

# Chapitre 4 : Analyse probabiliste des crues du Rhône à Beaucaire du XVI<sup>ème</sup> siècle à aujourd'hui

Mathieu

April 10, 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>4</b>
2.1	Concepts de base et hypothèses	4
2.2	Propagation de l'incertitude hydrométrique des mesures systématiques (modèle A)	5
2.3	Seuil de perception incertain (modèle B)	6
2.4	Durée de la période historique incertaine (modèle C)	6
2.5	Seuil de perception et durée de la période historique incertaines (modèle D)	7
2.6	Débit des crues historiques compris dans un intervalle (modèle E)	7
2.7	Distribution empirique d'un échantillon mixte	8
<b>3</b>	<b>Données disponibles</b>	<b>8</b>
3.1	Échantillon mixte de crues du Rhône à Beaucaire	8
3.2	Homogénéité des données	8
3.2.1	Données continues	9
3.2.2	Données historiques	10
<b>4</b>	<b>Application aux crues du Rhône à Beaucaire</b>	<b>11</b>
4.1	Résultats pour la période récente dégradée (1816-2020)	12
4.1.1	Quel est l'apport de l'utilisation des crues historiques pour une longueur de chronique "courante" ?	14
4.1.2	Quel est l'impact de la méconnaissance du seuil de perception sur l'estimation des quantiles ?	15
4.1.3	Quel est l'impact de la méconnaissance de la durée de la période historique sur l'estimation des quantiles ?	16
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 ?	17
4.1.5	Quel est l'apport de la connaissance du débit des crues historiques ?	18
4.1.6	Discussion intermédiaire	19
4.2	Application à la période 1500-2020	20
4.2.1	Quel est l'apport des crues historiques pour l'analyse fréquentielle à Beaucaire ?	21
4.3	L'estimation des quantiles à Beaucaire (1500-2020) avec des a priori plus informatifs rend-elle l'utilisation des données historique plus pertinente ?	21

## 1 Introduction du chapitre

Un des problèmes majeurs de l'analyse fréquentielle des crues provient de la difficulté à estimer précisément les paramètres de la distribution choisie, notamment à cause de la longueur limitée des chroniques disponibles (Kjeldsen et al. (2011); Apel et al. (2004)). Cette incertitude d'échantillonnage est d'autant plus grande que la période de retour du quantile estimé est grande devant la longueur de la chronique (i.e. estimer le débit de la crue millénale en se basant sur une chronique de débits maximum annuels de 20 ans). Cette situation est fréquente, notamment en France où la longueur des chroniques de débit mesurées en continu est en moyenne de l'ordre de 50 ans (Le Coz, 2017), alors que la plus forte crue connue ou la crue centennale si celle-ci est supérieure est la référence en terme de niveau de protection dans les communes. En Angleterre, il est déconseillé d'estimer des quantiles de période de retour supérieure à la moitié de la longueur de la chronique continue utilisée (WHS, 2008). Cependant, est possible de réduire l'incertitude d'échantillonnage en élargissant le jeu de données de diverses manières.

Tout d'abord, il existe des méthodes permettant d'élargir spatialement le jeu de données sous certaines conditions, en faisant par exemple l'hypothèse que la distribution des crues est homogène au sein d'une région définie à un facteur d'échelle près (Hosking and Wallis (1997); Gaume et al. (2010); Viglione et al. (2013); Nguyen et al. (2014)) ou bien en utilisant un modèle qui prend en compte les dépendances spatiales entre stations (Kjeldsen and Jones (2009); Renard (2011); Sun et al. (2014)). Néanmoins, il semble inapproprié d'appliquer une analyse régionale à la station du Rhône à Beaucaire, la notion de région homogène ayant peu de sens ici étant donné la taille du bassin versant et les nombreuses influences hydro-climatiques qui impactent la distribution des crues.

Une autre méthode populaire consiste à élargir temporellement le jeu de données disponible en utilisant des données anciennes, pouvant être d'origine et de nature variées. Le cas idéal, décrit dans le chapitre 1 (REF), illustre la possibilité d'exhumer des enregistrements continus, antérieurs aux données disponibles dans les bases de données. Dans cette situation, l'exhaustivité des débits maximum annuels est garantie et le principal défi méthodologique est de considérer les diverses sources d'incertitudes. Plus généralement, les données historiques disponibles ne sont pas continues et peuvent prendre des formes variées et donc nécessiter des traitements statistiques différents. Il peut s'agir de témoignages (Pichard (1995); Kjeldsen et al. (2014)), de repères de crues (Parkes and Demeritt (2016); Piotte et al. (2016); Engeland et al. (2020); MEDD (n.d.)), de reconstructions de crues pré-historiques ("paleofloods" en anglais) issues de divers proxys tels que les dépôts sédimentaires ou les cernes des espèces végétales ripariennes (Stedinger and Cohn (1986); Benito et al. (2004); Dezileau et al. (2014); St. George et al. (2020); Engeland et al. (2020)). L'utilisation des données historiques pour l'analyse fréquentielle n'est pas récente, en témoignent par exemple les travaux de Benson (1950) ou de Hirsch and Stedinger (1987) qui se consacraient principalement au placement des crues sur les courbes de fréquence. Les travaux de Stedinger and Cohn (1986) ont ensuite ouvert la voie à l'inclusion de ces données dans l'estimation de la distribution des crues via l'utilisation de l'estimateur du maximum de vraisemblance et les méthodes Monte-Carlo. De nombreuses méthodes existent à ce jour pour inclure les données historiques à l'analyse fréquentielle des crues. Parmi elles, les méthodes d'inférence bayésienne permettent d'inclure des données de nature diverse via l'utilisation de la fonction de vraisemblance (Stedinger and Cohn (1986); Kuczera (1999)) et l'utilisation d'algorithmes Monte Carlo Markov Chain (MCMC) (Reis and Stedinger (2005); Renard et al. (2006)). La plupart des

études récentes qui valorisent les données historiques soulignent qu’une prise en compte complète des incertitudes est nécessaire (Neppel et al. (2010); Parkes and Demeritt (2016)).

Les données de crues historiques ne sont généralement pas continues et concernent des crues d’une magnitude suffisante pour avoir par exemple laissé une trace dans les écrits ou pour avoir mérité l’installation d’un repère de crue. Le dénominateur commun du traitement statistique de ces données de forme variée est le seuil de perception (Gerard and Karpuk (1979); Stedinger and Cohn (1986)). Il s’agit ici de faire l’hypothèse que toutes les crues ayant dépassé ce seuil de perception ont laissé une trace, ce qui garantit l’exhaustivité du recensement des crues historiques. Le corolaire à cette idée est que, pour toutes les années de la période historique sans mention de crue, on fait l’hypothèse que le débit maximum annuel a été inférieur au seuil de perception. Il est parfois possible de reconstituer le débit des crues historiques supérieures au seuil de perception, notamment via l’utilisation de modèles hydrauliques (Neppel et al. (2010), Machado et al. (2015)). Il n’est cependant pas obligatoire de connaître le débit des crues historiques. La seule connaissance du nombre de crues ayant dépassé le seuil de perception est suffisante pour exploiter les données via l’utilisation de la loi binomiale, tel que décrit par Stedinger and Cohn (1986) ou Payraastre et al. (2011). 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, le seuil de perception est dans la grande majorité des cas supposé parfaitement connu, bien que la sensibilité des résultats au choix du seuil de perception soit parfois explorée (Stedinger and Cohn (1986), Viglione et al. (2013); Macdonald et al. (2014); Payraastre et al. (2011)). Seuls les travaux de Parkes and Demeritt (2016) semblent avoir considéré une incertitude du seuil de perception, et ce uniquement dans des cas où le débit des crues historiques est connu. Néanmoins, le seuil de perception n’apparaît pas dans la formulation de la vraisemblance, et les distributions a priori utilisées ne sont pas explicitées. Il est pourtant possible d’intégrer directement l’incertitude du seuil de perception, l’incertitude des quantiles estimés étant ainsi impactée en fonction de la méconnaissance du seuil de perception.

Le concept de seuil de perception est accompagné de la définition de la durée de la période historique. La date qui marque le début de cette période historique (date à partir de laquelle le seuil de perception est actif) est complexe et même parfois impossible à déterminer. Pourtant, à l’instar du seuil de perception, cette durée est généralement considérée comme étant parfaitement connue dans la littérature. Définir le début de la période historique à la date de première crue connue est dangereux, cela peut mener à une sous-estimation de la durée de la période historique et donc à une sur-estimation des quantiles de crue. Pour pallier à ce problème, Prosdocimi (2018) propose une comparaison de méthodes pour estimer la durée de période historique lorsque celle-ci n’est pas connue. La prise en compte de ce problème au sein même du modèle probabiliste semble possible mais n’a pas été étudié dans la littérature. Cependant, il semble légitime que l’incertitude des quantiles estimés soit impactée par la méconnaissance autour de cette durée.

Même si les débits de la période continue sont généralement bien mieux connus que ceux de la période historique, il existe une incertitude autour de ces données. Cette incertitude est pourtant souvent négligée dans le cas de l’utilisation de données historiques. Seuls quelques travaux proposent de la prendre

en compte : Parkes and Demeritt (2016) considèrent l’incertitude autour de la période continue via l’utilisation d’un pourcentage d’erreur fixe, et Neppel et al. (2010) via l’utilisation de modèles d’erreur plus élaborés. Dans le cas de la station du Rhône à Beaucaire, l’incertitude des débits de la période continue a été minutieusement déterminée pour chacune des sources existantes et propagée aux débits maximum annuels (chapitre 1 (REF)). Cette prise en compte des incertitudes hydrométriques semblait ici indispensable compte tenu de la longueur exceptionnelle de la chronique et de l’utilisation de données continues particulièrement anciennes. Cette incertitude pourra être propagée aux quantiles estimés dans ce chapitre via l’utilisation de procédures Monte-Carlo, de manière similaire à la propagation réalisée au chapitre 1 (REF).

Ce chapitre présente un modèle probabiliste qui utilise le nombre de dépassements d’un seuil de perception pour l’analyse fréquentielle et qui prend en compte l’incertitude des débits de la période continue. Un des objectifs majeurs de ce chapitre est de reconnaître la nature imparfaitement connue du seuil de perception et de la durée de la période historique en en faisant des paramètres à part entière du modèle probabiliste. Cet objectif est exploré grâce à l’utilisation d’un jeu de données continues de 205 ans à Beaucaire. Ce jeu de données est artificiellement dégradé afin de se replacer dans le contexte d’un échantillon mixte, tout en connaissant parfaitement les caractéristiques de l’échantillon historique afin de pouvoir évaluer les résultats des modèles. L’apport de la connaissance du débit des crues historiques est également exploré et comparé à la seule connaissance du nombre de dépassement du seuil de perception. Ces mêmes modèles sont ensuite appliqués à un échantillon mixte de 1500 à 2020 à Beaucaire pour lequel l’impact des différentes incertitudes sur les quantiles est discuté.

Les méthodes d’analyse fréquentielle des crues historiques sont présentées dans une première partie (section 2). Les données disponibles à Beaucaire sont présentées et l’homogénéité de ces échantillons est vérifiée (section 3). Les modèles sont ensuite appliqués à un échantillon dégradé, puis à l’intégralité de l’échantillon de crues du Rhône à Beaucaire (section 4). Les résultats sont ensuite discutés en section 5.

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

### 2.1 Concepts de base et hypothèses

**BEN : A EFFACER, DOUBLON AVEC INTRO?** L’utilisation d’occurrences de crues historiques pour l’analyse fréquentielle nécessite a minima l’hypothèse suivante : toutes les crues ayant dépassé une certaine magnitude ont laissé une trace dans les archives ou les mémoires. On nommera cette magnitude : ”seuil de perception”. Cela implique que pour toutes les années de la période historique sans mention de crue, on fait l’hypothèse que le débit inconnu a été inférieur au seuil de perception. Aussi, l’échantillon des crues supérieures au seuil de perception est supposé exhaustif. Il existe des méthodes d’analyse pour lesquelles il n’est pas nécessaire de connaître précisément le débit des événements supérieurs au seuil de perception (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 (loi de Gumbel). Ainsi, on a la fonction de répartition de la GEV :

$F(x; \boldsymbol{\theta}) = e^{-(1-\xi(\frac{x-\mu}{\sigma}))^{1/\xi}}$ . Lorsque le paramètre de forme est strictement positif ( $\xi > 0$ ), on se trouve dans le cas "loi de Weibull" avec une borne supérieure et des quantiles inférieurs à ceux d'une loi de Gumbel. Dans le cas contraire ( $\xi < 0$ ), il s'agit du cas "loi de Fréchet" de la distribution GEV. 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 dans un premier temps non 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  $n - k$  années restantes. 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 (équation 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) la vraisemblance 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 (Renard et al., 2006). 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 qui maximise la distribution a posteriori est appelé maxpost et s'écrit  $\hat{\boldsymbol{\theta}}$ . Pour l'ensemble des simulations du présent chapitre, les a priori des paramètres de la distribution GEV seront les suivants : distribution plate (uniforme et très peu informative) pour  $\mu$  et  $\sigma$ , et distribution gaussienne de moyenne zéro et d'écart type 0.2 pour  $\xi$ . Cet a priori du paramètre de forme est notamment cohérent avec les suggestions de Martins and Stedinger (2000).

