

Une (petite) Introduction à la Théorie des Valeurs Extrêmes

Quelles Applications pour l'Hydrologie ?

Mathieu Ribatet

CEMAGREF Lyon, Unité de Recherche Hydrologie-Hydraulique, FRANCE

INRS-ETE, Université de Québec, Chaire HydroStat, CANADA

Institut Polytechnique de Grenoble, FRANCE

Séminaire IMPEC – 11 Juin 2007

“Connaître ou prédire les propriétés statistiques d'événements qui n'ont pas été encore observé (ou très peu). . .”



L'Ouvèze à Vaison-la-Romaine, France. 22 Septembre 1992.

- On se place dans un contexte **probabiliste**
- La variable d'étude sera représentée par une **variable aléatoire**

“Connaître ou prédire les propriétés statistiques d'événements qui n'ont pas été encore observé (ou très peu). . . ”



L'Ouvèze à Vaison-la-Romaine, France. 22 Septembre 1992.

- On se place dans un contexte **probabiliste**
- La variable d'étude sera représentée par une **variable aléatoire**

“Connaître ou prédire les propriétés statistiques d'événements qui n'ont pas été encore observé (ou très peu). . .”



L'Ouvèze à Vaison-la-Romaine, France. 22 Septembre 1992.

- On se place dans un contexte **probabiliste**
- La variable d'étude sera représentée par une **variable aléatoire**

Notion de Variable Aléatoire

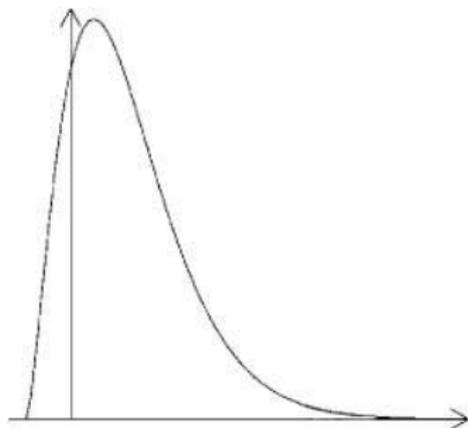
La variable d'étude X sera considérée comme :

- **aléatoire** i.e. la valeur **ne peut être prédite** avec certitude
- son comportement sera dirigé par une **densité f**
- ou par une **fonction de répartition (f.d.r) F**

Le but consiste donc à “connaître” F

Particularité des Valeurs Extrêmes

- Par définition, peu de données seront disponibles
- Les approches probabilistes sont-elles adaptées ?
(souvent on veut beaucoup de données pour “tirer des conclusions”)



Notion de Variable Aléatoire

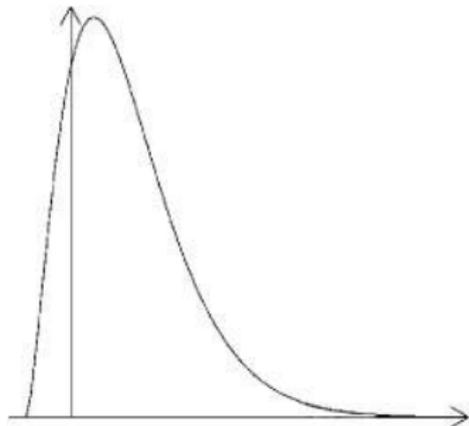
La variable d'étude X sera considérée comme :

- **aléatoire** *i.e.* la valeur **ne peut être prédite** avec certitude
- son comportement sera dirigé par une **densité** f
- ou par une **fonction de répartition** (*f.d.r*) F

Le but consiste donc à “connaître” F

Particularité des Valeurs Extrêmes

- Par définition, peu de données seront disponibles
- Les approches probabilistes sont-elles adaptées ?
(souvent on veut beaucoup de données pour “tirer des conclusions”)



Notion de Variable Aléatoire

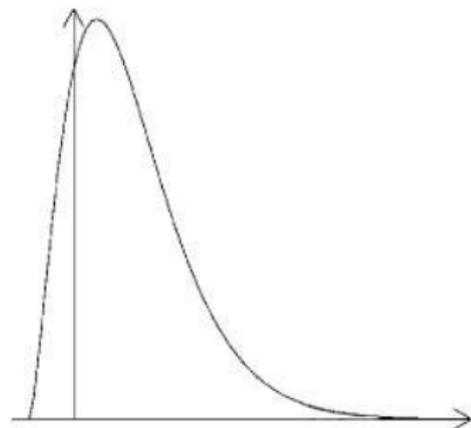
La variable d'étude X sera considérée comme :

- **aléatoire** *i.e.* la valeur **ne peut être prédite** avec certitude
- son comportement sera dirigé par une **densité** f
- ou par une **fonction de répartition** (*f.d.r*) F

Le but consiste donc à “connaître” F

Particularité des Valeurs Extrêmes

- Par définition, peu de données seront disponibles
- Les approches probabilistes sont-elles adaptées ?
(souvent on veut beaucoup de données pour “tirer des conclusions”)



Notion de Variable Aléatoire

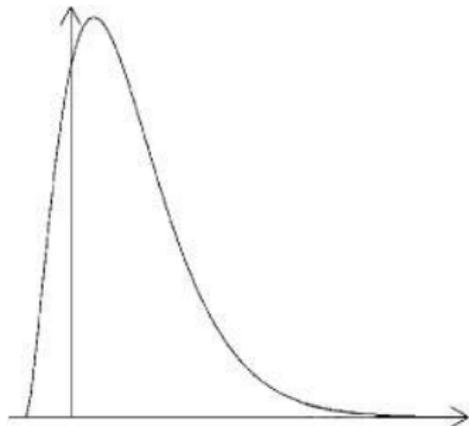
La variable d'étude X sera considérée comme :

- **aléatoire** *i.e.* la valeur **ne peut être prédite** avec certitude
- son comportement sera dirigé par une **densité** f
- ou par une **fonction de répartition** (*f.d.r*) F

Le but consiste donc à “connaître” F

Particularité des Valeurs Extrêmes

- Par définition, peu de données seront disponibles
- Les approches probabilistes sont-elles adaptées ?
(souvent on veut beaucoup de données pour “tirer des conclusions”)



Notion de Variable Aléatoire

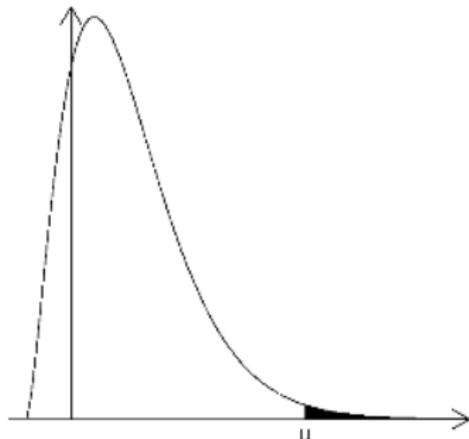
La variable d'étude X sera considérée comme :

- **aléatoire** *i.e.* la valeur **ne peut être prédite** avec certitude
- son comportement sera dirigé par une **densité** f
- ou par une **fonction de répartition** (*f.d.r*) F

Le but consiste donc à “connaître” F

Particularité des Valeurs Extrêmes

- Par définition, peu de données seront disponibles
- Les approches probabilistes sont-elles adaptées ?
(souvent on veut beaucoup de données pour “tirer des conclusions”)



Différences Avec les Statistiques “Classiques”

- Pourquoi n'utilise t'on pas les techniques classiques pour les extrêmes ?
- On pourrait trouver F pour toutes les observations, pas seulement les extrêmes

$$X \sim N(\mu, \sigma^2) \implies \hat{\Pr}[X > x] = 1 - \hat{F}(x)$$

Incohérences

- μ et σ sont **estimés** surtout par les **valeurs “centrales”**
- qualité prédictive du modèle jugée par ces valeurs
- **Différents modèles** conduiront a des **extrapolations très différentes**
- Si seul les **extrêmes sont d'intérêts** pourquoi s'occuper de la **partie centrale ???**

Différences Avec les Statistiques “Classiques”

- Pourquoi n'utilise t'on pas les techniques classiques pour les extrêmes ?
- On pourrait trouver F pour toutes les observations, pas seulement les extrêmes

$$X \sim N(\mu, \sigma^2) \implies \hat{\Pr}[X > x] = 1 - \hat{F}(x)$$

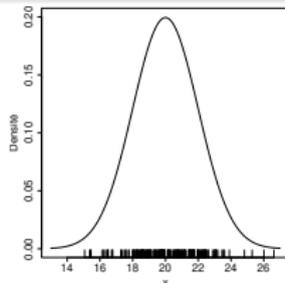
Incohérences

- μ et σ sont **estimés** surtout par les **valeurs “centrales”**
- qualité prédictive du modèle jugée par ces valeurs
- **Différents modèles** conduiront a des **extrapolations très différentes**
- Si seul les **extrêmes sont d'intérêts** pourquoi s'occuper de la **partie centrale** ???

Différences Avec les Statistiques “Classiques”

- Pourquoi n'utilise t'on pas les techniques classiques pour les extrêmes ?
- On pourrait trouver F pour toutes les observations, pas seulement les extrêmes

$$X \sim N(\mu, \sigma^2) \implies \hat{\Pr}[X > x] = 1 - \hat{F}(x)$$



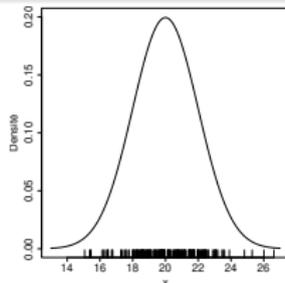
Incohérences

- μ et σ sont **estimés** surtout par les **valeurs “centrales”**
- qualité prédictive du modèle jugée par ces valeurs
- **Différents modèles** conduiront a des **extrapolations très différentes**
- Si seul les **extrêmes sont d'intérêts** pourquoi s'occuper de la **partie centrale ???**

Différences Avec les Statistiques “Classiques”

- Pourquoi n'utilise t'on pas les techniques classiques pour les extrêmes ?
- On pourrait trouver F pour toutes les observations, pas seulement les extrêmes

$$X \sim N(\mu, \sigma^2) \implies \hat{\Pr}[X > x] = 1 - \hat{F}(x)$$



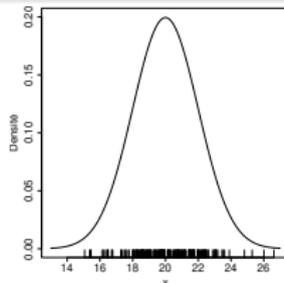
Incohérences

- μ et σ sont **estimés** surtout par les **valeurs “centrales”**
- qualité prédictive du modèle jugée par ces valeurs
- Différents modèles conduiront a des **extrapolations très différentes**
- Si seul les **extrêmes sont d'intérêts** pourquoi s'occuper de la **partie centrale ???**

Différences Avec les Statistiques “Classiques”

- Pourquoi n'utilise t'on pas les techniques classiques pour les extrêmes ?
- On pourrait trouver F pour toutes les observations, pas seulement les extrêmes

$$X \sim N(\mu, \sigma^2) \implies \hat{\Pr}[X > x] = 1 - \hat{F}(x)$$



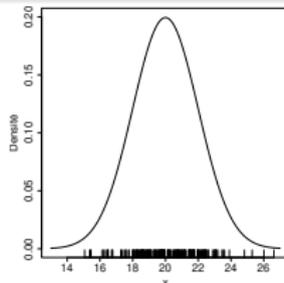
Incohérences

- μ et σ sont **estimés** surtout par les **valeurs “centrales”**
- qualité prédictive du modèle jugée par ces valeurs
- Différents modèles conduiront a des **extrapolations très différentes**
- Si seul les **extrêmes sont d'intérêts** pourquoi s'occuper de la **partie centrale ???**

Différences Avec les Statistiques “Classiques”

- Pourquoi n'utilise t'on pas les techniques classiques pour les extrêmes ?
- On pourrait trouver F pour toutes les observations, pas seulement les extrêmes

$$X \sim N(\mu, \sigma^2) \implies \hat{\Pr}[X > x] = 1 - \hat{F}(x)$$



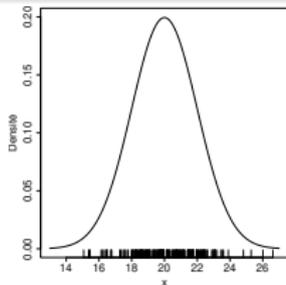
Incohérences

- μ et σ sont **estimés** surtout par les **valeurs “centrales”**
- qualité prédictive du modèle jugée par ces valeurs
- **Différents modèles** conduiront a des **extrapolations très différentes**
- Si seul les **extrêmes sont d'intérêts** pourquoi s'occuper de la **partie centrale ???**

Différences Avec les Statistiques “Classiques”

- Pourquoi n'utilise t'on pas les techniques classiques pour les extrêmes ?
- On pourrait trouver F pour toutes les observations, pas seulement les extrêmes

$$X \sim N(\mu, \sigma^2) \implies \hat{\Pr}[X > x] = 1 - \hat{F}(x)$$



Incohérences

- μ et σ sont **estimés** surtout par les **valeurs “centrales”**
- qualité prédictive du modèle jugée par ces valeurs
- **Différents modèles** conduiront a des **extrapolations très différentes**
- Si seul les **extrêmes sont d'intérêts** pourquoi s'occuper de la **partie centrale** ???

Différences Avec les Statistiques “Classiques”

- Pourquoi n'utilise t'on pas les techniques classiques pour les extrêmes ?
- On pourrait trouver F pour toutes les observations, pas seulement les extrêmes

La Théorie des Valeurs Extrêmes (**EVT**) nous donne un **cadre théorique** pour l'**estimation des queues de distributions**.

Incohérences

- μ et σ sont **estimés** surtout par les **valeurs “centrales”**
- qualité prédictive du modèle jugée par ces valeurs
- **Différents modèles** conduiront a des **extrapolations très différentes**
- Si seul les **extrêmes sont d'intérêts** pourquoi s'occuper de la **partie centrale** ???

Plan de la Présentation

- 1 La Théorie des Valeurs Extrêmes
 - Loi Généralisée des Valeurs Extrêmes
 - Loi Pareto Généralisée
- 2 Estimation Régionale des Débits de Pointes
 - Quelques Rappels
 - Incorporer l'Information Régionale
 - Définir la Loi a Priori
 - Résultats
- 3 Utilisation d'une Loi a Priori Plus Apte à l'Extrapolation
 - Justification
 - Les Sauts Réversibles
 - Résultats

Maxima par Blocs (Maxima Annuels)

- Soit X_1, X_2, \dots, X_n des v.a. indépendantes et identiquement distribuées
- Les X_i ont une f.d.r F
- Posons $M_n = \max \{X_1, X_2, \dots, X_n\}$

Alors

$$\begin{aligned}\Pr[M_n \leq x] &= \Pr[X_1 \leq x, X_2 \leq x, \dots, X_n \leq x] \\ &= \Pr[X_1 \leq x] \dots \Pr[X_n \leq x] \quad (\text{v.a. i.i.d.}) \\ &= [F(x)]^n \quad \text{v.a.i.i.d.}\end{aligned}$$

Souvent F est **inconnue**. Alors :

- on remplace F par \hat{F} mais
 - erreur ε sur $F \Rightarrow F(x)^n$
 - incohérence cf. remarque sur la partie centrale
- on va chercher la loi de M_n plutôt que celle des X_i

Maxima par Blocs (Maxima Annuels)

- Soit X_1, X_2, \dots, X_n des v.a. indépendantes et identiquement distribuées
- Les X_i ont une f.d.r F
- Posons $M_n = \max \{X_1, X_2, \dots, X_n\}$

Alors

$$\begin{aligned}
 \Pr[M_n \leq x] &= \Pr[X_1 \leq x, X_2 \leq x, \dots, X_n \leq x] \\
 &= \Pr[X_1 \leq x] \dots \Pr[X_n \leq x] \quad (\text{v.a. i.i.d.}) \\
 &= [F(x)]^n \quad \text{v.a.i.i.d.}
 \end{aligned}$$

Souvent F est **inconnue**. Alors :

- on remplace F par \hat{F} mais
 - erreur ε sur $F \Rightarrow F(x)^n$
 - incohérence cf. remarque sur la partie centrale
- on va chercher la loi de M_n plutôt que celle des X_i

Maxima par Blocs (Maxima Annuels)

- Soit X_1, X_2, \dots, X_n des v.a. indépendantes et identiquement distribuées
- Les X_i ont une f.d.r F
- Posons $M_n = \max \{X_1, X_2, \dots, X_n\}$

Alors

$$\begin{aligned}\Pr[M_n \leq x] &= \Pr[X_1 \leq x, X_2 \leq x, \dots, X_n \leq x] \\ &= \Pr[X_1 \leq x] \dots \Pr[X_n \leq x] \quad (\text{v.a. i.i.d.}) \\ &= [F(x)]^n \quad \text{v.a.i.i.d.}\end{aligned}$$

Souvent F est **inconnue**. Alors :

- on remplace F par \hat{F} mais
 - erreur ε sur $F \Rightarrow F(x)^n$
 - incohérence cf. remarque sur la partie centrale
- on va chercher la loi de M_n plutôt que celle des X_i

Maxima par Blocs (Maxima Annuels)

- Soit X_1, X_2, \dots, X_n des v.a. indépendantes et identiquement distribuées
- Les X_i ont une f.d.r F
- Posons $M_n = \max \{X_1, X_2, \dots, X_n\}$

Alors

$$\begin{aligned}
 \Pr[M_n \leq x] &= \Pr[X_1 \leq x, X_2 \leq x, \dots, X_n \leq x] \\
 &= \Pr[X_1 \leq x] \dots \Pr[X_n \leq x] \quad (\text{v.a. i.i.d.}) \\
 &= [F(x)]^n \quad \text{v.a.i.i.d.}
 \end{aligned}$$

Souvent F est **inconnue**. Alors :

- on remplace F par \hat{F} mais
 - erreur ε sur $F \Rightarrow F(x)^n$
 - incohérence cf. remarque sur la partie centrale
- on va chercher la loi de M_n plutôt que celle des X_i

Maxima par Blocs (Maxima Annuels)

- Soit X_1, X_2, \dots, X_n des v.a. indépendantes et identiquement distribuées
- Les X_i ont une f.d.r F
- Posons $M_n = \max \{X_1, X_2, \dots, X_n\}$

Alors

$$\begin{aligned}\Pr[M_n \leq x] &= \Pr[X_1 \leq x, X_2 \leq x, \dots, X_n \leq x] \\ &= \Pr[X_1 \leq x] \dots \Pr[X_n \leq x] \quad (\text{v.a. i.i.d.}) \\ &= [F(x)]^n \quad \text{v.a.i.i.d.}\end{aligned}$$

Souvent F est **inconnue**. Alors :

- on remplace F par \hat{F} mais
 - erreur ε sur $F \Rightarrow F(x)^n$
 - incohérence cf. remarque sur la partie centrale
- on va chercher la loi de M_n plutôt que celle des X_i

Maxima par Blocs (Maxima Annuels)

- Soit X_1, X_2, \dots, X_n des v.a. indépendantes et identiquement distribuées
- Les X_i ont une f.d.r F
- Posons $M_n = \max \{X_1, X_2, \dots, X_n\}$

Alors

$$\begin{aligned}\Pr[M_n \leq x] &= \Pr[X_1 \leq x, X_2 \leq x, \dots, X_n \leq x] \\ &= \Pr[X_1 \leq x] \dots \Pr[X_n \leq x] \quad (\text{v.a. i.i.d.}) \\ &= [F(x)]^n \quad \text{v.a.i.i.d.}\end{aligned}$$

Souvent F est **inconnue**. Alors :

- on remplace F par \hat{F} mais
 - erreur ε sur $F \Rightarrow F(x)^n$
 - incohérence cf. remarque sur la partie centrale
- on va chercher la loi de M_n plutôt que celle des X_i

Extremal Types Theorem [Leadbetter et al., 1983]

S'il existe des suites de constantes $\{a_n\}_{n \geq 0}$ et $\{b_n\}_{n \geq 0}$ telles que :

$$\Pr \left[\frac{M_n - b_n}{a_n} \leq x \right] \longrightarrow G(x), \quad n \rightarrow +\infty \quad (1)$$

où G est une f.d.r. (non dégénérée).

Alors G est l'une des 3 f.d.r. suivantes :

Gumbel	$G(x) = \exp[-\exp(-x)],$	$-\infty < x < +\infty$
Fréchet	$G(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0, \alpha > 0 \end{cases}$	
Weibull Négatif	$G(x) = \begin{cases} \exp(-(-x)^\alpha), & x < 0, \alpha > 0 \\ 1, & x \geq 0 \end{cases}$	

Extremal Types Theorem [Leadbetter et al., 1983]

S'il existe des suites de constantes $\{a_n\}_{n \geq 0}$ et $\{b_n\}_{n \geq 0}$ telles que :

$$\Pr \left[\frac{M_n - b_n}{a_n} \leq x \right] \longrightarrow G(x), \quad n \rightarrow +\infty \quad (1)$$

où G est une f.d.r. (non dégénérée).

Alors G est l'une des 3 f.d.r. suivantes :

Gumbel $G(x) = \exp[-\exp(-x)], \quad -\infty < x < +\infty$

Fréchet $G(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0, \alpha > 0 \end{cases}$

Weibull Négatif $G(x) = \begin{cases} \exp(-(-x)^\alpha), & x < 0, \alpha > 0 \\ 1, & x \geq 0 \end{cases}$

Extremal Types Theorem [Leadbetter et al., 1983]

S'il existe des suites de constantes $\{a_n\}_{n \geq 0}$ et $\{b_n\}_{n \geq 0}$ telles que :

$$\Pr \left[\frac{M_n - b_n}{a_n} \leq x \right] \longrightarrow G(x), \quad n \rightarrow +\infty \quad (1)$$

où G est une f.d.r. (non dégénérée).

Alors G est l'une des 3 f.d.r. suivantes :

Gumbel $G(x) = \exp[-\exp(-x)], \quad -\infty < x < +\infty$

Fréchet $G(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0, \alpha > 0 \end{cases}$

Weibull Négatif $G(x) = \begin{cases} \exp(-(-x)^\alpha), & x < 0, \alpha > 0 \\ 1, & x \geq 0 \end{cases}$

Extremal Types Theorem [Leadbetter et al., 1983]

S'il existe des suites de constantes $\{a_n\}_{n \geq 0}$ et $\{b_n\}_{n \geq 0}$ telles que :

$$\Pr \left[\frac{M_n - b_n}{a_n} \leq x \right] \longrightarrow G(x), \quad n \rightarrow +\infty \quad (1)$$

où G est une f.d.r. (non dégénérée).

Alors G est l'une des 3 f.d.r. suivantes :

Gumbel $G(x) = \exp[-\exp(-x)], \quad -\infty < x < +\infty$

Fréchet $G(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0, \alpha > 0 \end{cases}$

Weibull Négatif $G(x) = \begin{cases} \exp(-(-x)^\alpha), & x < 0, \alpha > 0 \\ 1, & x \geq 0 \end{cases}$

La distribution Généralisée des Valeurs Extrêmes

- 3 distributions possibles pour M_n
- Laquelle des 3 faut il considérer ?
- La GEV nous aide...

La Loi GEV [Von Mises, 1954 ; Jenkinson, 1955]

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

où $z_+ = \max(0, x)$ et $\sigma > 0$.

