance du seuil par rapport à l'a priori renseigné qui est ici très large ( $\mathcal{N}(9000, 2000)$ ). La valeur maxpost est de  $9163 m^3/s$  soit une erreur relative de 2%. Néanmoins, une telle méconnaissance de ce paramètre impacte grandement l'estimation des quantiles. Dans une situation plus réaliste, un a priori plus précis aurait pu être choisi afin de limiter cet impact.

#### 4.1.3 Quel est l'impact de la méconnaissance de la durée de la période historique sur l'estimation des quantiles ?

Le modèle C permet de représenter la méconnaissance de la durée de la période historique, qui est l'un des deux paramètres de la loi Binomiale utilisée ici pour modéliser le nombre d'occurrences de crues sup-seuil. Sur la figure 4, les estimations maxpost des quantiles pour le modèle C ont des valeurs légèrement supérieures aux estimations du modèle A. Cette légère sur-estimation provient d'une durée de période

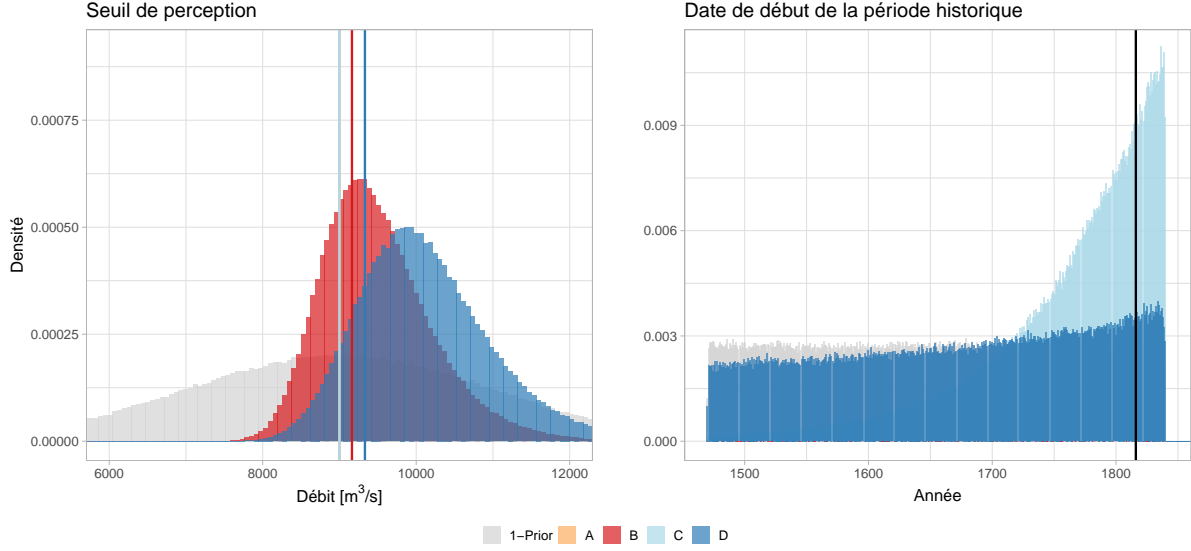


Figure 7: Distributions a priori et a posteriori pour le seuil de perception (gauche) et la date de début de la période historique (droite). Les droites verticales représentent l'estimation maxpost du paramètre de forme pour chacun des modèles. La droite verticale noire représente la vraie date de début de la période historique. Il s'agit ici de l'année 1816.

historique sous-estimée par le modèle, visible sur la figure 7 (droite). En effet, la date maxpost est l'année 1833, alors que la chronique débute réellement en 1816. Cette sur-estimation peut s'expliquer par une fréquence des crues sup-seuil plus importante au cours de la période continue (4 crues / 50 ans = 0.08 crues/an) qu'au cours de la période historique (10 crues / 153 ans = 0.065 crues/an). Ce déséquilibre reflète l'existence de phases durant lesquelles la fréquence d'occurrence des crues sup-seuil oscille malgré le fait qu'aucune non-stationnarité des données n'ait été détectée par les tests à la section (REF). L'impact de ces phases est exacerbé par des longueurs de chronique trop petites devant la durée de ces oscillations.