## 2.2 Propagation de l'incertitude hydrométrique des mesures systématiques (modèle A)

Dans la section précédente, l'incertitude 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 Seuil de perception incertain (modèle B)

Afin de considérer dans le modèle probabiliste 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. Un seul seuil de perception est ici considéré pour l'ensemble de l'échantillon et 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, ce qui n'est ici plus le cas. La vraisemblance s'écrit alors :

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

On peut alors écrire une nouvelle distribution a posteriori :

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

La distribution des paramètres a posteriori de ce modèle reflètent l'incertitude hydrométrique de la période continue, l'incertitude d'échantillonnage, ainsi que l'incertitude du seuil de perception. Ce modèle sera nommé "modèle B" dans les sections suivantes. Il est ici nécessaire de spécifier une distribution a priori du seuil de perception qui reflète la connaissance, qu'elle soit très partielle ou relativement précise, de ce paramètre.

### 2.4 Durée de la période historique incertaine (modèle C)

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. Généralement, la date qui marque la fin de la période historique est parfaitement connue, car elle correspond également au début des enregistrements continus. En revanche, la date du début de l'échantillon historique (que l'on appellera  $t^*$ ), à partir de laquelle toutes les crues supérieures au seuil de perception seront connues, correspond à une période lointaine et relativement mal connue de l'échantillon historique. Ainsi, nous proposons ici de considérer la durée de la période historique  $n$  comme étant un paramètre à part entière du modèle probabiliste. Le seuil de perception est en revanche supposé parfaitement connu dans ce cas de figure. 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 :

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

La méconnaissance de la durée de la période historique  $n$  est donc prise en compte dans le modèle et a un impact sur l'incertitude des résultats. Ce modèle sera nommé "modèle C" dans les sections suivantes.

Une distribution a priori reflétant la connaissance partielle autour de la durée de la période historique devra être élicitée.

## 2.5 Seuil de perception et durée de la période historique incertains (modèle D)

Le seuil de perception  $S$  et la durée de la période historique  $n$  étant par définition reliés (un seuil de perception étant valide sur une durée donnée) on peut construire un modèle qui représente en même temps la méconnaissance autour de ces deux paramètres. La vraisemblance de ce modèle s'écrit :

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

On peut exprimer la vraisemblance du modèle telle que :

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

Ce modèle pour lequel  $S$  et  $n$  sont incertains sera nommé "modèle D" dans les sections suivantes.

## 2.6 Débit des crues historiques compris dans un intervalle (modèle E)

Dans certains cas, le débit des crues historiques supérieures au seuil de perception est connu. De façon similaire au modèle binomial décrit précédemment, on peut faire l'hypothèse que le débit maximum annuel de toutes les années de la période historique sans mention de crues est inférieur au seuil de perception. Le débit des crues historiques peut ensuite être pris en compte dans le modèle probabiliste décrit par Stedinger and Cohn (1986). Il est également possible de considérer que le débit des crues historiques n'est pas parfaitement connu, mais qu'il est compris dans un intervalle de confiance. Plusieurs exemples d'un tel modèle existent dans la littérature (par exemple : Payraastre et al. (2011) ou Parkes and Demeritt (2016)). La vraisemblance d'un tel modèle peut s'écrire :

$$L(\boldsymbol{\theta}; \mathbf{q}, \mathbf{y}) = \prod_{t=1}^j f(q_t; \boldsymbol{\theta}) \prod_{i=1}^k \left[ F(y_i^{sup}, \boldsymbol{\theta}) - F(y_i^{inf}, \boldsymbol{\theta}) \right] F(S; \boldsymbol{\theta})^{n-k} \quad (10)$$

où  $q_t$  correspond aux  $j$  crues de la période continue et  $y_i$  aux  $k$  crues de la période historique dont le débit est compris dans l'intervalle d'incertitude à 95%  $[y_i^{inf}; y_i^{sup}]$ . La distribution a posteriori du modèle s'écrit alors :

$$p(\boldsymbol{\theta} \mid \mathbf{q}, \mathbf{y}) \propto L(\boldsymbol{\theta}; \mathbf{q}, \mathbf{y}) p(\boldsymbol{\theta}) \quad (11)$$

Ici, seuil de perception et durée de la période historique sont supposés parfaitement connus. Ce modèle sera nommé "modèle E" dans les sections suivantes. Ici aussi, l'incertitude des débits de la période continue est propagée tel que décrit dans la section 2. On pourra comparer les quantiles obtenus avec le modèle E pour lequel le débit des crues historiques est connu (dans un intervalle), avec les résultats des modèles binomiaux pour lesquels seul le nombre de dépassements  $k$  du seuil de perception est connu.



## 2.7 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 :  $f_i = \frac{i-\alpha}{j+1-2\alpha}$ . Nous prenons ici  $\alpha = 0.5$  (Hazen, 1914). 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  $N_S$ , et la durée de la période complète correspond à  $j + n$  années. On a d'après Hirsch (1987) :

$$f_i = \begin{cases} \frac{i - \alpha}{N_S + 1 - 2\alpha} \frac{N_S}{j + n}, & i = 1, \dots, N_S \\ \frac{N_S}{j + n} + \frac{j + n - N_S}{j + n} \frac{(i - N_S - \alpha)}{(j - N_S + 1 - 2\alpha)}, & i = N_S + 1, \dots, j + k \end{cases} \quad (12)$$

Lorsque le débit des crues historiques est connu, on peut classer l'ensemble des crues supérieures au 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 cet échantillon. Une manière de contourner ce problème est de tirer aléatoirement le rang  $i$  de l'ensemble des crues supérieures au seuil dans l'intervalle entier  $[1; N_S]$ . Ce classement est aléatoire, mais il 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

L'échantillon de crues du Rhône à Beaucaire est constitué de deux types de données : 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, tirés de la base HISTRHÔNE et classées en deux catégories tel que décrit au chapitre 2 (REF). Les seuils de perception correspondant aux deux échantillons 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  $9000 \text{ m}^3/\text{s}$  (VALEURS DEFINIES AU CHAP 2). 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 *iid* (indépendantes et identiquement distribuées). C'est-à-dire que la distribution des crues ne change pas dans le temps et qu'une même distribution peut être utilisée pour modéliser les crues du XVI<sup>ème</sup> et du XXI<sup>ème</sup> siècle. Étant donné que deux types d'échantillons sont ici utilisés, deux types de tests statistiques sont appliqués dans les sections suivantes pour étudier l'homogénéité de ces données.



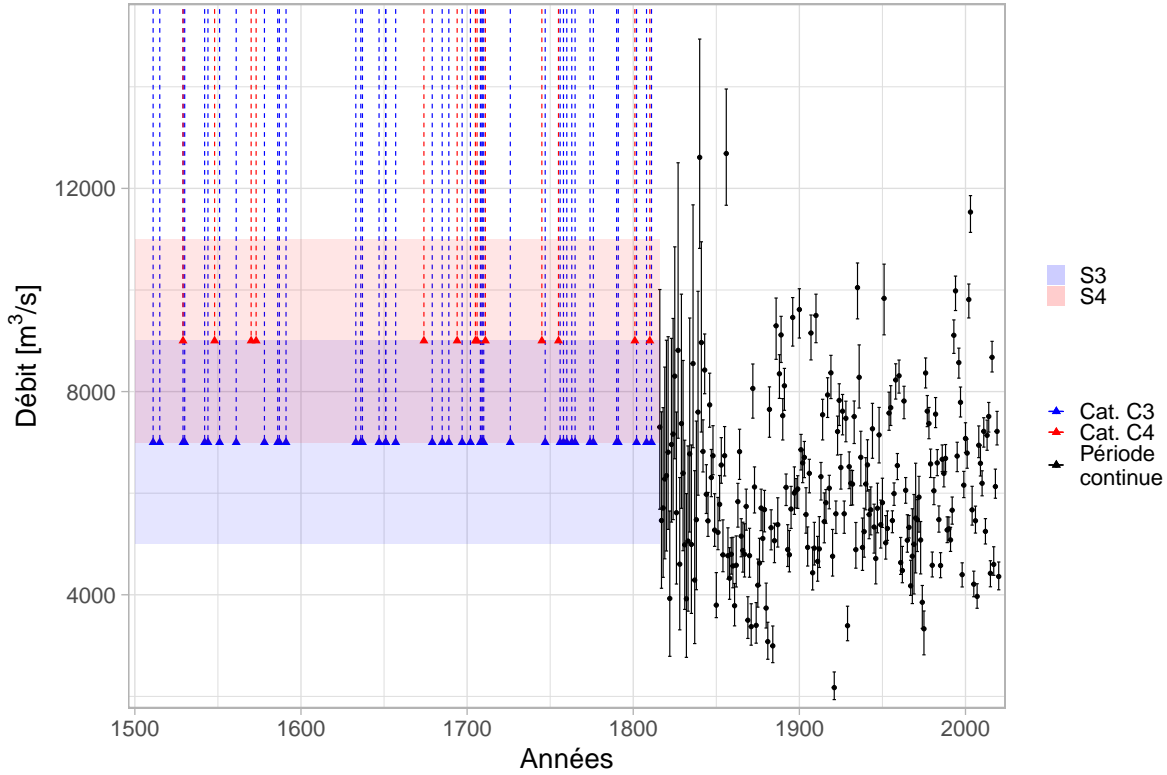


Figure 1: Échantillon de crues du Rhône à Beaucaire. L'incertitude à 95% autour des seuils de perception est représentée par les bandeaux bleu et rouge ("S3" et "S4")

### 3.2.1 Données continues

Trois tests seront utilisés pour qualifier l'homogénéité de l'échantillon de données continues : le test de Pettitt (1979) et la procédure de segmentation développée par Darienzo (2021) qui permettent de détecter des ruptures dans les séries temporelles, ainsi que le test de Mann-Kendall (Mann (1945); Kendall (1948)) qui permet de détecter l'existence de tendances. Les ruptures sont des changements soudains (i.e. les données ont une distribution différente avant et après un instant  $t$ , par exemple suite à un changement d'instrumentation), tandis que les tendances représentent des changements progressifs dans la distribution des données au cours du temps (par exemple : un changement progressif des conditions d'écoulement du bassin versant). Parmi ces trois tests, seule la procédure de segmentation de Darienzo (2021) permet de considérer l'incertitude des données d'entrée (déterminée au Chapitre 1 (REF)).

La p-value des tests de Pettitt et Mann-Kendall appliqués à la série maxpost des débits maximum annuels à Beaucaire est respectivement de 0.15 et 0.41. Au risque d'erreur 5%, on peut conclure qu'il n'existe pas de tendance ou de rupture dans la série. Il faut cependant s'assurer que ce résultat est toujours vrai lorsque l'on considère les incertitudes de la série.

L'application de la procédure de segmentation de Darienzo (2021) à la série de débits maximum annuels avec incertitude a conclu que le nombre optimal de segments pour la chronique de Beaucaire était de 1, et ce quel que soit le critère de segmentation considéré (AIC, BIC, HQC ou DIC). On peut ainsi conclure qu'aucune rupture n'existe dans les données. L'échantillon de données continues peut être considéré

homogène suite aux tests statistiques réalisés.

### 3.2.2 Données historiques

Les données pré-enregistrements continus (ou historiques) utilisées ici prennent la forme d'occurrences de crues supposées supérieures à un seuil de perception. Comme décrit par Lang et al. (1999), la fréquence des occurrences de crues supérieures à un seuil est supposée suivre un processus de Poisson. Afin de vérifier l'homogénéité des occurrences de crues, 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. La période historique est supposée 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  $S3 < S4$ .

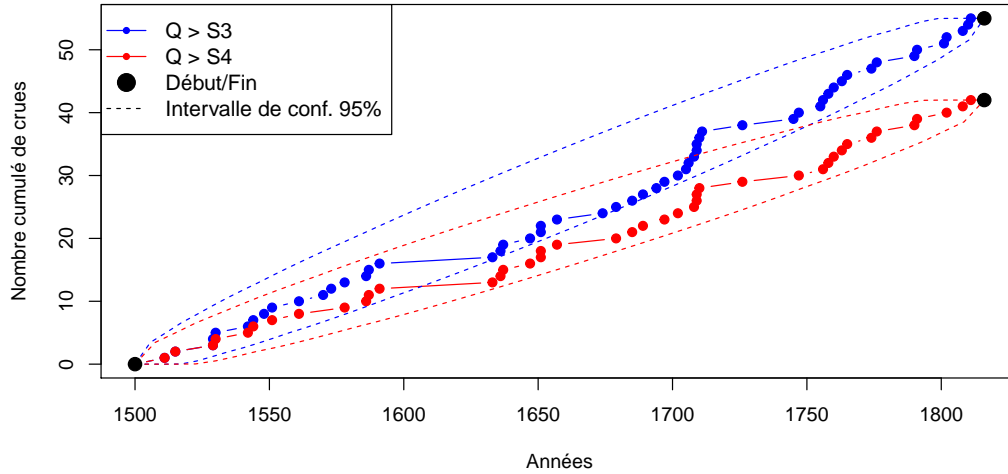


Figure 2: Nombre de crues cumulé et intervalles de confiance à 95% du processus de Poisson, pour deux échantillons d'occurrences de crues supérieures aux seuils  $S3$  (catégories C3 et C4) ou  $S4$  (catégorie C4) à Beaucaire (1500-1816)

Sur la figure 2, on remarque que les nombres cumulé de crues des deux échantillons sont compris dans les intervalles de confiance à 95% des processus de Poisson, ils peuvent donc être tous deux considérés 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 supérieures au 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  $9000 \text{ m}^3/\text{s}$ . L'homogénéité de ces deux échantillons "dégradés" est testée dans la figure 3. Les deux échantillons de crues cumulés sont compris dans les intervalles de confiance à 95%, ils sont donc tous deux homogènes.