Ainsi, les 3 lois sont englobées au sein de cette GEV, puisque :

Gumbel	$\xi = 0$	$GEV(0, 1, 0)$
Fréchet	$\xi > 0$	$GEV(1, \alpha^{-1}, \alpha^{-1})$
Weibull Négatif	$\xi < 0$	$GEV(-1, \alpha^{-1}, -\alpha^{-1})$

⇒ La loi de M_n converge vers une GEV

La distribution Généralisée des Valeurs Extrêmes

- 3 distributions possibles pour M_n
- Laquelle des 3 faut il considérer ?
- La GEV nous aide...

La Loi GEV [Von Mises, 1954 ; Jenkinson, 1955]

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

où $z_+ = \max(0, x)$ et $\sigma > 0$.

Ainsi, les 3 lois sont englobées au sein de cette GEV, puisque :

Gumbel	$\xi = 0$	$GEV(0, 1, 0)$
Fréchet	$\xi > 0$	$GEV(1, \alpha^{-1}, \alpha^{-1})$
Weibull Négatif	$\xi < 0$	$GEV(-1, \alpha^{-1}, -\alpha^{-1})$

⇒ La loi de M_n converge vers une GEV

La distribution Généralisée des Valeurs Extrêmes

- 3 distributions possibles pour M_n
- Laquelle des 3 faut il considérer ?
- La GEV nous aide...

La Loi GEV [Von Mises, 1954 ; Jenkinson, 1955]

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

où $z_+ = \max(0, x)$ et $\sigma > 0$.

Ainsi, les 3 lois sont englobées au sein de cette GEV, puisque :

Gumbel	$\xi = 0$	$GEV(0, 1, 0)$
Fréchet	$\xi > 0$	$GEV(1, \alpha^{-1}, \alpha^{-1})$
Weibull Négatif	$\xi < 0$	$GEV(-1, \alpha^{-1}, -\alpha^{-1})$

⇒ La loi de M_n converge vers une GEV

La distribution Généralisée des Valeurs Extrêmes

- 3 distributions possibles pour M_n
- Laquelle des 3 faut il considérer ?
- La GEV nous aide...

La Loi GEV [Von Mises, 1954 ; Jenkinson, 1955]

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

où $z_+ = \max(0, x)$ et $\sigma > 0$.

Ainsi, les 3 lois sont englobées au sein de cette GEV, puisque :

Gumbel	$\xi = 0$	$GEV(0, 1, 0)$
Fréchet	$\xi > 0$	$GEV(1, \alpha^{-1}, \alpha^{-1})$
Weibull Négatif	$\xi < 0$	$GEV(-1, \alpha^{-1}, -\alpha^{-1})$

⇒ La loi de M_n converge vers une GEV

La distribution Généralisée des Valeurs Extrêmes

- 3 distributions possibles pour M_n
- Laquelle des 3 faut il considérer ?
- La GEV nous aide...

La Loi GEV [Von Mises, 1954 ; Jenkinson, 1955]

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

où $z_+ = \max(0, x)$ et $\sigma > 0$.

Ainsi, les 3 lois sont englobées au sein de cette GEV, puisque :

Gumbel	$\xi = 0$	$GEV(0, 1, 0)$
Fréchet	$\xi > 0$	$GEV(1, \alpha^{-1}, \alpha^{-1})$
Weibull Négatif	$\xi < 0$	$GEV(-1, \alpha^{-1}, -\alpha^{-1})$

⇒ La loi de M_n converge vers une GEV

La distribution Généralisée des Valeurs Extrêmes

- 3 distributions possibles pour M_n
- Laquelle des 3 faut il considérer ?
- La GEV nous aide...

La Loi GEV [Von Mises, 1954 ; Jenkinson, 1955]

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

où $z_+ = \max(0, x)$ et $\sigma > 0$.

Ainsi, les 3 lois sont englobées au sein de cette GEV, puisque :

Gumbel	$\xi = 0$	$GEV(0, 1, 0)$
Fréchet	$\xi > 0$	$GEV(1, \alpha^{-1}, \alpha^{-1})$
Weibull Négatif	$\xi < 0$	$GEV(-1, \alpha^{-1}, -\alpha^{-1})$

\implies La loi de M_n converge vers une GEV

Que faire avec la GEV ?

- L'EVT nous donne une convergence **asymptotique**
- On ne sera jamais dans le contexte asymptotique. . .

Principe d'une Analyse des Valeurs Extrêmes

Les résultats asymptotiques sont supposés valides pour des niveaux finis (mais assez élevés)

Au final, on ajuste cette GEV à
notre échantillon

(maximum de vraisemblance, moments
pondérés, . . .)

Que faire avec la GEV ?

- L'EVT nous donne une convergence **asymptotique**
- On ne sera jamais dans le contexte asymptotique...

Principe d'une Analyse des Valeurs Extrêmes

Les résultats asymptotiques sont supposés valides pour des niveaux finis (mais assez élevés)

Au final, on ajuste cette GEV à
notre échantillon

(maximum de vraisemblance, moments pondérés, ...)

Que faire avec la GEV ?

- L'EVT nous donne une convergence **asymptotique**
- On ne sera jamais dans le contexte asymptotique. . .

Principe d'une Analyse des Valeurs Extrêmes



Les résultats asymptotiques sont supposés valides pour des niveaux finis (mais assez élevés)

Au final, on ajuste cette GEV à
notre échantillon

(maximum de vraisemblance, moments
pondérés, . . .)

Que faire avec la GEV ?

- L'EVT nous donne une convergence **asymptotique**
- On ne sera jamais dans le contexte asymptotique. . .

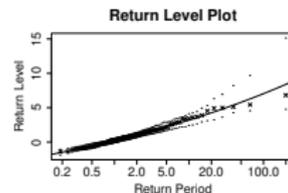
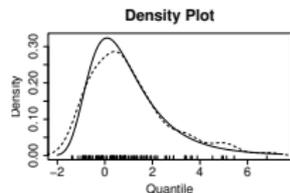
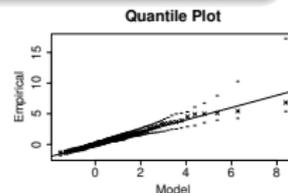
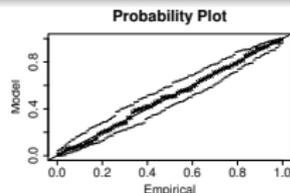
Principe d'une Analyse des Valeurs Extrêmes



Les résultats asymptotiques sont supposés valides pour des niveaux fins (mais assez élevés)

Au final, on ajuste cette GEV à notre échantillon

(maximum de vraisemblance, moments pondérés, ...)



Une Application (synthétique)

- 1 Simuler 365 réalisations i.id. $N(0, 1)$ ($x < -rnorm(365)$)
- 2 Garder le max de cet échantillon ($y < -max(x)$)
- 3 Recommencer (1-2) cent fois

⇒ On a un échantillon de 100 maxima avec lequel on ajuste la GEV

Une Application (synthétique)

- 1 Simuler 365 réalisations i.id. $N(0, 1)$ ($x < -rnorm(365)$)
- 2 Garder le max de cet échantillon ($y < -max(x)$)
- 3 Recommencer (1-2) cent fois

⇒ On a un échantillon de 100 maxima avec lequel on ajuste la GEV

Une Application (synthétique)

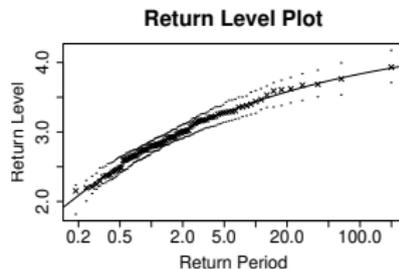
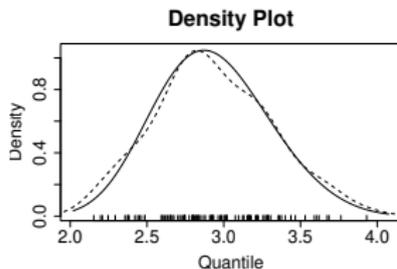
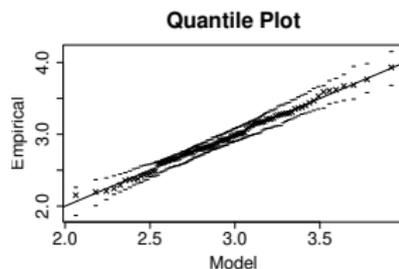
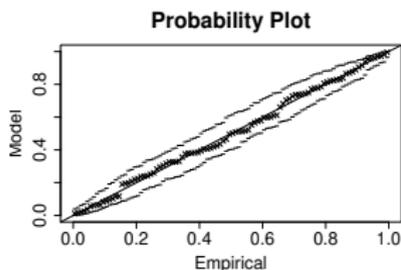
- 1 Simuler 365 réalisations i.id. $N(0, 1)$ ($x < -rnorm(365)$)
- 2 Garder le max de cet échantillon ($y < -max(x)$)
- 3 Recommencer (1-2) cent fois

⇒ On a un échantillon de 100 maxima avec lequel on ajuste la GEV

Une Application (synthétique)

- 1 Simuler 365 réalisations i.id. $N(0, 1)$ ($x < -rnorm(365)$)
- 2 Garder le max de cet échantillon ($y < -max(x)$)
- 3 Recommencer (1-2) cent fois

⇒ On a un échantillon de 100 maxima avec lequel on ajuste la GEV

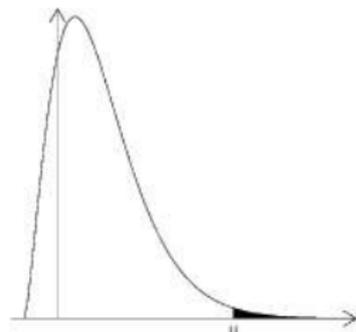


Plan de la Présentation

- 1 La Théorie des Valeurs Extrêmes
 - Loi Généralisée des Valeurs Extrêmes
 - Loi Pareto Généralisée
- 2 Estimation Régionale des Débits de Pointes
 - Quelques Rappels
 - Incorporer l'Information Régionale
 - Définir la Loi a Priori
 - Résultats
- 3 Utilisation d'une Loi a Priori Plus Apte à l'Extrapolation
 - Justification
 - Les Sauts Réversibles
 - Résultats

- Nous avons considéré

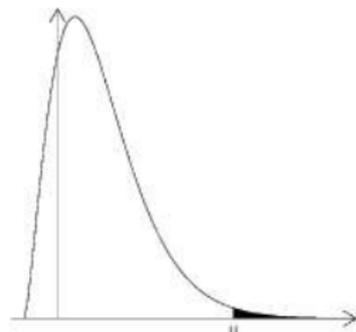
$$M_n = \max \{X_i, i = 1, \dots, n\}$$
- Ces M_n caractérisent la queue de distribution mais
- d'autres v.a. sont aussi représentatives



En particulier,

- $X | X > u$ représente aussi le caractère extrême
- A t on un résultat analogue à la GEV pour $X | X > u$?
- Oui, grâce à la loi **Pareto Généralisée (GPD)**

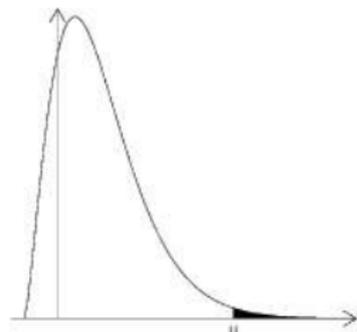
- Nous avons considéré
 $M_n = \max \{X_i, i = 1, \dots, n\}$
- Ces M_n caractérisent la queue de distribution mais
- d'autres v.a. sont aussi représentatives



En particulier,

- $X | X > u$ représente aussi le caractère extrême
- A t on un résultat analogue à la GEV pour $X | X > u$?
- Oui, grâce à la loi **Pareto Généralisée (GPD)**

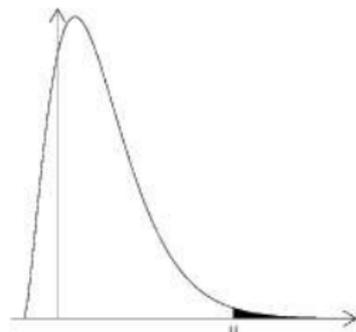
- Nous avons considéré
 $M_n = \max \{X_i, i = 1, \dots, n\}$
- Ces M_n caractérisent la queue de distribution mais
- d'autres v.a. sont aussi représentatives



En particulier,

- $X | X > u$ représente aussi le caractère extrême
- A t on un résultat analogue à la GEV pour $X | X > u$?
- Oui, grâce à la loi **Pareto Généralisée (GPD)**

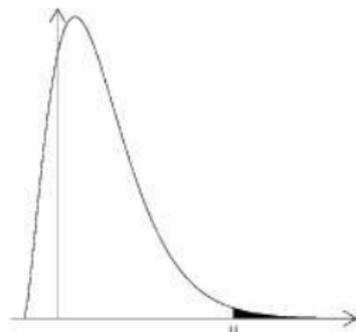
- Nous avons considéré
 $M_n = \max \{X_i, i = 1, \dots, n\}$
- Ces M_n caractérisent la queue de distribution mais
- d'autres v.a. sont aussi représentatives



En particulier,

- $X | X > u$ représente aussi le caractère extrême
- A t on un résultat analogue à la GEV pour $X | X > u$?
- Oui, grâce à la loi **Pareto Généralisée** (GPD)

- Nous avons considéré
 $M_n = \max \{X_i, i = 1, \dots, n\}$
- Ces M_n caractérisent la queue de distribution mais
- d'autres v.a. sont aussi représentatives



En particulier,

- $X | X > u$ représente aussi le caractère extrême
- A t on un résultat analogue à la GEV pour $X | X > u$?
- Oui, grâce à la loi **Pareto Généralisée (GPD)**

Loi Paréto Généralisée (GPD) [Pickands, 1975]

Nous avons vu que $\Pr[(M_n - b_n)/a_n \leq x] \rightarrow G(x), \quad n \rightarrow +\infty$

De même la loi de la v.a. $X | X > u$ vérifie

$$\Pr[X \leq x | X > u] \rightarrow H(x), \quad u \rightarrow u^+$$

où $H(x) = 1 - (1 + \xi \frac{x-u}{\sigma})_+^{-1/\xi}$ pour $x > u$ et $\sigma > 0$.

u seuil, paramètre de position

σ paramètre d'échelle

ξ paramètre de forme

$\xi < 0$ Observations bornées

$\xi = 0$ Loi exponentielle

$\xi > 0$ Queue lourde

Loi Paréto Généralisée (GPD) [Pickands, 1975]

Nous avons vu que $\Pr[(M_n - b_n)/a_n \leq x] \rightarrow G(x), \quad n \rightarrow +\infty$

De même la loi de la v.a. $X | X > u$ vérifie

$$\Pr[X \leq x | X > u] \rightarrow H(x), \quad u \rightarrow u^+$$

où $H(x) = 1 - (1 + \xi \frac{x-u}{\sigma})_+^{-1/\xi}$ pour $x > u$ et $\sigma > 0$.

u seuil, paramètre de position

σ paramètre d'échelle

ξ paramètre de forme

$\xi < 0$ Observations bornées

$\xi = 0$ Loi exponentielle

$\xi > 0$ Queue lourde

Loi Paréto Généralisée (GPD) [Pickands, 1975]

Nous avons vu que $\Pr[(M_n - b_n)/a_n \leq x] \rightarrow G(x), \quad n \rightarrow +\infty$

De même la loi de la v.a. $X | X > u$ vérifie

$$\Pr[X \leq x | X > u] \rightarrow H(x), \quad u \rightarrow u^+$$

où $H(x) = 1 - (1 + \xi \frac{x-u}{\sigma})_+^{-1/\xi}$ pour $x > u$ et $\sigma > 0$.

u seuil, paramètre de position

σ paramètre d'échelle

ξ paramètre de forme

$\xi < 0$ Observations bornées

$\xi = 0$ Loi exponentielle

$\xi > 0$ Queue lourde

Loi Paréto Généralisée (GPD) [Pickands, 1975]

Nous avons vu que $\Pr[(M_n - b_n)/a_n \leq x] \rightarrow G(x), \quad n \rightarrow +\infty$

De même la loi de la v.a. $X | X > u$ vérifie

$$\Pr[X \leq x | X > u] \rightarrow H(x), \quad u \rightarrow u^+$$

où $H(x) = 1 - (1 + \xi \frac{x-u}{\sigma})_+^{-1/\xi}$ pour $x > u$ et $\sigma > 0$.

u seuil, paramètre de position

σ paramètre d'échelle

ξ paramètre de forme

$\xi < 0$ Observations bornées

$\xi = 0$ Loi exponentielle

$\xi > 0$ Queue lourde

Une Application (synthétique)

- 1 Simuler 5000 réalisations i.id. $N(0, 1)$ (`x <- rnorm(5000)`)
- 2 Trier ces réalisations (`x <- sort(x, decreasing = TRUE)`)
- 3 Conserver les 100 plus fortes valeurs (`y <- x[1 : 100]`)

⇒ On a un échantillon de 100 excès au dessus du seuil $x[101]$ avec lequel on ajuste la GPD

Une Application (synthétique)

- 1 Simuler 5000 réalisations i.id. $N(0, 1)$ ($x < - rnorm(5000)$)
- 2 Trier ces réalisations ($x < - sort(x, decreasing = TRUE)$)
- 3 Conserver les 100 plus fortes valeurs ($y < - x[1 : 100]$)

⇒ On a un échantillon de 100 excès au dessus du seuil $x[101]$ avec lequel on ajuste la GPD

Une Application (synthétique)

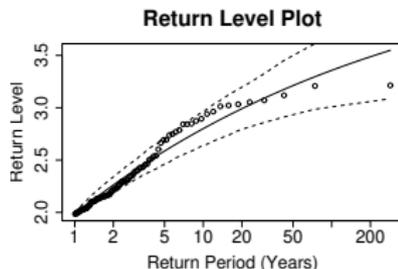
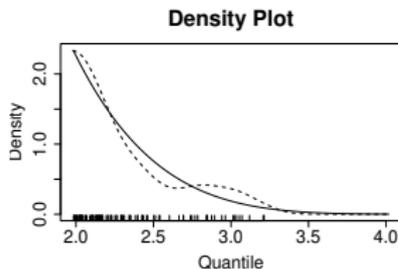
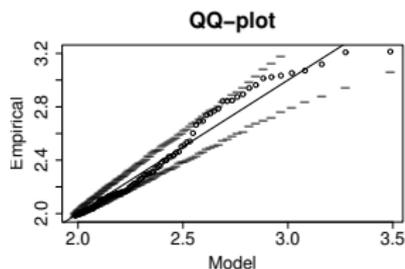
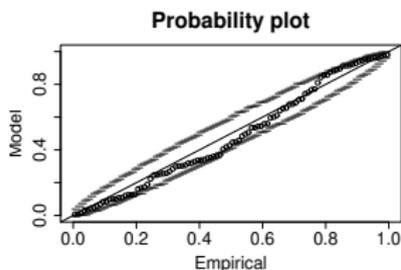
- 1 Simuler 5000 réalisations i.id. $N(0, 1)$ (`x <- rnorm(5000)`)
- 2 Trier ces réalisations (`x <- sort(x, decreasing = TRUE)`)
- 3 Conserver les 100 plus fortes valeurs (`y <- x[1 : 100]`)

⇒ On a un échantillon de 100 excès au dessus du seuil `x[101]` avec lequel on ajuste la GPD

Une Application (synthétique)

- 1 Simuler 5000 réalisations i.id. $N(0, 1)$ ($x < -rnorm(5000)$)
- 2 Trier ces réalisations ($x < -sort(x, decreasing = TRUE)$)
- 3 Conserver les 100 plus fortes valeurs ($y < -x[1 : 100]$)

⇒ On a un échantillon de 100 excès au dessus du seuil $x[101]$ avec lequel on ajuste la GPD



De la Théorie à la Pratique

- Théorie \implies distributions sont (asymptotiquement) connues
- Peu de données \implies estimations trop incertaines
- On veut des estimations ... Que faire ???

De manière globale, il y a deux approches :

- Utiliser de l'information additionnelle dans la procédure d'estimation
- Travail direct sur les estimateurs [Juarez, 2004], [Zhang, 2007]

Dans cette présentation, nous allons distinguer le premier point uniquement¹ :

- Utilisation d'info supplémentaire : Information Régionale
- Travail sur les estimateurs : Modéliser tous les excès

¹Deuxième point sera traité demain

De la Théorie à la Pratique

- Théorie \implies distributions sont (asymptotiquement) connues
- Peu de données \implies estimations trop incertaines
- On veut des estimations ... Que faire ???

De manière globale, il y a deux approches :

- Utiliser de l'information additionnelle dans la procédure d'estimation
- Travail direct sur les estimateurs [Juarez, 2004], [Zhang, 2007]

Dans cette présentation, nous allons distinguer le premier point uniquement¹ :

- Utilisation d'info supplémentaire : Information Régionale
- Travail sur les estimateurs : Modéliser tous les excès

¹Deuxième point sera traité demain

De la Théorie à la Pratique

- Théorie \implies distributions sont (asymptotiquement) connues
- Peu de données \implies estimations trop incertaines
- On veut des estimations ... Que faire ???

De manière globale, il y a deux approches :

- Utiliser de l'information additionnelle dans la procédure d'estimation
- Travail direct sur les estimateurs [Juarez, 2004], [Zhang, 2007]

Dans cette présentation, nous allons distinguer le premier point uniquement¹ :

- Utilisation d'info supplémentaire : Information Régionale
- Travail sur les estimateurs : Modéliser tous les excès

¹Deuxième point sera traité demain

De la Théorie à la Pratique

- Théorie \implies distributions sont (asymptotiquement) connues
- Peu de données \implies estimations trop incertaines
- On veut des estimations ... Que faire ???

De manière globale, il y a deux approches :

- Utiliser de l'information additionnelle dans la procédure d'estimation
- Travail direct sur les estimateurs [Juarez, 2004], [Zhang, 2007]

Dans cette présentation, nous allons distinguer le premier point uniquement¹ :

- Utilisation d'info supplémentaire : Information Régionale
- Travail sur les estimateurs : Modéliser tous les excès

¹Deuxième point sera traité demain

De la Théorie à la Pratique

- Théorie \implies distributions sont (asymptotiquement) connues
- Peu de données \implies estimations trop incertaines
- On veut des estimations ... Que faire ???

De manière globale, il y a deux approches :

- Utiliser de l'information additionnelle dans la procédure d'estimation
- Travail direct sur les estimateurs [Juarez, 2004], [Zhang, 2007]

Dans cette présentation, nous allons distinguer le premier point uniquement¹ :

- Utilisation d'info supplémentaire : Information Régionale
- Travail sur les estimateurs : Modéliser tous les excès

¹Deuxième point sera traité demain

De la Théorie à la Pratique

- Théorie \implies distributions sont (asymptotiquement) connues
- Peu de données \implies estimations trop incertaines
- On veut des estimations ... Que faire ???

De manière globale, il y a deux approches :

- Utiliser de l'information additionnelle dans la procédure d'estimation
- Travail direct sur les estimateurs [Juarez, 2004], [Zhang, 2007]

Dans cette présentation, nous allons distinguer le premier point uniquement¹ :

- Utilisation d'info supplémentaire : Information Régionale
- Travail sur les estimateurs : Modéliser tous les excès

¹Deuxième point sera traité demain

De la Théorie à la Pratique

- Théorie \implies distributions sont (asymptotiquement) connues
- Peu de données \implies estimations trop incertaines
- On veut des estimations ... Que faire ???

De manière globale, il y a deux approches :

- Utiliser de l'information additionnelle dans la procédure d'estimation
- Travail direct sur les estimateurs [Juarez, 2004], [Zhang, 2007]

Dans cette présentation, nous allons distinguer le premier point uniquement¹ :

- Utilisation d'info supplémentaire : **Information Régionale**
- Travail sur les estimateurs : **Modéliser tous les excès**

¹Deuxième point sera traité demain

De la Théorie à la Pratique

- Théorie \implies distributions sont (asymptotiquement) connues
- Peu de données \implies estimations trop incertaines
- On veut des estimations ... Que faire ???

