

# COMPORTEMENT ASYMPTOTIQUE DE LA DISTRIBUTION DES PLUIES EXTRÊMES EN FRANCE

Aurélie MULLER

Thèse présentée pour obtenir le diplôme de docteur de l'Université Montpellier 2

24 novembre 2006



Patrick ARNAUD	Cemagref Aix-en-Provence
Jean-Noël BACRO	Université Montpellier 2
Michel LANG	Cemagref Lyon

# Travaux interdisciplinaires

## Domaines

- Hydrologie
- Statistiques des valeurs extrêmes

## Problématiques Hydrologique et Statistique

- Gestion des risques d'inondation :
  - dimensionnement d'ouvrages hydrauliques de protection
  - établissement de Plans de Prévention des Risques liés aux Inondations
  - définition d'une échelle d'intensités de pluie (faible, moyenne, forte)
- Diverses méthodes en hydrologie (Gradex, Agregée, Shypre, etc.) : des résultats différents. Que choisir ?
- Application de la théorie des valeurs extrêmes au cas des données de pluie.

# Comment analyser les valeurs extrêmes ?

## Un point de vue local pour décrire les valeurs extrêmes de pluie

- 1 Analyse des valeurs extrêmes **indépendantes** (e.g. des maxima annuels ou des dépassements de seuil fixé)
  
- 2 Analyse stochastique du processus des valeurs extrêmes **dépendantes** (e.g. averses fortes consécutives d'un même événement pluvieux)

# Comment analyser les valeurs extrêmes ?

## Un point de vue local pour décrire les valeurs extrêmes de pluie

- 1 Analyse des valeurs extrêmes **indépendantes** (e.g. des maxima annuels ou des dépassements de seuil fixé)
  - analyse uni-variée des extrêmes de pluie à un pas de temps de mesure donné
  - analyse multi-variée des extrêmes de pluie à différents pas de temps
- 2 Analyse stochastique du processus des valeurs extrêmes **dépendantes** (e.g. averses fortes consécutives d'un même événement pluvieux)

# Comment analyser les valeurs extrêmes ?

## Un point de vue local pour décrire les valeurs extrêmes de pluie

- 1 Analyse des valeurs extrêmes **indépendantes** (e.g. des maxima annuels ou des dépassements de seuil fixé)
  - analyse uni-variée des extrêmes de pluie à un pas de temps de mesure donné
  - analyse multi-variée des extrêmes de pluie à différents pas de temps
- 2 Analyse stochastique du processus des valeurs extrêmes **dépendantes** (e.g. averses fortes consécutives d'un même événement pluvieux)
  - générateurs stochastiques de pluie (e.g. Shypre)
  - modélisation d'une série i.d. de valeurs extrêmes

# Loi des valeurs extrêmes (Fisher et Tippett, 1928)

Soient  $M_n = \max\{X_1, \dots, X_n\}$  un échantillon de loi  $F$  inconnue.

Sous certaines hypothèses,

la loi de  $M_n$  est

asymptotiquement une loi

**GEV** (Generalized Extreme Value) :

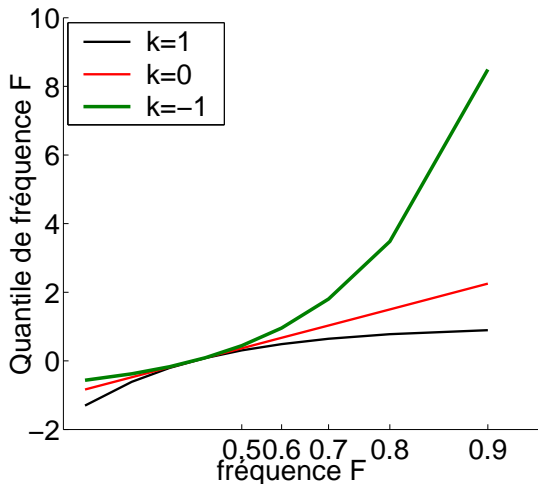
$$\mathbb{P}(M_n \leq x) \approx$$

$$\begin{cases} \exp(-[1 - k(x - \beta)/\alpha]^{1/k}), & k \neq 0, & x \text{ tel que } 1 - k(x - \beta)/\alpha > 0 \\ \exp(-\exp(-(x - \beta)/\alpha)), & k = 0, & x \in \mathbb{R}, \text{ 'loi Gumbel' } \end{cases}$$

$$\text{où } \begin{cases} \alpha > 0, & \text{paramètre d'échelle} \\ \beta \in \mathbb{R}, & \text{paramètre de position} \\ k \in \mathbb{R}, & \text{paramètre de forme} \end{cases}$$

Soient  $M_n = \max\{X_1, \dots, X_n\}$  un échantillon de loi  $F$  inconnue.

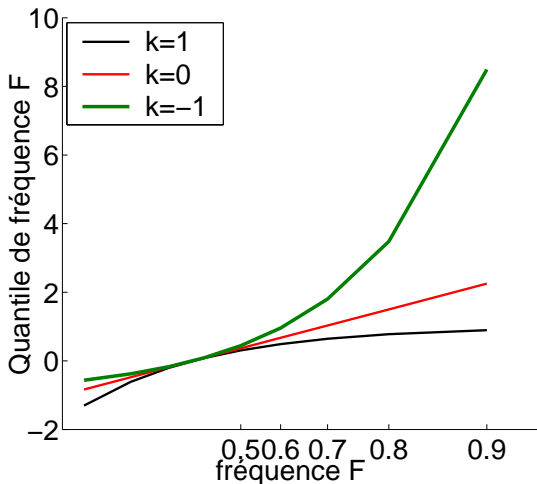
Sous certaines hypothèses,  
la loi de  $M_n$  est  
asymptotiquement une loi  
**GEV** (Generalized Extreme  
Value) :



Soient  $M_n = \max\{X_1, \dots, X_n\}$  un échantillon de loi  $F$  inconnue.

Sous certaines hypothèses, la loi de  $M_n$  est asymptotiquement une loi **GEV** (Generalized Extreme Value) :

**Loi Gumbel** : usage fréquent par les hydrologues français





# Techniques d'estimation

## des paramètres des lois de valeurs extrêmes

### Méthodes les plus courantes

maximum de vraisemblance, méthodes des moments, moments pondérés (Hosking et al.1985), méthode bayésienne

### Choix du cadre bayésien

# Techniques d'estimation

## des paramètres des lois de valeurs extrêmes

### Méthodes les plus courantes

maximum de vraisemblance, méthodes des moments, moments pondérés (Hosking et al.1985), méthode **bayésienne**

### Choix du cadre bayésien



- Inclusion d'informations exogènes aux données
- Accès direct à la distribution des quantiles ...
- ... sans recours à des théorèmes asymptotiques
- Applicable avec des vraisemblances complexes

# Techniques d'estimation

## des paramètres des lois de valeurs extrêmes

### Méthodes les plus courantes

maximum de vraisemblance, méthodes des moments, moments pondérés (Hosking et al.1985), méthode **bayésienne**

### Choix du cadre bayésien



- Inclusion d'informations exogènes aux données
- Accès direct à la distribution des quantiles ...
- ... sans recours à des théorèmes asymptotiques
- Applicable avec des vraisemblances complexes
- Nécessite des algorithmes sophistiqués (Monte Carlo par Chaînes de Markov ou échantillonnage préférentiel)



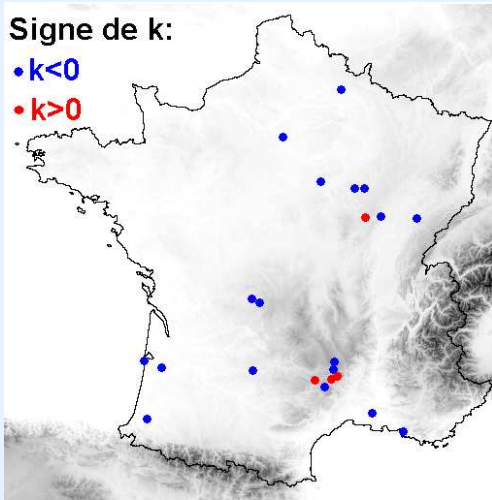
# Remise en question de la loi Gumbel ?