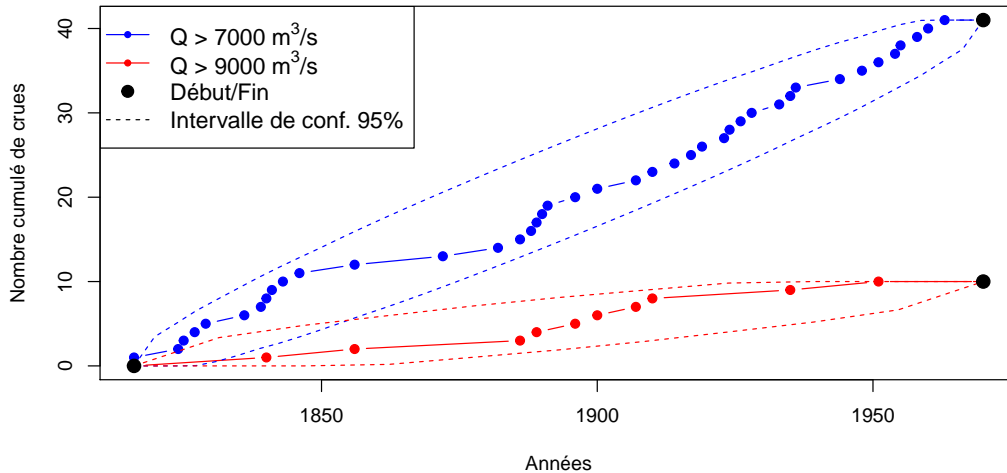


Figure 3: Nombre de crues cumulé et intervalles de confiance à 95% du processus de Poisson, pour deux échantillons d'occurrences de crues supérieures aux seuils  $S3$  ou  $S4$  à Pont de Beaucaire (1816-1969)

Les échantillons de crues des périodes systématique et historique ont indépendamment été jugés homogènes. Il faut cependant garder à l'esprit que ces échantillons peuvent ne pas être homogènes entre eux pour diverses raisons. Par exemple, le seuil de perception utilisé pour l'échantillon systématique ( $S3$  ou  $S4$ ) peut ne pas être cohérent avec le seuil de perception de l'échantillon historique dont la valeur exacte est inconnue.

## 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. Le modèle A fait l'hypothèse que le seuil de perception  $S$  et la durée de la période historique  $n$  sont parfaitement connus, tandis que dans le cas du modèle D, ces deux valeurs sont considérées incertaines. Le modèle B fait l'hypothèse que seul le seuil de perception  $S$  est incertain, alors que pour le modèle C, seule la durée de la période historique  $n$  est incertaine. Le modèle E suppose quant à lui que le débit des crues historique est connu et est contenu dans un intervalle d'incertitude,  $S$  et  $n$  sont alors supposés parfaitement connus. Pour les cinq modèles, l'incertitude des débits de la période continue est propagée.

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 estimer des valeurs acceptables.

Table 1: Caractéristiques des échantillons de crues du Rhône à Beaucaire.  $S$  désigne le seuil de perception et  $t^*$  la date de début de la période historique.

n°	Période historique	Période continue	Seuil $S$ [m <sup>3</sup> /s]	Nb. de crues $> S$		A priori $S$ [m <sup>3</sup> /s]	A priori $t^*$
				per. hist.	per. cont.		
1	1500-1815	1816-2020	7000	55	57	$\mathcal{N}(7000, 2000)$	$\mathcal{U}(1111, 1511)$
2	1500-1815	1816-2020	9000	13	14	$\mathcal{N}(9000, 2000)$	$\mathcal{U}(1129, 1529)$
3	1816-1969	1970-2020	7000	41	16	$\mathcal{N}(7000, 2000)$	$\mathcal{U}(1316, 1816)$
4	1816-1969	1970-2020	9000	10	4	$\mathcal{N}(9000, 2000)$	$\mathcal{U}(1340, 1840)$

Les modèles B, C et D font l’hypothèse que  $S$  et/ou  $n$  sont inconnus, il faudra donc affecter une distribution a priori à ces paramètres. Ces distributions sont présentées dans le tableau 1 pour chacun des échantillons. Le but étant ici d’explorer les performances des modèles, les a priori seront très peu informatifs. L’a priori du seuil de perception  $S$  (modèles B et D) est supposé Gaussien, avec pour moyenne la valeur connue ou supposée du seuil (soit  $S3$  ou  $S4$ ) et pour écart type 2000 m<sup>3</sup>/s. Par souci de clarté, on ne parlera pas ici de la durée de la période historique  $n$  mais de la date de début de la période historique  $t^*$  (la date de fin de la période historique étant ici parfaitement connue pour les 4 échantillons). La distribution a priori de  $t^*$  (modèles C et D) est supposée uniforme avec pour borne supérieure la date de la première crue de l’échantillon historique considéré, appelée  $t_{k=1}$ . Par définition, la période historique débute au plus tard à la date de cette première crue. La borne inférieure de la distribution uniforme sera fixée 400 ans avant la date de la première crue  $t_{k=1}$  afin de représenter la méconnaissance de  $t^*$ .

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

Les 4 modèles décrits dans la section 2 ont été appliqués à l’échantillon 4 du tableau 1. Les estimations pour les crues centennales et millénales sont présentées dans la figure 4, dans laquelle les 4 modèles GEV-Binomiale sont comparés au modèle GEV (chapitre 1 (REF)) appliqué successivement à la chronique continue totale (1816-2020) et à la chronique continue dégradée (1970-2020).

Table 2: Résultats maxpost et incertitudes des 6 modèles pour l’échantillon 4. Q100 et Q1000 représentent respectivement le débit des crues centennales et millénales,  $\xi$  le paramètre de forme de la distribution GEV,  $S$  le seuil de perception et  $t^*$  la date de début de la période historique. Les écarts type des distributions a posteriori sont représentés par les colonnes débutant par la lettre ”u”.

Modèle	Q100 [m <sup>3</sup> /s]	uQ100 [m <sup>3</sup> /s]	Q1000 [m <sup>3</sup> /s]	uQ1000 [m <sup>3</sup> /s]	$\xi$	$u\xi$	$S$ [m <sup>3</sup> /s]	$uS$ [m <sup>3</sup> /s]	$t^*$	$ut^*$
GEV 1816-2020	11451	687	13919	1351	0.058	0.044	X	X	X	X
GEV 1970-2020	11076	2560	13154	6159	0.077	0.102	X	X	X	X
A	11132	1189	13367	3019	0.062	0.088	9000	X	1816	X
B	11302	2381	13622	5823	0.058	0.102	9163	729	1816	X
C	11517	779	14069	2057	0.041	0.083	9000	X	1833	71
D	11147	2018	13262	4837	0.074	0.096	9332	883	1785	107

**PARAGRAPHE A DANS LES SECTIONS SUIVANTES EN 4.1.X ?** 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. On observe que

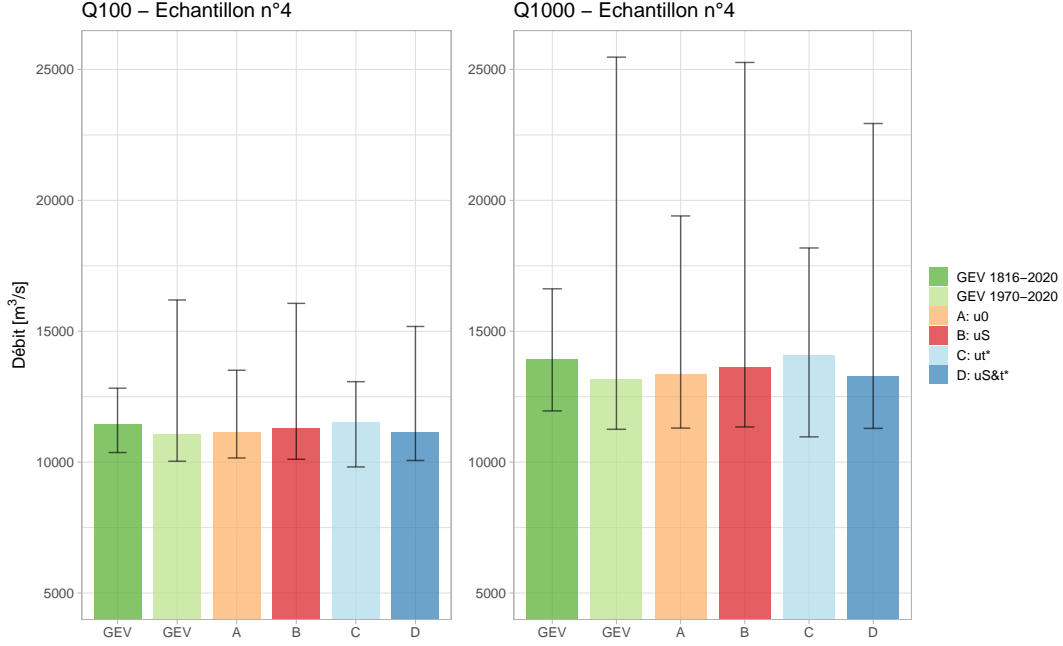


Figure 4: Estimations maxpost et incertitudes à 95% pour Q100 et Q1000 pour 6 modèles appliqués à l'échantillon 4 (1816-2020 dégradé, S4)

l'incertitude du modèle GEV pour la période 1970-2020 est la plus importante de tous les modèles (avec un écart type de 6159 [m³/s] contre 3019 [m³/s] pour le modèle A dans le cas de la crue millénale). 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, puisqu'elles permettent de se rapprocher de l'incertitude que l'on obtiendrait avec une longue période continue (GEV 1816-2020), sans pour autant l'atteindre.

La figure 5 présente l'ensemble des quantiles jusqu'à la crue décamillénale pour les 6 modèles. Ici, le rang des crues des deux périodes a été tiré aléatoirement comme décrit dans la section 2.7. 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.

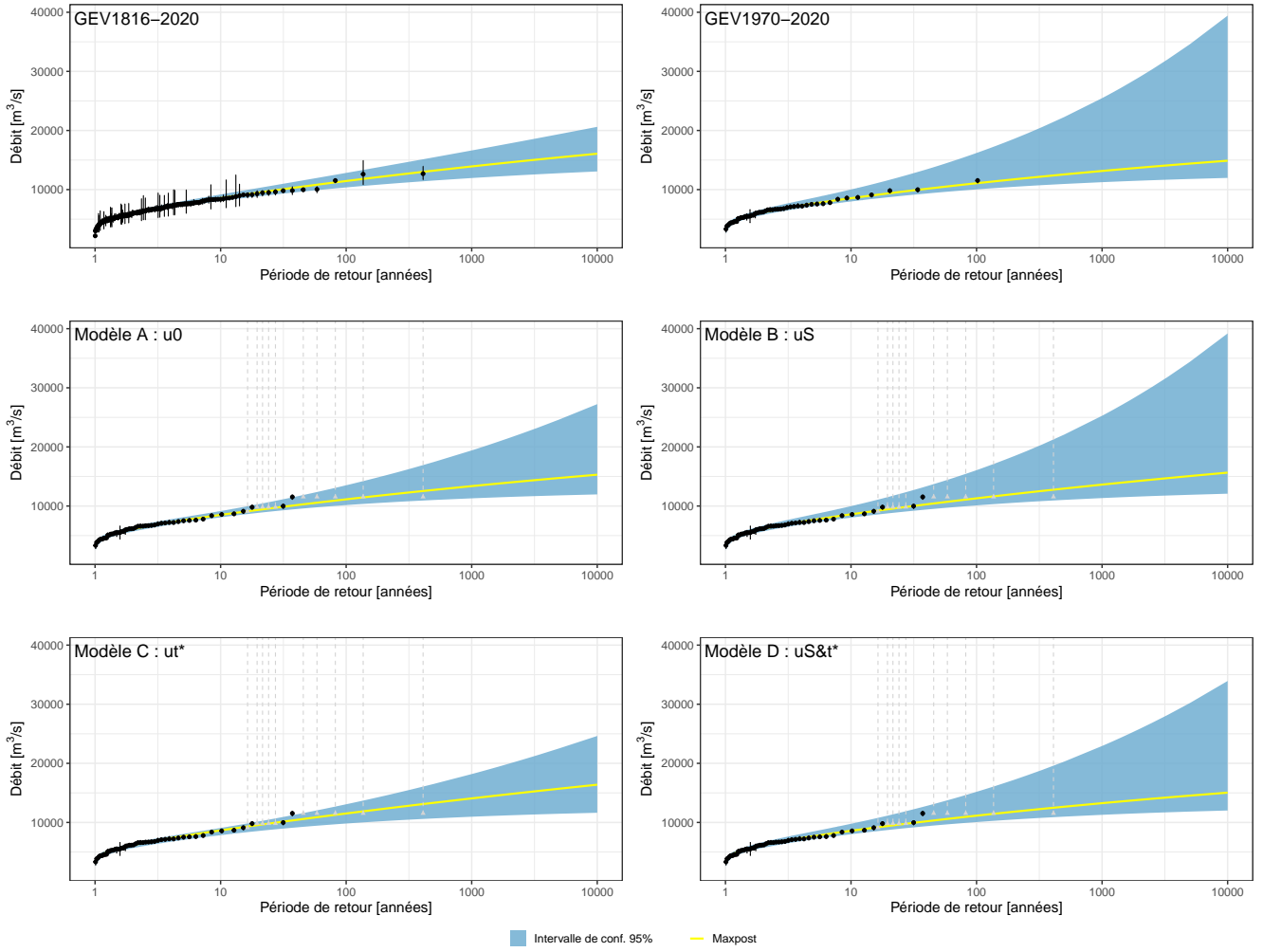


Figure 5: Quantiles de débit maximum annuel estimés par 6 modèles pour l'échantillon 4 (1816-2020 dégradé,  $S_4$ ). Les observations sont en noir pour la période continue (l'incertitude est également représentée) et en gris pour la période historique.