De manière globale, il y a deux approches :

- Utiliser de l'information additionnelle dans la procédure d'estimation
- Travail direct sur les estimateurs [Juarez, 2004], [Zhang, 2007]

Dans cette présentation, nous allons distinguer le premier point uniquement¹ :

- Utilisation d'info supplémentaire : **Information Régionale**
- Travail sur les estimateurs : **Modéliser tous les excès**

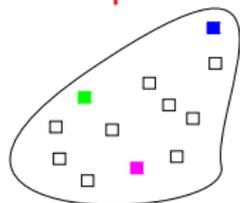
¹Deuxième point sera traité demain

Plan de la Présentation

- 1 La Théorie des Valeurs Extrêmes
 - Loi Généralisée des Valeurs Extrêmes
 - Loi Pareto Généralisée
- 2 Estimation Régionale des Débits de Pointes
 - Quelques Rappels
 - Incorporer l'Information Régionale
 - Définir la Loi a Priori
 - Résultats
- 3 Utilisation d'une Loi a Priori Plus Apte à l'Extrapolation
 - Justification
 - Les Sauts Réversibles
 - Résultats

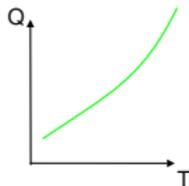
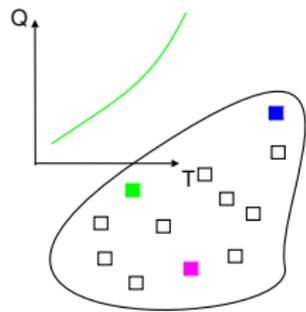
Concept de l'Indice de Crue [Dalrymple, 1960]

Au sein d'une **région homogène**, toutes les distributions sont **identiques à un facteur d'échelle près** (propre au site)



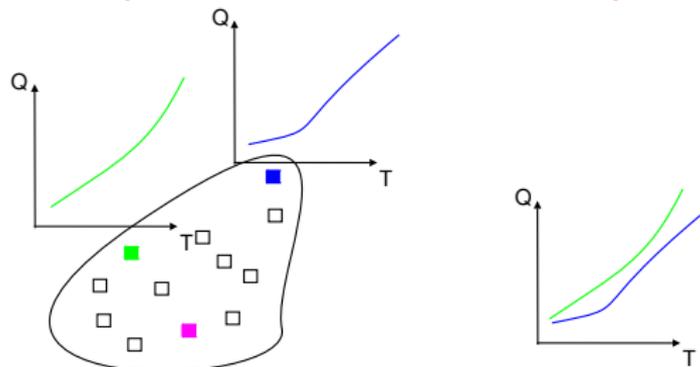
Concept de l'Indice de Crue [Dalrymple, 1960]

Au sein d'une **région homogène**, toutes les distributions sont **identiques à un facteur d'échelle près** (propre au site)



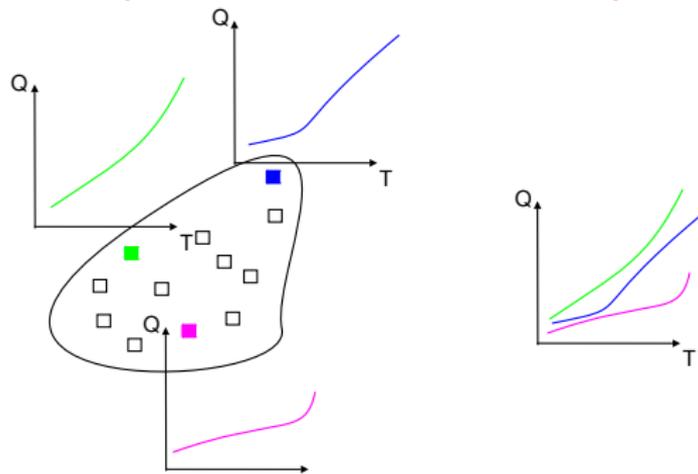
Concept de l'Indice de Crue [Dalrymple, 1960]

Au sein d'une **région homogène**, toutes les distributions sont **identiques à un facteur d'échelle près** (propre au site)



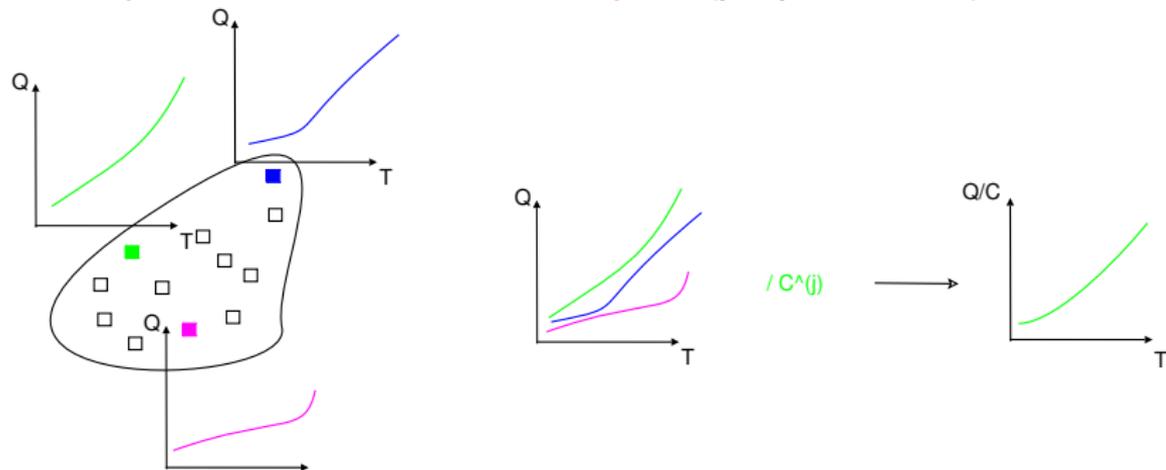
Concept de l'Indice de Crue [Dalrymple, 1960]

Au sein d'une **région homogène**, toutes les distributions sont **identiques à un facteur d'échelle près** (propre au site)



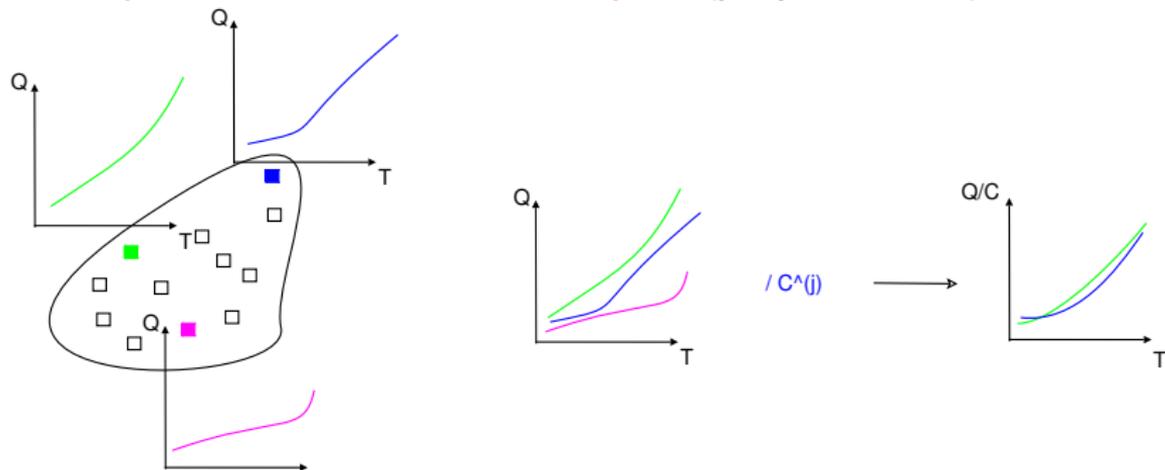
Concept de l'Indice de Crue [Dalrymple, 1960]

Au sein d'une **région homogène**, toutes les distributions sont **identiques à un facteur d'échelle près** (propre au site)



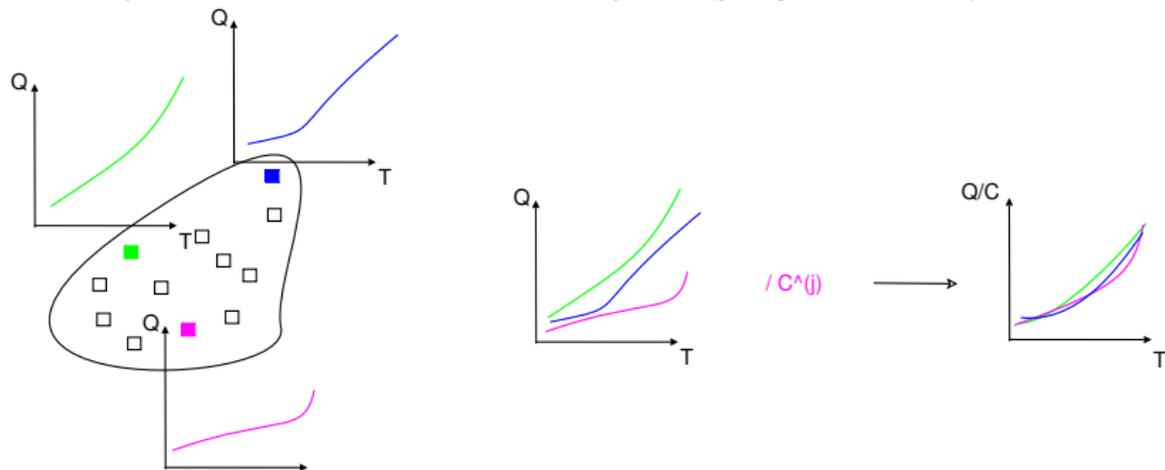
Concept de l'Indice de Crue [Dalrymple, 1960]

Au sein d'une **région homogène**, toutes les distributions sont **identiques à un facteur d'échelle près** (propre au site)



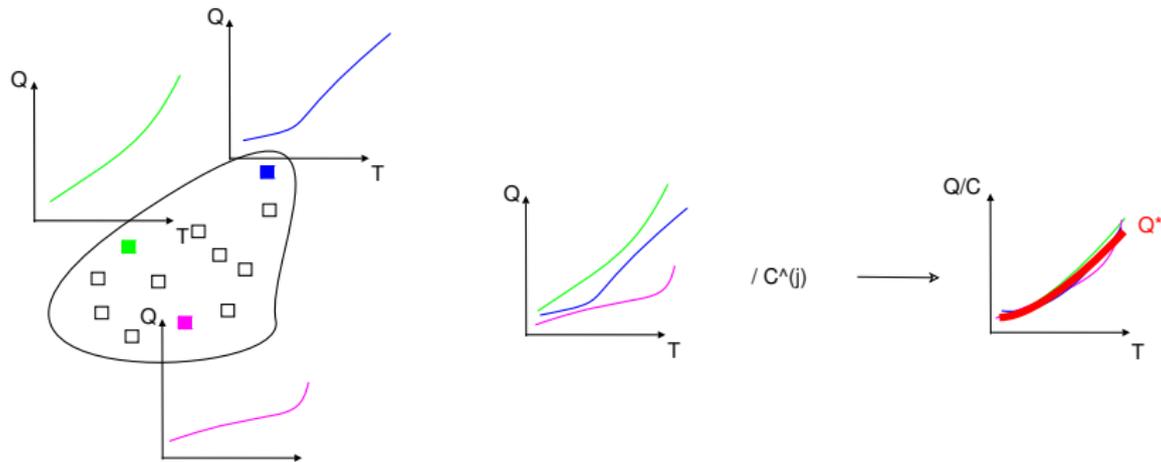
Concept de l'Indice de Crue [Dalrymple, 1960]

Au sein d'une **région homogène**, toutes les distributions sont **identiques à un facteur d'échelle près** (propre au site)



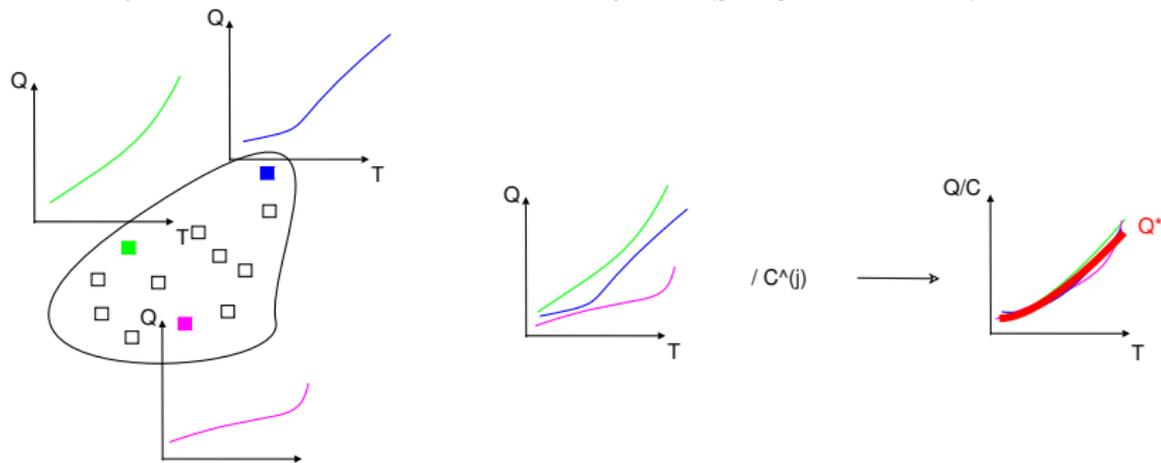
Concept de l'Indice de Crue [Dalrymple, 1960]

Au sein d'une **région homogène**, toutes les distributions sont **identiques à un facteur d'échelle près** (propre au site)



Concept de l'Indice de Crue [Dalrymple, 1960]

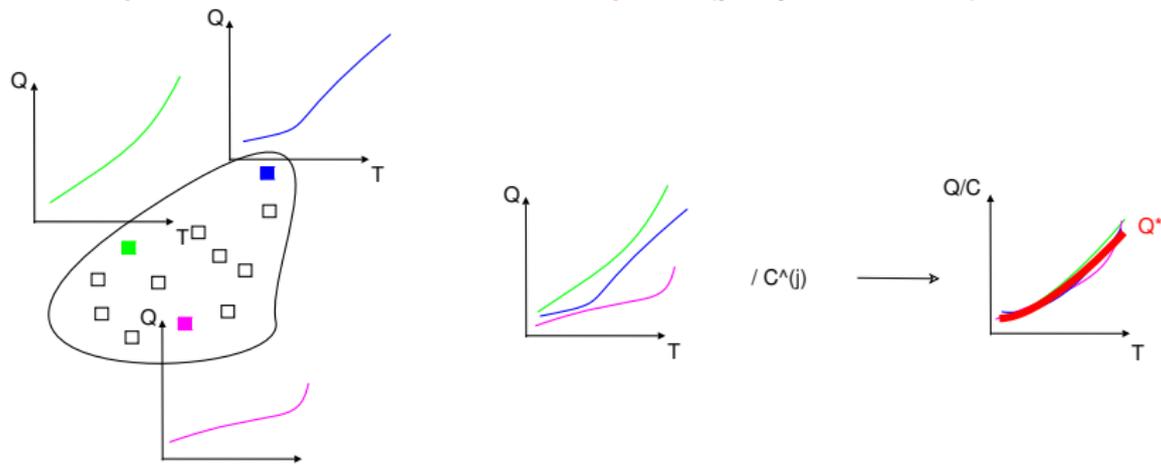
Au sein d'une **région homogène**, toutes les distributions sont **identiques à un facteur d'échelle près** (propre au site)



Q^* est cette unique loi dite **loi régionale**

Concept de l'Indice de Crue [Dalrymple, 1960]

Au sein d'une **région homogène**, toutes les distributions sont **identiques à un facteur d'échelle près** (propre au site)



Q^* est cette unique loi dite **loi régionale**

Il faut donc construire ces **régions homogènes**

Tester le degré d'homogénéité

La statistique H_1 [Hosking and Wallis, 1997]

- On dispose d'information régionale : Est elle pertinente ?
- Nécessité de tester le degré d'homogénéité de la région
- Utilisation de la statistique H_1 [Hosking and Wallis, 1997]

Test d'Hosking and Wallis [1997]

- 1 Ajuster une loi Kappa (4 paramètres) régionale f_{reg}
- 2 Simuler N régions stochastiques (même caractéristiques) sous l'hypothèse de l'indice de crue selon f_{reg}
- 3 "Comparer" distribution des L -moments simulés et réels

Tester le degré d'homogénéité

La statistique H_1 [Hosking and Wallis, 1997]

- On dispose d'information régionale : Est elle pertinente ?
- Nécessité de tester le degré d'homogénéité de la région
- Utilisation de la statistique H_1 [Hosking and Wallis, 1997]

Test d'Hosking and Wallis [1997]

- 1 Ajuster une loi Kappa (4 paramètres) régionale f_{reg}
- 2 Simuler N régions stochastiques (même caractéristiques) sous l'hypothèse de l'indice de crue selon f_{reg}
- 3 "Comparer" distribution des L -moments simulés et réels

Tester le degré d'homogénéité

La statistique H_1 [Hosking and Wallis, 1997]

- On dispose d'information régionale : Est elle pertinente ?
- Nécessité de tester le degré d'homogénéité de la région
- Utilisation de la statistique H_1 [Hosking and Wallis, 1997]

Test d'Hosking and Wallis [1997]

- 1 Ajuster une loi Kappa (4 paramètres) régionale f_{reg}
- 2 Simuler N régions stochastiques (même caractéristiques) sous l'hypothèse de l'indice de crue selon f_{reg}
- 3 "Comparer" distribution des L -moments simulés et réels

Tester le degré d'homogénéité

La statistique H_1 [Hosking and Wallis, 1997]

- On dispose d'information régionale : Est elle pertinente ?
- Nécessité de tester le degré d'homogénéité de la région
- Utilisation de la statistique H_1 [Hosking and Wallis, 1997]

Test d'Hosking and Wallis [1997]

- 1 Ajuster une loi Kappa (4 paramètres) régionale f_{reg}
- 2 Simuler N régions stochastiques (même caractéristiques) sous l'hypothèse de l'indice de crue selon f_{reg}
- 3 "Comparer" distribution des L -moments simulés et réels

Tester le degré d'homogénéité

La statistique H_1 [Hosking and Wallis, 1997]

- On dispose d'information régionale : Est elle pertinente ?
- Nécessité de tester le degré d'homogénéité de la région
- Utilisation de la statistique H_1 [Hosking and Wallis, 1997]

Test d'Hosking and Wallis [1997]

- 1 Ajuster une loi Kappa (4 paramètres) régionale f_{reg}
- 2 Simuler N régions stochastiques (même caractéristiques) sous l'hypothèse de l'indice de crue selon f_{reg}
- 3 "Comparer" distribution des L -moments simulés et réels

Tester le degré d'homogénéité

La statistique H_1 [Hosking and Wallis, 1997]

- On dispose d'information régionale : Est elle pertinente ?
- Nécessité de tester le degré d'homogénéité de la région
- Utilisation de la statistique H_1 [Hosking and Wallis, 1997]

Test d'Hosking and Wallis [1997]

- 1 Ajuster une loi Kappa (4 paramètres) régionale f_{reg}
- 2 Simuler N régions stochastiques (même caractéristiques) sous l'hypothèse de l'indice de crue selon f_{reg}
- 3 "Comparer" distribution des L -moments simulés et réels

Critique de l'Indice de Crue

- Modèle **sans vraisemblance** [Katz et al., 2002]
⇒ **Connaissance des incertitudes** d'estimations ???
- Propriétés d'invariance **non vérifiée** [Gupta et al., 1994]
[Robinson and Sivapalan, 1997]
⇒ **Biais, réduction artificielle** de la variance
- Utilisation **non optimale** de l'information (poids identiques pour toutes les observations)

Critique de l'Indice de Crue

- Modèle **sans vraisemblance** [Katz et al., 2002]
 - ⇒ **Connaissance des incertitudes** d'estimations ???
- Propriétés d'invariance **non vérifiée** [Gupta et al., 1994] [Robinson and Sivapalan, 1997]
 - ⇒ **Biais, réduction artificielle** de la variance
- Utilisation **non optimale** de l'information (poids identiques pour toutes les observations)

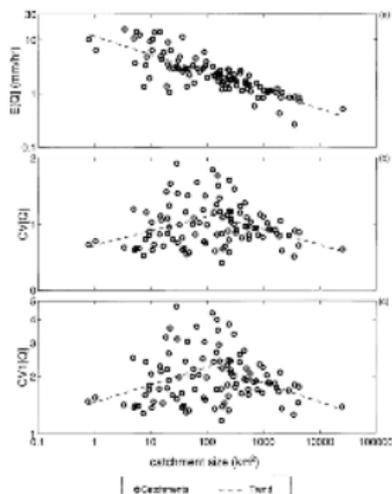


Figure tirée de Robinson and Sivapalan [1997]

Critique de l'Indice de Crue

- Modèle **sans vraisemblance** [Katz et al., 2002]
 - ⇒ **Connaissance des incertitudes** d'estimations ???
- Propriétés d'invariance **non vérifiée** [Gupta et al., 1994] [Robinson and Sivapalan, 1997]
 - ⇒ **Biais, réduction artificielle** de la variance
- Utilisation **non optimale** de l'information (poids identiques pour toutes les observations)

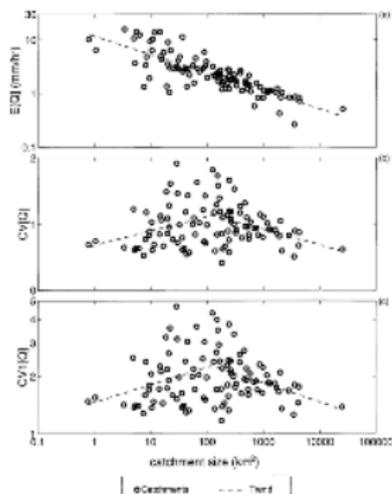


Figure tirée de Robinson and Sivapalan [1997]

Critique de l'Indice de Crue

- Modèle **sans vraisemblance** [Katz et al., 2002]
⇒ **Connaissance des incertitudes** d'estimations ???
- Propriétés d'invariance **non vérifiée** [Gupta et al., 1994]
[Robinson and Sivapalan, 1997]
⇒ **Biais, réduction artificielle** de la variance
- Utilisation **non optimale** de l'information (poids identiques pour toutes les observations)

Nos objectifs

La notion d'information régionale reste toujours pertinente, utilisons la... mais

- Nécessité de relaxer les hypothèses
- Nécessité d'introduire un modèle statistique bien fondé
- Distinguer l'information locale et régionale

Critique de l'Indice de Crue

- Modèle **sans vraisemblance** [Katz et al., 2002]
⇒ **Connaissance des incertitudes** d'estimations ???
- Propriétés d'invariance **non vérifiée** [Gupta et al., 1994]
[Robinson and Sivapalan, 1997]
⇒ **Biais, réduction artificielle** de la variance
- Utilisation **non optimale** de l'information (poids identiques pour toutes les observations)

Nos objectifs

La notion d'information régionale reste toujours pertinente, utilisons la... mais

- Nécessité de relaxer les hypothèses
- Nécessité d'introduire un modèle statistique bien fondé
- Distinguer l'information locale et régionale

Critique de l'Indice de Crue

- Modèle **sans vraisemblance** [Katz et al., 2002]
⇒ **Connaissance des incertitudes** d'estimations ???
- Propriétés d'invariance **non vérifiée** [Gupta et al., 1994]
[Robinson and Sivapalan, 1997]
⇒ **Biais, réduction artificielle** de la variance
- Utilisation **non optimale** de l'information (poids identiques pour toutes les observations)

Nos objectifs

La notion d'information régionale reste toujours pertinente, utilisons la... mais

- Nécessité de relaxer les hypothèses
- Nécessité d'introduire un modèle statistique bien fondé
- Distinguer l'information locale et régionale

Plan de la Présentation

- 1 La Théorie des Valeurs Extrêmes
 - Loi Généralisée des Valeurs Extrêmes
 - Loi Pareto Généralisée
- 2 Estimation Régionale des Débits de Pointes
 - Quelques Rappels
 - Incorporer l'Information Régionale
 - Définir la Loi a Priori
 - Résultats
- 3 Utilisation d'une Loi a Priori Plus Apte à l'Extrapolation
 - Justification
 - Les Sauts Réversibles
 - Résultats

Incorporer l'Information Supplémentaire

Théorème de Bayes (1763)

$$\pi(\theta|x) = \frac{\pi(\theta)\pi(x; \theta)}{\int_{\Theta} \pi(\theta)\pi(x; \theta)d\theta} \quad (2)$$

où $\pi(\theta|x)$ et $\pi(\theta)$ sont les distributions a posteriori et a priori respectivement et $\pi(x; \theta)$ est la fonction de vraisemblance.