22 longues séries de plus de 100 ans de mesures.  
Estimations des maxima annuels journaliers.

Signe de  $k$ :

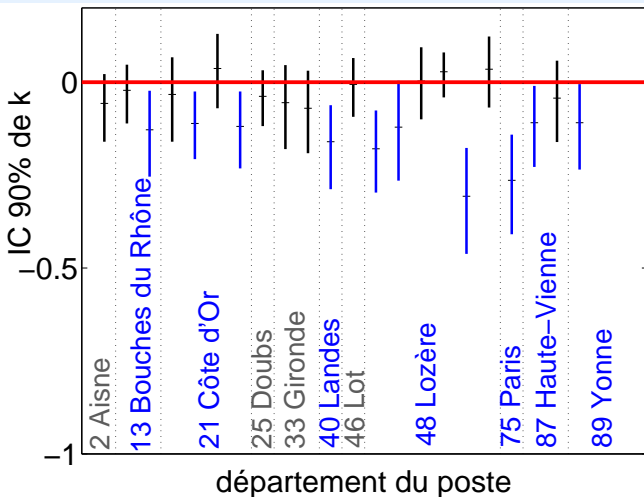
•  $k < 0$

•  $k > 0$



# Remise en question de la loi Gumbel ?

Mais difficile de rejeter la loi Gumbel...



# Plan de l'exposé

- 1 **Maxima annuels de pluie : une analyse multi-durées**
- 2 **Extrêmes simulés par un générateur stochastique de pluie**
- 3 **Processus de persistance des averses**
- 4 **Conclusions et perspectives**

# Plan de l'exposé

- 1 **Maxima annuels de pluie : une analyse multi-durées**
- 2 Extrêmes simulés par un générateur stochastique de pluie
- 3 Processus de persistance des averses
- 4 Conclusions et perspectives

# Information multi-durées

*Mesures de pluie*    *Intérêt*

---

pas de temps 1h

journalière



# Information multi-durées

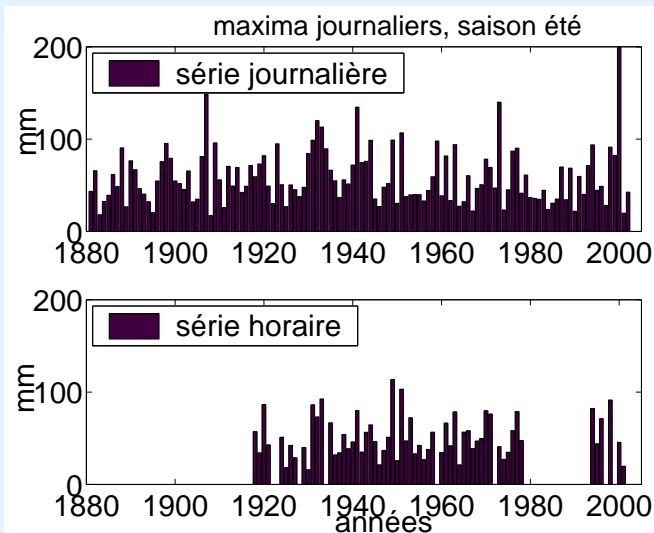
*Mesures de pluie*    *Intérêt*

---

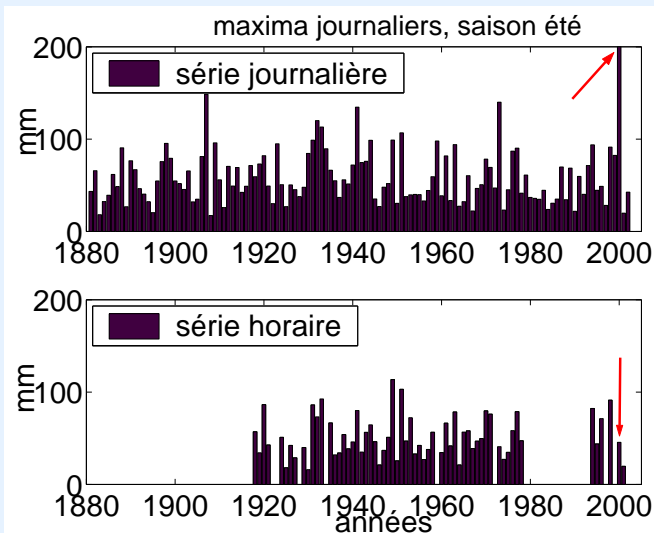
pas de temps 1h    agglomération sur différentes durées (1h-72h),  
⇒ **meilleure description des extrêmes**,  
important en milieu urbain ou pour des petits  
bassins versants qui réagissent vite

journalière    **réseau dense de mesures, données plus  
renseignées**

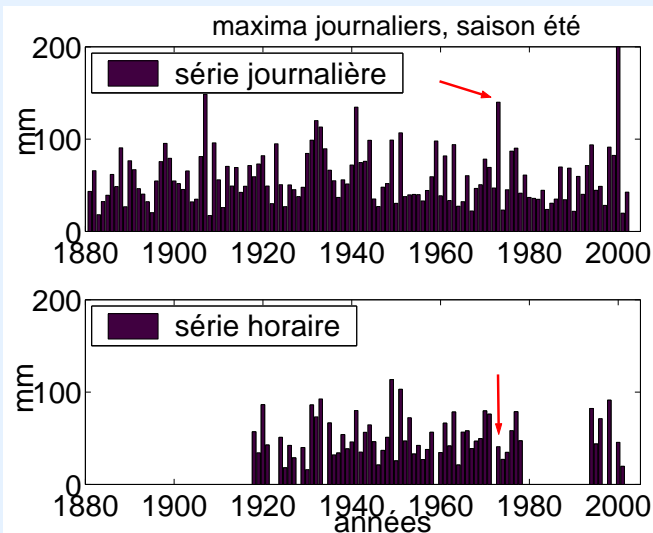
# Exemple : information multi-durées à Marseille



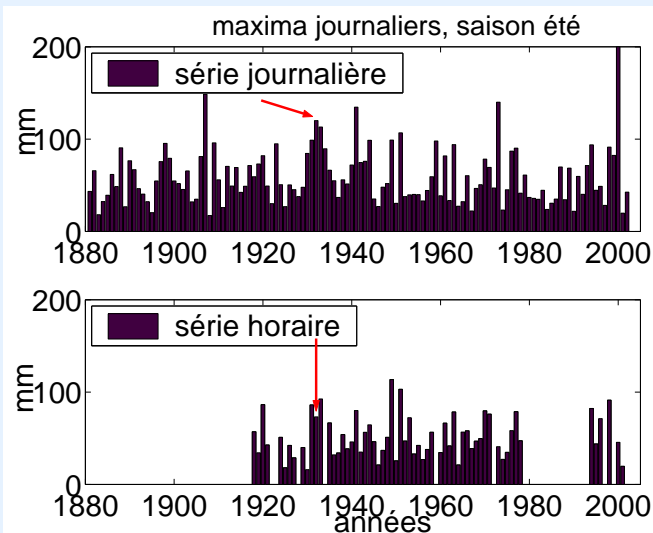
# Exemple : information multi-durées à Marseille