#### 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 la période de retour du quantile visé entraîne des résultats très incertains (GEV 1970-2020 dans les figures 4 et 5). La durée de la chronique continue (50 ans) est ici très petite devant la période de retour visée (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 d'occurrences 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 lorsque  $S$  et  $n$  sont connus. Pour les 6 modèles, une part de l'incertitude provient 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 (cas "loi de Weibull"). Comme l'on peut

s’y attendre, l’estimation de ce paramètre est plus précise dans le cas GEV 1816-2020. Les distributions a posteriori sont très proches pour les modèles A et C (tableau 2).

**RMQ BEN** D’ailleurs comme on ne réduit que faiblement l’incertitude sur  $\xi$  par rapport à GEV 1970-2020, d’où vient la réduction d’incertitude des quantiles? De position et échelle j’imagine? Comme tu as de la place en figure 6 est-ce que ça ne vaudrait pas le coup de mettre les 3 paramètres?

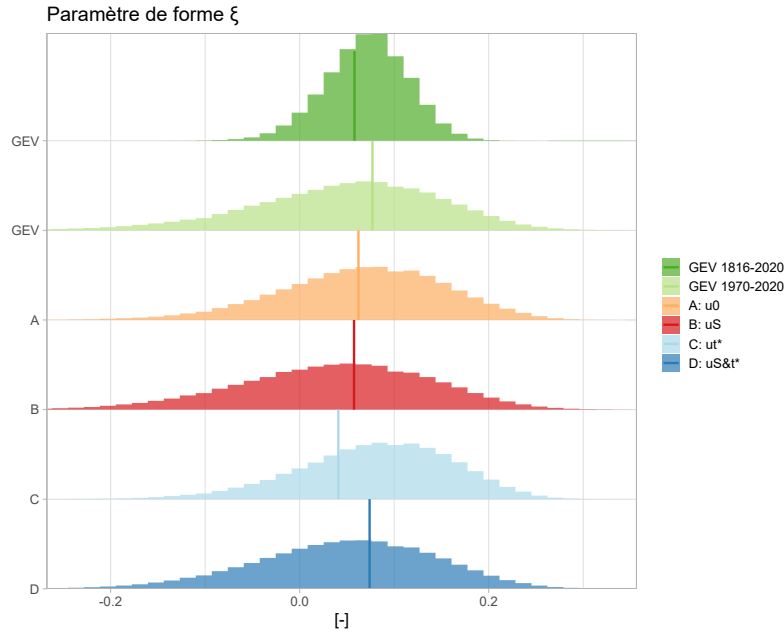


Figure 6: Distributions a posteriori du paramètre de forme de la distribution GEV des débits maximum annuels pour les 4 modèles estimés sur l’échantillon 4. Les estimations des modèles GEV sont également indiquées. Les droites verticales représentent les valeurs maxpost.

#### 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, et se rapproche de celle obtenue avec les données continues seulement (GEV 1970-2020). La méconnaissance du seuil a donc des conséquences importantes sur les estimations puisqu’elle réduit fortement l’intérêt d’utiliser les occurrences historiques. La vraie valeur du seuil de perception pour l’échantillon 4 est  $S4 = 9000 \text{ m}^3/\text{s}$ . 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 incertain :  $\mathcal{N}(9000, 2000)$ . La valeur maxpost est de 9163  $\text{m}^3/\text{s}$  soit une erreur relative de 2%. Dans une situation plus réaliste, un a priori du seuil de perception plus précis aurait pu être choisi afin de limiter cet impact. Il faut également noter que l’incertitude a posteriori du paramètre de forme pour le modèle B (figure 6) est plus grande que celle du modèle A et devient alors quasi-identique à celle du modèle GEV 1970-2020.



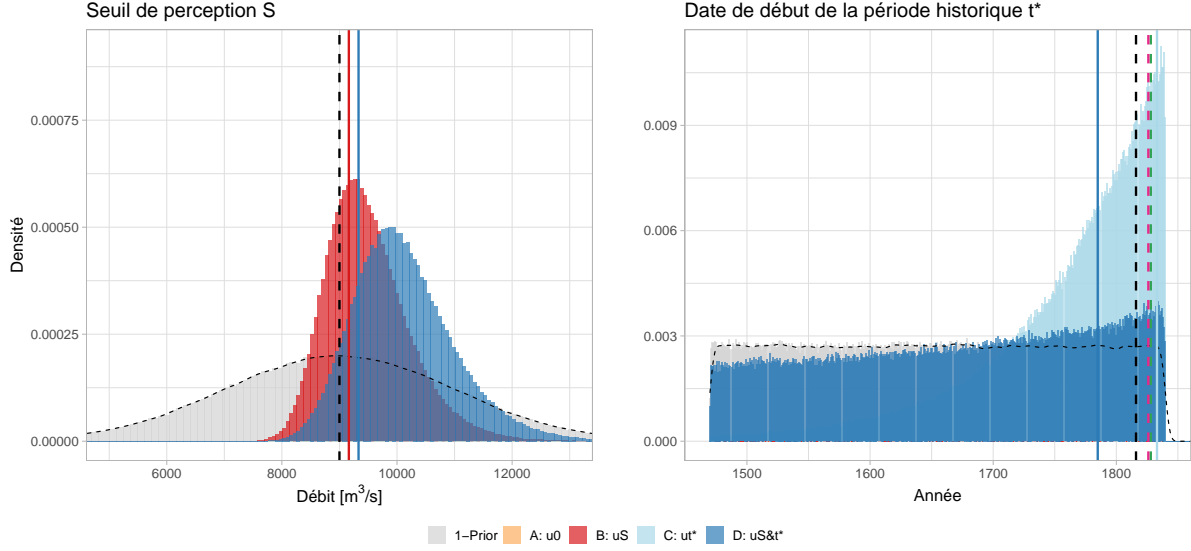


Figure 7: Distributions a priori et a posteriori pour le seuil de perception (gauche) et la date de début de la période historique (droite). Les droites verticales pleines représentent l’estimation maxpost du paramètre pour chacun des modèles et les droites en pointillés noirs représentent les valeurs de référence. Les droites verticales en pointillés vert et rose représentent respectivement les estimations de  $t^*$  par la méthode de Prosdocimi (2018) et la méthode de la période de retour du seuil  $S$ .

#### 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 supérieures à un 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 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 sous-estimation de 16 ans peut s’expliquer par une fréquence des crues supérieures au seuil  $S4$  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 oscille malgré le fait qu’aucune non-stationnarité des données n’ait été détectée par les tests à la section 3.2. 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 :  $\mathcal{U}(1340, 1840)$ . 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 autour de  $S$  et  $n$  en même temps dans le modèle probabiliste paraît être la solution la plus raisonnable dans certains cas, notamment pour des événements très anciens et mal connus. 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 C. 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 m<sup>3</sup>/s), est très large (écart type = 883 m<sup>3</sup>/s). Le seuil de perception paraît ici un peu moins bien estimé que par le modèle B (écart type = 729 m<sup>3</sup>/s). La date de début de la période historique est elle encore plus difficilement estimée, notamment en comparaison avec l'estimation du modèle C. On remarque que la distribution a posteriori est très similaire à celle de l'a priori, même si elle est légèrement asymétrique et marque un maximum non loin de la vraie valeur (l'année 1816). Cependant, les quantiles présentent une plus faible incertitude pour le modèle D que pour le modèle B. Cela provient de corrélations entre les paramètres qui peuvent être observées sur la figure 8. On remarque notamment une assez bonne corrélation entre la durée de la période historique  $n$  et le seuil de perception  $S$ , ainsi qu'entre le seuil de perception et le paramètre de forme  $\xi$ . Il est donc complexe d'identifier ces paramètres séparément. Une élicitation plus précise de leurs a priori sera certainement nécessaire.

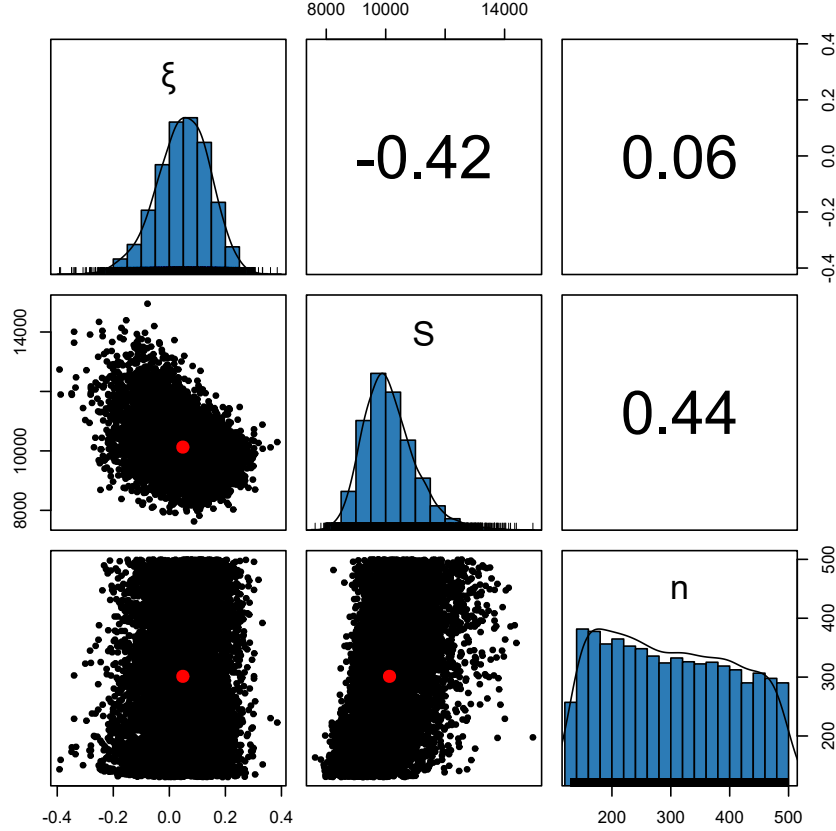


Figure 8: Scatterplots des distributions a posteriori de trois paramètres du modèle D pour l'échantillon 4 : le paramètre de forme  $\xi$  (sans unité), le seuil de perception  $S$  (en  $\text{m}^3/\text{s}$ ) et la durée de la période historique  $n$  (en années). Les nombres inscrits dans les cases supérieures correspondent aux corrélations entre les paramètres.

#### 4.1.5 Quel est l'apport de la connaissance du débit des crues historiques ?

Les modèles A, B, C et D n'utilisent que l'information du nombre de dépassements  $k$  d'un seuil de perception  $S$  pendant une durée  $n$ . Le débit des crues historiques ayant dépassé le seuil est donc ignoré. Le modèle E permet de prendre en compte cette donnée de débit ainsi que l'incertitude correspondant à chaque crue. Il a été appliqué à l'échantillon 4 du tableau 1 en considérant l'incertitude des débits de la période historique (1816-1969) calculée au chapitre 1 (REF).