- Information Régionale $\implies \pi(\theta)$
- Données site cible $\implies \pi(x; \theta)$
- Estimation $\implies \pi(\theta|x)$

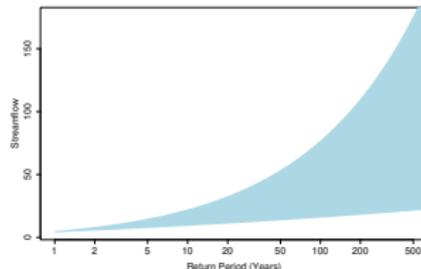
Incorporer l'Information Supplémentaire

Théorème de Bayes (1763)

$$\pi(\theta|x) = \frac{\pi(\theta)\pi(x; \theta)}{\int_{\Theta} \pi(\theta)\pi(x; \theta)d\theta} \quad (2)$$

où $\pi(\theta|x)$ et $\pi(\theta)$ sont les distributions a posteriori et a priori respectivement et $\pi(x; \theta)$ est la fonction de vraisemblance.

- Information Régionale $\implies \pi(\theta)$
- Données site cible $\implies \pi(x; \theta)$
- Estimation $\implies \pi(\theta|x)$



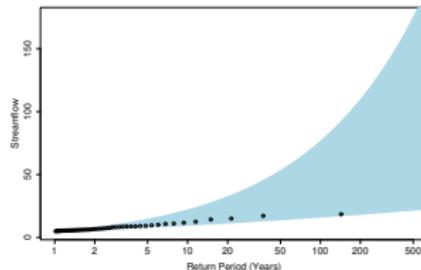
Incorporer l'Information Supplémentaire

Théorème de Bayes (1763)

$$\pi(\theta|x) = \frac{\pi(\theta)\pi(x; \theta)}{\int_{\Theta} \pi(\theta)\pi(x; \theta)d\theta} \quad (2)$$

où $\pi(\theta|x)$ et $\pi(\theta)$ sont les distributions a posteriori et a priori respectivement et $\pi(x; \theta)$ est la fonction de vraisemblance.

- Information Régionale $\implies \pi(\theta)$
- Données site cible $\implies \pi(x; \theta)$
- Estimation $\implies \pi(\theta|x)$



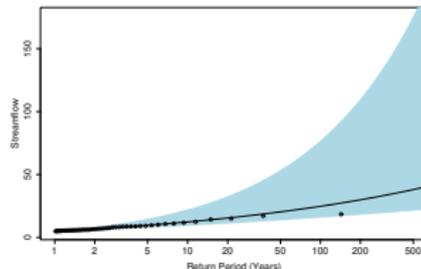
Incorporer l'Information Supplémentaire

Théorème de Bayes (1763)

$$\pi(\theta|x) = \frac{\pi(\theta)\pi(x; \theta)}{\int_{\Theta} \pi(\theta)\pi(x; \theta)d\theta} \quad (2)$$

où $\pi(\theta|x)$ et $\pi(\theta)$ sont les distributions a posteriori et a priori respectivement et $\pi(x; \theta)$ est la fonction de vraisemblance.

- Information Régionale $\implies \pi(\theta)$
- Données site cible $\implies \pi(x; \theta)$
- Estimation $\implies \pi(\theta|x)$



Plan de la Présentation

- 1 La Théorie des Valeurs Extrêmes
 - Loi Généralisée des Valeurs Extrêmes
 - Loi Pareto Généralisée
- 2 Estimation Régionale des Débits de Pointes
 - Quelques Rappels
 - Incorporer l'Information Régionale
 - Définir la Loi a Priori
 - Résultats
- 3 Utilisation d'une Loi a Priori Plus Apte à l'Extrapolation
 - Justification
 - Les Sauts Réversibles
 - Résultats

Construction de la loi a priori

On considère ici des modèles Bayésiens **paramétriques**
⇒ loi a priori connue via ses **hyperparamètres**

Que veut-on d'une loi a priori ?

- Adaptée à notre information régionale
- Adaptée aux données d'études

Contraintes sur la loi a priori

- Débit > 0 ⇒ asymétrie sur le paramètre de position μ
- Paramètre d'échelle $\sigma > 0$ ⇒ asymétrie sur σ

Construction de la loi a priori

On considère ici des modèles Bayésiens **paramétriques**
⇒ loi a priori connue via ses **hyperparamètres**

Que veut-on d'une loi a priori ?

- Adaptée à notre information régionale
- Adaptée aux données d'études

Contraintes sur la loi a priori

- Débit > 0 ⇒ asymétrie sur le paramètre de position μ
- Paramètre d'échelle $\sigma > 0$ ⇒ asymétrie sur σ

Construction de la loi a priori

On considère ici des modèles Bayésiens **paramétriques**
⇒ loi a priori connue via ses **hyperparamètres**

Que veut-on d'une loi a priori ?

- Adaptée à notre information régionale
- Adaptée aux données d'études

Contraintes sur la loi a priori

- Débit > 0 ⇒ asymétrie sur le paramètre de position μ
- Paramètre d'échelle $\sigma > 0$ ⇒ asymétrie sur σ

Construction de la loi a priori (2)

Formulation du modèle

Dans un premier travail [Ribatet et al., 2007], nous avons introduit la loi suivante :

$$\pi(\theta) \propto \frac{1}{u\sigma} \exp \left[\frac{1}{2}(\theta' - \gamma)^T \Sigma^{-1}(\theta' - \gamma) \right] \quad (3)$$

où $\theta' = (\log u, \log \sigma, \xi)$.

γ et Σ sont les hyperparamètres.

Construction de la loi a priori (2)

Formulation du modèle

Dans un premier travail [Ribatet et al., 2007], nous avons introduit la loi suivante :

$$\pi(\theta) \propto \frac{1}{u\sigma} \exp \left[\frac{1}{2}(\theta' - \gamma)^T \Sigma^{-1}(\theta' - \gamma) \right] \quad (3)$$

où $\theta' = (\log u, \log \sigma, \xi)$.

γ et Σ sont les hyperparamètres.

- $\gamma = (\gamma_1, \gamma_2, \gamma_3)$, $\Sigma = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}$
- γ_i et d_i sont estimés grâce à l'information régionale.

Construction de la loi a priori (2)

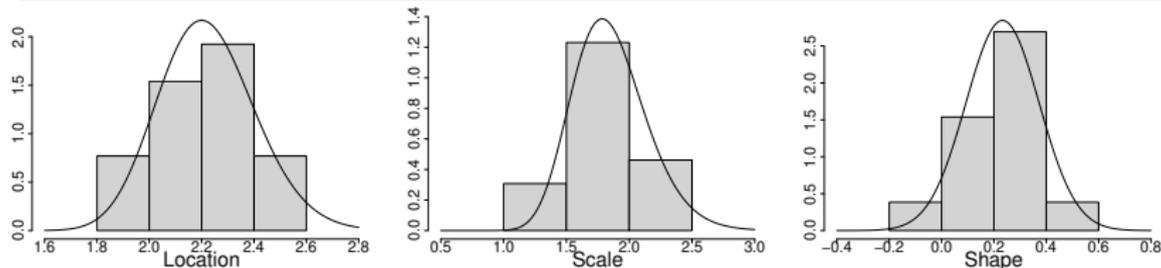
Formulation du modèle

Dans un premier travail [Ribatet et al., 2007], nous avons introduit la loi suivante :

$$\pi(\theta) \propto \frac{1}{u\sigma} \exp \left[\frac{1}{2} (\theta' - \gamma)^T \Sigma^{-1} (\theta' - \gamma) \right] \quad (3)$$

où $\theta' = (\log u, \log \sigma, \xi)$.

γ et Σ sont les hyperparamètres.



Construction de la loi a priori (2)

Formulation du modèle

Dans un premier travail [Ribatet et al., 2007], nous avons introduit la loi suivante :

$$\pi(\theta) \propto \frac{1}{u\sigma} \exp \left[\frac{1}{2}(\theta' - \gamma)^T \Sigma^{-1}(\theta' - \gamma) \right] \quad (3)$$

où $\theta' = (\log u, \log \sigma, \xi)$.

γ et Σ sont les hyperparamètres.

- **Amélioration** : Gestion de la dépendance entre u , σ et ξ
⇒ réduction variance de la CM ⇒ d'estimation
- **Difficulté** : Grande dimension ⇒ Fiabilité des Copules ???

Construction de la loi a priori (2)

Formulation du modèle

Dans un premier travail [Ribatet et al., 2007], nous avons introduit la loi suivante :

$$\pi(\theta) \propto \frac{1}{u\sigma} \exp \left[\frac{1}{2}(\theta' - \gamma)^T \Sigma^{-1}(\theta' - \gamma) \right] \quad (3)$$

où $\theta' = (\log u, \log \sigma, \xi)$.

γ et Σ sont les hyperparamètres.

- **Amélioration** : Gestion de la dépendance entre u , σ et ξ
⇒ réduction variance de la CM ⇒ d'estimation
- **Difficulté** : Grande dimension ⇒ Fiabilité des Copules ???

Ajuster la loi a priori selon l'information régionale

Définition : Pseudo paramètres au site cible

$$\tilde{u}^{(i)} = C^{(j)} u_*^{(i)}, \quad \tilde{\sigma}^{(i)} = C^{(j)} \sigma_*^{(i)}, \quad \tilde{\xi}^{(i)} = \xi_*^{(i)}$$

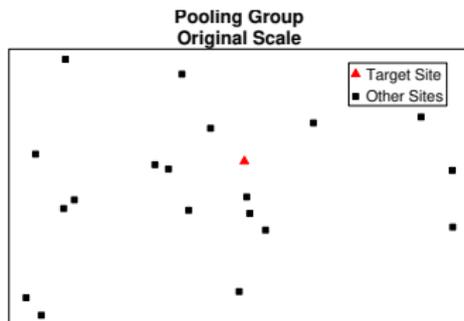
sont appelés pseudo paramètres au site cible pour les paramètres de position, d'échelle et de forme respectivement.

Ajuster la loi a priori selon l'information régionale

Définition : Pseudo paramètres au site cible

$$\tilde{u}^{(i)} = C^{(j)} u_*^{(i)}, \quad \tilde{\sigma}^{(i)} = C^{(j)} \sigma_*^{(i)}, \quad \tilde{\xi}^{(i)} = \xi_*^{(i)}$$

sont appelés pseudo paramètres au site cible pour les paramètres de position, d'échelle et de forme respectivement.

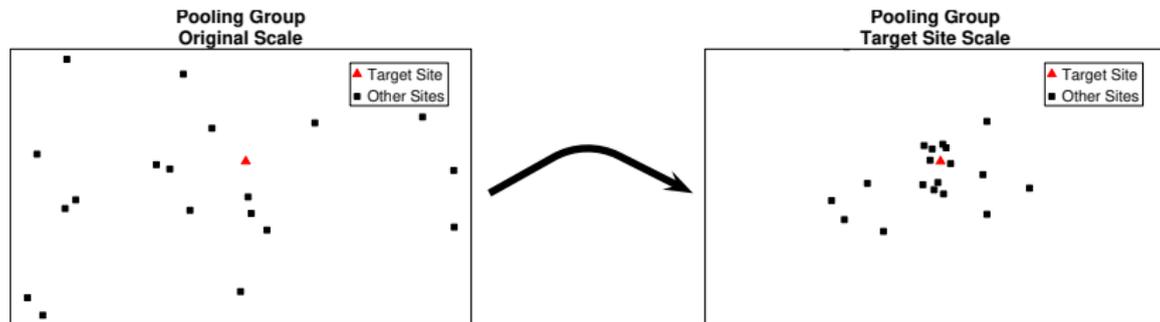


Ajuster la loi a priori selon l'information régionale

Définition : Pseudo paramètres au site cible

$$\tilde{u}^{(i)} = C^{(j)} u_*^{(i)}, \quad \tilde{\sigma}^{(i)} = C^{(j)} \sigma_*^{(i)}, \quad \tilde{\xi}^{(i)} = \xi_*^{(i)}$$

sont appelés pseudo paramètres au site cible pour les paramètres de position, d'échelle et de forme respectivement.



Ajuster la loi a priori selon l'information régionale

Définition : Pseudo paramètres au site cible

$$\tilde{u}^{(i)} = C^{(j)} u_*^{(i)}, \quad \tilde{\sigma}^{(i)} = C^{(j)} \sigma_*^{(i)}, \quad \tilde{\xi}^{(i)} = \xi_*^{(i)}$$

sont appelés pseudo paramètres au site cible pour les paramètres de position, d'échelle et de forme respectivement.

Propriété

Si les observations sont GPD/GEV et sous l'hypothèse de l'Indice de Crue alors,

$$(\tilde{u}^{(i)}, \tilde{\sigma}^{(i)}, \tilde{\xi}^{(i)}) \sim (u^{(j)}, \sigma^{(j)}, \xi^{(j)}), \quad i \neq j$$

où j correspond à l'indice du site cible.

Ajuster la loi a priori selon l'information régionale (2)

Par la méthode des moments, les hyperparamètres sont donc définis suivant :

$$\hat{\gamma}_1 = \frac{1}{N-1} \sum_{i \neq j} \log \tilde{u}^{(i)},$$

$$\hat{d}_1 = \frac{1}{N-1} \sum_{i \neq j} \text{Var}[\log \tilde{u}^{(i)}],$$

$$\hat{\gamma}_2 = \frac{1}{N-1} \sum_{i \neq j} \log \tilde{\sigma}^{(i)},$$

$$\hat{d}_2 = \frac{1}{N-1} \sum_{i \neq j} \text{Var}[\log \tilde{\sigma}^{(i)}],$$

$$\hat{\gamma}_3 = \frac{1}{N-1} \sum_{i \neq j} \xi^{(i)} = \bar{\xi},$$

$$\hat{d}_3 = \frac{1}{N-2} \sum_{i \neq j} \left(\xi^{(i)} - \bar{\xi} \right)^2$$

Ajuster la loi a priori selon l'information régionale (2)

Par la méthode des moments, les hyperparamètres sont donc définis suivant :

$$\hat{\gamma}_1 = \frac{1}{N-1} \sum_{i \neq j} \log \tilde{u}^{(i)},$$

$$\hat{d}_1 = \frac{1}{N-1} \sum_{i \neq j} \text{Var}[\log \tilde{u}^{(i)}],$$

$$\hat{\gamma}_2 = \frac{1}{N-1} \sum_{i \neq j} \log \tilde{\sigma}^{(i)},$$

$$\hat{d}_2 = \frac{1}{N-1} \sum_{i \neq j} \text{Var}[\log \tilde{\sigma}^{(i)}],$$

$$\hat{\gamma}_3 = \frac{1}{N-1} \sum_{i \neq j} \xi^{(i)} = \bar{\xi},$$

$$\hat{d}_3 = \frac{1}{N-2} \sum_{i \neq j} \left(\xi^{(i)} - \bar{\xi} \right)^2$$

Remarque

- γ_i sont semblables aux estimations par le modèle de l'indice de crue
- d_i contrôlent le "degré de précision" relatif à ces estimations

Ajuster la loi a priori selon l'information régionale (2)

Par la méthode des moments, les hyperparamètres sont donc définis suivant :

$$\hat{\gamma}_1 = \frac{1}{N-1} \sum_{i \neq j} \log \tilde{u}^{(i)},$$

$$\hat{d}_1 = \frac{1}{N-1} \sum_{i \neq j} \text{Var}[\log \tilde{u}^{(i)}],$$

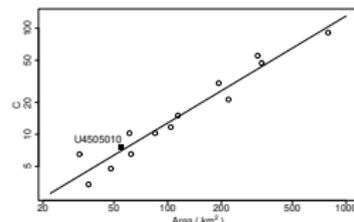
$$\hat{\gamma}_2 = \frac{1}{N-1} \sum_{i \neq j} \log \tilde{\sigma}^{(i)},$$

$$\hat{d}_2 = \frac{1}{N-1} \sum_{i \neq j} \text{Var}[\log \tilde{\sigma}^{(i)}],$$

$$\hat{\gamma}_3 = \frac{1}{N-1} \sum_{i \neq j} \xi^{(i)} = \bar{\xi},$$

$$\hat{d}_3 = \frac{1}{N-2} \sum_{i \neq j} \left(\xi^{(i)} - \bar{\xi} \right)^2$$

- Facteur “normalisant” $C^{(j)}$ = médiane
- Contexte Bayésien \Rightarrow **ne pas utiliser données site cible pour la loi a priori**
- Estimation $C^{(j)}$ par régression



Plan de la Présentation

- 1 La Théorie des Valeurs Extrêmes
 - Loi Généralisée des Valeurs Extrêmes
 - Loi Pareto Généralisée
- 2 Estimation Régionale des Débits de Pointes
 - Quelques Rappels
 - Incorporer l'Information Régionale
 - Définir la Loi a Priori
 - Résultats
- 3 Utilisation d'une Loi a Priori Plus Apte à l'Extrapolation
 - Justification
 - Les Sauts Réversibles
 - Résultats

La Région Homogène

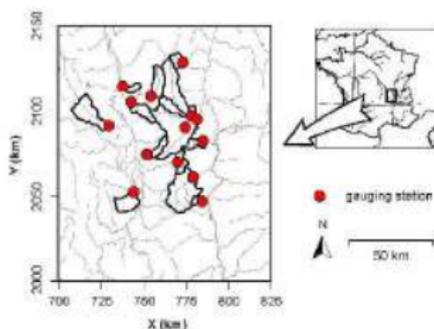
- Initialement plusieurs régions ont été délimitées suivant différents critères physiographiques/climatiques, puis modifiées afin de respecter l'homogénéité H_1
- Au final, un groupe de 14 stations a été retenu

La Région Homogène

- Initialement plusieurs régions ont été délimitées suivant différents critères physiographiques/climatiques, puis modifiées afin de respecter l'homogénéité H_1
- Au final, un groupe de 14 stations a été retenu

La Région Homogène

- Initialement plusieurs régions ont été délimitées suivant différents critères physiographiques/climatiques, puis modifiées afin de respecter l'homogénéité H_1
- Au final, un groupe de 14 stations a été retenu

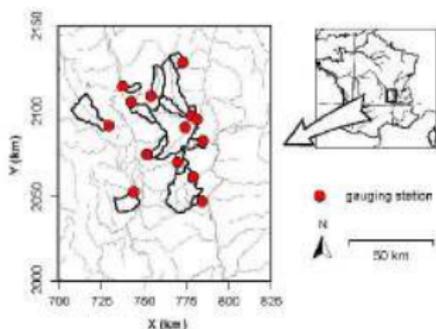


Caractéristiques des Données

- Enregistrement : 22–37 ans
- Superficie : 32–792 km²
- Bassins premier ordre (disjoints)
- 3 sites d'intérêts (37 ans)
- Sélection \approx 2 événements/an

La Région Homogène

- Initialement plusieurs régions ont été délimitées suivant différents critères physiographiques/climatiques, puis modifiées afin de respecter l'homogénéité H_1
- Au final, un groupe de 14 stations a été retenu

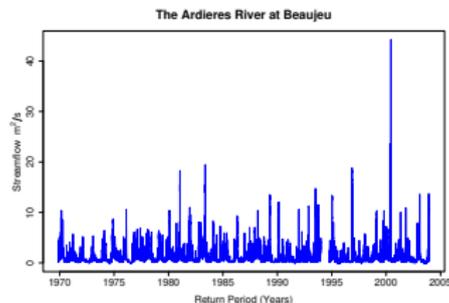


Caractéristiques des Données

- Enregistrement : 22–37 ans
- Superficie : 32–792 km²
- Bassins premier ordre (disjoints)
- 3 sites d'intérêts (37 ans)
- Sélection \approx 2 événements/an

La Région Homogène

- Initialement plusieurs régions ont été délimitées suivant différents critères physiographiques/climatiques, puis modifiées afin de respecter l'homogénéité H_1
- Au final, un groupe de 14 stations a été retenu

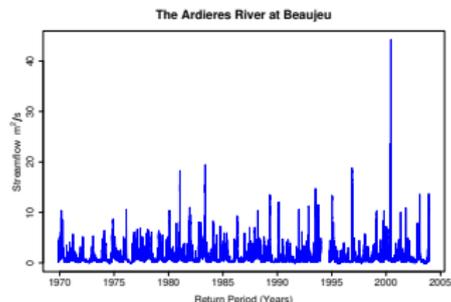


Caractéristiques des Données

- Enregistrement : 22–37 ans
- Superficie : 32–792 km^2
- Bassins premier ordre (disjoints)
- 3 sites d'intérêts (37 ans)
- Sélection ≈ 2 événements/an

La Région Homogène

- Initialement plusieurs régions ont été délimitées suivant différents critères physiographiques/climatiques, puis modifiées afin de respecter l'homogénéité H_1
- Au final, un groupe de 14 stations a été retenu

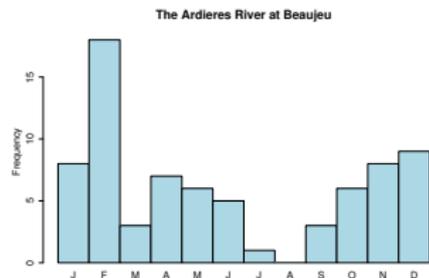


Caractéristiques des Données

- Enregistrement : 22–37 ans
- Superficie : 32–792 km^2
- Bassins premier ordre (disjoints)
- 3 sites d'intérêts (37 ans)
- Sélection ≈ 2 événements/an

La Région Homogène

- Initialement plusieurs régions ont été délimitées suivant différents critères physiographiques/climatiques, puis modifiées afin de respecter l'homogénéité H_1
- Au final, un groupe de 14 stations a été retenu



Caractéristiques des Données

- Enregistrement : 22–37 ans
- Superficie : 32–792 km²
- Bassins premier ordre (disjoints)
- 3 sites d'intérêts (37 ans)
- Sélection ≈ 2 événements/an

Sensibilité à la Période d'Enregistrement

Méthodologie

But de l'étude

- **Cas partiellement jaugés** : échantillons de petites tailles
 - Comparaison de l'approche proposée avec les estimateurs utilisés classiquement
 - Etude "globale" : étude comportement sur différentes périodes de retour
-
- Valeurs de références : MLE sur les séries entières
 - Estimateurs étudiés : MLE, PWU, PWB, IFL et BAY
 - Inférences sur des sous échantillons de taille (5, 10, 15, 20, 25, 30, 37) ans pour les 3 sites d'intérêts
 - Quantiles "cibles" : débits max instantané de période de retour 2, 5, 10 et 20 ans