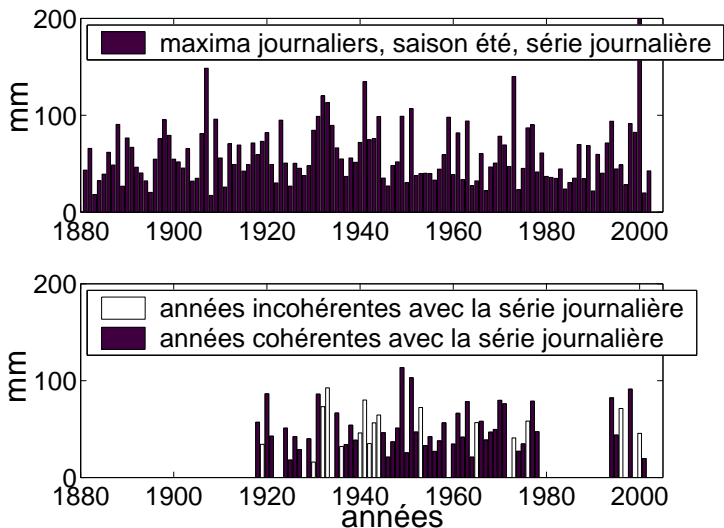
# Exemple : information multi-durées à Marseille



# Exemple : information multi-durées à Marseille



# Exemple : information multi-durées à Marseille



# Modèle multi-durées pour les maxima annuels

## Notation

$P_d$  : cumul maximal annuel ou saisonnier de pluie précipitée en  $d$  heures

$P_J$  : cumul maximal annuel ou saisonnier de pluie journalière  
( $P_J \neq P_{24}$ )

## Pour définir le **modèle multi-varié** ( $P_d, d \in \{1, \dots, 72, J\}$ )

- **Loi marginale.** Pour chaque durée  $d$  : une loi GEV de paramètres  $\alpha_d, \beta_d, k_d$
- Définir les **relations entre les paramètres**  $(\alpha_d, \beta_d, k_d)_{d \in \{1, 2, \dots, 72, J\}}$  des lois marginales.
- Définir la **loi jointe** du vecteur  $\{(P_d)_{d \in \{1, 2, \dots, 72, J\}}\}$

# Modèle multi-durées pour les maxima annuels

Relations entre paramètres des lois GEV marginales de  $\{(P_d)_{d \in \{1, 2, \dots, 72, J\}}\}$

## 1 Relations entre les paramètres (Koutsoyiannis et al. 1998) :

$$\text{pour } d \in \{1, \dots, 72\} \begin{cases} \alpha_d = d\alpha / (d + \vartheta)^\varepsilon \\ \beta_d = d\beta / (d + \vartheta)^\varepsilon \\ k_d = k, \end{cases} \text{ avec } \vartheta > 0, \varepsilon \in (0, 1).$$



# Modèle multi-durées pour les maxima annuels

Relations entre paramètres des lois GEV marginales de  $\{(P_d)_{d \in \{1, 2, \dots, 72, J\}}\}$

- 1 Relations entre les paramètres (Koutsoyiannis et al. 1998) :

$$\text{pour } d \in \{1, \dots, 72\} \begin{cases} \alpha_d = d\alpha / (d + \vartheta)^\varepsilon \\ \beta_d = d\beta / (d + \vartheta)^\varepsilon \\ k_d = k, \end{cases} \text{ avec } \vartheta > 0, \varepsilon \in (0, 1).$$

- 2 Relation d'**ordre** entre les cumulés :  $d \leq d' \Rightarrow P_d \leq P_{d'} \leq \frac{d'}{d} P_d$ .

# Modèle multi-durées pour les maxima annuels

Relations entre paramètres des lois GEV marginales de  $\{(P_d)_{d \in \{1, 2, \dots, 72, J\}}\}$

- 1 Relations entre les paramètres (Koutsoyiannis et al. 1998) :

$$\text{pour } d \in \{1, \dots, 72\} \begin{cases} \alpha_d = d\alpha / (d + \vartheta)^\varepsilon \\ \beta_d = d\beta / (d + \vartheta)^\varepsilon \\ k_d = k, \end{cases} \text{ avec } \vartheta > 0, \varepsilon \in (0, 1).$$

- 2 Relation d'**ordre** entre les cumulés :  $d \leq d' \Rightarrow P_d \leq P_{d'} \leq \frac{d'}{d} P_d$ .  
(Nadarajah et al. 1998) : si  $P_d$  et  $P_{d'}$  suivent des lois GEV telles que  $k_d \leq 0, k_{d'} \leq 0$  alors  $k_d = k_{d'}$ .

# Modèle multi-durées pour les maxima annuels

Relations entre paramètres des lois GEV marginales de  $\{(P_d)_{d \in \{1,2,\dots,72,J\}}\}$

- 1 Relations entre les paramètres (Koutsoyiannis et al.1998) :

$$\text{pour } d \in \{1, \dots, 72\} \begin{cases} \alpha_d = d\alpha / (d + \vartheta)^\varepsilon \\ \beta_d = d\beta / (d + \vartheta)^\varepsilon \\ k_d = k, \end{cases} \text{ avec } \vartheta > 0, \varepsilon \in (0, 1).$$

- 2 Relation d'ordre entre les cumulés :  $d \leq d' \Rightarrow P_d \leq P_{d'} \leq \frac{d'}{d} P_d$ .  
(Nadarajah et al. 1998) : si  $P_d$  et  $P_{d'}$  suivent des lois GEV telles que  $k_d \leq 0, k_{d'} \leq 0$  alors  $k_d = k_{d'}$ .

- 3  $P_J$  et  $P_{24}$  : étude d'un **même processus, i.d., mesuré à différentes fréquences**. (Robinson et Tawn 2000) :

$$\mathbb{P}(P_{24} \leq x) \approx \mathbb{P}(P_J \leq x)^\Theta \quad \text{avec } \Theta \in [1, 24]$$

$$\Rightarrow \begin{cases} \alpha_{24} = \alpha_J \Theta^{-k_J} \\ \beta_{24} = \beta_J + \alpha_J / k_J (1 - \Theta^{-k_J}) \\ k_{24} = k_J \end{cases}$$

# Modèle multi-durées pour les maxima annuels

En résumé, les paramètres du modèle sont :

$\alpha_J, \beta_J, k_J, \Theta, \vartheta, \varepsilon$

# Modèle multi-durées pour les maxima annuels

## En résumé, les paramètres du modèle sont :

$\alpha_J, \beta_J, k_J, \Theta, \vartheta, \varepsilon$

On peut retrouver les paramètres des GEV de chaque série de durée  $d$  :