L'incertitude autour des quantiles estimée par le modèle C est très similaire à celle estimée par le modèle A (figure 4), de même que la distribution du paramètre de forme (figure 6), et ce malgré un a priori très peu informatif pour la date de début de la période historique : ( $\mathcal{U}(1840, 1340)$ ). Une forte méconnaissance de la durée de la période historique n'a donc que peu d'impact sur la précision de l'estimation des quantiles, contrairement à la méconnaissance du seuil de perception.

#### 4.1.4 Quel est l'impact de la méconnaissance du seuil de perception et de la durée de la période historique sur l'estimation des quantiles ?

Représenter la méconnaissance de ces deux données en même temps paraît être la solution la plus raisonnable dans le cas d'événements très anciens et pouvant être peu précisément décrits dans les témoignages. Le modèle D est ici utilisé à cet effet. Les quantiles maxpost estimés dans la figure 4 pour le modèle D semblent cohérents avec les valeurs de référence. En revanche, la largeur de l'intervalle de confiance est importante et se situe entre celle du modèle B et du modèle D. Même si l'estimation est plus précise que celle du modèle GEV sur l'échantillon 1970-2020, elle reste imprécise pour la crue millénale. L'observation des paramètres a posteriori sur la figure 7 permet de comprendre l'origine de cette large incertitude. La distribution du seuil de perception, bien que centrée à proximité de la vraie valeur (maxpost à  $9331 \text{ m}^3/\text{s}$ ), est très large. Le seuil de perception paraît ici un peu moins bien estimé

que par le modèle B. La date de début de la période historique est elle encore plus difficilement estimée. On remarque que la distribution a posteriori est pratiquement aussi large que celle de l'a priori, même si elle marque un maximum non loin de la vraie valeur (l'année 1816). La plus faible incertitude autour des quantiles pour le modèle D par rapport au modèle B provient de corrélations entre les paramètres qui peut être observée sur la figure 8. On remarque notamment une assez bonne corrélation entre la durée de la période historique et le seuil de perception, ainsi qu'entre le seuil de perception et le paramètre de forme.

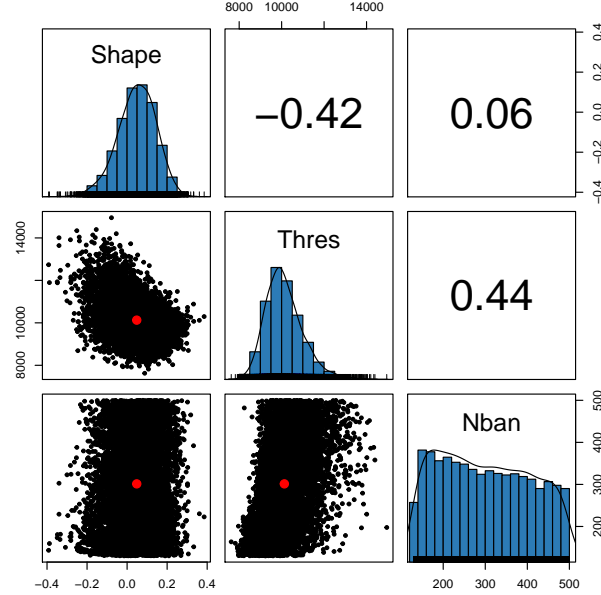


Figure 8: Scatterplot des distribution a posteriori de trois paramètres du modèle D pour l'échantillon 4. Shape correspond au paramètre de forme  $\xi$  (sans unité), Thres au seuil de perception  $S$  (en  $m^3/s$ ) et Nban à la durée de la période historique  $n$  (en années). Les nombres dans les cases supérieures correspondent aux corrélations entre les paramètres.

#### 4.1.5 Ajout d'un modèle pour lequel le débit des crues sup-seuil historiques sont connus ? A FAIRE ?

Déjà codé dans les grandes lignes mais environ une journée de travail pour l'adapter au cas de Beaucaire avec propagation des incertitudes de la période continue.

Dans le but de déterminer s'il y'a un vrai intérêt à estimer le débit des crues historique VS utiliser uniquement le nombre de dépassements du seuil.