Sensibilité à la Période d'Enregistrement

Méthodologie

But de l'étude

- **Cas partiellement jaugés** : échantillons de petites tailles
 - **Comparaison** de l'approche proposée **avec les estimateurs utilisés classiquement**
 - Etude "globale" : **étude** comportement sur **différentes périodes de retour**
-
- Valeurs de références : MLE sur les séries entières
 - Estimateurs étudiés : MLE, PWU, PWB, IFL et BAY
 - Inférences sur des sous échantillons de taille (5, 10, 15, 20, 25, 30, 37) ans pour les 3 sites d'intérêts
 - Quantiles "cibles" : débits max instantané de période de retour 2, 5, 10 et 20 ans

Sensibilité à la Période d'Enregistrement

Méthodologie

But de l'étude

- **Cas partiellement jaugés** : échantillons de petites tailles
 - **Comparaison** de l'approche proposée **avec les estimateurs utilisés classiquement**
 - Etude "globale" : **étude** comportement sur **différentes périodes de retour**
-
- Valeurs de références : MLE sur les séries entières
 - Estimateurs étudiés : MLE, PWU, PWB, IFL et BAY
 - Inférences sur des sous échantillons de taille (5, 10, 15, 20, 25, 30, 37) ans pour les 3 sites d'intérêts
 - Quantiles "cibles" : débits max instantané de période de retour 2, 5, 10 et 20 ans

Sensibilité à la Période d'Enregistrement

Méthodologie

But de l'étude

- **Cas partiellement jaugés** : échantillons de petites tailles
 - **Comparaison** de l'approche proposée **avec les estimateurs utilisés classiquement**
 - Etude "globale" : **étude** comportement sur **différentes périodes de retour**
-
- Valeurs de références : MLE sur les séries entières
 - Estimateurs étudiés : MLE, PWU, PWB, IFL et BAY
 - Inférences sur des sous échantillons de taille (5, 10, 15, 20, 25, 30, 37) ans pour les 3 sites d'intérêts
 - Quantiles "cibles" : débits max instantané de période de retour 2, 5, 10 et 20 ans

Sensibilité à la Période d'Enregistrement

Méthodologie

But de l'étude

- **Cas partiellement jaugés** : échantillons de petites tailles
 - **Comparaison** de l'approche proposée **avec les estimateurs utilisés classiquement**
 - Etude "globale" : **étude** comportement sur **différentes périodes de retour**
-
- Valeurs de références : MLE sur les séries entières
 - Estimateurs étudiés : MLE, PWU, PWB, IFL et BAY
 - Inférences sur des sous échantillons de taille (5, 10, 15, 20, 25, 30, 37) ans pour les 3 sites d'intérêts
 - Quantiles "cibles" : débits max instantané de période de retour 2, 5, 10 et 20 ans

Sensibilité à la Période d'Enregistrement

Méthodologie

But de l'étude

- **Cas partiellement jaugés** : échantillons de petites tailles
 - **Comparaison** de l'approche proposée **avec les estimateurs utilisés classiquement**
 - Etude "globale" : **étude** comportement sur **différentes périodes de retour**
-
- Valeurs de références : MLE sur les séries entières
 - Estimateurs étudiés : MLE, PWU, PWB, IFL et BAY
 - Inférences sur des sous échantillons de taille (5, 10, 15, 20, 25, 30, 37) ans pour les 3 sites d'intérêts
 - Quantiles "cibles" : débits max instantané de période de retour 2, 5, 10 et 20 ans

Sensibilité à la Période d'Enregistrement

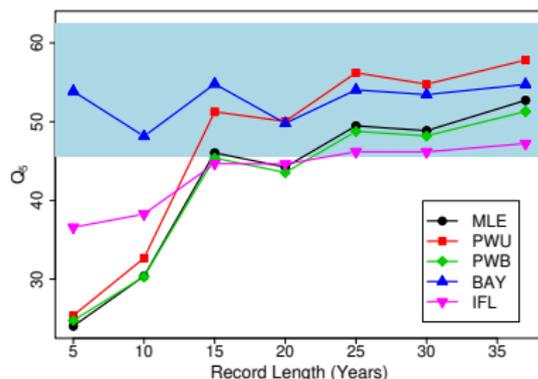
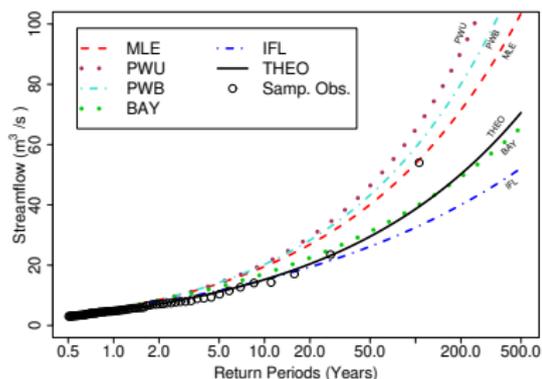
Méthodologie

But de l'étude

- **Cas partiellement jaugés** : échantillons de petites tailles
 - **Comparaison** de l'approche proposée **avec les estimateurs utilisés classiquement**
 - Etude "globale" : **étude** comportement sur **différentes périodes de retour**
-
- Valeurs de références : MLE sur les séries entières
 - Estimateurs étudiés : MLE, PWU, PWB, IFL et BAY
 - Inférences sur des sous échantillons de taille (5, 10, 15, 20, 25, 30, 37) ans pour les 3 sites d'intérêts
 - Quantiles "cibles" : débits max instantané de période de retour 2, 5, 10 et 20 ans

Sensibilité à la Période d'Enregistrement

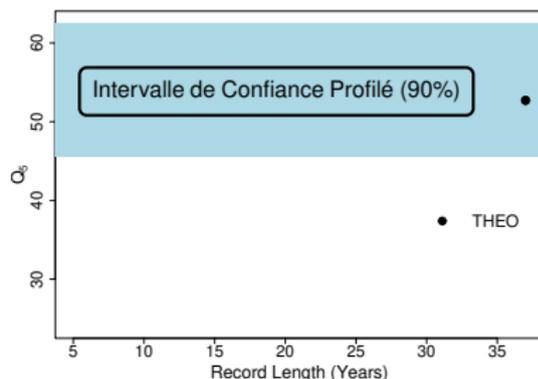
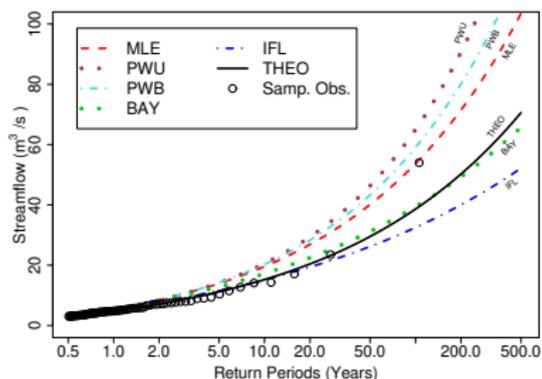
Résultats



- Pour des périodes d'enregistrement < 15 ans, **sous estimations permanentes** des approches **locales** et **IFL** :
 - Approches locales : les plus "forts" extrêmes n'ont pas encore été mesurés
 - estimateur IFL : sous estimation de l'indice de crue $C^{(j)}$
- L'approche **Bayésienne** est la plus **robuste** :
Toujours dans l'intervalle de confiance profilé (90%)

Sensibilité à la Période d'Enregistrement

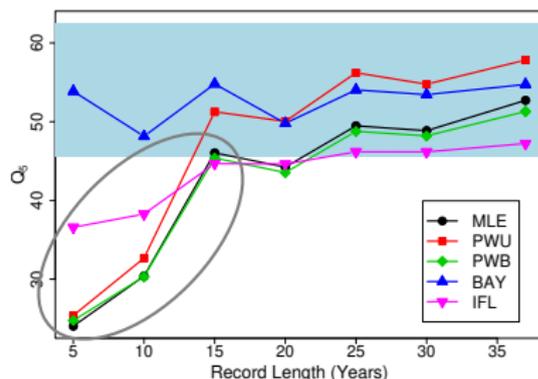
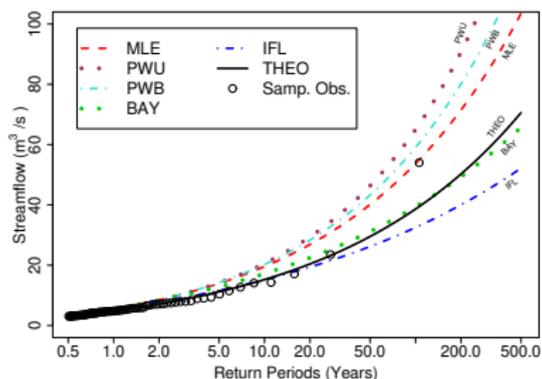
Résultats



- Pour des périodes d'enregistrement < 15 ans, **sous estimations permanentes** des approches **locales et IFL** :
 - Approches locales : les plus "forts" extrêmes n'ont pas encore été mesurés
 - estimateur IFL : sous estimation de l'indice de crue $C^{(j)}$
- L'approche **Bayésienne** est la plus **robuste** :
Toujours dans l'intervalle de confiance profilé (90%)

Sensibilité à la Période d'Enregistrement

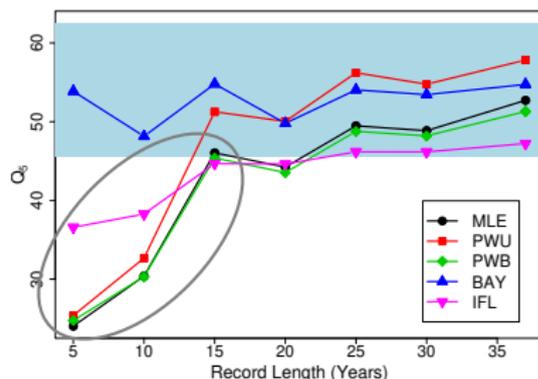
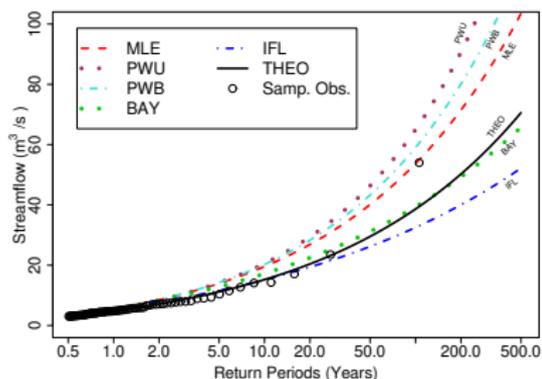
Résultats



- Pour des périodes d'enregistrement < 15 ans, **sous estimations permanentes** des approches **locales** et **IFL** :
 - Approches locales : les plus "forts" extrêmes n'ont pas encore été mesurés
 - estimateur IFL : sous estimation de l'indice de crue $C^{(j)}$
- L'approche **Bayésienne** est la plus **robuste** :
Toujours dans l'intervalle de confiance profilé (90%)

Sensibilité à la Période d'Enregistrement

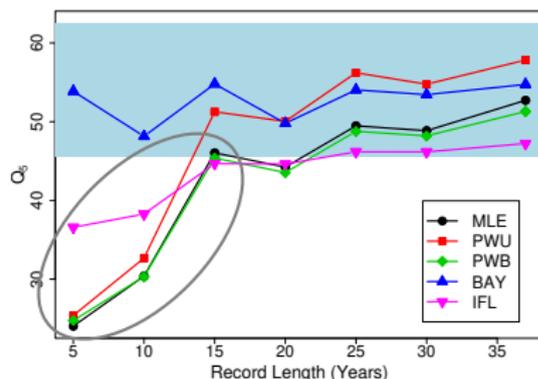
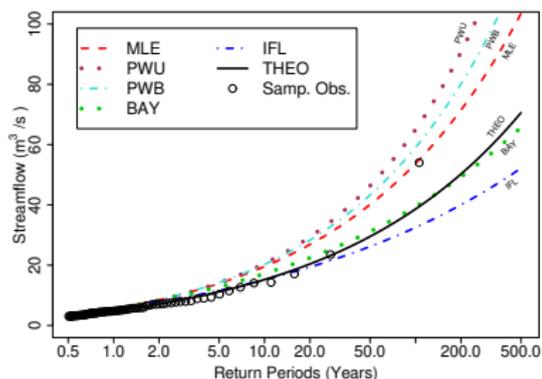
Résultats



- Pour des périodes d'enregistrement < 15 ans, **sous estimations permanentes** des approches **locales et IFL** :
 - Approches locales : les plus “forts” extrêmes n'ont pas encore été mesurés
 - estimateur IFL : sous estimation de l'indice de crue $C^{(j)}$
- L'approche **Bayésienne** est la plus **robuste** :
Toujours dans l'intervalle de confiance profilé (90%)

Sensibilité à la Période d'Enregistrement

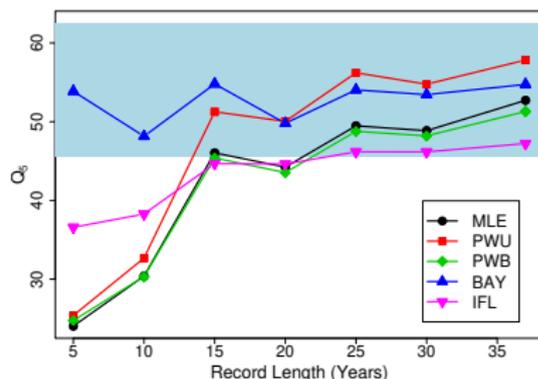
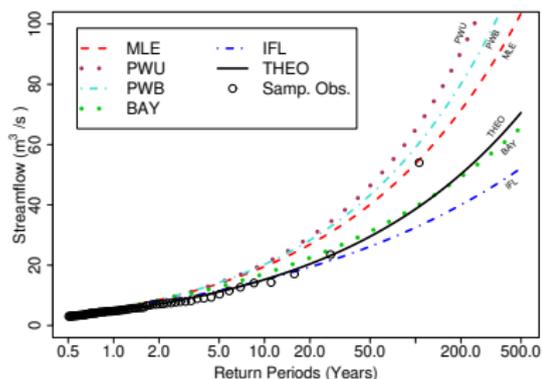
Résultats



- Pour des périodes d'enregistrement < 15 ans, **sous estimations permanentes** des approches **locales et IFL** :
 - Approches locales : les plus "forts" extrêmes n'ont pas encore été mesurés
 - estimateur IFL : sous estimation de l'indice de crue $C^{(j)}$
- L'approche **Bayésienne** est la plus **robuste** :
Toujours dans l'intervalle de confiance profilé (90%)

Sensibilité à la Période d'Enregistrement

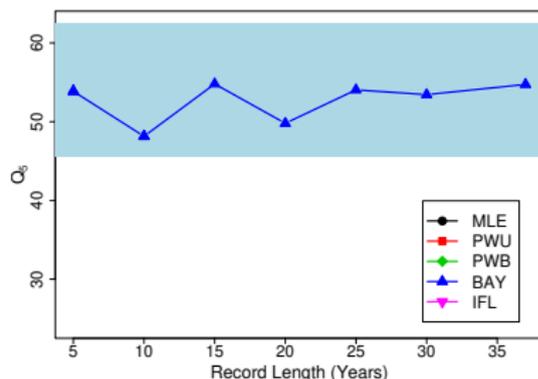
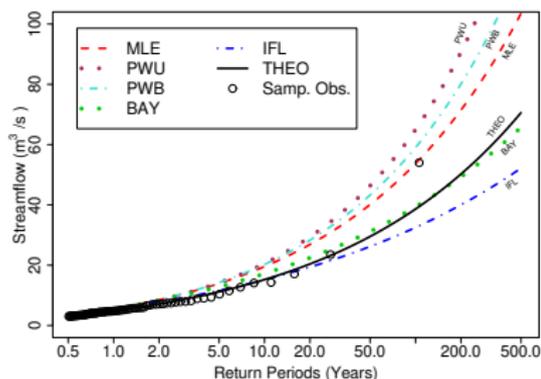
Résultats



- Pour des périodes d'enregistrement < 15 ans, **sous estimations permanentes** des approches **locales et IFL** :
 - Approches locales : les plus "forts" extrêmes n'ont pas encore été mesurés
 - estimateur IFL : sous estimation de l'indice de crue $C^{(j)}$
- L'approche **Bayésienne** est la plus **robuste** :
Toujours dans l'intervalle de confiance profilé (90%)

Sensibilité à la Période d'Enregistrement

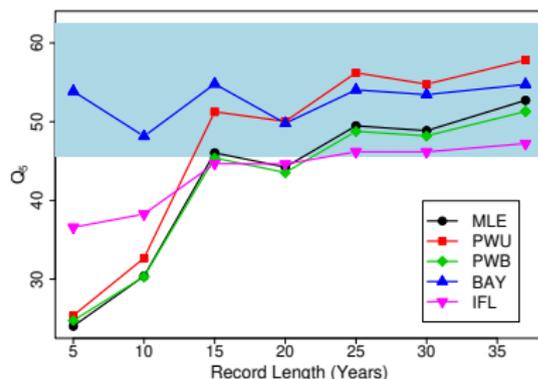
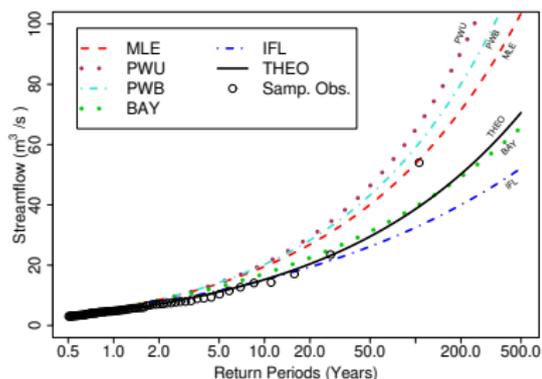
Résultats



- Pour des périodes d'enregistrement < 15 ans, **sous estimations permanentes** des approches **locales et IFL** :
 - Approches locales : les plus “forts” extrêmes n'ont pas encore été mesurés
 - estimateur IFL : sous estimation de l'indice de crue $C^{(j)}$
- L'approche **Bayésienne** est la plus **robuste** :
Toujours dans l'intervalle de confiance profilé (90%)

Sensibilité à la Période d'Enregistrement

Résultats



- Pour des périodes d'enregistrement < 15 ans, **sous estimations permanentes** des approches **locales et IFL** :
 - Approches locales : les plus “forts” extrêmes n'ont pas encore été mesurés
 - estimateur IFL : sous estimation de l'indice de crue $C^{(j)}$
- L'approche **Bayésienne** est la plus **robuste** :
Toujours dans l'intervalle de confiance profilé (90%)

Sensibilité au Degré d'Homogénéité

Méthodologie

- On considère maintenant **4 régions différentes** :

He⁺ Région complètement hétérogène

He Région probablement hétérogène

Ho Région probablement homogène

Ho⁺ Région complètement homogène

- Inférences sont faites sur tous les sous échantillons de tailles (5, 10, 15, 20, 25, 30) ans
- Estimateurs étudiés : MLE, IFL et BAY
- Comparaison à l'aide d'un **score** (mesure du biais et de la variance de toutes les estimations)
 - Score ≈ 1 : Bon estimateur
 - Score ≈ 0 : Mauvais estimateur

Sensibilité au Degré d'Homogénéité

Méthodologie

- On considère maintenant **4 régions différentes** :
 - He^+ Région complètement hétérogène
 - He Région probablement hétérogène
 - Ho Région probablement homogène
 - Ho^+ Région complètement homogène
- Inférences sont faites sur tous les sous échantillons de tailles (5, 10, 15, 20, 25, 30) ans
- Estimateurs étudiés : MLE, IFL et BAY
- Comparaison à l'aide d'un **score** (mesure du biais et de la variance de toutes les estimations)
 - Score ≈ 1 : Bon estimateur
 - Score ≈ 0 : Mauvais estimateur

Sensibilité au Degré d'Homogénéité

Méthodologie

- On considère maintenant **4 régions différentes** :
 - He^+ Région complètement hétérogène
 - He Région probablement hétérogène
 - Ho Région probablement homogène
 - Ho^+ Région complètement homogène
- Inférences sont faites sur tous les sous échantillons de tailles (5, 10, 15, 20, 25, 30) ans
- Estimateurs étudiés : MLE, IFL et BAY
- Comparaison à l'aide d'un **score** (mesure du biais et de la variance de toutes les estimations)
 - Score ≈ 1 : Bon estimateur
 - Score ≈ 0 : Mauvais estimateur

Sensibilité au Degré d'Homogénéité

Méthodologie

- On considère maintenant **4 régions différentes** :
 - He^+ Région complètement hétérogène
 - He Région probablement hétérogène
 - Ho Région probablement homogène
 - Ho^+ Région complètement homogène
- Inférences sont faites sur tous les sous échantillons de tailles (5, 10, 15, 20, 25, 30) ans
- Estimateurs étudiés : MLE, IFL et BAY
- Comparaison à l'aide d'un **score** (mesure du biais et de la variance de toutes les estimations)
 - Score ≈ 1 : Bon estimateur
 - Score ≈ 0 : Mauvais estimateur

Sensibilité au Degré d'Homogénéité

Méthodologie

- On considère maintenant **4 régions différentes** :
 - He^+ Région complètement hétérogène
 - He Région probablement hétérogène
 - Ho Région probablement homogène
 - Ho^+ Région complètement homogène
- Inférences sont faites sur tous les sous échantillons de tailles (5, 10, 15, 20, 25, 30) ans
- Estimateurs étudiés : MLE, IFL et BAY
- Comparaison à l'aide d'un **score** (mesure du biais et de la variance de toutes les estimations)
 - Score ≈ 1 : Bon estimateur
 - Score ≈ 0 : Mauvais estimateur

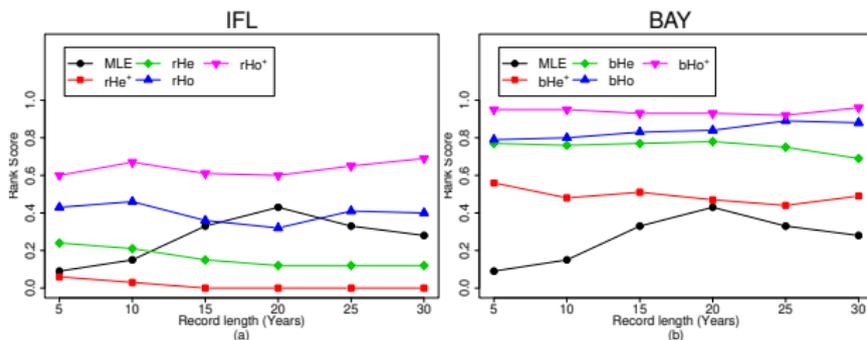
Sensibilité au Degré d'Homogénéité

Méthodologie

- On considère maintenant **4 régions différentes** :
 - He^+ Région complètement hétérogène
 - He Région probablement hétérogène
 - Ho Région probablement homogène
 - Ho^+ Région complètement homogène
- Inférences sont faites sur tous les sous échantillons de tailles (5, 10, 15, 20, 25, 30) ans
- Estimateurs étudiés : MLE, IFL et BAY
- Comparaison à l'aide d'un **score** (mesure du biais et de la variance de toutes les estimations)
 - Score ≈ 1 : Bon estimateur
 - Score ≈ 0 : Mauvais estimateur