- $\alpha_d = \frac{d(24+\vartheta)^\varepsilon}{24(d+\vartheta)^\varepsilon} \alpha_J \Theta^{-k_J}$
- $\beta_d = \frac{d(24+\vartheta)^\varepsilon}{24(d+\vartheta)^\varepsilon} (\beta_J + \alpha_J (1 - \Theta^{-k_J}) / k_J)$
- $k_d = k_J$

# Modèle multi-durées pour les maxima annuels

## En résumé, les paramètres du modèle sont :

$$\alpha_J, \beta_J, k_J, \Theta, \vartheta, \varepsilon$$

On peut retrouver les paramètres des GEV de chaque série de durée  $d$  :

- $\alpha_d = \frac{d(24+\vartheta)^\varepsilon}{24(d+\vartheta)^\varepsilon} \alpha_J \Theta^{-k_J}$
- $\beta_d = \frac{d(24+\vartheta)^\varepsilon}{24(d+\vartheta)^\varepsilon} (\beta_J + \alpha_J(1 - \Theta^{-k_J})/k_J)$
- $k_d = k_J$

## En fait,

deux types de pluies (convectives/frontales)

⇒ deux paramétrisations (petites durées/grandes durées)

⇒ 8 paramètres au total :

$$\alpha_J, \beta_J, k_J, \Theta, \vartheta, \varepsilon, \alpha', \vartheta'.$$

# Modèles multi-durées pour les maxima annuels

## La loi jointe

3 modèles :

- 1  $V_1$  : on utilise les durées 1h, 6h, 12h, 24h, 72h et  $J$ , en les supposant **indépendantes**.

# Modèles multi-durées pour les maxima annuels

## La loi jointe

3 modèles :

- 1  $V_1$  : on utilise les durées 1h, 6h, 12h, 24h, 72h et  $J$ , en les supposant **indépendantes**.
- 2  $V_2$  : hypothèse d'**indépendance**, avec les pluies de 1h, 24h, 72h,



# Modèles multi-durées pour les maxima annuels

## La loi jointe

3 modèles :

- 1  $V_1$  : on utilise les durées 1h, 6h, 12h, 24h, 72h et  $J$ , en les supposant **indépendantes**.
- 2  $V_2$  : hypothèse d'**indépendance**, avec les pluies de 1h, 24h, 72h, et  $J$  sur une **période d'observation disjointe** des pluies horaires.

# Modèles multi-durées pour les maxima annuels

## La loi jointe

3 modèles :

- 1  $V_1$  : on utilise les durées 1h, 6h, 12h, 24h, 72h et  $J$ , en les supposant **indépendantes**.
- 2  $V_2$  : hypothèse d'**indépendance**, avec les pluies de 1h, 24h, 72h, et  $J$  sur une **période d'observation disjointe** des pluies horaires.
- 3  $V_3$  : on reprend les durées précédentes et on introduit une **loi bi-variée extrême pour les pluies de 24h,72h**. On utilise la loi bi-variée logistique :

$$G(x, y) = \exp(-(x^{-1/\phi} + y^{-1/\phi})^\phi), \quad 0 < \phi \leq 1.$$

$\phi \rightarrow 0$  : **dépendance totale**

$\phi = 1$  : **indépendance**

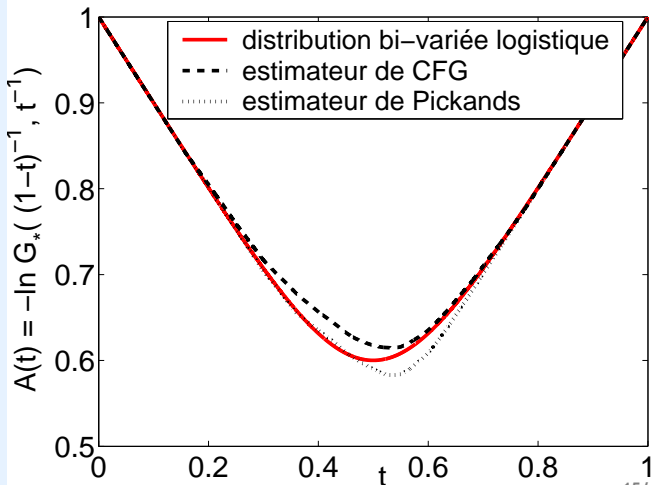
# Loi de $(P_{24}, P_{72})$ : loi bi-variée logistique ?

Vérification sur données de Marseille

## Test de déviance

⇒ Le modèle logistique est meilleur que le modèle d'indépendance avec une  $p$ -value  $< 0.1\%$ .

## Fonction de dépendance de Pickands



# Estimation bayésienne

## Loi a priori

Choix de lois a priori peu informatives :

$\alpha_J, \alpha'$	loi lognormale (0,100)
$\beta_J$	loi normale (0,100)
$k_J$	loi uniforme sur [-1,1]
$\Theta$	loi uniforme sur [1,24]
$\vartheta, \vartheta'$	loi normale (0,100), tronquée en 0
$\varepsilon$	loi lognormale (0,100)
$\phi$ (dans $V_3$ )	loi uniforme sur [0,1]

complété par des restrictions dues à la physique des phénomènes :

- $d \leq d' \Rightarrow P_d \leq P_{d'} \leq \frac{d'}{d} P_d$
- $P_J \leq P_{24} \leq 2P_J$
- $P_1(T = 10ans) \in [21, 60], P_6(T = 10ans) \in [51, 95]$  (information régionale sur zone méditerranéenne)

# Application à Marseille : Résultats

## Rappel

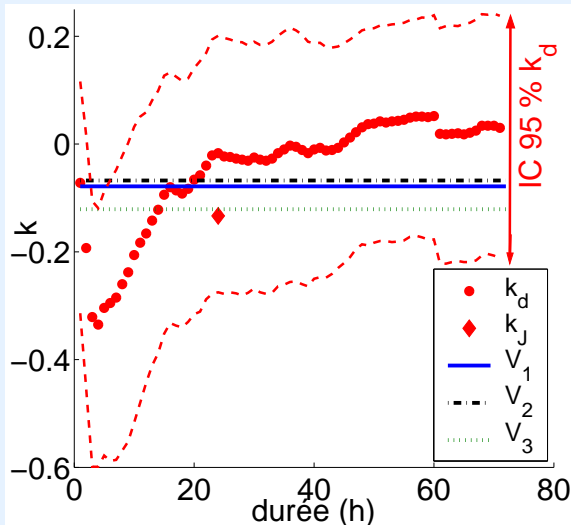
- 1  $V_1$  : indépendance, données de 1h,6h,12h,24h,72h et journalières
- 2  $V_2$  : indépendance, données de 1h,24h,72h, journalières sur période disjointe des données horaires
- 3  $V_3$  : idem  $V_2$  avec modélisation extrême bi-variée de (24h,72h)

## Modélisation bi-variée extrême de $V_3$

IC à 90% du paramètre  $\phi$  : [0.2,0.3]  $\Rightarrow$  ne contient pas 1  $\Rightarrow P_{24}$  et  $P_{72}$  significativement dépendantes.

# Comportement des extrêmes ?

## Estimation du paramètre $k$



Résultats cohérents avec les observations : les estimations du modèle restent dans les IC 95% des observations.