#### 4.1.6 Discussion intermédiaire

L'utilisation des modèles décrits dans la section (REF) sur un échantillon artificiellement dégradé et dont les caractéristiques sont parfaitement connues permet d'évaluer la performance des modèles et l'impact de la méconnaissance du seuil de perception et de la durée de la période historique sur l'estimation des quantiles. On constate en observant les résultats que la méconnaissance du seuil de perception a un plus

fort impact sur l'incertitude des résultats que la méconnaissance de la durée de la période historique. Même si les distributions a posteriori du seuil de perception pour les modèles B et D sont globalement centrées sur la vraie valeur ( $9000 \text{ m}^3/\text{s}$ ), leur incertitude a un fort impact sur l'incertitude des quantiles. En revanche, l'estimation de la durée de la période historique dans le cas des modèles C et D semble elle aussi relativement peu précise, mais n'impacte que peu l'incertitude des résultats si on compare ces derniers à ceux du modèle A. Cela est dû à des corrélations entre paramètres (figure 8) qui entraînent une réduction de l'incertitude finale.

Ces premiers résultats démontrent que l'incertitude des quantiles peut être sous-estimée lorsque l'on considère seuil de perception et durée de période historique comme étant bien connus alors que ce n'est pas le cas. Les modèles utilisés ici permettent de prendre en compte cette méconnaissance dans l'estimation des quantiles extrêmes. On pourra par la suite appliquer ces modèles au cas réel de l'échantillon de crues de la période 1500-2020 à Beaucaire, dont le seuil de perception et la durée de période historique ne sont pas connus précisément.

## 4.2 Application à la période 1500-2020

### 4.2.1 BASELINE VS A pour l'échantillon 1 ?. Quel est l'apport des crues historiques pour l'analyse fréquentielle à Beaucaire ?

### 4.2.2 Modèle A pour les échantillons 1 et 2 : Quel est l'impact du choix de l'échantillon de crues historiques (S3 vs S4)

## 4.3 Discussion

## 5 Conclusion du chapitre

Conclusions sur l'intérêt des crues historique pour l'estimation des quantiles extrêmes à Beaucaire.

Est-ce qu'on observe une réelle amélioration par rapport aux résultats du chapitre 1 ?

Selon les résultats, ajouter quelque chose comme : "l'utilisation des crues historiques peut mener à faire de fortes hypothèses (seuil et durée de la période), il faut être pragmatique sur la considération des incertitudes, même si elles mènent à des estimations incertaines des quantiles

Ces conclusions sont valables uniquement à Beaucaire (période continue très longue, paramètre de forme positif, pas de tendance observée due au changement climatique ou autre). Mais que nous à appris l'application du modèle à un échantillon dégradé plus représentatif des longueurs de chronique habituelles ?

Les a priori données ici sont très peu informatifs. Des a priori plus informatifs peuvent être élicités (comment ? Courbe de tarage Beaucaire / Arles, modélisation ,...)

Perspectives sur les modèles d'analyse fréquentielle en contexte non-stationnaire et sur les modèles régionaux.

## References

- Coles, S. (2001). "Classical Extreme Value Theory and Models". In: *An Introduction to Statistical Modeling of Extreme Values*. Ed. by S. Coles. Springer Series in Statistics. London: Springer, pp. 45–73. ISBN: 978-1-4471-3675-0. DOI: [10.1007/978-1-4471-3675-0\\_3](https://doi.org/10.1007/978-1-4471-3675-0_3).
- Hirsch, R. M. (1987). "Probability plotting position formulas for flood records with historical information". In: *Journal of Hydrology* 96.1, pp. 185–199. ISSN: 00221694. DOI: [10.1016/0022-1694\(87\)90152-1](https://doi.org/10.1016/0022-1694(87)90152-1).

- Lang, M., T. Ouarda, and B. Bobée (1999). “Towards operational guidelines for over-threshold modeling”. In: *Journal of Hydrology* 225.3, pp. 103–117. ISSN: 00221694. DOI: [10.1016/S0022-1694\(99\)00167-5](https://doi.org/10.1016/S0022-1694(99)00167-5).
- Stedinger, J. R. and T. A. Cohn (1986). “Flood Frequency Analysis With Historical and Paleoflood Information”. In: *Water Resources Research* 22.5, pp. 785–793. ISSN: 1944-7973. DOI: [10.1029/WR022i005p00785](https://doi.org/10.1029/WR022i005p00785).