Sensibilité au Degré d'Homogénéité

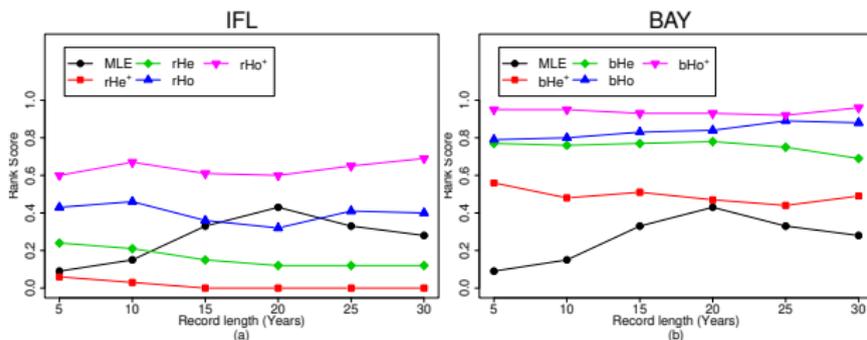
Résultats



- Pour IFL et BAY : faible sensibilité du score à la taille d'échantillon
- IFL est plus sensible au degré d'homogénéité
- Pire score de BAY \approx meilleur score de IFL
- BAY semble plus performant indépendamment la période d'enregistrement
- Pour une pér. d'enr. > 15 ans et les autres stations : MLE plus performant que IFL

Sensibilité au Degré d'Homogénéité

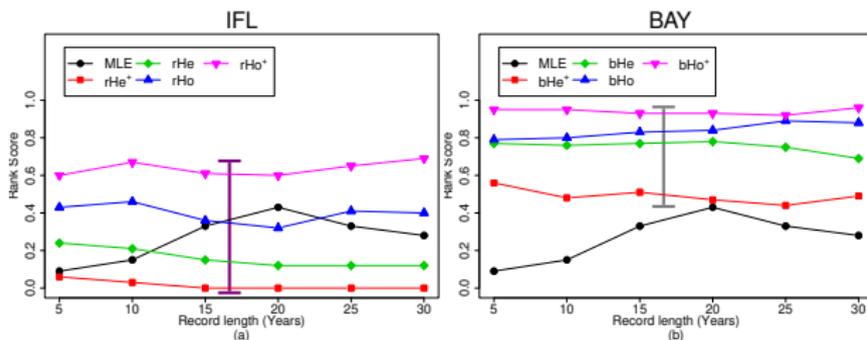
Résultats



- Pour **IFL et BAY** : faible sensibilité du score à la **taille d'échantillon**
- IFL est plus sensible au degré d'homogénéité
- Pire score de BAY \approx meilleur score de IFL
- BAY semble plus performant indépendamment la période d'enregistrement
- Pour une pér. d'enr. > 15 ans et les autres stations : MLE plus performant que IFL

Sensibilité au Degré d'Homogénéité

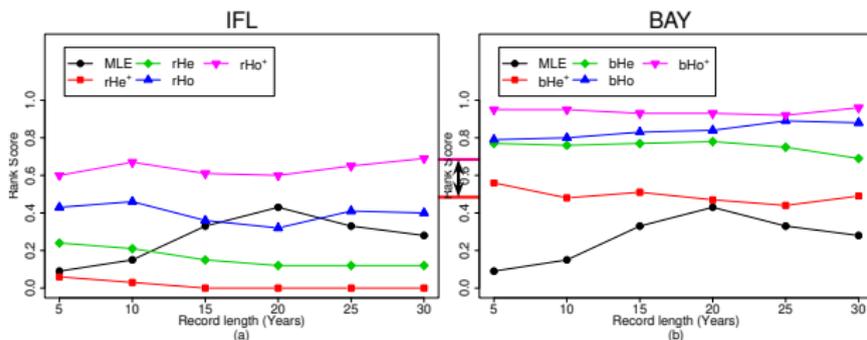
Résultats



- Pour IFL et BAY : faible sensibilité du score à la taille d'échantillon
- IFL est plus sensible au degré d'homogénéité
- Pire score de BAY \approx meilleur score de IFL
- BAY semble plus performant indépendamment la période d'enregistrement
- Pour une pér. d'enr. > 15 ans et les autres stations : MLE plus performant que IFL

Sensibilité au Degré d'Homogénéité

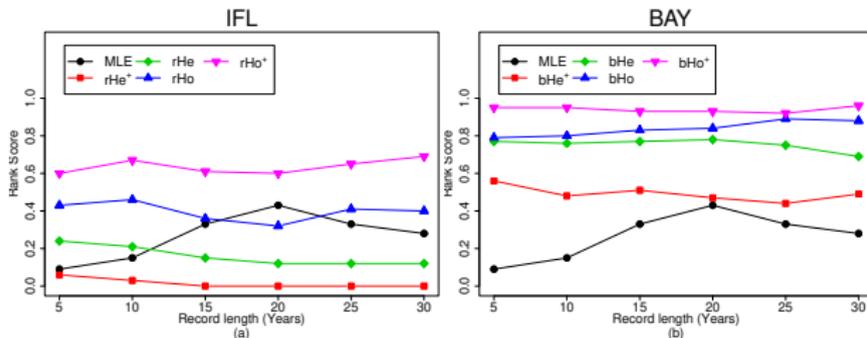
Résultats



- Pour **IFL** et **BAY** : faible sensibilité du score à la taille d'échantillon
- **IFL** est plus sensible au degré d'homogénéité
- Pire score de **BAY** \approx meilleur score de **IFL**
- **BAY** semble plus performant indépendamment la période d'enregistrement
- Pour une pér. d'enr. > 15 ans et les autres stations : MLE plus performant que **IFL**

Sensibilité au Degré d'Homogénéité

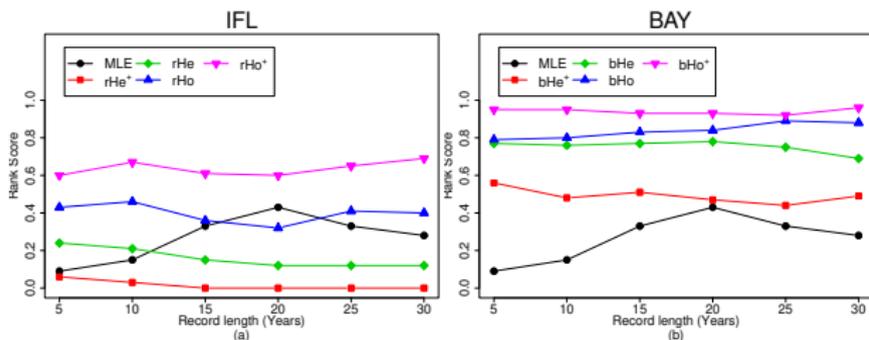
Résultats



- Pour IFL et BAY : faible sensibilité du score à la taille d'échantillon
- IFL est plus sensible au degré d'homogénéité
- Pire score de BAY \approx meilleur score de IFL
- BAY semble plus performant indépendamment la période d'enregistrement
- Pour une pér. d'enr. > 15 ans et les autres stations : MLE plus performant que IFL

Sensibilité au Degré d'Homogénéité

Résultats



- Pour IFL et BAY : faible sensibilité du score à la taille d'échantillon
- IFL est plus sensible au degré d'homogénéité
- Pire score de BAY \approx meilleur score de IFL
- BAY semble plus performant indépendamment la période d'enregistrement
- Pour une pér. d'enr. > 15 ans et les autres stations : MLE plus performant que IFL

Conclusion sur ce modèle

- 1 Introduction d'un modèle statistique rigoureux
- 2 Relâcher les hypothèses trop sévères de l'indice de crue
- 3 BAY plus performant (toutes configurations confondues)
- 4 BAY plus robuste au degré d'homogénéité
- 5 Ne pas travailler avec des régions très petites et très homogènes
- 6 Mais plus grandes et homogènes (selon terminologie Hosking)

Conclusion sur ce modèle

- 1 Introduction d'un modèle statistique rigoureux
- 2 Relâcher les hypothèses trop sévères de l'indice de crue
- 3 BAY plus performant (toutes configurations confondues)
- 4 BAY plus robuste au degré d'homogénéité
- 5 Ne pas travailler avec des régions très petites et très homogènes
- 6 Mais plus grandes et homogènes (selon terminologie Hosking)

Conclusion sur ce modèle

- 1 Introduction d'un modèle statistique rigoureux
- 2 Relâcher les hypothèses trop sévères de l'indice de crue
- 3 BAY plus performant (toutes configurations confondues)
- 4 BAY plus robuste au degré d'homogénéité
- 5 Ne pas travailler avec des régions très petites et très homogènes
- 6 Mais plus grandes et homogènes (selon terminologie Hosking)

Conclusion sur ce modèle

- 1 Introduction d'un modèle statistique rigoureux
- 2 Relâcher les hypothèses trop sévères de l'indice de crue
- 3 BAY plus performant (toutes configurations confondues)
- 4 BAY plus robuste au degré d'homogénéité
- 5 Ne pas travailler avec des régions très petites et très homogènes
- 6 Mais plus grandes et homogènes (selon terminologie Hosking)

Conclusion sur ce modèle

- 1 Introduction d'un modèle statistique rigoureux
- 2 Relâcher les hypothèses trop sévères de l'indice de crue
- 3 BAY plus performant (toutes configurations confondues)
- 4 BAY plus robuste au degré d'homogénéité
- 5 Ne pas travailler avec des régions très petites et très homogènes
- 6 Mais plus grandes et homogènes (selon terminologie Hosking)

Conclusion sur ce modèle

- 1 Introduction d'un modèle statistique rigoureux
- 2 Relâcher les hypothèses trop sévères de l'indice de crue
- 3 BAY plus performant (toutes configurations confondues)
- 4 BAY plus robuste au degré d'homogénéité
- 5 Ne pas travailler avec des régions très petites et très homogènes
- 6 Mais plus grandes et homogènes (selon terminologie Hosking)

Plan de la Présentation

- 1 La Théorie des Valeurs Extrêmes
 - Loi Généralisée des Valeurs Extrêmes
 - Loi Pareto Généralisée
- 2 Estimation Régionale des Débits de Pointes
 - Quelques Rappels
 - Incorporer l'Information Régionale
 - Définir la Loi a Priori
 - Résultats
- 3 Utilisation d'une Loi a Priori Plus Apte à l'Extrapolation
 - Justification
 - Les Sauts Réversibles
 - Résultats

- Section précédente \implies méthodologie pour définir une loi a priori améliorant les estimations
- Étude préliminaire a montré que ce modèle n'était pas assez robuste pour les plus forts quantiles $T > 20$ ans

- Section précédente \implies méthodologie pour définir une loi a priori améliorant les estimations
- Étude préliminaire a montré que ce modèle n'était pas assez robuste pour les plus forts quantiles $T > 20$ ans

Model	SD					NRMSE				
	Q_{10}	Q_{20}	Q_{50}	Q_{100}	Q_{1000}	Q_{10}	Q_{20}	Q_{50}	Q_{100}	Q_{1000}
<i>MLE</i>	30.80	57.19	129.1	242.1	2445.4	0.17	0.31	0.75	1.14	5.83
<i>PWU</i>	25.54	42.81	83.2	135.4	665.2	0.16	0.26	0.48	0.63	1.61
<i>PWB</i>	25.98	42.77	80.7	128.6	589.3	0.16	0.27	0.47	0.60	1.40
<i>MOM</i>	6.36	0.22	0.8	6.4	0.2	0.72	0.85	0.72	0.85	1.00
<i>BAY</i>	8.61	14.58	26.7	40.2	130.8	0.04	0.09	0.16	0.19	0.31
<i>IFL</i>	12.50	16.17	22.2	28.0	57.4	0.13	0.13	0.15	0.15	0.19

- Section précédente \implies méthodologie pour définir une loi a priori améliorant les estimations
- Étude préliminaire a montré que ce modèle n'était pas assez robuste pour les plus forts quantiles $T > 20$ ans

Model	SD					NRMSE				
	Q ₁₀	Q ₂₀	Q ₅₀	Q ₁₀₀	Q ₁₀₀₀	Q ₁₀	Q ₂₀	Q ₅₀	Q ₁₀₀	Q ₁₀₀₀
MLE	30.80	57.19	129.1	242.1	2445.4	0.17	0.31	0.75	1.14	5.83
PWU	25.54	42.81	83.2	135.4	665.2	0.16	0.26	0.48	0.63	1.61
PWB	25.98	42.77	80.7	128.6	589.3	0.16	0.27	0.47	0.60	1.40
MOM	6.36	0.22	0.8	6.4	0.2	0.72	0.85	0.72	0.85	1.00
BAY	8.61	14.58	26.7	40.2	130.8	0.04	0.09	0.16	0.19	0.31
IFL	12.50	16.17	22.2	28.0	57.4	0.13	0.13	0.15	0.15	0.19

- Section précédente \implies méthodologie pour définir une loi a priori améliorant les estimations
- Étude préliminaire a montré que ce modèle n'était pas assez robuste pour les plus forts quantiles $T > 20$ ans

Model	SD					NRMSE				
	Q ₁₀	Q ₂₀	Q ₅₀	Q ₁₀₀	Q ₁₀₀₀	Q ₁₀	Q ₂₀	Q ₅₀	Q ₁₀₀	Q ₁₀₀₀
MLE	30.80	57.19	129.1	242.1	2445.4	0.17	0.31	0.75	1.14	5.83
PWU	25.54	42.81	83.2	135.4	665.2	0.16	0.26	0.48	0.63	1.61
PWB	25.98	42.77	80.7	128.6	589.3	0.16	0.27	0.47	0.60	1.40
MOM	6.36	0.22	0.8	6.4	0.2	0.72	0.85	0.72	0.85	1.00
BAY	8.61	14.58	26.7	40.2	130.8	0.04	0.09	0.16	0.19	0.31
IFL	12.50	16.17	22.2	28.0	57.4	0.13	0.13	0.15	0.15	0.19

- Section précédente \implies méthodologie pour définir une loi a priori améliorant les estimations
- Étude préliminaire a montré que ce modèle n'était pas assez robuste pour les plus forts quantiles $T > 20$ ans

Model	SD					NRMSE				
	Q_{10}	Q_{20}	Q_{50}	Q_{100}	Q_{1000}	Q_{10}	Q_{20}	Q_{50}	Q_{100}	Q_{1000}
MLE	30.80	57.19	129.1	242.1	2445.4	0.17	0.31	0.75	1.14	5.83
PWU	25.54	42.81	83.2	135.4	665.2	0.16	0.26	0.48	0.63	1.61
PWB	25.98	42.77	80.7	128.6	589.3	0.16	0.27	0.47	0.60	1.40
MOM	6.36	0.22	0.8	6.4	0.2	0.72	0.85	0.72	0.85	1.00
BAY	8.61	14.58	26.7	40.2	130.8	0.04	0.09	0.16	0.19	0.31
IFL	12.50	16.17	22.2	28.0	57.4	0.13	0.13	0.15	0.15	0.19

\implies **Modèle plus performant pour les forts quantiles**

- Section précédente \implies méthodologie pour définir une loi a priori améliorant les estimations
- Étude préliminaire a montré que ce modèle n'était pas assez robuste pour les plus forts quantiles $T > 20$ ans

Model	NBIAS			SD			NRMSE		
	loc	scale	shape	loc	scale	shape	loc	scale	shape
<i>MLE</i>	0.039	0.217	-0.624	0.842	8.424	0.371	0.054	0.546	1.558
<i>PWU</i>	0.039	0.073	-0.292	0.842	6.685	0.285	0.054	0.405	1.132
<i>PWB</i>	0.039	0.133	-0.458	0.842	6.782	0.282	0.054	0.425	1.177
<i>MOM</i>	0.039	0.249	-0.753	0.842	6.359	0.222	0.054	0.453	1.138
<i>BAY_{mean}</i>	0.005	0.021	-0.067	0.724	0.220	0.063	0.033	0.025	0.252
<i>BAY_{med}</i>	0.014	0.019	-0.083	0.746	0.217	0.062	0.036	0.023	0.254
<i>BAY_{mod}</i>	0.031	0.015	-0.111	0.812	0.280	0.064	0.048	0.023	0.269
<i>IFL</i>	0.003	-0.001	0.103	2.711	2.100	0.004	0.121	0.125	0.104

\implies **Modèle plus performant pour les forts quantiles**

- Section précédente \implies méthodologie pour définir une loi a priori améliorant les estimations
- Étude préliminaire a montré que ce modèle n'était pas assez robuste pour les plus forts quantiles $T > 20$ ans

Model	NBIAS			SD			NRMSE		
	loc	scale	shape	loc	scale	shape	loc	scale	shape
<i>MLE</i>	0.039	0.217	-0.624	0.842	8.424	0.371	0.054	0.546	1.558
<i>PWU</i>	0.039	0.073	-0.292	0.842	6.685	0.285	0.054	0.405	1.132
<i>PWB</i>	0.039	0.133	-0.458	0.842	6.782	0.282	0.054	0.425	1.177
<i>MOM</i>	0.039	0.249	-0.753	0.842	6.359	0.222	0.054	0.453	1.138
<i>BAY_{mean}</i>	0.005	0.021	-0.067	0.724	0.220	0.063	0.033	0.025	0.252
<i>BAY_{med}</i>	0.014	0.019	-0.083	0.746	0.217	0.062	0.036	0.023	0.254
<i>BAY_{mod}</i>	0.031	0.015	-0.111	0.812	0.280	0.064	0.048	0.023	0.269
<i>IFL</i>	0.003	-0.001	0.103	2.711	2.100	0.004	0.121	0.125	0.104

\implies Modèle plus performant pour les forts quantiles

- Section précédente \implies méthodologie pour définir une loi a priori améliorant les estimations
- Étude préliminaire a montré que ce modèle n'était pas assez robuste pour les plus forts quantiles $T > 20$ ans

Model	NBIAS			SD			NRMSE		
	loc	scale	shape	loc	scale	shape	loc	scale	shape
<i>MLE</i>	0.039	0.217	-0.624	0.842	8.424	0.371	0.054	0.546	1.558
<i>PWU</i>	0.039	0.073	-0.292	0.842	6.685	0.285	0.054	0.405	1.132
<i>PWB</i>	0.039	0.133	-0.458	0.842	6.782	0.282	0.054	0.425	1.177
<i>MOM</i>	0.039	0.249	-0.753	0.842	6.359	0.222	0.054	0.453	1.138
<i>BAY_{mean}</i>	0.005	0.021	-0.067	0.724	0.220	0.063	0.033	0.025	0.252
<i>BAY_{med}</i>	0.014	0.019	-0.083	0.746	0.217	0.062	0.036	0.023	0.254
<i>BAY_{mod}</i>	0.031	0.015	-0.111	0.812	0.280	0.064	0.048	0.023	0.269
<i>IFL</i>	0.003	-0.001	0.103	2.711	2.100	0.004	0.121	0.125	0.104

\implies Modèle plus performant pour les forts quantiles

Idée

Proposer (pas imposer) un paramètre de forme régional ζ_{Fix}

Formulation nouvelle de la loi a priori : Mixture

Un hyperparamètre de plus : p_ξ

$$\pi(\theta) = \begin{cases} (1 - p_\xi)\pi_{\text{in}}(\theta), & \text{pour } \theta \in \Theta \setminus \Theta_0 \\ p_\xi\pi_{\xi_{\text{Fix}}}(\theta), & \text{pour } \theta \in \Theta_0 \end{cases}$$

- l'hyperparamètre p_ξ supplémentaire contrôle le “degré de croyance” en ξ_{Fix}
- Il paraît donc intuitif de lier p_ξ à la statistique d'homogénéité H_1 :

$$p_\xi = \frac{\exp(-H_1)}{1 + \exp(-H_1)}$$

Quelques Propriétés pour p_ξ

- $\lim_{H_1 \rightarrow +\infty} p_\xi = 0$
 - $\lim_{H_1 \rightarrow -\infty} p_\xi = 1$
 - $H_1 = 0$ lorsque $p_\xi = 0.5$
- [Stephenson, 2004]

Formulation nouvelle de la loi a priori : Mixture

Un hyperparamètre de plus : ρ_ξ

$$\pi(\theta) = \begin{cases} (1 - \rho_\xi)\pi_{\text{in}}(\theta), & \text{pour } \theta \in \Theta \setminus \Theta_0 \\ \rho_\xi\pi_{\xi_{\text{Fix}}}(\theta), & \text{pour } \theta \in \Theta_0 \end{cases}$$

- l'hyperparamètre ρ_ξ supplémentaire contrôle le “degré de croyance” en ξ_{Fix}
- Il paraît donc intuitif de lier ρ_ξ à la statistique d'homogénéité H_1 :

$$\rho_\xi = \frac{\exp(-H_1)}{1 + \exp(-H_1)}$$

Quelques Propriétés pour ρ_ξ

- $\lim_{H_1 \rightarrow +\infty} \rho_\xi = 0$
 - $\lim_{H_1 \rightarrow -\infty} \rho_\xi = 1$
 - $H_1 = 0$ lorsque $\rho_\xi = 0.5$
- [Stephenson, 2004]

Formulation nouvelle de la loi a priori : Mixture

Un hyperparamètre de plus : ρ_ξ

$$\pi(\theta) = \begin{cases} (1 - \rho_\xi)\pi_{\text{in}}(\theta), & \text{pour } \theta \in \Theta \setminus \Theta_0 \\ \rho_\xi\pi_{\xi_{\text{Fix}}}(\theta), & \text{pour } \theta \in \Theta_0 \end{cases}$$

- l'hyperparamètre ρ_ξ supplémentaire contrôle le “degré de croyance” en ξ_{Fix}
- Il paraît donc intuitif de lier ρ_ξ à la statistique d'homogénéité H_1 :

$$\rho_\xi = \frac{\exp(-H_1)}{1 + \exp(-H_1)}$$

Quelques Propriétés pour ρ_ξ

- $\lim_{H_1 \rightarrow +\infty} \rho_\xi = 0$
 - $\lim_{H_1 \rightarrow -\infty} \rho_\xi = 1$
 - $H_1 = 0$ lorsque $\rho_\xi = 0.5$
- [Stephenson, 2004]

Formulation nouvelle de la loi a priori : Mixture

Un hyperparamètre de plus : p_ξ

$$\pi(\theta) = \begin{cases} (1 - p_\xi)\pi_{\text{in}}(\theta), & \text{pour } \theta \in \Theta \setminus \Theta_0 \\ p_\xi\pi_{\xi_{\text{Fix}}}(\theta), & \text{pour } \theta \in \Theta_0 \end{cases}$$

- l'hyperparamètre p_ξ supplémentaire contrôle le “degré de croyance” en ξ_{Fix}
- Il paraît donc intuitif de lier p_ξ à la statistique d'homogénéité H_1 :

$$p_\xi = \frac{\exp(-H_1)}{1 + \exp(-H_1)}$$

Quelques Propriétés pour p_ξ

- $\lim_{H_1 \rightarrow +\infty} p_\xi = 0$
 - $\lim_{H_1 \rightarrow -\infty} p_\xi = 1$
 - $H_1 = 0$ lorsque $p_\xi = 0.5$
- [Stephenson, 2004]

Plan de la Présentation

- 1 La Théorie des Valeurs Extrêmes
 - Loi Généralisée des Valeurs Extrêmes
 - Loi Pareto Généralisée
- 2 Estimation Régionale des Débits de Pointes
 - Quelques Rappels
 - Incorporer l'Information Régionale
 - Définir la Loi a Priori
 - Résultats
- 3 Utilisation d'une Loi a Priori Plus Apte à l'Extrapolation
 - Justification
 - Les Sauts Réversibles
 - Résultats