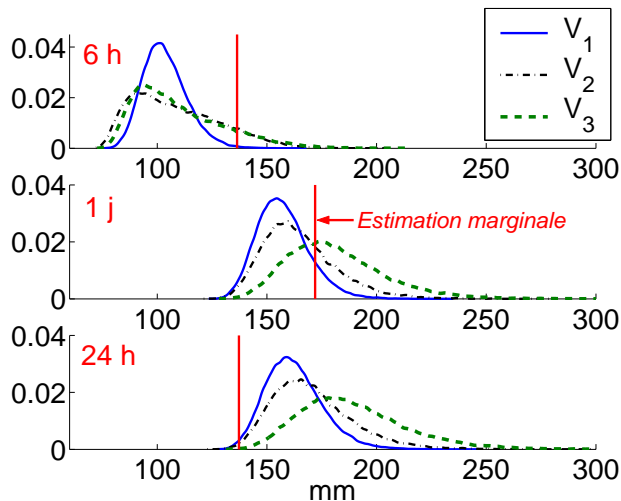
Queues de distribution de type **GEV** : IC à 90% :

$$IC_1(k_J) = [-0.16, -0.01]$$

$$IC_2(k_J) = [-0.18, 0.01]$$

$$IC_3(k_J) = [-0.24, -0.03]$$

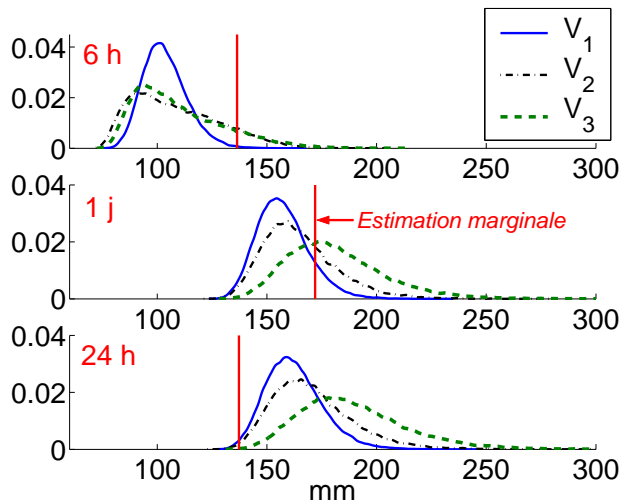
# Distributions a posteriori des quantiles de période de retour 100 ans à Marseille



## Limites de l'étude marginale

- Incohérences :  
 $P_6(100) > P_{12}(100)$ ,  
 $P_J(100) > P_{24}(100)$
- Sous-estimation dues aux lacunes en valeurs extrêmes

# Distributions a posteriori des quantiles de période de retour 100 ans à Marseille



## Apport du modèle

- Contraintes physiques respectées
- Meilleure estimation des extrêmes
- Modèle  $V_3 \Rightarrow$  des quantiles plus élevés



# Comportement des extrêmes

## Bilan : apport du modèle multi-durées

- La loi GEV est significativement plus adaptée aux données de valeurs extrêmes que la loi Gumbel.
- Le modèle permet d'estimer des quantiles pour différentes durées
  - avec un nombre relativement petit de paramètres,
  - en réduisant la sensibilité à l'échantillonnage.

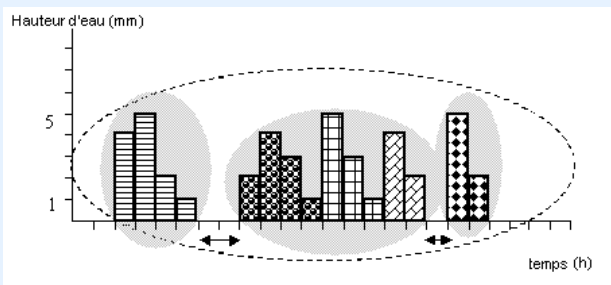
# Plan de l'exposé

- 1 Maxima annuels de pluie : une analyse multi-durées
- 2 Extrêmes simulés par un générateur stochastique de pluie**
- 3 Processus de persistance des averses
- 4 Conclusions et perspectives

# Événements pluvieux

## Définition d'un événement pluvieux

- décrit au pas de temps journalier : ensemble de jours avec une pluie  $\geq 4$  mm et au moins un jour avec une pluie  $\geq 20$  mm
- puis décrit plus finement en horaire : un événement est constitué d'averses et de périodes sèches.



# Un générateur d'événements pluvieux : Shypre

- Shypre décrit le **processus des événements** via des variables aléatoires (nombre d'événements par an, d'averses par événement, durée des averses, volume des averses, etc.)
- Shypre a 3 paramètres :
  - Ne* nombre moyen d'événements par an
  - Pjx* moyenne de la pluie journalière maximale par événement
  - Dtot* durée moyenne d'un événement.
- Shypre génère de longues chroniques d'événements ⇒ permet d'estimer des distributions empiriques de quantiles de pluie pour différentes durées entre 1h et 72h.

# Incertitudes de Shypre et de la loi GPD

## Comparaison de Shypre et de la loi GPD

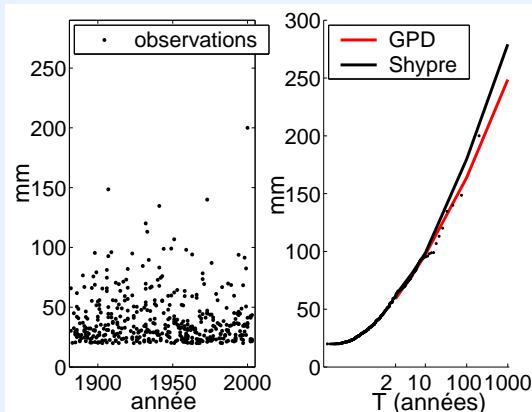
- Shypre permet d'estimer la loi des dépassements de 20 mm.
- D'autre part, la loi **GPD** modélise les dépassements de 20 mm :

$$\mathbb{P}(X \leq x | X > u) = 1 - (1 - k(x - u)/\alpha)^{1/k},$$

$$\left\{ \begin{array}{l} X = \text{pluie journalière supérieure à 20 mm (variable aléatoire)} \\ x > u \\ u = 20 \text{ mm} \\ \alpha, k = \text{paramètres d'échelle et de forme.} \end{array} \right.$$

⇒ On compare les estimations et incertitudes des deux modèles.

# Cas d'étude : pluie journalière supérieure à 20 mm à Marseille



122 années de mesures,  
444 événements

Distribution des quantiles

# Incertitudes

## Précision des estimations ?

Estimation **ponctuelle** de quantiles  $\Rightarrow$  peu informatif, si la précision de la valeur estimée est inconnue.

$\Rightarrow$  On cherche à quantifier l'**incertitude des estimations**.

# Incertitudes

## Précision des estimations ?

Estimation ponctuelle de quantiles  $\Rightarrow$  peu informatif, si la précision de la valeur estimée est inconnue.

$\Rightarrow$  On cherche à quantifier l'incertitude des estimations.

## Sources d'incertitudes

Incertitudes dues aux **données** d'entrées : échantillonnage, erreurs de mesures

Incertitudes dues au **modèle** : choix des paramètres, structure du modèle



# Incertitudes

## Précision des estimations ?

Estimation ponctuelle de quantiles  $\Rightarrow$  peu informatif, si la précision de la valeur estimée est inconnue.

$\Rightarrow$  On cherche à quantifier l'incertitude des estimations.

## Sources d'incertitudes

Incertitudes dues aux **données** d'entrées : échantillonnage, erreurs de mesures

Incertitudes dues au **modèle** : choix des paramètres, structure du modèle

On s'intéresse à :

- la **sensibilité des modèles à l'échantillonnage**,
- à l'**incertitude de paramétrisation**.