Les résultats sont présentés dans la figure 9 et le tableau 3. On constate une diminution de l'incertitude d'environ 25% autour du Q1000 pour le modèle E (avec un écart type de 2255  $[\text{m}^3/\text{s}]$  pour Q1000) par rapport au modèle binomial A (écart type de 3019  $[\text{m}^3/\text{s}]$  pour Q1000). En revanche, l'incertitude du modèle E reste environ 65% supérieure à celle du modèle GEV 1816-2020 pour Q1000. La connaissance du débit des crues historiques, bien qu'elle ne soit pas une condition nécessaire à l'utilisation de données pré-enregistrements systématiques, permet donc de réduire l'incertitude autour des quantiles extrêmes. Cette incertitude reste en revanche supérieure à celle du modèle GEV pour lequel le débit maximum

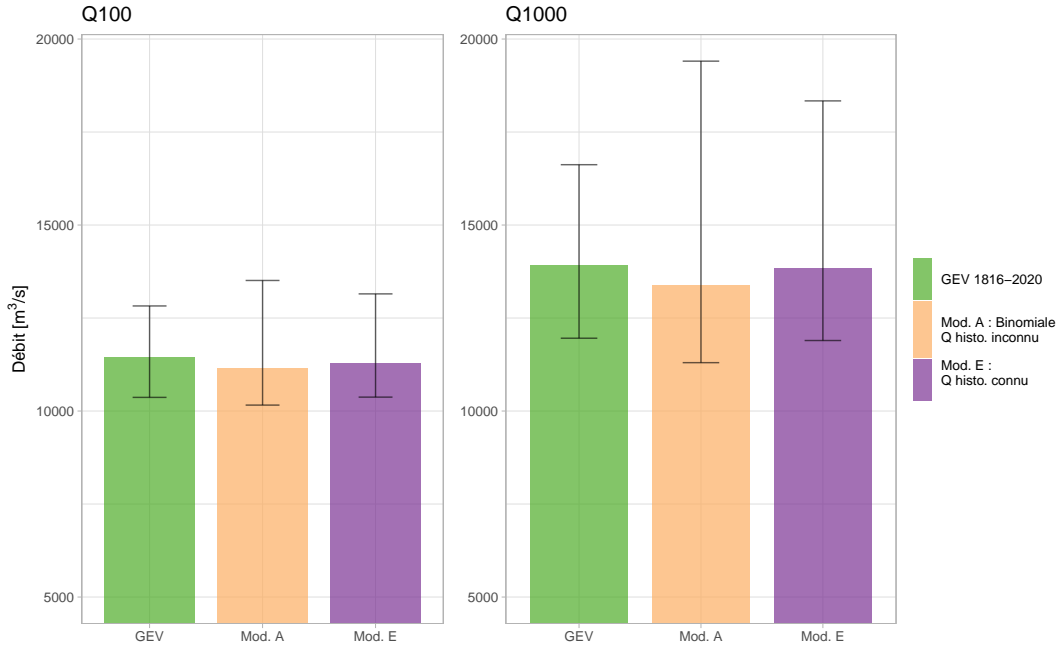


Figure 9: Estimations maxpost et incertitudes à 95% pour Q100 et Q1000 pour 3 modèles appliqués à l'échantillon 4 (1816-2020 dégradé,  $S_4$ ). Le modèle A n'utilise que le nombre de dépassements du seuil de perception, alors que le modèle E considère le débit des crues historiques (et son incertitude) ayant dépassé le seuil  $S$ .

Table 3: Résultats maxpost et incertitudes pour trois modèles appliqués à l'échantillon 4. Q100 et Q1000 représentent respectivement le débit des crues centennales et millénales. Les écarts type des distributions a posteriori sont représentés par les colonnes débutant par la lettre "u".

Modèle	Q100 [m³/s]	uQ100 [m³/s]	Q1000 [m³/s]	uQ1000 [m³/s]
GEV 1816-2020	11451	687	13919	1351
A	11132	1189	13367	3019
E	11286	932	13827	2255

annuel de toutes les années de l'échantillon (historique + continu) est connu.

#### 4.1.6 Discussion intermédiaire

L'utilisation des modèles décrits dans la section 2 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  $S$  et de la durée de la période historique  $n$  sur l'estimation des quantiles extrêmes. 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 centrées à proximité de la vraie valeur (9000 m³/s), l'incertitude résultant de la détermination du seuil 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 les appliquer 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. Si les a priori utilisés jusqu’ici étaient peu informatifs afin de comprendre les performances du modèle, ils devront être élicités plus précisément par la suite afin d’obtenir des résultats qui reflètent davantage la connaissance/méconnaissance du seuil de perception et de la durée de la période historique.

Enfin, la comparaison du modèle A pour lequel seul le nombre de dépassements du seuil de perception est connu avec le modèle E pour lequel le débit des crues historiques est connu (ainsi que l’incertitude autour de ces débits) a permis de démontrer l’intérêt de reconstituer le débit des crues historiques. Néanmoins, ces résultats ne sont valables que pour le seuil de perception  $S_4$  utilisé ici. Plusieurs études (notamment Stedinger and Cohn (1986) et Payraastre et al. (2011)) ont démontré que l’écart d’incertitude entre les résultats de ces deux types de modèles tendait à se réduire à mesure que la période de retour du seuil de perception tendait vers environ 50 ans, jusqu’à devenir nul au delà de cette magnitude. Cela encourage donc l’utilisation du nombre de dépassements d’un seuil de perception lorsqu’il n’est pas possible d’avoir de meilleure information sur la période historique. Au demeurant, les résultats de cette section mettent en garde sur le fait de supposer que le seuil de perception est parfaitement connu et dans une moindre mesure que la durée de la période historique est connue lorsque ce n’est pas le cas.

## 4.2 Application à la période 1500-2020

Les modèles A, B, C et D de la section 2 ont été appliqués à l’échantillon 2 (1500-2020, seuil  $S_4$ ), les résultats sont présentés dans la figure 10 et sont comparés au modèle GEV sur l’échantillon continu 1816-2020. Sans surprise, on observe des résultats moins incertains que pour l’échantillon dégradé (figure 4). L’incertitude des résultats des modèles GEV-Binomiale est au moins équivalente à celle de la référence (GEV 1816-2020), voire plus faible pour les modèles supposant un seuil de perception connu (A et C).

Table 4: Résultats maxpost et incertitudes des 5 modèles pour l’échantillon 2.  $Q_{100}$  et  $Q_{1000}$  représentent respectivement le débit des crues centennales et millénales,  $\xi$  le paramètre de forme de la distribution GEV,  $S$  le seuil de perception et  $t^*$  la date de début de la période historique. Les écarts type des distributions a posteriori sont représentés par les colonnes débutant par la lettre ”u”.

Modèle	$Q_{100}$ [m <sup>3</sup> /s]	u $Q_{100}$ [m <sup>3</sup> /s]	$Q_{1000}$ [m <sup>3</sup> /s]	u $Q_{1000}$ [m <sup>3</sup> /s]	$\xi$	$u\xi$	$S$ [m <sup>3</sup> /s]	u $S$ [m <sup>3</sup> /s]	$t^*$	u $t^*$
GEV 1816-2020	11451	687	13919	1351	0,058	0,044	X	X	X	X
A	10977	391	13149	816	0,073	0,038	9000	X	1500	X
B	11438	698	13875	1391	0,06	0,044	9628	504	1500	X
C	10975	395	13139	809	0,074	0,038	9000	X	1527	44
D	11336	745	13721	1467	0,061	0,044	9613	551	1529	62

Les distributions a posteriori de  $S$  et  $t^*$  sont présentées dans la figure 11. Ces distributions sont similaires à celles observées pour l’échantillon 4 pour lequel les a priori étaient les mêmes. Le modèle estime un seuil de perception supérieur à la valeur de 9000 m<sup>3</sup>/s (modèles B et D, tableau 4), ainsi qu’une date de début de la période historique plus récente que la valeur supposée de 1500 (modèles C

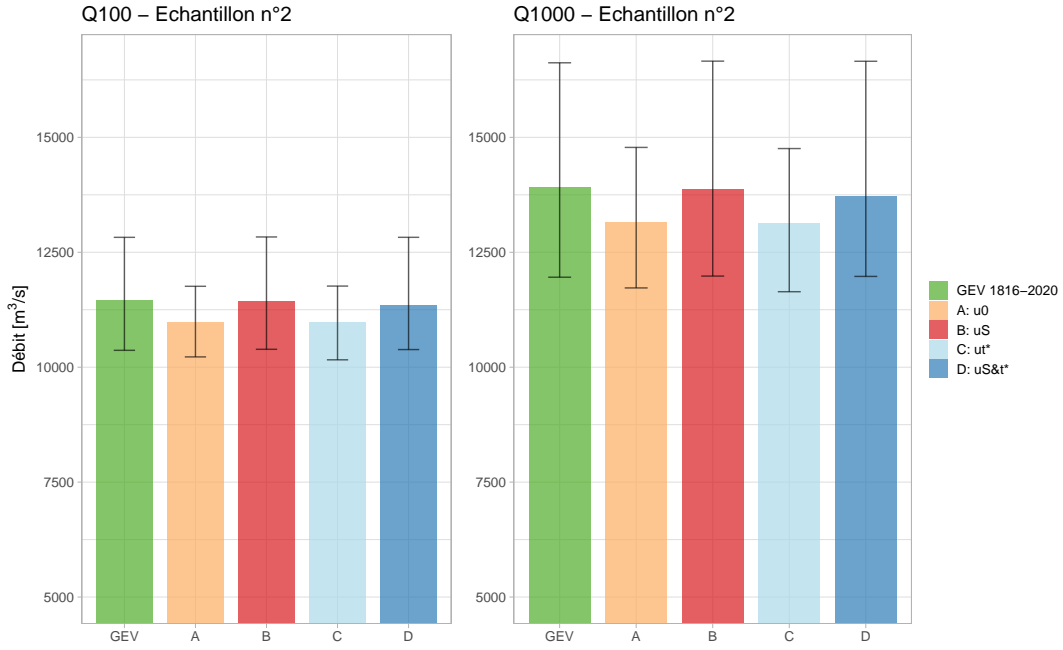


Figure 10: Estimations maxpost et incertitudes à 95% pour Q100 et Q1000 par 5 modèles différents pour l'échantillon 2 (1500-2020,  $S_4$ )

et D, tableau 4). Cette tendance exprimée par les modèles vers un seuil plus important ou une période historique plus courte pourrait être le symptôme d'une non-exhaustivité des crues dans l'échantillon C4 de la base HISTRHÔNE, et ce malgré qu'aucune non-homogénéité de la fréquence des crues supérieures au seuil  $S_4$  n'ait été détectée (section 3.2).

#### 4.2.1 Quel est l'apport des crues historiques pour l'analyse fréquentielle à Beaucaire ?

De la même manière qu'avec l'échantillon dégradé, on observe ici que la méconnaissance du seuil de perception (B et D) a plus d'impact sur l'incertitude des résultats que la méconnaissance de la durée de la période historique (C et D). De plus, l'incertitude autour des résultats est plus faible que celle de la référence uniquement dans les cas où le seuil de perception est supposé connu (modèles A et C). L'utilisation du nombre d'occurrences de crues historiques supérieures à un seuil ne permet de réduire l'incertitude autour des quantiles estimés que lorsque le seuil de perception est supposé parfaitement connu. L'utilisation d'a priori plus informatifs pourrait probablement donner des résultats moins incertains et plus réalistes quant à la connaissance du seuil de perception et de la durée de la période historique.

#### 4.3 L'estimation des quantiles à Beaucaire (1500-2020) avec des a priori plus informatifs rend-elle l'utilisation des données historique plus pertinente ?