Loi a Priori Revisitée

Présentation Théorique

Soit $\pi_{\text{in}}(\theta)$ la loi a priori du modèle introduit précédemment, alors nous définissons une nouvelle loi a priori selon :

$$\pi(\theta) = \begin{cases} (1 - p_{\xi})\pi_{\text{in}}(\theta), & \text{pour } \theta \in \Theta \setminus \Theta_0 \\ p_{\xi}\pi_{\xi_{\text{Fix}}}(\theta), & \text{pour } \theta \in \Theta_0 \end{cases}$$

où $p_{\xi} \in [0, 1]$ et $\pi_{\xi_{\text{Fix}}}(\theta) = \frac{\pi_{\text{in}}(\mu, \sigma, \xi_{\text{Fix}})}{\int_{\mu, \sigma} \pi_{\text{in}}(\mu, \sigma, \xi_{\text{Fix}}) d\mu d\sigma}$

Présentation Moins Formelle

On se souvient que $\pi(\theta) \approx$ “intuition”. Ici, “intuition double” :

- même intuition π_{in} que modèle initial [Ribatet et al., 2006]
- donner la “chance” à un paramètre de forme ξ_{Fix} précis

Loi a Priori Revisitée

Présentation Théorique

Soit $\pi_{\text{in}}(\theta)$ la loi a priori du modèle introduit précédemment, alors nous définissons une nouvelle loi a priori selon :

$$\pi(\theta) = \begin{cases} (1 - p_{\xi})\pi_{\text{in}}(\theta), & \text{pour } \theta \in \Theta \setminus \Theta_0 \\ p_{\xi}\pi_{\xi_{\text{Fix}}}(\theta), & \text{pour } \theta \in \Theta_0 \end{cases}$$

où $p_{\xi} \in [0, 1]$ et $\pi_{\xi_{\text{Fix}}}(\theta) = \frac{\pi_{\text{in}}(\mu, \sigma, \xi_{\text{Fix}})}{\int_{\mu, \sigma} \pi_{\text{in}}(\mu, \sigma, \xi_{\text{Fix}}) d\mu d\sigma}$

Présentation Moins Formelle

On se souvient que $\pi(\theta) \approx$ “intuition”. Ici, “**intuition double**” :

- même **intuition** π_{in} que **modèle initial** [Ribatet et al., 2006]
- donner la “**chance**” à un **paramètre de forme** ξ_{Fix} **précis**

Loi a Priori Revisitée

Présentation Théorique

Soit $\pi_{\text{in}}(\theta)$ la loi a priori du modèle introduit précédemment, alors nous définissons une nouvelle loi a priori selon :

$$\pi(\theta) = \begin{cases} (1 - p_{\xi})\pi_{\text{in}}(\theta), & \text{pour } \theta \in \Theta \setminus \Theta_0 \\ p_{\xi}\pi_{\xi_{\text{Fix}}}(\theta), & \text{pour } \theta \in \Theta_0 \end{cases}$$

où $p_{\xi} \in [0, 1]$ et $\pi_{\xi_{\text{Fix}}}(\theta) = \frac{\pi_{\text{in}}(\mu, \sigma, \xi_{\text{Fix}})}{\int_{\mu, \sigma} \pi_{\text{in}}(\mu, \sigma, \xi_{\text{Fix}}) d\mu d\sigma}$

Présentation Moins Formelle

On se souvient que $\pi(\theta) \approx$ “intuition”. Ici, “**intuition double**” :

- même **intuition** π_{in} que **modèle initial** [Ribatet et al., 2006]
- donner la “**chance**” à un **paramètre de forme** ξ_{Fix} **précis**

Principe et Originalité

- Difficulté : Simulations Chaînes de Markov dans **2 espaces**
 - Espace habituel $\Theta = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi \in \mathbb{R}\}$
 - Sous espace $\Theta_0 = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi = \xi_{\text{Fix}}\}$
- Utilisation de chaînes de Markov à **sauts réversibles**
- Les “sauts” permettent de **passer d'un espace à l'autre** [Green, 1995]

Intérêts

- Propose un “bon candidat” pour $\xi \Rightarrow$ **réduit la variance** pour l'estimation de ξ
- On n'impose pas $\xi = \xi_{\text{Fix}} \Rightarrow$ on ne **néglige pas les incertitudes** sur ξ
- Seules les données au **site cible valideront** le “bon candidat” ξ_{Fix}

Principe et Originalité

- Difficulté : Simulations Chaînes de Markov dans **2 espaces**
 - Espace habituel $\Theta = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi \in \mathbb{R}\}$
 - Sous espace $\Theta_0 = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi = \xi_{\text{Fix}}\}$
- Utilisation de chaînes de Markov à sauts réversibles
- Les “sauts” permettent de passer d'un espace à l'autre [Green, 1995]

Intérêts

- Propose un “bon candidat” pour $\xi \Rightarrow$ réduit la variance pour l'estimation de ξ
- On n'impose pas $\xi = \xi_{\text{Fix}} \Rightarrow$ on ne néglige pas les incertitudes sur ξ
- Seules les données au site cible valideront le “bon candidat” ξ_{Fix}

Principe et Originalité

- Difficulté : Simulations Chaînes de Markov dans **2 espaces**
 - Espace habituel $\Theta = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi \in \mathbb{R}\}$
 - Sous espace $\Theta_0 = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi = \xi_{\text{Fix}}\}$
- Utilisation de chaînes de Markov à **sauts réversibles**
- Les “sauts” permettent de **passer d'un espace à l'autre**
[Green, 1995]

Intérêts

- Propose un “bon candidat” pour $\xi \Rightarrow$ **réduit la variance** pour l'estimation de ξ
- On n'impose pas $\xi = \xi_{\text{Fix}} \Rightarrow$ on ne **néglige pas les incertitudes** sur ξ
- Seules les données au **site cible valideront** le “bon candidat” ξ_{Fix}

Principe et Originalité

- Difficulté : Simulations Chaînes de Markov dans **2 espaces**
 - Espace habituel $\Theta = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi \in \mathbb{R}\}$
 - Sous espace $\Theta_0 = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi = \xi_{\text{Fix}}\}$
- Utilisation de chaînes de Markov à **sauts réversibles**
- Les “sauts” permettent de **passer d'un espace à l'autre**
[Green, 1995]

Intérêts

- Propose un “bon candidat” pour $\xi \Rightarrow$ **réduit la variance** pour l'estimation de ξ
- On n'impose pas $\xi = \xi_{\text{Fix}} \Rightarrow$ on ne **néglige pas les incertitudes** sur ξ
- Seules les données au **site cible valideront** le “bon candidat” ξ_{Fix}

Principe et Originalité

- Difficulté : Simulations Chaînes de Markov dans **2 espaces**
 - Espace habituel $\Theta = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi \in \mathbb{R}\}$
 - Sous espace $\Theta_0 = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi = \xi_{\text{Fix}}\}$
- Utilisation de chaînes de Markov à **sauts réversibles**
- Les “sauts” permettent de **passer d'un espace à l'autre**
[Green, 1995]

Intérêts

- Propose un “bon candidat” pour $\xi \Rightarrow$ **réduit** la **variance** pour l'estimation de ξ
- On n'impose pas $\xi = \xi_{\text{Fix}} \Rightarrow$ on ne **néglige pas** les **incertitudes** sur ξ
- Seules les données au **site cible valideront** le “bon candidat” ξ_{Fix}

Principe et Originalité

- Difficulté : Simulations Chaînes de Markov dans **2 espaces**
 - Espace habituel $\Theta = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi \in \mathbb{R}\}$
 - Sous espace $\Theta_0 = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi = \xi_{\text{Fix}}\}$
- Utilisation de chaînes de Markov à **sauts réversibles**
- Les “sauts” permettent de **passer d'un espace à l'autre**
[Green, 1995]

Intérêts

- Propose un “bon candidat” pour $\xi \Rightarrow$ **réduit** la **variance** pour l'estimation de ξ
- On n'impose pas $\xi = \xi_{\text{Fix}} \Rightarrow$ on ne **néglige pas** les **incertitudes** sur ξ
- Seules les données au **site cible** valideront le “bon candidat” ξ_{Fix}

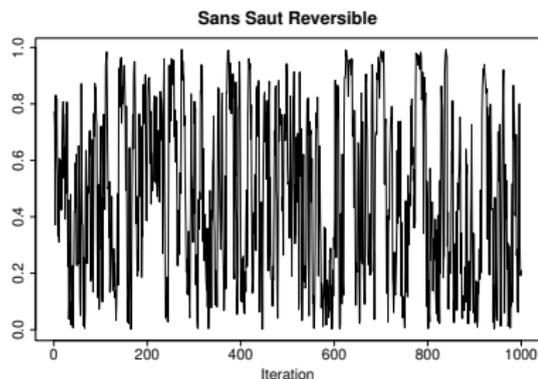
Principe et Originalité

- Difficulté : Simulations Chaînes de Markov dans **2 espaces**
 - Espace habituel $\Theta = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi \in \mathbb{R}\}$
 - Sous espace $\Theta_0 = \{(u, \sigma, \xi) : u > 0, \sigma > 0, \xi = \xi_{\text{Fix}}\}$
- Utilisation de chaînes de Markov à **sauts réversibles**
- Les “sauts” permettent de **passer d'un espace à l'autre**
[Green, 1995]

Intérêts

- Propose un “bon candidat” pour $\xi \Rightarrow$ **réduit** la **variance** pour l'estimation de ξ
- On n'impose pas $\xi = \xi_{\text{Fix}} \Rightarrow$ on ne **néglige pas** les **incertitudes** sur ξ
- Seules les données au **site cible valideront** le “bon candidat” ξ_{Fix}

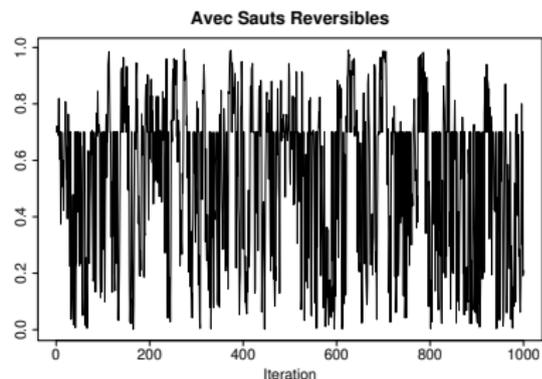
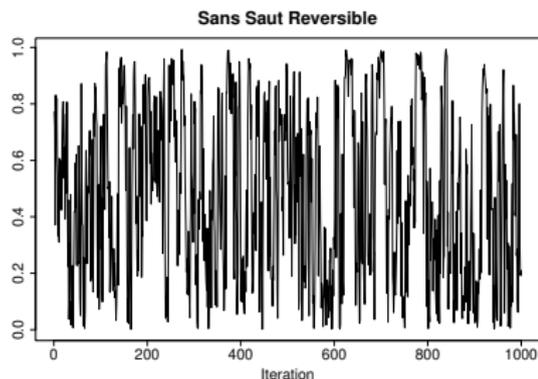
Les Sauts Réversibles en Images



Remarque

Pour une Chaîne de Markov continue, la probabilité de tomber sur ce fameux candidat est **toujours** nulle... pas pour les Sauts Réversibles !!!

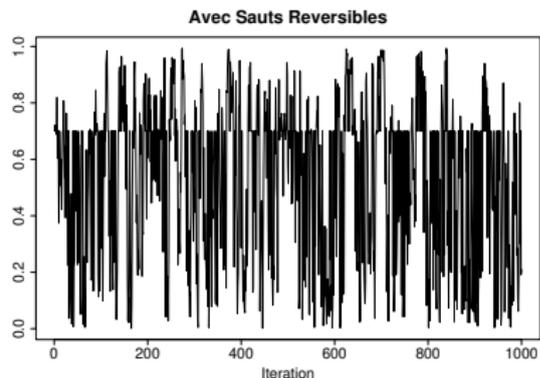
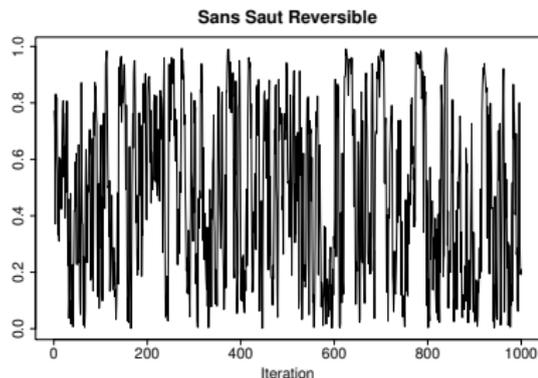
Les Sauts Réversibles en Images



Remarque

Pour une Chaîne de Markov continue, la probabilité de tomber sur ce fameux candidat est **toujours** nulle... pas pour les Sauts Réversibles !!!

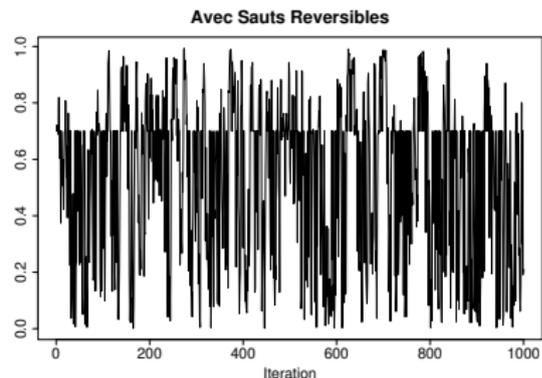
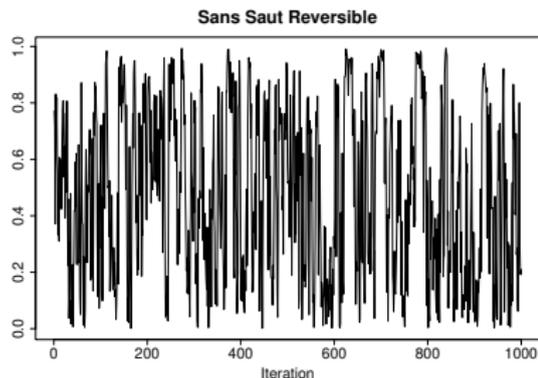
Les Sauts Réversibles en Images



Remarque

Pour une Chaîne de Markov continue, la probabilité de tomber sur ce fameux candidat est **toujours** nulle. . . pas pour les Sauts Réversibles !!!

Les Sauts Réversibles en Images



Remarque

Pour une Chaîne de Markov continue, la probabilité de tomber sur ce fameux candidat est **toujours** nulle. . . **pas pour les Sauts Réversibles !!!**

Plan de la Présentation

- 1 La Théorie des Valeurs Extrêmes
 - Loi Généralisée des Valeurs Extrêmes
 - Loi Pareto Généralisée
- 2 Estimation Régionale des Débits de Pointes
 - Quelques Rappels
 - Incorporer l'Information Régionale
 - Définir la Loi a Priori
 - Résultats
- 3 Utilisation d'une Loi a Priori Plus Apte à l'Extrapolation
 - Justification
 - Les Sauts Réversibles
 - Résultats

Méthodologie

- Deux types de lois régionales : **fortement GP** (Conf 1–3) et **proche Expo** (Conf 4–6)
- Plusieurs configurations étudiées :
 - Conf1, Conf4 : **petites régions** (10 sites) avec beaucoup de données (450 obs. au total)
 - Conf2, Conf5 : **grandes régions** (20 sites) avec peu de données (450 obs. au total)
 - Conf3, Conf6 : **régions intermédiaires** (15 sites) avec beaucoup de données (700 obs. au total)
- Comparaison entre **3 estimateurs régionaux** :
 - IFL : Indice de Crue de Dalrymple (1960)
 - BAY : estimateur proposée par Ribatet et al. (2006)
 - REV : estimateur avec Sauts Réversibles
- Etude sur 3 quantiles $Q_{0.75}$, $Q_{0.95}$ et $Q_{0.995}$ (i.e. périodes de retour $\approx 2, 10$ et 100 ans)
- Taille échantillon au site cible : **10, 25, 40**

Méthodologie

- Deux types de lois régionales : **fortement GP** (Conf 1–3) et **proche Expo** (Conf 4–6)
- Plusieurs configurations étudiées :
 - Conf1, Conf4 : **petites régions** (10 sites) avec beaucoup de données (450 obs. au total)
 - Conf2, Conf5 : **grandes régions** (20 sites) avec peu de données (450 obs. au total)
 - Conf3, Conf6 : **régions intermédiaires** (15 sites) avec beaucoup de données (700 obs. au total)
- Comparaison entre **3 estimateurs régionaux** :
 - IFL : Indice de Crue de Dalrymple (1960)
 - BAY : estimateur proposée par Ribatet et al. (2006)
 - REV : estimateur avec Sauts Réversibles
- Etude sur 3 quantiles $Q_{0.75}$, $Q_{0.95}$ et $Q_{0.995}$ (i.e. périodes de retour $\approx 2, 10$ et 100 ans)
- Taille échantillon au site cible : **10, 25, 40**

Méthodologie

- Deux types de lois régionales : **fortement GP** (Conf 1–3) et **proche Expo** (Conf 4–6)
- Plusieurs configurations étudiées :
 - Conf1, Conf4 : **petites régions** (10 sites) avec beaucoup de données (450 obs. au total)
 - Conf2, Conf5 : **grandes régions** (20 sites) avec peu de données (450 obs. au total)
 - Conf3, Conf6 : **régions intermédiaires** (15 sites) avec beaucoup de données (700 obs. au total)
- Comparaison entre **3 estimateurs régionaux** :
 - IFL : Indice de Crue de Dalrymple (1960)
 - BAY : estimateur proposée par Ribatet et al. (2006)
 - REV : estimateur avec Sauts Réversibles
- Etude sur 3 quantiles $Q_{0.75}$, $Q_{0.95}$ et $Q_{0.995}$ (i.e. périodes de retour $\approx 2, 10$ et 100 ans)
- Taille échantillon au site cible : **10, 25, 40**

Méthodologie

- Deux types de lois régionales : **fortement GP** (Conf 1–3) et **proche Expo** (Conf 4–6)
- Plusieurs configurations étudiées :
 - Conf1, Conf4 : **petites régions** (10 sites) avec beaucoup de données (450 obs. au total)
 - Conf2, Conf5 : **grandes régions** (20 sites) avec peu de données (450 obs. au total)
 - Conf3, Conf6 : **régions intermédiaires** (15 sites) avec beaucoup de données (700 obs. au total)
- Comparaison entre **3 estimateurs régionaux** :
 - IFL : Indice de Crue de Dalrymple (1960)
 - BAY : estimateur proposée par Ribatet et al. (2006)
 - REV : estimateur avec Sauts Réversibles
- Etude sur 3 quantiles $Q_{0.75}$, $Q_{0.95}$ et $Q_{0.995}$ (i.e. périodes de retour $\approx 2, 10$ et 100 ans)
- Taille échantillon au site cible : **10, 25, 40**

Méthodologie

- Deux types de lois régionales : **fortement GP** (Conf 1–3) et **proche Expo** (Conf 4–6)
- Plusieurs configurations étudiées :
 - Conf1, Conf4 : **petites régions** (10 sites) avec beaucoup de données (450 obs. au total)
 - Conf2, Conf5 : **grandes régions** (20 sites) avec peu de données (450 obs. au total)
 - Conf3, Conf6 : **régions intermédiaires** (15 sites) avec beaucoup de données (700 obs. au total)
- Comparaison entre **3 estimateurs régionaux** :
 - IFL : Indice de Crue de Dalrymple (1960)
 - BAY : estimateur proposée par Ribatet et al. (2006)
 - REV : estimateur avec Sauts Réversibles
- Etude sur 3 quantiles $Q_{0.75}$, $Q_{0.95}$ et $Q_{0.995}$ (i.e. périodes de retour $\approx 2, 10$ et 100 ans)
- Taille échantillon au site cible : **10, 25, 40**

Méthodologie

- Deux types de lois régionales : **fortement GP** (Conf 1–3) et **proche Expo** (Conf 4–6)
- Plusieurs configurations étudiées :
 - Conf1, Conf4 : **petites régions** (10 sites) avec beaucoup de données (450 obs. au total)
 - Conf2, Conf5 : **grandes régions** (20 sites) avec peu de données (450 obs. au total)
 - Conf3, Conf6 : **régions intermédiaires** (15 sites) avec beaucoup de données (700 obs. au total)
- Comparaison entre **3 estimateurs régionaux** :
 - IFL : Indice de Crue de Dalrymple (1960)
 - BAY : estimateur proposée par Ribatet et al. (2006)
 - REV : estimateur avec Sauts Réversibles
- Etude sur 3 quantiles $Q_{0.75}$, $Q_{0.95}$ et $Q_{0.995}$ (i.e. périodes de retour $\approx 2, 10$ et 100 ans)
- Taille échantillon au site cible : **10, 25, 40**

Méthodologie

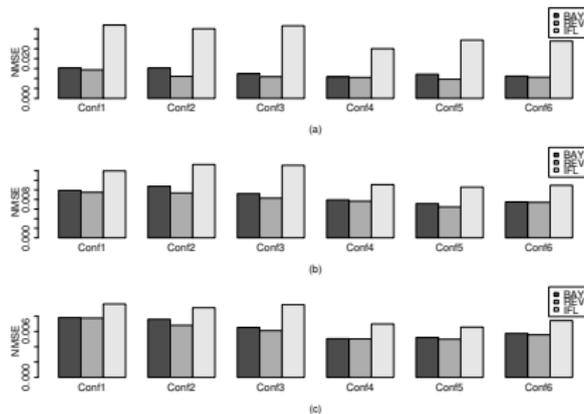
- Deux types de lois régionales : **fortement GP** (Conf 1–3) et **proche Expo** (Conf 4–6)
- Plusieurs configurations étudiées :
 - Conf1, Conf4 : **petites régions** (10 sites) avec beaucoup de données (450 obs. au total)
 - Conf2, Conf5 : **grandes régions** (20 sites) avec peu de données (450 obs. au total)
 - Conf3, Conf6 : **régions intermédiaires** (15 sites) avec beaucoup de données (700 obs. au total)
- Comparaison entre **3 estimateurs régionaux** :
 - IFL : Indice de Crue de Dalrymple (1960)
 - BAY : estimateur proposée par Ribatet et al. (2006)
 - REV : estimateur avec Sauts Réversibles
- Etude sur 3 quantiles $Q_{0.75}$, $Q_{0.95}$ et $Q_{0.995}$ (i.e. périodes de retour $\approx 2, 10$ et 100 ans)
- Taille échantillon au site cible : **10, 25, 40**

Comparaison BAY / IFL

Model	$Q_{0.75}$			$Q_{0.95}$			$Q_{0.995}$		
	NBIAS	SD	NMSE	NBIAS	SD	NMSE	NBIAS	SD	NMSE
					Conf1				
BAY	0.015	0.123	0.015	0.001	0.187	0.035	-0.006	0.318	0.101
IFL	0.037	0.189	0.037	0.025	0.195	0.038	-0.004	0.230	0.053
					Conf2				
BAY	0.019	0.122	0.015	0.030	0.249	0.063	0.110	0.561	0.326
IFL	0.041	0.183	0.035	0.025	0.191	0.037	-0.022	0.221	0.049
					Conf3				
BAY	0.019	0.110	0.012	0.006	0.174	0.030	-0.003	0.292	0.085
IFL	0.035	0.188	0.037	0.025	0.195	0.039	-0.002	0.222	0.049
					Conf4				
BAY	0.009	0.104	0.011	-0.007	0.149	0.022	-0.021	0.233	0.054
IFL	0.023	0.157	0.025	0.022	0.163	0.027	0.022	0.192	0.037
					Conf5				
BAY	0.018	0.109	0.012	0.012	0.193	0.037	0.033	0.378	0.144
IFL	0.036	0.168	0.029	0.033	0.173	0.031	0.024	0.197	0.039
					Conf6				
BAY	0.024	0.103	0.011	0.001	0.151	0.023	-0.038	0.222	0.050
IFL	0.028	0.168	0.029	0.028	0.177	0.032	0.028	0.202	0.042