# Incertitudes d'échantillonnage

## Sensibilité des estimateurs aux valeurs extrêmes

<i>Modèle</i>	<i>Paramètres estimés par ...</i>
<i>SHYPRE</i>	les moyennes de trois variables ( $N_e$ , $P_{jx}$ , $D_{tot}$ )
<i>GPD</i>	les moments pondérés d'ordre 0 et 1 de la variable des dépassements.

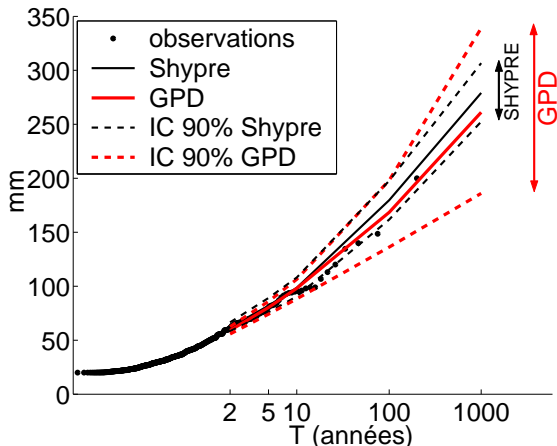
⇒ On peut déjà prévoir la faible sensibilité de Shypre à l'échantillonnage.

### En effet :

<i>Modèle</i>	<i>Paramètre</i>	<i>Estimation</i>	<i>Ecart-type</i>
<i>SHYPRE</i>	$N_e$	3.7	0.17
	$P_{jx}$ (mm)	39.64	1.03
	$D_{tot}$ (jours)	1.72	0.05
<i>GPD</i>	$\alpha$	17.3	1.3
	$k$	-0.119	0.22

# Incertitudes d'échantillonnage

Distribution des pluies journalières supérieures à 20 mm

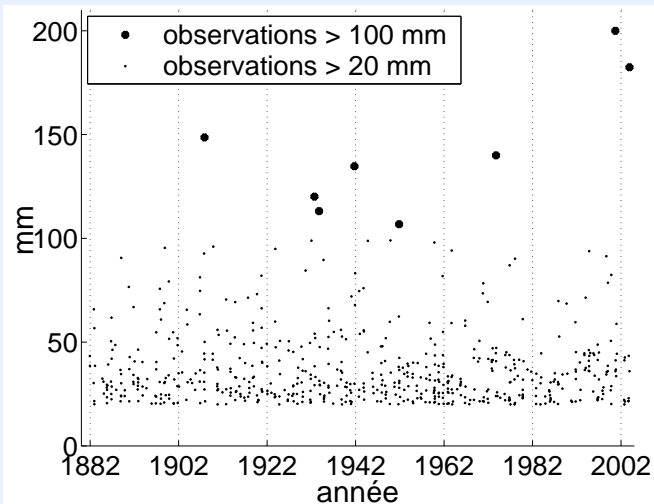


La GPD n'est pas une loi exponentielle : l'IC à 90% de  $k$  est  $[-0.21, -0.03]$

# Incertitudes

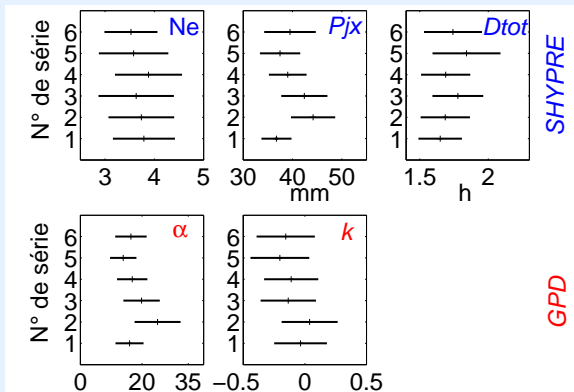
Effet de la taille de la série et de la présence ou non de valeurs extrêmes ?

## Découpage de la série de 122 ans en six séries de 20 ans



# Incertitudes des paramètres

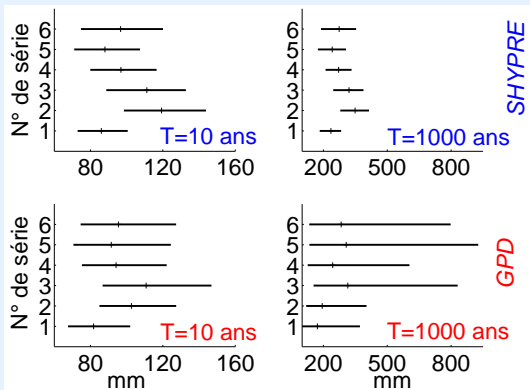
## Résultats: (a) IC 90% des paramètres



Plus grande incertitude avec 20 années : les IC de  $k$  contiennent 0

# Incertitude des quantiles

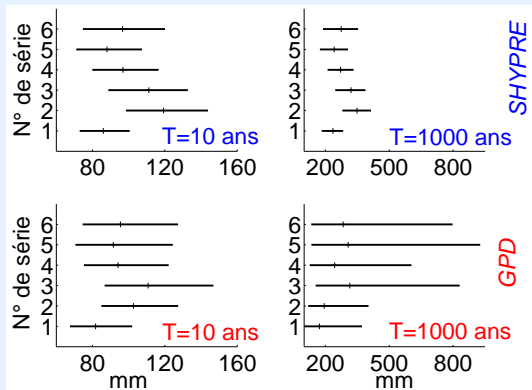
## Résultats : (b) IC 90% des quantiles



**T = 10 ans** : quantiles peu sensibles à l'échantillonnage.  
 Résultats similaires dans les deux approches.

# Incertitude des quantiles

## Résultats : (b) IC 90% des quantiles



$T = 1000$  ans : les deux modèles diffèrent significativement. Shypre semble plus stable que la GPD en terme de médiane des quantiles.

# Comportement des extrêmes

## Bilan : apport du générateur de pluie Shypre

Shypre exploite plus d'information que la seule variable du cumul de pluie : estimations relativement précises

## Bilan : loi GPD ou exponentielle ?

- Avec plus de 120 ans de données : la loi exponentielle n'est pas adaptée pour modéliser les dépassements de seuil
- Avec 20 années de données : difficile de choisir, l'incertitude est trop importante sur  $k$



# Plan de l'exposé

- 1 Maxima annuels de pluie : une analyse multi-durées
- 2 Extrêmes simulés par un générateur stochastique de pluie
- 3 Processus de persistance des averses**
- 4 Conclusions et perspectives

# Persistence des averses

## Présentation du phénomène : exemple (Arnaud 2004)

Sur un poste des Alpes, *les 3 plus fortes averses observées de la série (pour la saison hiver) sont dans le même événement.*

Ce cas se retrouve dans de nombreux événements extrêmes observés (souvent en climat méditerranéen ou tropical)

# Modélisation de la persistance des averses

## Points à prendre en compte :

- Modélisation des averses au delà d'un seuil donné  $\implies$  **Théorie des valeurs extrêmes**.
- Les averses des événements extrêmes ne sont pas indépendantes  $\implies$  **Chaîne de Markov**, à l'ordre 1.

# Modélisation de la persistance des averses

## Points à prendre en compte :

- Modélisation des averses au delà d'un seuil donné  $\implies$  Théorie des valeurs extrêmes.
- Les averses des événements extrêmes ne sont pas indépendantes  $\implies$  Chaîne de Markov, à l'ordre 1.