Après analyse des résultats des modèles pour différents échantillons, il apparaît que l'estimation la plus prudente des quantiles de crues à Beaucaire (1500-2020) consiste à utiliser le modèle D ( $S$  et  $n$  incertains) en parallèle d'une élicitation d'a priori relativement informatifs. L'utilisation de l'échantillon correspondant au seuil de perception le plus grand est également plus légitime. Il s'agit ici du seuil  $S_4$ . Sa distribution a priori est à nouveau supposée gaussienne et centrée sur la valeur supposée de  $9000 \text{ m}^3/\text{s}$ , en accord avec les estimations de Pichard et al. (2017) et des résultats du chapitre 2 (REF). L'écart type de

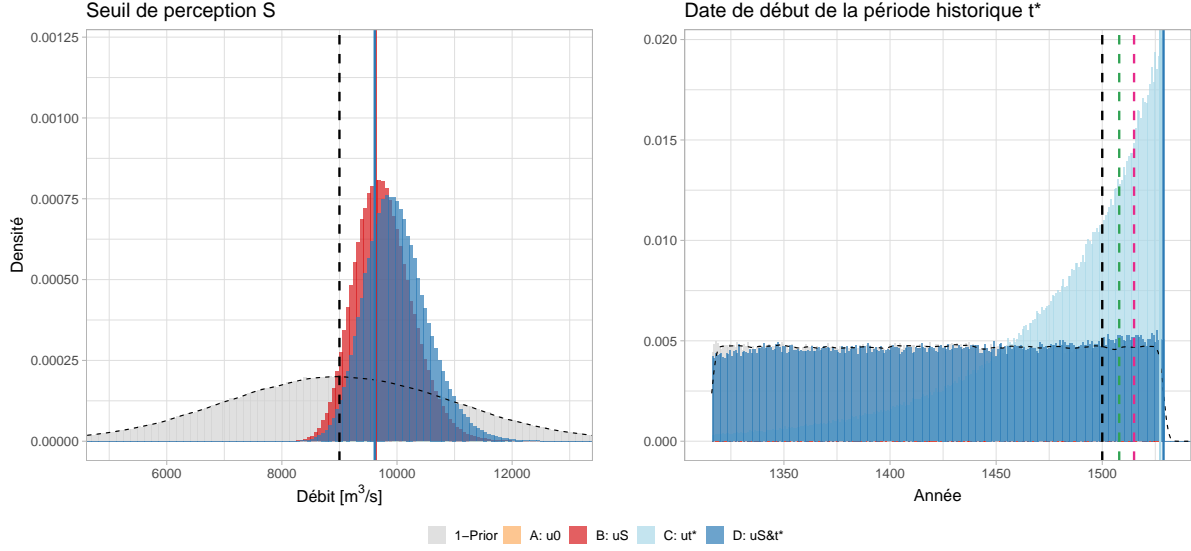


Figure 11: Distributions a priori et a posteriori pour le seuil de perception (gauche) et la date de début de la période historique (droite). Les droites verticales pleines représentent l'estimation maxpost du paramètre pour chacun des modèles et les droites en pointillés noirs représentent les valeurs de référence. Les droites verticales en pointillés vert et rose représentent respectivement les estimations de  $t^*$  par la méthode de Prosdocimi (2018) et par la méthode de la période de retour du seuil  $S$ .

la distribution est fixé à  $500 \text{ m}^3/\text{s}$ , ce qui le rend l'a priori plus informatif que lors des calculs précédents. En ce qui concerne la date de début de la période historique  $t^*$ , une distribution uniforme est utilisée. La borne supérieure de la distribution reste fixée à la date de la première crue de la série (en 1529). En revanche, la borne inférieure est affectée à deux fois la durée entre la date de la première crue (1529) et la valeur supposée de  $t^*$  (1500), soit 58 ans. La borne inférieure de la distribution a priori est donc l'année 1471. La valeur théorique calculée en utilisant la méthode proposée par Prosdocimi (2018) est 1510, elle est bien comprise dans la distribution a priori. Il en est de même pour la valeur théorique de  $t^*$  correspondant à la différence entre la période de retour du seuil de perception  $S_4$  (environ 15 ans) et la date de la première crue, ce qui correspond à l'année 1515.

Les résultats du modèle D avec ces nouveaux a priori sont présentés dans la figure 12 ainsi que le tableau 5 et comparés avec les estimations GEV 1816-2020. Les quantiles du modèle D sont moins incertains que ceux de la référence. L'utilisation des crues historiques apparait pertinente pour réduire l'incertitude des quantiles, même dans le cas où  $S$  et  $n$  sont incertains. L'utilisation d'a priori plus informatifs a permis de réduire d'environ 25% l'écart type de la distribution a posteriori de la crue millénale par rapport à des a priori très peu informatifs.

Les distributions a posteriori de  $S$  et  $t^*$  sont présentées dans la figure 13. Une fois de plus, la distribution a posteriori du seuil de perception est décalée vers des valeurs plus élevées que la valeur supposée de  $9000 \text{ m}^3/\text{s}$ , avec un seuil maxpost à  $9386 \text{ m}^3/\text{s}$ . La distribution a posteriori de  $t^*$  est à nouveau très proche de la distribution a priori, avec une densité un peu plus élevée pour les années proches de la date de la première crue. L'estimation maxpost de  $t^*$  est ici 1526, soit une durée de la période historique  $n$  26 ans plus courte qu'attendu. Cette sous-estimation de  $n$  est non-seulement due au fait que les valeurs théoriques de la durée de la période historique sont plus petites que la durée supposée mais aussi à la



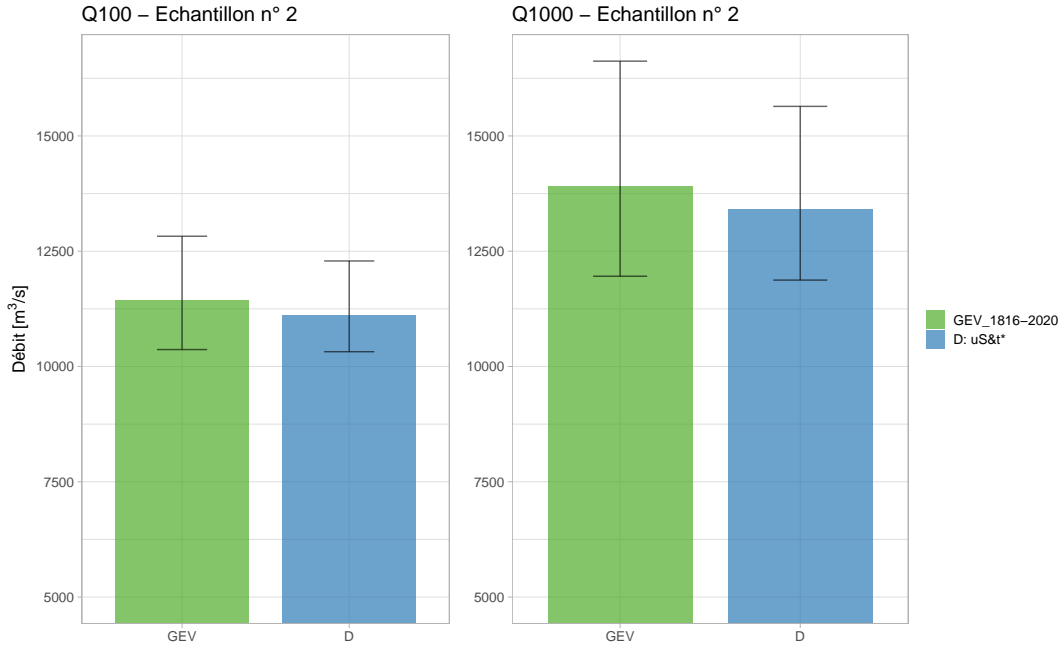


Figure 12: Estimations maxpost et incertitudes à 95% pour Q100 et Q1000 par 2 modèles différents pour l'échantillon 2 (1500-2020,  $S4$ ) après révision des distributions a priori de  $S4$  et  $n$ .

Table 5: Résultats maxpost et incertitudes de 2 modèles pour l'échantillon 2 après révision des distributions a priori de  $S4$  et  $n$ . Q100 et Q1000 représentent respectivement le débit des crues centennales et millénales,  $\xi$  le paramètre de forme de la distribution GEV,  $S$  le seuil de perception et  $t^*$  la date de début de la période historique. Les écarts type des distributions a posteriori sont représentés par les colonnes débutant par la lettre "u".

Modèle	Q100 [m³/s]	uQ100 [m³/s]	Q1000 [m³/s]	uQ1000 [m³/s]	$\xi$	$u\xi$	$S$ [m³/s]	$uS$ [m³/s]	$t^*$	$ut^*$
GEV 1816-2020	11451	687	13919	1351	0.058	0.044	X	X	X	X
D	11118	585	13421	1110	0.063	0.040	9386	334	1526	17

probable non-exhaustivité des données historiques décrite dans la section précédente.

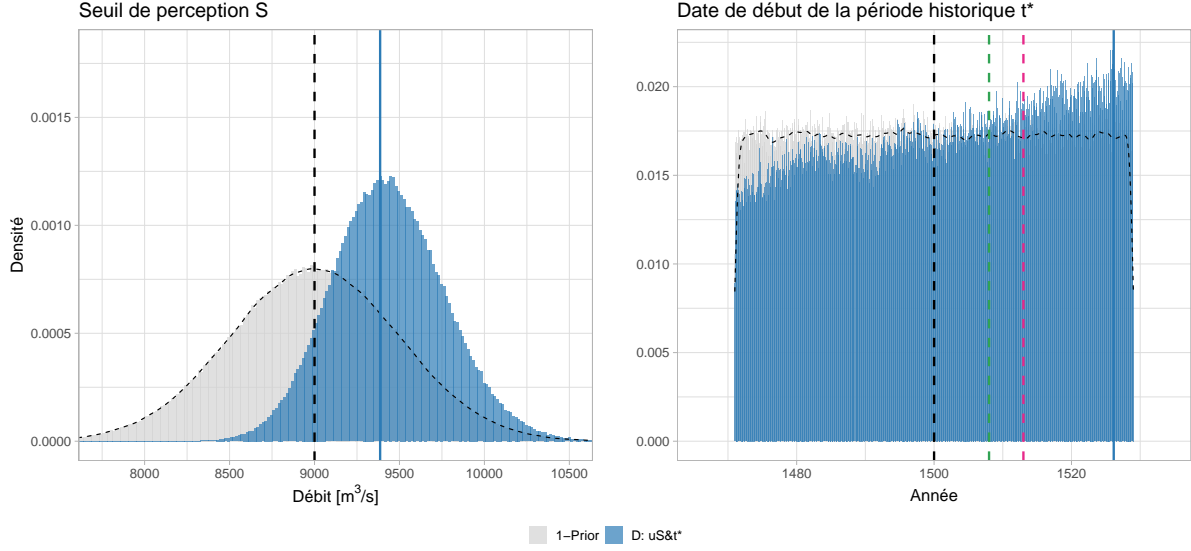


Figure 13: Distributions a priori et a posteriori pour le seuil de perception (gauche) et la date de début de la période historique (droite). Les droites verticales représentent l'estimation maxpost du paramètre de forme pour chacun des modèles. La droite verticale noire représente la date de début supposée de la chronique historique, ici 1500.

## 5 Discussion

L'intérêt de la valorisation des données historiques pour l'analyse fréquentielle des crues est connu et étudié depuis longtemps (Benson (1950); Stedinger and Cohn (1986)). L'utilisation des données anciennes doit être accompagnée d'une estimation complète des incertitudes (Kjeldsen et al., 2014). Lors de l'utilisation d'un échantillon d'occurrences de crues supérieures à un seuil (le débit des crues historiques supérieures au seuil de perception n'est pas connu), la méconnaissance du seuil de perception et de la durée de la période historique est souvent négligée. Seule l'incertitude provenant de l'estimation des paramètres de la distribution choisie est généralement considérée. Plusieurs modèles sont proposés dans ce chapitre permettant de prendre en compte la méconnaissance autour de ces deux paramètres. La propagation de l'incertitude des débits de la période récente est également effectuée.

Les modèles ont dans un premier temps été testés avec des a priori très peu informatifs et sur un échantillon continu artificiellement dégradé afin de se replacer dans un contexte historique. Ainsi, seuil de perception et durée de la période historique sont parfaitement connus. Les quantiles estimés ont été comparés avec les estimations d'un modèle GEV pour l'entièreté de la période (figure 4). Il est apparu que considérer le seuil de perception comme étant incertain avait bien plus d'impact sur l'incertitude des résultats que considérer une méconnaissance sur la durée de la période historique. En revanche, quand ces deux paramètres sont considérés incertains en même temps, l'incertitude autour des quantiles est réduite par rapport au cas où seul le seuil est incertain. Une corrélation entre les distributions a posteriori de ces deux paramètres a été mise en évidence. Cette corrélation n'est pas surprenante étant donné que seuil et durée de la période historique sont par définition reliés : le nombre de dépassements du seuil de perception  $k$  est à la fois dépendant du seuil  $S$  et de la durée  $n$ . L'éllicitation d'a priori plus informatifs pour ces deux paramètres est donc nécessaire. Bien que le seuil de perception soit un objet parfois difficile à cerner, il est donc possible de représenter sa méconnaissance à l'intérieur même du modèle probabiliste. Considérer

que le seuil de perception est parfaitement connu lorsque ce n'est pas le cas peut mener à une importante sous-estimation de l'incertitude des quantiles, de même que pour la durée de la période historique, dans une moindre mesure. Dans le cas de cet échantillon dégradé, l'utilisation des données historiques permet de réduire l'incertitude des quantiles par rapport à la seule utilisation de l'échantillon continu (dégradé), et ce quelque soit l'hypothèse retenue concernant la méconnaissance de  $S$  et  $n$ .

L'estimation du modèle binomial pour  $S$  et  $n$  connus (modèle A) a ensuite été comparée à une estimation pour laquelle le débit des crues historiques est connu et situé dans un intervalle (modèle E) pour l'échantillon dégradé (figure 9). L'incertitude des résultats du modèle E se situe entre celle du modèle A et celle du modèle GEV pour l'ensemble de la période. Dans ce cas précis, l'utilisation du débit des crues historiques s'est donc avérée légèrement plus informative que l'utilisation du seul nombre de dépassements du seuil de perception. D'après Stedinger and Cohn (1986) et Payraastre et al. (2011), la magnitude du seuil de perception utilisé influe directement sur l'incertitude des résultats dans le cas du modèle binomial. Ainsi, pour un seuil de perception suffisamment grand, l'incertitude des estimations du modèle binomial est identique à celle du modèle pour lequel le débit des crues historique est connu. L'intérêt de connaître précisément le débit des crues historique est dans ce cas moins légitime. Payraastre et al. (2011) estiment que l'on se trouve dans cette situation lorsque la période de retour du seuil de perception est supérieur à environ 50 ans (dans les cas testés). A Beaucaire, la période de retour du seuil de perception  $S4$  est d'environ 15 ans, ce qui explique l'avantage du modèle E sur le modèle A dans ce cas précis.