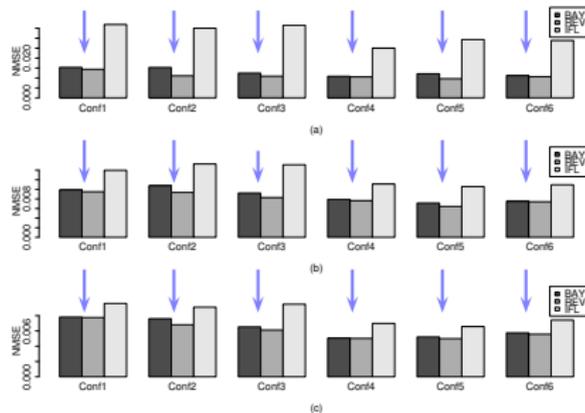
- $Q_{0.75}$: BAY est plus efficace (confirme étude précédente)
- $Q_{0.95}$:
 - BAY meilleur pour petites et moyennes régions
 - IFL meilleur sur les grandes régions
- $Q_{0.995}$: IFL est plus efficace

Comparaison Globale



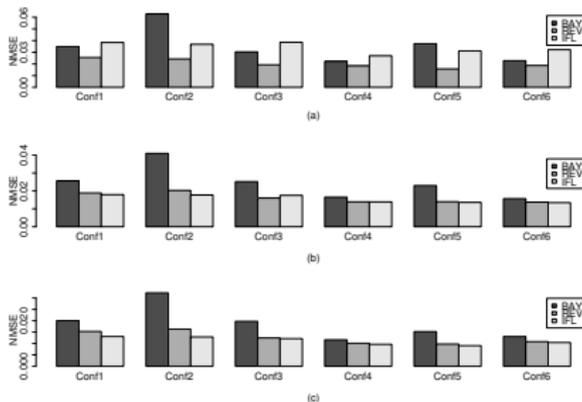
- $Q_{0.75}$:
- $Q_{0.95}$:
- $Q_{0.995}$:

Comparaison Globale



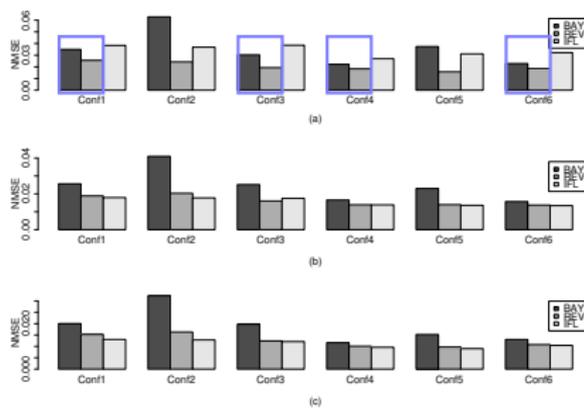
- $Q_{0.75}$: approches Bayésiennes meilleures
- $Q_{0.95}$:
- $Q_{0.995}$:

Comparaison Globale



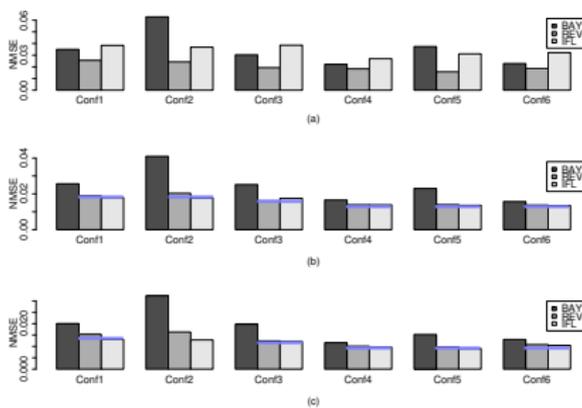
- $Q_{0.75}$: approches Bayésiennes meilleures
- $Q_{0.95}$:
- $Q_{0.995}$:

Comparaison Globale



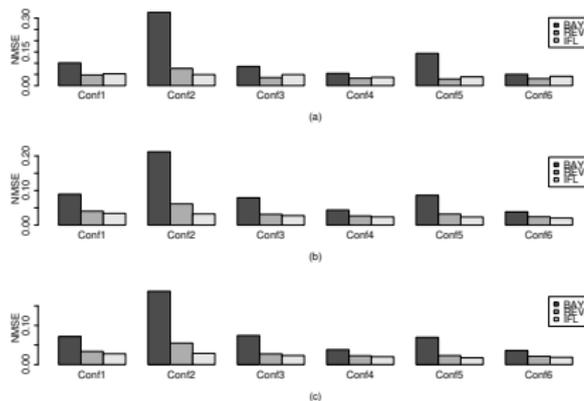
- $Q_{0.75}$: approches Bayésiennes meilleures
- $Q_{0.95}$:
 - Peu données au site cible : modèles Bayésiens meilleurs
 - Sinon REV et IFL aussi performants
- $Q_{0.995}$:

Comparaison Globale



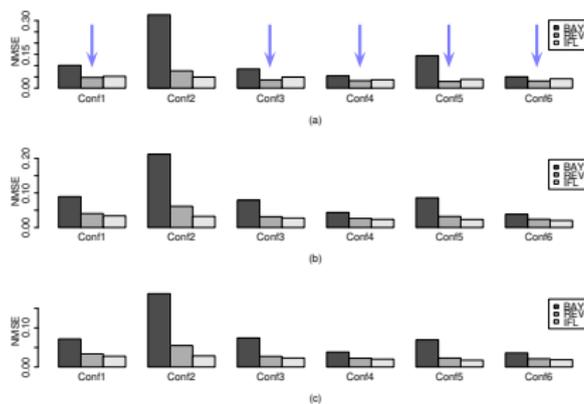
- $Q_{0.75}$: approches Bayésiennes meilleures
- $Q_{0.95}$:
 - Peu données au site cible : modèles Bayésiens meilleurs
 - Sinon REV et IFL aussi performants
- $Q_{0.995}$:

Comparaison Globale



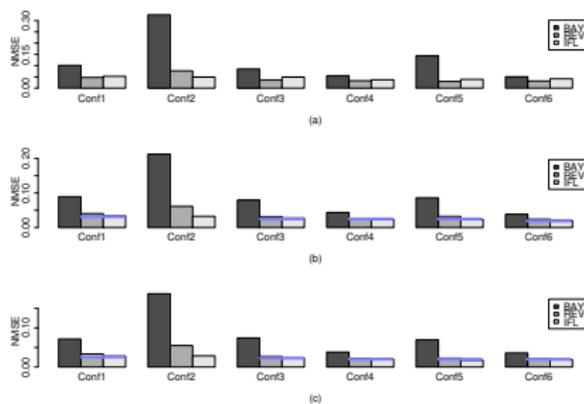
- $Q_{0.75}$: approches Bayésiennes meilleures
- $Q_{0.95}$: REV et IFL meilleurs
- $Q_{0.995}$:

Comparaison Globale



- $Q_{0.75}$: approches Bayésiennes meilleures
- $Q_{0.95}$: REV et IFL meilleurs
- $Q_{0.995}$:
 - Peu données au site cible : REV meilleur
 - Sinon REV et IFL aussi performants

Comparaison Globale



- $Q_{0.75}$: approches Bayésiennes meilleures
- $Q_{0.95}$: REV et IFL meilleurs
- $Q_{0.995}$:
 - Peu données au site cible : REV meilleur
 - Sinon REV et IFL aussi performants

Influence de l'Indice de Crue

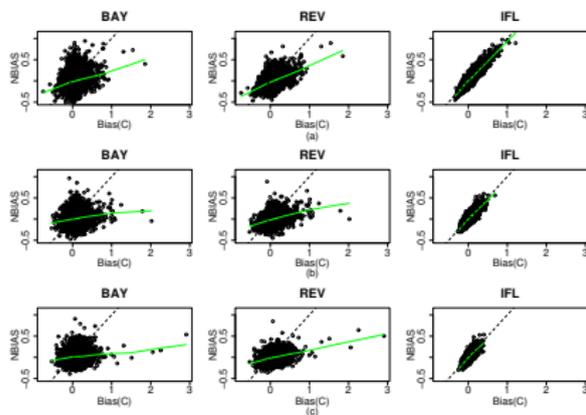
- Les 3 estimateurs font appel à l'estimation de l'indice de Crue
- Comment se répercute un biais sur ce paramètre selon les modèles ?
 - Approches Bayésiennes moins sensibles aux forts biais
 - Réponse linéaire pour IFL par construction du modèle
 - Différence entre REV et BAY : REV moins variable (par définition des SR)

Influence de l'Indice de Crue

- Les 3 estimateurs font appel à l'estimation de l'indice de Crue
- Comment se répercute un biais sur ce paramètre selon les modèles ?
 - Approches Bayésiennes moins sensibles aux forts biais
 - Réponse linéaire pour IFL par construction du modèle
 - Différence entre REV et BAY : REV moins variable (par définition des SR)

Influence de l'Indice de Crue

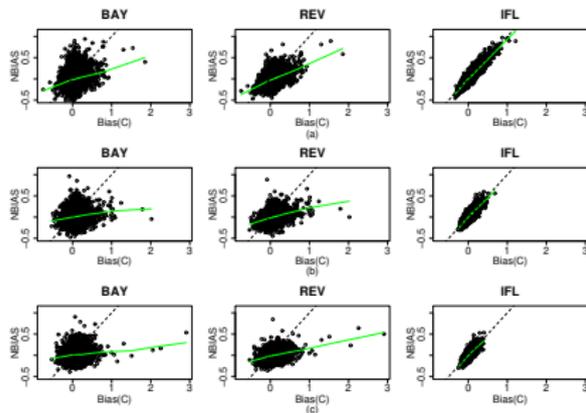
- Les 3 estimateurs font appel à l'estimation de l'indice de Crue
- Comment se répercute un biais sur ce paramètre selon les modèles ?



- Approches Bayésiennes moins sensibles aux forts biais
- Réponse linéaire pour IFL par construction du modèle
- Différence entre REV et BAY : REV moins variable (par définition des SR)

Influence de l'Indice de Crue

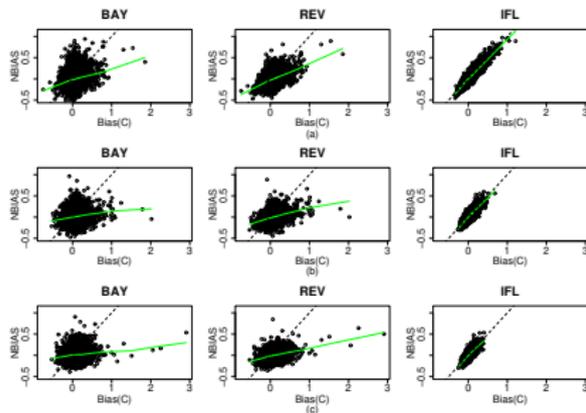
- Les 3 estimateurs font appel à l'estimation de l'indice de Crue
- Comment se répercute un biais sur ce paramètre selon les modèles ?



- Approches Bayésiennes moins sensibles aux forts biais
- Réponse linéaire pour IFL par construction du modèle
- Différence entre REV et BAY : REV moins variable (par définition des SR)

Influence de l'Indice de Crue

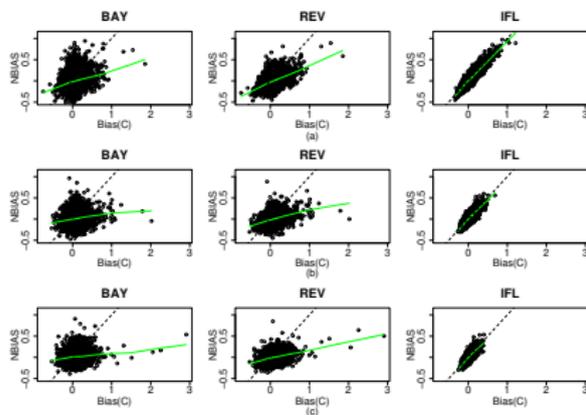
- Les 3 estimateurs font appel à l'estimation de l'indice de Crue
- Comment se répercute un biais sur ce paramètre selon les modèles ?



- Approches Bayésiennes moins sensibles aux forts biais
- Réponse linéaire pour IFL par construction du modèle
- Différence entre REV et BAY : REV moins variable (par définition des SR)

Influence de l'Indice de Crue

- Les 3 estimateurs font appel à l'estimation de l'indice de Crue
- Comment se répercute un biais sur ce paramètre selon les modèles ?



- Approches Bayésiennes moins sensibles aux forts biais
- Réponse linéaire pour IFL par construction du modèle
- Différence entre REV et BAY : REV moins variable (par définition des SR)

Synthèse des Résultats

- Modèle considérant avec une **probabilité** > 0 un paramètre de **forme régional**
- Ce **paramètre régional** n'est pas imposé mais **validé ou infirmé** par les **données** au **site cible**
- **Performance** du modèle Bayésien initial pour $T < 20$ ans est **conservée**
- Pour $T > 20$ ans, deux cas de figures
 - **Peu de données** au site cible : **REV meilleur** que IFL
 - **Sinon** : **REV = IFL**
- **Approches Bayésiennes** sont **moins sensibles** à l'**erreur** d'estimation sur l'**indice de crue**
⇒ Points essentiel car erreur incontrôlable
- **Conseil** : **Incorporer** des **sites bien instrumentés** dans la **région homogène**

Synthèse des Résultats

- Modèle considérant avec une **probabilité** > 0 un paramètre de **forme régional**
- Ce **paramètre régional** n'est pas imposé mais **validé ou infirmé** par les **données** au **site cible**
- **Performance** du modèle Bayésien initial pour $T < 20$ ans est **conservée**
- Pour $T > 20$ ans, deux cas de figures
 - Peu de données au site cible : REV meilleur que IFL
 - Sinon : REV = IFL
- **Approches Bayésiennes** sont **moins sensibles** à l'erreur d'estimation sur l'**indice de crue**
⇒ Points essentiel car erreur incontrôlable
- **Conseil** : Incorporer des **sites bien instrumentés** dans la région homogène

Synthèse des Résultats

- Modèle considérant avec une **probabilité** > 0 un paramètre de **forme régional**
- Ce **paramètre régional** n'est pas imposé mais **validé ou infirmé** par les **données** au **site cible**
- **Performance** du modèle Bayésien initial pour $T < 20$ ans est **conservée**
- Pour $T > 20$ ans, deux cas de figures
 - Peu de données au site cible : REV meilleur que IFL
 - Sinon : REV = IFL
- **Approches Bayésiennes** sont **moins sensibles** à l'erreur d'estimation sur l'**indice de crue**
⇒ Points essentiel car erreur incontrôlable
- **Conseil** : Incorporer des **sites bien instrumentés** dans la région homogène

Synthèse des Résultats

- Modèle considérant avec une **probabilité** > 0 un paramètre de **forme régional**
- Ce **paramètre régional** n'est pas imposé mais **validé ou infirmé** par les **données** au **site cible**
- **Performance** du modèle Bayésien initial pour $T < 20$ ans est **conservée**
- Pour $T > 20$ ans, deux cas de figures
 - **Peu de données** au site cible : **REV meilleur** que IFL
 - **Sinon** : **REV = IFL**
- **Approches Bayésiennes** sont **moins sensibles** à l'**erreur d'estimation** sur l'**indice de crue**
⇒ Points essentiel car erreur incontrôlable
- **Conseil** : **Incorporer** des **sites bien instrumentés** dans la **région homogène**

Synthèse des Résultats

- Modèle considérant avec une **probabilité** > 0 un paramètre de **forme régional**
- Ce **paramètre régional** n'est pas imposé mais **validé ou infirmé** par les **données** au **site cible**
- **Performance** du modèle Bayésien initial pour $T < 20$ ans est **conservée**
- Pour $T > 20$ ans, deux cas de figures
 - **Peu de données** au site cible : **REV meilleur** que IFL
 - Sinon : **REV = IFL**
- **Approches Bayésiennes** sont **moins sensibles** à l'erreur d'estimation sur l'**indice de crue**
⇒ Points essentiel car erreur incontrôlable
- **Conseil** : **Incorporer** des **sites bien instrumentés** dans la **région homogène**

Synthèse des Résultats

- Modèle considérant avec une **probabilité** > 0 un paramètre de **forme régional**
- Ce **paramètre régional** n'est pas imposé mais **validé ou infirmé** par les **données** au **site cible**
- **Performance** du modèle Bayésien initial pour $T < 20$ ans est **conservée**
- Pour $T > 20$ ans, deux cas de figures
 - **Peu de données** au site cible : **REV meilleur** que IFL
 - **Sinon** : **REV = IFL**
- **Approches Bayésiennes** sont **moins sensibles** à l'erreur d'estimation sur l'**indice de crue**
⇒ Points essentiel car erreur incontrôlable
- **Conseil** : **Incorporer** des **sites bien instrumentés** dans la **région homogène**

Synthèse des Résultats

- Modèle considérant avec une **probabilité** > 0 un paramètre de **forme régional**
- Ce **paramètre régional** n'est pas imposé mais **validé ou infirmé** par les **données** au **site cible**
- **Performance** du modèle Bayésien initial pour $T < 20$ ans est **conservée**
- Pour $T > 20$ ans, deux cas de figures
 - **Peu de données** au site cible : **REV meilleur** que IFL
 - **Sinon** : **REV = IFL**
- **Approches Bayésiennes** sont **moins sensibles** à l'**erreur** d'estimation sur l'**indice de crue**
⇒ Points essentiel car erreur incontrôlable
- **Conseil** : **Incorporer** des **sites bien instrumentés** dans la **région homogène**

Synthèse des Résultats

- Modèle considérant avec une **probabilité** > 0 un paramètre de **forme régional**
- Ce **paramètre régional** n'est pas imposé mais **validé ou infirmé** par les **données** au **site cible**
- **Performance** du modèle Bayésien initial pour $T < 20$ ans est **conservée**
- Pour $T > 20$ ans, deux cas de figures
 - **Peu de données** au site cible : **REV meilleur** que IFL
 - **Sinon** : **REV = IFL**
- **Approches Bayésiennes** sont **moins sensibles** à l'**erreur** d'estimation sur l'**indice de crue**
⇒ Points essentiel car erreur incontrôlable
- **Conseil** : **Incorporer** des **sites bien instrumentés** dans la **région homogène**

Logiciels

Tous les analyses ont été réalisées sous le logiciel libre
 disponible librement à l'adresse

<http://www.R-project.org>

Nous avons implémenté 3 packages pour ce logiciel

- **RFA** Outils pour l'Analyse Fréquentielle Régionale des Extrêmes¹
- **POT** Modélisation (multivariée) des Sup-Seuil¹
- **evdbayes** Analyse Bayésienne pour la Théorie des Valeurs Extrêmes²

Ils sont librement disponibles sur le site de [ou](#) sur ma page.
Des guides d'utilisateurs et des démos sont disponibles également.

¹Ribatet, M.

²Stephenson, A. et Ribatet, M.

Logiciels

Tous les analyses ont été réalisées sous le logiciel libre
 disponible librement à l'adresse

<http://www.R-project.org>

Nous avons implémenté 3 packages pour ce logiciel

- **RFA** Outils pour l'Analyse Fréquentielle Régionale des Extrêmes¹
- **POT** Modélisation (multivariée) des Sup-Seuil¹
- **evdbayes** Analyse Bayésienne pour la Théorie des Valeurs Extrêmes²

Ils sont librement disponibles sur le site de [ou](#) sur ma page.
Des guides d'utilisateurs et des démos sont disponibles également.

¹Ribatet, M.

²Stephenson, A. et Ribatet, M.

Logiciels

Tous les analyses ont été réalisées sous le logiciel libre
 disponible librement à l'adresse

<http://www.R-project.org>

Nous avons implémenté 3 packages pour ce logiciel

- **RFA** Outils pour l'Analyse Fréquentielle Régionale des Extrêmes¹
- **POT** Modélisation (multivariée) des Sup-Seuil¹
- **evdbayes** Analyse Bayésienne pour la Théorie des Valeurs Extrêmes²

Ils sont librement disponibles sur le site de [ou](#) sur ma page.
Des guides d'utilisateurs et des démos sont disponibles également.

¹Ribatet, M.

²Stephenson, A. et Ribatet, M.

Logiciels

Tous les analyses ont été réalisées sous le logiciel libre  disponible librement à l'adresse

<http://www.R-project.org>

Nous avons implémenté 3 packages pour ce logiciel

- **RFA** Outils pour l'Analyse Fréquentielle Régionale des Extrêmes¹
- **POT** Modélisation (multivariée) des Sup-Seuil¹
- **evdbayes** Analyse Bayésienne pour la Théorie des Valeurs Extrêmes²

Ils sont librement disponibles sur le site de [ou](#) sur ma page. Des guides d'utilisateurs et des démos sont disponibles également.

¹Ribatet, M.

²Stephenson, A. et Ribatet, M.

Logiciels

Tous les analyses ont été réalisées sous le logiciel libre  disponible librement à l'adresse

<http://www.R-project.org>

Nous avons implémenté 3 packages pour ce logiciel

- **RFA** Outils pour l'Analyse Fréquentielle Régionale des Extrêmes¹
- **POT** Modélisation (multivariée) des Sup-Seuil¹
- **evdbayes** Analyse Bayésienne pour la Théorie des Valeurs Extrêmes²

Ils sont librement disponibles sur le site de  ou sur ma page. Des guides d'utilisateurs et des démos sont disponibles également.

¹Ribatet, M.

²Stephenson, A. et Ribatet, M.

Logiciels

Tous les analyses ont été réalisées sous le logiciel libre  disponible librement à l'adresse

<http://www.R-project.org>

Nous avons implémenté 3 packages pour ce logiciel

- **RFA** Outils pour l'Analyse Fréquentielle Régionale des Extrêmes¹
- **POT** Modélisation (multivariée) des Sup-Seuil¹
- **evdbayes** Analyse Bayésienne pour la Théorie des Valeurs Extrêmes²

Ils sont librement disponibles sur le site de  ou sur ma page. Des guides d'utilisateurs et des démos sont disponibles également.

¹Ribatet, M.

²Stephenson, A. et Ribatet, M.

For Further Reading



Coles, S. (2001).

An Introduction to Statistical Modelling of Extreme Values.
Springer Series in Statistics. Springer Series in Statistics, London.



Green, P. (1995).

Reversible jump markov chain monte carlo computation and bayesian model determination.
Biometrika, 82 :711–732.



Hosking, J. R. M. and Wallis, J. R. (1997).

Regional Frequency Analysis.
Cambridge University Press.



M. Ribatet, E. Sauquet, J-M. Grésillon, and T. Ouarda.

A Regional Bayesian POT Model for Flood Frequency Analysis.
SERRA, 21(4), 327–339, 2007.



M. Ribatet, E. Sauquet, J-M. Grésillon, and T. Ouarda.

Usefulness of the Reversible Markov Chain Monte Carlo Model in Regional Flood Frequency Analysis.
Water Resources Research, (in press), 2007.

A scenic landscape of a rocky river flowing through a forested valley. The river is filled with numerous large, smooth, brown rocks of various sizes. The water is a deep blue, reflecting the clear sky. The surrounding area is lush with green vegetation, including tall evergreen trees and dense shrubs. In the background, rolling green hills and mountains are visible under a bright blue sky with a few wispy clouds. The overall atmosphere is peaceful and natural.

Merci Beaucoup !