## Notion de dépendance asymptotique

Soit  $X_t$  le processus des averses, supposé stationnaire. On pose  $x^* = \sup\{x \in \mathbb{R} : \mathbb{P}(X_t \leq x) < 1\}$ . Le processus est **asymptotiquement indépendant** (resp. **asymptotiquement dépendant**) si

$$\lim_{x \rightarrow x^*} \mathbb{P}(X_{t+1} > x | X_t > x) = 0 \quad (\text{resp. } > 0)$$

# Modélisation des extrêmes de la série des averses

La théorie ne permet de modéliser que les dépassements de seuil  $u$

- Loi marginale  $F$  loi GPD

# Modélisation des extrêmes de la série des averses

La théorie ne permet de modéliser que les dépassements de seuil  $u$

- Loi marginale  $F$  loi GPD
- Loi de la **chaîne de Markov**  $G(x, y) = \mathbb{P}(X_t \leq x, X_{t+1} \leq y)$ ,  
 $x > u, y > u$

# Modélisation des extrêmes de la série des averses

La théorie ne permet de modéliser que les dépassements de seuil  $u$

- Loi marginale  $F$  loi GPD
- Loi de la **chaîne de Markov**  $G(x, y) = \mathbb{P}(X_t \leq x, X_{t+1} \leq y)$ ,  $x > u, y > u$ 
  - 1 loi **bi-variée des valeurs extrêmes**.
  - 2 autre formalisation (**Ledford et Tawn, 1996**) :

$$\mathbb{P}(X_t > x, X_{t+1} > y) \approx \mathcal{L}(-\log F(x)^{-1}, -\log F(y)^{-1}) \{-\log F(x)\}^{c_1} \{-\log F(y)\}^{c_2}$$

$c_i > 0, c_1 + c_2 \geq 1$ , et  $\mathcal{L}$  une fonction à variation lente.

# Choix d'une loi bi-variée pour la chaîne de Markov

*Problème*



- **Lois bi-variées extrêmes** : les extrêmes sont asymptotiquement dépendants, sauf dans le cas de l'indépendance stricte.



# Choix d'une loi bi-variée pour la chaîne de Markov

*Problème*



- **Lois bi-variées extrêmes** : les extrêmes sont asymptotiquement dépendants, sauf dans le cas de l'indépendance stricte.

*Bien*



- **Formalisation de Ledford et Tawn (1996)** : modélise des processus asymptotiquement dépendants ou asymptotiquement indépendants.

# Choix d'une loi bi-variée pour la chaîne de Markov

Problème



- **Lois bi-variées extrêmes** : les extrêmes sont asymptotiquement dépendants, sauf dans le cas de l'indépendance stricte.

Bien



- **Formalisation de Ledford et Tawn (1996)** : modélise des processus asymptotiquement dépendants ou asymptotiquement indépendants.

$$\Rightarrow \text{Choix : } \mathbb{P}(X_{t+1} > y | X_t > x) = \frac{a\{\log F(x) \log F(y)\}^{1/\eta}}{1-F(x)}, \quad 0 < \eta \leq 1.$$

- $\eta = 1$  : dépendance asymptotique des extrêmes
- $0.5 < \eta < 1$  : indépendance asymptotique + association positive des extrêmes
- $0 < \eta < 0.5$  : indépendance asymptotique + association négative des extrêmes

# Cas d'étude : averses de Marseille, 1918-2002

## Loi marginale

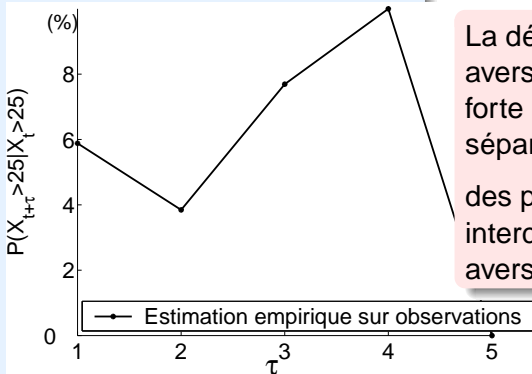
Choix du seuil  $u$  : 25 mm.

# Cas d'étude : averses de Marseille, 1918-2002

## Loi marginale

Choix du seuil  $u$  : 25 mm.

## Loi jointe



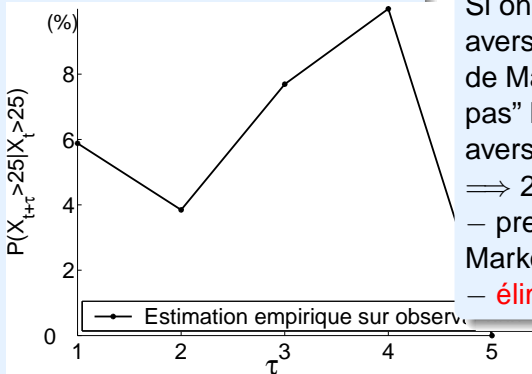
La dépendance entre deux averses consécutives est moins forte qu'entre deux averses séparées de 2 ou 3 averses : des petites averses sont intercalées entre les fortes averses.

# Cas d'étude : averses de Marseille, 1918-2002

## Loi marginale

Choix du seuil  $u$  : 25 mm.

## Loi jointe



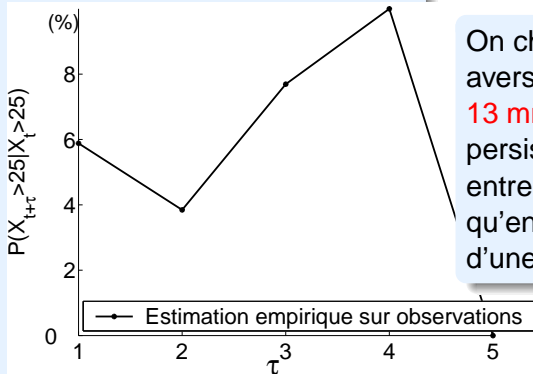
Si on garde la série des averses telle quelle, la chaîne de Markov à l'ordre 1 "ne verra pas" la persistance des fortes averses consécutives  
 $\implies$  2 possibilités :  
 – prendre une chaîne de Markov d'ordre 4 ou 5  
 – **éliminer les petites averses.**

# Cas d'étude : averses de Marseille, 1918-2002

## Loi marginale

Choix du seuil  $u$  : 25 mm.

## Loi jointe



On choisit d'éliminer les averses qui sont inférieures à **13 mm** : à partir de ce seuil, la persistance semble plus forte entre averses consécutives qu'entre averses séparées d'une (ou plus) averse(s).