Pour l'échantillon de crues de Beaucaire de 1500 à 2020, l'utilisation des données anciennes reste, en l'absence d'estimations du débit des crues historiques, conditionnée à l'utilisation des seuils de perception de la base HISTRHÔNE :  $S3$  ou  $S4$ . L'estimation des quantiles étant plus précise pour un seuil de perception correspondant à une période de retour importante, il est naturel de préférer le seuil  $S4$ . Il paraît également plus prudent d'utiliser le modèle qui fait l'hypothèse d'une méconnaissance du seuil de perception et de la durée de la période historique (modèle D), en parallèle d'une élicitation réaliste des a priori de ces deux paramètres. Les résultats du modèle D sous les conditions décrites précédemment sont légèrement plus informatifs que l'utilisation de la seule chronique continue de 1816 à 2020 (figure 12). Cela permet donc de conclure sur l'intérêt de l'utilisation des données historiques dans le cas de Beaucaire. Néanmoins, des doutes subsistent sur l'exhaustivité de l'échantillon historique.

L'application des quatre modèles à l'échantillon complet de 1500 à 2020 à Beaucaire avec des a priori peu informatifs a montré que les estimations de la durée historique et du seuil de perception étaient différentes des valeurs supposées. Les modèles B et D estiment un seuil de perception plus grand que supposé et les modèles C et D estiment une durée de la période historique légèrement plus courte que celle supposée (figure 11). Cette tendance pourrait être le symptôme d'une non-exhaustivité des crues dans les échantillons historiques de la base HISTRHÔNE, et ce bien qu'aucune non-homogénéité de la fréquence des crues n'ait été détectée (section 3.2). Le nombre  $k$  de dépassements du seuil de perception est donc possiblement sous-estimé dans les données disponibles. Cette sous-estimation pourrait provenir de la nature même des données, qui ne sont pas des données de crues supérieures à un seuil au sens physique. Les catégories de la base HISTRHÔNE sont définies sur la perception des dommages par les populations ripariennes et non un seuil de perception physique directement lié au dépassement d'un débit. Or, comme décrit au chapitre 2 (REF), la perception des dommages par les populations a probablement évolué au cours du temps. Cette variabilité dans la perception des dommages et la conséquence de nombreux facteurs, qu'ils soient physiques ou anthropiques. Ainsi, relier directement les conséquences d'une crue à son seul débit de pointe est un raccourci dangereux. Par exemple, pour un même débit de pointe, la perception de la gravité d'une crue peut varier selon la durée de la crue, le niveau de protection, la rupture

de digues, la densité de population... De plus, la qualité et la quantité des témoignages de crues historiques qui parviennent jusqu'à nous sont dépendants du contexte sociétal, qui est lui aussi très variable d'une époque à l'autre : contexte politique, religieux, médiatique, culture de la conservation d'archives... Par exemple, l'impressionnante quantité de données recueillie dans la base HISTRHÔNE est notamment la conséquence de la présence d'instances religieuses dans la région provençale, au sein desquelles l'érudition et la tradition écrite était importante. Cet ensemble de paramètres est important à garder en tête lors de l'utilisation de données historique pour l'analyse fréquentielle, et plus particulièrement lors de la phase d'enquête historique. La méthodologie de la recherche de données est tout aussi importante que l'analyse statistique, afin de garantir l'exhaustivité des échantillons et la connaissance du contextes hydrologique, hydraulique et social contemporain aux crues étudiées.

Dans les études historiques menées sur des bassins versants français de petite taille, la détermination du seuil de perception  $S$  et l'exhaustivité du nombre de crues  $k$  supérieures au seuil semblent être les principales limites. Par exemple, Payraastre (2005) écrit : *"De façon à disposer d'une information fiable nous avons parfois dû limiter la durée de la reconstitution historique à moins de deux siècles (durée visée au départ), un seuil de perception ne pouvant raisonnablement être défini"*. Il en est de même pour les données de la base HISTRHÔNE utilisées ici et dont les plus anciens témoignages remontent au XIII<sup>ème</sup> siècle. L'avis d'expert de Geoges PICHARD, l'historien à l'origine de cette base de données, était que l'exhaustivité des crues était probablement atteinte à partir du XVI<sup>ème</sup> siècle, ce qui nous a conduits à utiliser les données de la base à partir de l'année 1500. Ce constat était sans doute optimiste et nous avons été confortés dans ce choix par les résultats l'analyse de stationnarité réalisée à la section 3.2. Un début de période historique plus tardif aurait pu mener à de meilleurs résultats. Il est risqué de vouloir à tout prix utiliser le plus grand nombre de données historiques possible dans un contexte d'analyse fréquentielle. L'exhaustivité doit être le premier critère, devant la quantité de données utilisée. Par exemple, Gaume et al. (2010) écrivent dans un contexte légèrement différent d'analyse historique et régionale des crues : *"comprehensive and dense inventories of ungauged extremes with a controlled and predefined number of target watersheds not selected on the basis of the magnitude of the past floods (French example), should be preferred to loose collations of isolated extreme values with unknown equivalent coverage. For a given series, if there is an uncertainty on the threshold values, the highest guesses should be selected to obtain the pessimistic result concerning the credibility bounds."*

Suite à ces différents constats, on pourrait penser à l'utilisation d'un modèle qui considère non-seulement le seuil de perception  $S$  et la durée de la période historique  $n$  comme étant incertains, mais également pour lequel le nombre de dépassements  $k$  du seuil de perception est lui aussi incertain. Néanmoins, les résultats de ce chapitre ont permis de mettre en évidence que la méconnaissance de  $S$  avait à elle seule un fort impact sur les résultats. De plus, lorsqu'il existe un tel doute sur l'exhaustivité des données historiques, discuter de l'incertitude du seuil de perception ou de la durée de la période historique paraît secondaire. Ici, l'utilisation même des données historiques pourrait être remise en cause, l'échantillon d'enregistrements continus étant exceptionnellement long (205 ans) et permettant une estimation satisfaisante des quantiles de crue. Une des façons de compléter cette étude pourrait être d'estimer le débit des événements historiques à l'aide de modèles hydrauliques avec une prise en compte complète des incertitudes inhérentes à cet exercice (bathymétrie historique, rupture de digues...) listées dans le chapitre 2 (REF). Une des pistes possibles pour obtenir une estimation réaliste des incertitudes dans ce contexte pourrait être l'utilisation de méthodes Monte Carlo couplées au modèle hydraulique. Il faudra garder à l'esprit qu'il n'est pas nécessaire de remonter jusqu'au XVI<sup>ème</sup> siècle pour que l'intérêt de ces reconstitutions soit utile. De plus, il est également possible d'utiliser plusieurs seuils de perception dans le but de garantir l'exhaustivité des données pour les périodes les plus anciennes. Cette possibilité est fréquemment

décrite dans la littérature mais n’a jamais été exploitée dans le cas d’un seuil de perception incertain. Considérer plusieurs seuils de perception incertains peut néanmoins nécessiter un traitement statistique bien plus complexe que celui décrit dans ce chapitre.

Bien que l’homogénéité des données ait été vérifiée (section 3), il est probable que les données utilisées dans ce chapitre soient impactées par une variabilité climatique et/ou anthropique qui pourrait fragiliser l’hypothèse de stationnarité nécessaire à l’analyse fréquentielle. L’hypothèse de stationnarité est de nos jours fréquemment remise en cause (Milly et al., 2008). Des tendances autour de la magnitude des crues ont été identifiées dans plusieurs régions en Europe (Hall et al. (2014); Blöschl et al. (2019)) et en France Giuntoli et al. (2019), mais aucune règle n’existe en France à ce jour pour prendre en compte l’impact du changement climatique sur l’estimation du risque inondation Madsen et al. (2014). Néanmoins, la non-prise en compte de ces sources de variabilité a probablement moins d’impact sur les résultats que la méconnaissance autour des données historiques. Il est tout de même possible d’intégrer les changements temporels des processus climatiques ou des caractéristiques du bassin versant au sein même du modèle probabiliste comme cela est de plus en plus fréquemment décrit dans la littérature (voir Salas et al. (2018) pour une revue complète des différentes possibilités).

## 6 Conclusion du chapitre

La méconnaissance du seuil de perception et de la durée de la période historique est quasi-systématiquement négligée lors de l’utilisation de données historiques. Les modèles mis au point dans ce chapitre permettent de considérer cette méconnaissance au sein même du modèle probabiliste et de la propager aux résultats. Il est apparu en testant les modèles sur une chronique artificiellement dégradée que la méconnaissance du seuil de perception avait plus d’impact sur l’incertitude des résultats que la méconnaissance de la durée de la période historique. De plus, une corrélation entre ces deux paramètres a été mise en évidence. L’utilisation de la chronique complète (1500-2020) a ensuite permis d’améliorer les estimations des quantiles extrêmes réalisées au chapitre 1 (REF), même dans le cas où seuil de perception et durée de la période historiques sont incertains, à condition d’utiliser des a priori suffisamment informatifs. Ces résultats soulignent l’intérêt de l’utilisation des données historiques pour l’analyse fréquentielle, ainsi que l’importance d’une prise en compte complète des incertitudes autour de ces données. Ces conclusions ne sont pas systématiquement généralisables en dehors du cas de Beaucaire, où la chronique d’enregistrements continus est particulièrement longue et où le paramètre de forme est positif. De plus, aucune tendance pouvant être attribuée aux changements des caractéristiques du bassin versant ou au changement climatique n’a pu être mise en évidence à Beaucaire. Néanmoins, pour aller au delà de l’analyse fréquentielle des crues, la perspective de modéliser la variabilité climatique semble intéressante avec un jeu de données aussi fourni.

L’application des modèles à Beaucaire de 1500-2020 a également permis de mettre en lumière une probable sous-estimation du nombre de dépassements du seuil de perception au cours de la période historique. Considérer une incertitude autour de  $S$  ou  $n$  est probablement mineur devant le problème de la non-exhaustivité des données historiques. Il est certainement possible de tirer un meilleur profit de ces données historiques en n’utilisant qu’une partie plus récente et mieux connue, ou bien en tentant d’estimer le débit des crues historiques, à condition de s’assurer de leur exhaustivité.

## Bibliographie

- Apel, H., A. H. Thielen, B. Merz, and G. Blöschl (2004). “Flood risk assessment and associated uncertainty”. In: *Natural Hazards and Earth System Sciences* 4.2. Publisher: Copernicus GmbH, pp. 295–308. ISSN: 1561-8633. DOI: [10.5194/nhess-4-295-2004](https://doi.org/10.5194/nhess-4-295-2004).
- Benito, G., M. Lang, M. Barriendos, M. C. Llasat, F. Francés, T. Ouarda, V. Thorndycraft, Y. Enzel, A. Bardossy, D. Coeur, and B. Bobée (2004). “Use of Systematic, Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation. Review of Scientific Methods”. In: *Natural Hazards* 31.3, pp. 623–643. ISSN: 1573-0840. DOI: [10.1023/B:NHAZ.0000024895.48463.eb](https://doi.org/10.1023/B:NHAZ.0000024895.48463.eb).
- Benson, M. A. (1950). “Use of historical data in flood-frequency analysis”. In: *Transactions, American Geophysical Union* 31.3, p. 419. ISSN: 0002-8606. DOI: [10.1029/TR031i003p00419](https://doi.org/10.1029/TR031i003p00419).
- Blöschl, G., J. Hall, A. Viglione, R. A. P. Perdigão, J. Parajka, B. Merz, D. Lun, B. Arheimer, G. T. Aronica, A. Bilibashi, M. Boháč, O. Bonacci, M. Borga, I. Čanjevac, A. Castellarin, G. B. Chirico, P. Claps, N. Frolova, D. Ganora, L. Gorbachova, A. Gül, J. Hannaford, S. Harrigan, M. Kireeva, A. Kiss, T. R. Kjeldsen, S. Kohnová, J. J. Koskela, O. Ledvinka, N. Macdonald, M. Mavrova-Guirguinova, L. Mediero, R. Merz, P. Molnar, A. Montanari, C. Murphy, M. Osuch, V. Ovcharuk, I. Radevski, J. L. Salinas, E. Sauquet, M. Šraj, J. Szolgay, E. Volpi, D. Wilson, K. Zaimi, and N. Živković (2019). “Changing climate both increases and decreases European river floods”. In: *Nature* 573.7772, pp. 108–111. ISSN: 1476-4687. DOI: [10.1038/s41586-019-1495-6](https://doi.org/10.1038/s41586-019-1495-6).
- Dariento, M. (2021). “Detection and estimation of stage-discharge rating shifts for retrospective and real-time streamflow quantification”. These de doctorat. Université Grenoble Alpes.
- Dezileau, L., B. Terrier, J. Berger, P. Blanchemanche, A. Latapie, R. Freydier, L. Bremond, A. Paquier, M. Lang, and J. Delgado (2014). “A multdating approach applied to historical slackwater flood deposits of the Gardon River, SE France”. In: *Geomorphology* 214, pp. 56–68. ISSN: 0169555X. DOI: [10.1016/j.geomorph.2014.03.017](https://doi.org/10.1016/j.geomorph.2014.03.017).
- Engeland, K., A. Aano, I. Steffensen, E. Støren, and Ø. Paasche (2020). *New flood frequency estimates for the largest river in Norway based on the combination of short and long time series*. preprint. Catchment hydrology/Instruments and observation techniques. DOI: [10.5194/hess-2020-269](https://doi.org/10.5194/hess-2020-269).
- Gaume, E., L. Gaál, A. Viglione, J. Szolgay, S. Kohnová, and G. Blöschl (2010). “Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites”. In: *Journal of Hydrology* 394.1, pp. 101–117. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2010.01.008](https://doi.org/10.1016/j.jhydrol.2010.01.008).
- Gerard, R. and E. W. Karpuk (1979). “Probability Analysis of Historical Flood Data”. In: *Journal of the Hydraulics Division* 105.9. Publisher: American Society of Civil Engineers, pp. 1153–1165. DOI: [10.1061/JYCEAJ.0005273](https://doi.org/10.1061/JYCEAJ.0005273).
- Giuntoli, I., B. Renard, and M. Lang (2019). “Floods in France”. In: *Changes in Flood Risk in Europe*. 1st ed. CRC Press, p. 13. ISBN: 978-0-203-09809-7.
- Hall, J., B. Arheimer, M. Borga, R. Brázdil, P. Claps, A. Kiss, T. R. Kjeldsen, J. Kriaučiūnienė, Z. W. Kundzewicz, M. Lang, M. C. Llasat, N. Macdonald, N. McIntyre, L. Mediero, B. Merz, R. Merz, P. Molnar, A. Montanari, C. Neuhold, J. Parajka, R. A. P. Perdigão, L. Plavcová, M. Rogger, J. L. Salinas, E. Sauquet, C. Schär, J. Szolgay, A. Viglione, and G. Blöschl (2014). “Understanding flood regime changes in Europe: a state-of-the-art assessment”. In: *Hydrology and Earth System Sciences* 18.7, pp. 2735–2772. ISSN: 1607-7938. DOI: [10.5194/hess-18-2735-2014](https://doi.org/10.5194/hess-18-2735-2014).
- Hazen, A. (1914). “Storage to be Provided in Impounding Municipal Water Supply”. In: *Transactions of the American Society of Civil Engineers* 77.1. Publisher: American Society of Civil Engineers, pp. 1539–1640. DOI: [10.1061/taceat.0002563](https://doi.org/10.1061/taceat.0002563).
- 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).



- Hirsch, R. M. and J. R. Stedinger (1987). “Plotting positions for historical floods and their precision”. In: *Water Resources Research* 23.4, pp. 715–727. ISSN: 00431397. DOI: [10.1029/WR023i004p00715](https://doi.org/10.1029/WR023i004p00715).
- Hosking, J. R. M. and J. R. Wallis (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge: Cambridge University Press. ISBN: 978-0-521-43045-6. DOI: [10.1017/CB09780511529443](https://doi.org/10.1017/CB09780511529443).
- Kendall, M. (1948). *Rank Correlation Methods*. 1. London: Charles Griffin & Company.
- Kjeldsen, T., N. Macdonald, M. Lang, L. Mediero, T. Albuquerque, E. Bogdanowicz, R. Brázdil, A. Castellarin, V. David, A. Fleig, G. Gül, J. Kriauciuniene, S. Kohnová, B. Merz, O. Nicholson, L. Roald, J. Salinas, D. Sarauskiene, M. Šraj, W. Strupczewski, J. Szolgay, A. Toumazis, W. Vanneuville, N. Veijalainen, and D. Wilson (2014). “Documentary evidence of past floods in Europe and their utility in flood frequency estimation”. In: *Journal of Hydrology* 517, pp. 963–973. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2014.06.038](https://doi.org/10.1016/j.jhydrol.2014.06.038).
- Kjeldsen, T. R. and D. A. Jones (2009). “An exploratory analysis of error components in hydrological regression modeling”. In: *Water Resources Research* 45.2. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2007WR006283>. ISSN: 1944-7973. DOI: [10.1029/2007WR006283](https://doi.org/10.1029/2007WR006283).
- Kjeldsen, T. R., R. Lamb, and S. D. Blazkova (2011). “Uncertainty in Flood Frequency Analysis”. In: *Applied Uncertainty Analysis for Flood Risk Management*. Imperial college press, pp. 153–197. ISBN: 978-1-84816-270-9.
- Kuczera, G. (1999). “Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference”. In: *Water Resources Research* 35.5. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/1999WR90001>. pp. 1551–1557. ISSN: 1944-7973. DOI: [10.1029/1999WR900012](https://doi.org/10.1029/1999WR900012).
- 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).
- Le Coz, J. (2017). *Quantifying discharges and fluxes of matters in rivers*. Habilitation à Diriger des Recherches. Université Grenoble Alpes, p. 93.
- Macdonald, N., T. R. Kjeldsen, I. Prosdoci, and H. Sangster (2014). “Reassessing flood frequency for the Sussex Ouse, Lewes: the inclusion of historical flood information since AD 1650”. In: *Natural Hazards and Earth System Sciences* 14.10, pp. 2817–2828. ISSN: 1684-9981. DOI: [10.5194/nhess-14-2817-2014](https://doi.org/10.5194/nhess-14-2817-2014).
- Machado, M. J., B. A. Botero, J. López, F. Francés, A. Díez-Herrero, and G. Benito (2015). “Flood frequency analysis of historical flood data under stationary and non-stationary modelling”. In: *Hydrology and Earth System Sciences* 19.6, pp. 2561–2576. ISSN: 1607-7938. DOI: [10.5194/hess-19-2561-2015](https://doi.org/10.5194/hess-19-2561-2015).
- Madsen, H., D. Lawrence, M. Lang, M. Martinkova, and T. R. Kjeldsen (2014). “Review of trend analysis and climate change projections of extreme precipitation and floods in Europe”. In: *Journal of Hydrology* 519, pp. 3634–3650. ISSN: 0022-1694. DOI: [10.1016/j.jhydrol.2014.11.003](https://doi.org/10.1016/j.jhydrol.2014.11.003).
- Mann, H. B. (1945). “Nonparametric Tests Against Trend”. In: *Econometrica* 13.3. Publisher: [Wiley, Econometric Society], pp. 245–259. ISSN: 0012-9682. DOI: [10.2307/1907187](https://doi.org/10.2307/1907187).
- Martins, E. S. and J. R. Stedinger (2000). “Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data”. In: *Water Resources Research* 36.3, pp. 737–744. ISSN: 00431397. DOI: [10.1029/1999WR900330](https://doi.org/10.1029/1999WR900330).
- MEDD (n.d.). *Repères de crues, plateforme collaborative de référence pour le recensement des repères de crues en France*. URL: <https://www.reperesdecruces.developpement-durable.gouv.fr/>.
- Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008). “Stationarity Is Dead: Whither Water Management?” In: *Science* 319.5863, pp. 573–574. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.1151915](https://doi.org/10.1126/science.1151915).
- Neppel, L., B. Renard, M. Lang, P.-A. Ayrat, 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”. In: *Hydrological Sciences Journal* 55.2, pp. 192–208. ISSN: 0262-6667, 2150-3435. DOI: [10.1080/02626660903546092](https://doi.org/10.1080/02626660903546092).
- Nguyen, C., E. Gaume, and O. Payrastre (2014). “Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations”. In: *Journal of Hydrology* 508, pp. 385–396. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2013.09.058](https://doi.org/10.1016/j.jhydrol.2013.09.058).
- Parkes, B. and D. Demeritt (2016). “Defining the hundred year flood: A Bayesian approach for using historic data to reduce uncertainty in flood frequency estimates”. In: *Journal of Hydrology* 540, pp. 1189–1208. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2016.07.025](https://doi.org/10.1016/j.jhydrol.2016.07.025).
- Payrastre, O., E. Gaume, and H. Andrieu (2011). “Usefulness of historical information for flood frequency analyses: Developments based on a case study”. In: *Water Resources Research* 47.8. ISSN: 00431397. DOI: [10.1029/2010WR009812](https://doi.org/10.1029/2010WR009812).
- Payrastre, O. (2005). “Possibility and usefulness flood series reconstruction, for the study of extreme floods in small watersheds. Case study of four small tributaries of the Aude river, France.” PhD thesis. Ecole nationale des ponts et chaussées. 392 pp.
- Pettitt, A. N. (1979). “A Non-Parametric Approach to the Change-Point Problem”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28.2, pp. 126–135. ISSN: 1467-9876. DOI: [10.2307/2346729](https://doi.org/10.2307/2346729).
- Pichard, G., G. Arnaud-Fassetta, V. Moron, and E. Roucaute (2017). “Hydro-climatology of the Lower Rhône Valley: historical flood reconstruction (AD 1300–2000) based on documentary and instrumental sources”. In: *Hydrological Sciences Journal* 62.11, pp. 1772–1795. ISSN: 0262-6667, 2150-3435. DOI: [10.1080/02626667.2017.1349314](https://doi.org/10.1080/02626667.2017.1349314).
- Pichard, G. (1995). “Les crues sur le bas Rhône de 1500 à nos jours. Pour une histoire hydro-climatique”. In: *Méditerranée* 82.3, pp. 105–116. ISSN: 0025-8296. DOI: [10.3406/medit.1995.2908](https://doi.org/10.3406/medit.1995.2908).
- Piotte, O., C. Boura, A. Cazaubon, C. Chaléon, D. Chambon, G. Guillevic, F. Pasquet, C. Perherin, and E. Rimbault (2016). “Collection, storage and management of high-water marks data: praxis and recommendations”. In: *E3S Web of Conferences* 7. Publisher: EDP Sciences, p. 16003. ISSN: 2267-1242. DOI: [10.1051/e3sconf/20160716003](https://doi.org/10.1051/e3sconf/20160716003).
- Prosdocimi, I. (2018). “German tanks and historical records: the estimation of the time coverage of ungauged extreme events”. In: *Stochastic Environmental Research and Risk Assessment* 32.3, pp. 607–622. ISSN: 1436-3240, 1436-3259. DOI: [10.1007/s00477-017-1418-8](https://doi.org/10.1007/s00477-017-1418-8).
- Reis, D. S. and J. R. Stedinger (2005). “Bayesian MCMC flood frequency analysis with historical information”. In: *Journal of Hydrology* 313.1, pp. 97–116. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2005.02.028](https://doi.org/10.1016/j.jhydrol.2005.02.028).
- Renard, B. (2011). “A Bayesian hierarchical approach to regional frequency analysis”. In: *Water Resources Research* 47.11. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2010WR010089>. ISSN: 1944-7973. DOI: [10.1029/2010WR010089](https://doi.org/10.1029/2010WR010089).
- 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”. In: *Water Resources Research* 42.12. ISSN: 00431397. DOI: [10.1029/2005WR004591](https://doi.org/10.1029/2005WR004591).
- Salas, J. D., J. Obeysekera, and R. M. Vogel (2018). “Techniques for assessing water infrastructure for nonstationary extreme events: a review”. In: *Hydrological Sciences Journal* 63.3, pp. 325–352. ISSN: 0262-6667. DOI: [10.1080/02626667.2018.1426858](https://doi.org/10.1080/02626667.2018.1426858).
- St. George, S., A. M. Hefner, and J. Avila (2020). “Paleofloods stage a comeback”. In: *Nature Geoscience* 13.12, pp. 766–768. ISSN: 1752-0894, 1752-0908. DOI: [10.1038/s41561-020-00664-2](https://doi.org/10.1038/s41561-020-00664-2).
- 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).

- Sun, X., M. Thyer, B. Renard, and M. Lang (2014). “A general regional frequency analysis framework for quantifying local-scale climate effects: A case study of ENSO effects on Southeast Queensland rainfall”. In: *Journal of Hydrology* 512, pp. 53–68. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2014.02.025](https://doi.org/10.1016/j.jhydrol.2014.02.025).
- Viglione, A., R. Merz, J. L. Salinas, and G. Blöschl (2013). “Flood frequency hydrology: 3. A Bayesian analysis”. In: *Water Resources Research* 49.2, pp. 675–692. ISSN: 1944-7973. DOI: [10.1029/2011WR010782](https://doi.org/10.1029/2011WR010782).
- WHS (2008). *Flood Estimation Handbook*. WHS. Section: Software. URL: [https://www.hydrosolutions.co.uk/software/flood\\_estimation\\_handbook/](https://www.hydrosolutions.co.uk/software/flood_estimation_handbook/) (visited on 03/10/2023).