# Analyse des dépendances des extrêmes

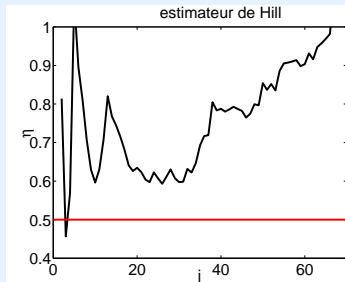
## Estimation du paramètre de dépendance des extrêmes $\eta$

- Estimateur de Hill :

$$\hat{H}_\eta(j) = \left( \frac{1}{j} \sum_{i=1}^j \log T_{(i)} / T_{(j)} \right)^{-1}$$

où

$$T_i = \min(-1/\log F(X_i), -1 \log F(X_{i+1}))$$



- Maximum de vraisemblance :  $\hat{\eta} \approx 0.54$
- Bayésien : moyenne  $\approx 0.5$ , IC 90%  $\approx [0.3, 0.8]$

# Analyse des dépendances des extrêmes

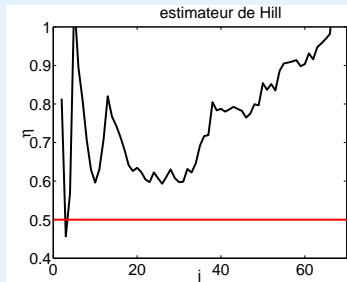
## Estimation du paramètre de dépendance des extrêmes $\eta$

- Estimateur de Hill :

$$\hat{H}_\eta(j) = \left( \frac{1}{j} \sum_{i=1}^j \log T_{(i)} / T_{(j)} \right)^{-1}$$

où

$$T_i = \min(-1/\log F(X_i), -1 \log F(X_{i+1}))$$



- Maximum de vraisemblance :  $\hat{\eta} \approx 0.54$
- Bayésien : moyenne  $\approx 0.5$ , IC 90%  $\approx [0.3, 0.8]$

⇒ Il semble que :

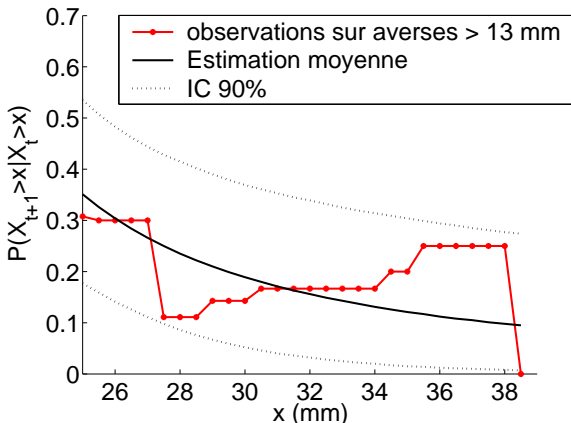
$\eta \neq 1$  : un modèle bi-varié extrême aurait donc été inadéquat.



# Résultats du modèle de persistance

Estimation bayésienne avec des lois a priori larges sur les paramètres

## Dépendance entre averses consécutives

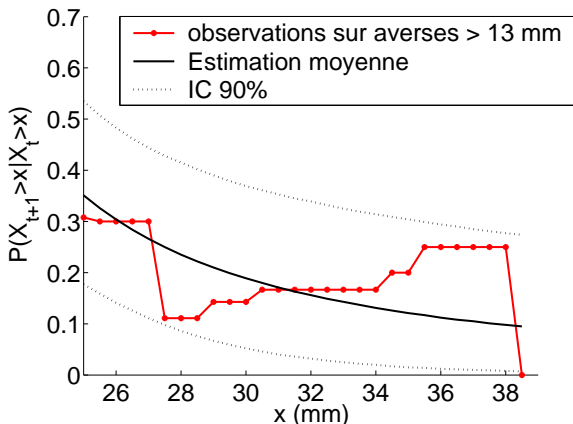


- Les résultats des observations sont à l'intérieur de l'IC à 90% du modèle

# Résultats du modèle de persistance

Estimation bayésienne avec des lois a priori larges sur les paramètres

## Dépendance entre averses consécutives

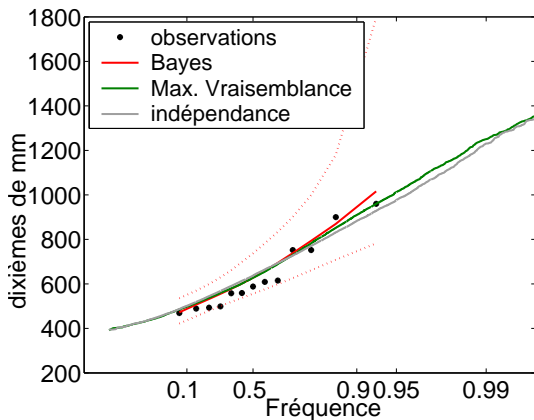


- Les résultats des observations sont à l'intérieur de l'IC à 90% du modèle

- Selon le modèle,  $\mathbb{P}(X_{t+1} > x | X_t > x)$  décroît avec  $x$ , qui n'est pas rendu par les observations (mais les observations sont calculées avec très peu de données)

# Capacité du modèle à représenter la persistance

Distribution du cumul  $X_1 + X_2 + X_3$  de 3 averses consécutives :



Les modèles d'indépendance et de persistance ne semblent pas significativement différents

# Comportement des extrêmes

## Bilan : modélisation de la persistance des averses

- Entité considérée ici : les averses
- Une estimation du  $k$  assez imprécise : intervalle de crédibilité à 90% de  $k$  :  $IC=[-0.52,0.17]$  mais favorisant les valeurs négatives.
- Sur le cas étudié : la persistance des averses n'aggrave pas les événements extrêmes.

# Plan de l'exposé

- 1 Maxima annuels de pluie : une analyse multi-durées
- 2 Extrêmes simulés par un générateur stochastique de pluie
- 3 Processus de persistance des averses
- 4 Conclusions et perspectives**

# Conclusions

## Comportement des pluies extrêmes ?

- Cas des longues séries et Marseille : le paramètre de forme des lois GEV et GPD semble être significativement **non nul, et négatif**.
- Il peut donc être **risqué d'utiliser les lois Gumbel et exponentielle** pour estimer les pluies extrêmes ...

# Conclusions

## Comportement des pluies extrêmes ?

- Cas des longues séries et Marseille : le paramètre de forme des lois GEV et GPD semble être significativement **non nul, et négatif**.
- Il peut donc être **risqué d'utiliser les lois Gumbel et exponentielle** pour estimer les pluies extrêmes ...

## Mais le paramètre de forme est difficile à estimer ...

⇒ Intérêt d'apporter de l'information, via :

- la modélisation multi-variée des pluies,
- la prise en compte d'autres variables descriptives des événements extrêmes (cf. Shypre),
- la compréhension du phénomène de persistance des averses.

# Perspectives

- Analyse **régionale** et **spatiale** des pluies (géostatistique, modèles multi-variés des extrêmes appliqués aux cadres spatial et régional).
- Mieux comprendre la modélisation **multi-variée** des pluies :
  - choix de la loi,
  - dimension supérieure à deux.



# Perspectives

- Mieux caractériser l'effet de la persistance des averses sur les événements extrêmes :
  - application du modèle proposé à des données présentant plus de persistance (poste des Cévennes, ou données simulées avec une loi bi-variée impliquant une forte dépendance des variables successives),
  - prendre en compte l'incertitude liée à la définition des averses.
- Mieux comprendre le phénomène de persistance : à l'aide d'une chaîne de Markov d'ordre supérieur à 1.
- Inclusion du modèle de persistance proposé dans le modèle **Shypre** de simulation de pluie.

# COMPORTEMENT ASYMPTOTIQUE DE LA DISTRIBUTION DES PLUIES EXTRÊMES EN FRANCE

Aurélie MULLER

Thèse présentée pour obtenir le diplôme de docteur de l'Université Montpellier 2

24 novembre 2006



Patrick ARNAUD

Cemagref Aix-en-Provence

Jean-Noël BACRO

Université Montpellier 2

Michel LANG

Cemagref Lyon