

BAYESIAN COMPARISON OF DIFFERENT RAINFALL DEPTH-DURATION-FREQUENCY

Aurélie MULLER (ENS Cachan),
Jean-Noël BACRO (Univ. Montpellier 2),
Michel LANG (Cemagref Lyon)

Sous presse dans Stochastic Environmental Research and Risk Assessment

mai 2007, groupe de travail SAMA



Travaux interdisciplinaires

Domaines

- Statistiques des valeurs extrêmes
- Hydrologie

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)
- 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)
 - 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)
 - 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)
 - 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) :

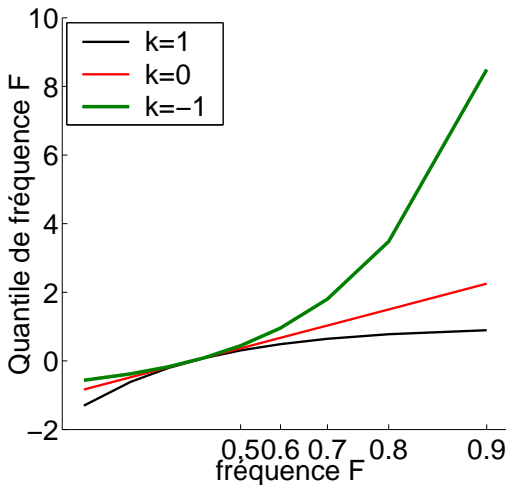
$$\mathbb{P}(M_n \leq x) \approx$$

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

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

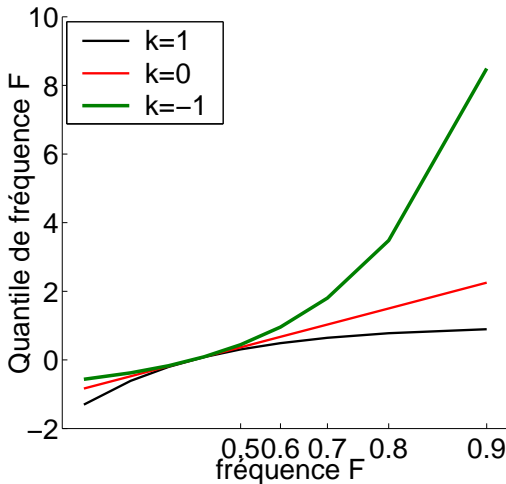
Soient $M_n = \max\{X_1, \dots, X_n\}$ un échantillon de loi F inconnue.

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.

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

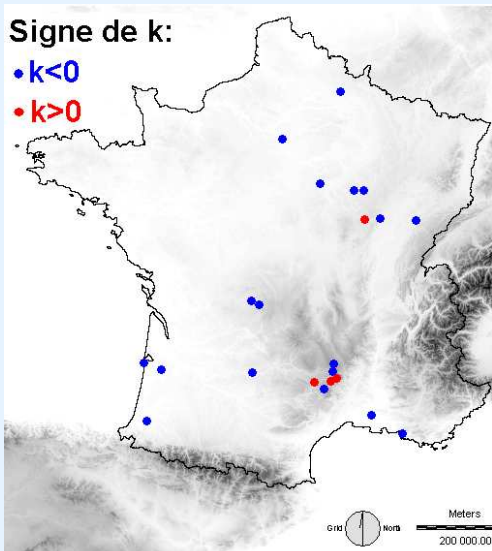


- Inclusion d'informations exogènes aux données
- Accès direct à la distribution des quantiles ...
- ... sans recours à des théorèmes asymptotiques
- Applicable avec des vraisemblances complexes
- Nécessite des algorithmes sophistiqués (Monte Carlo par Chaînes de Markov ou échantillonnage préférentiel)



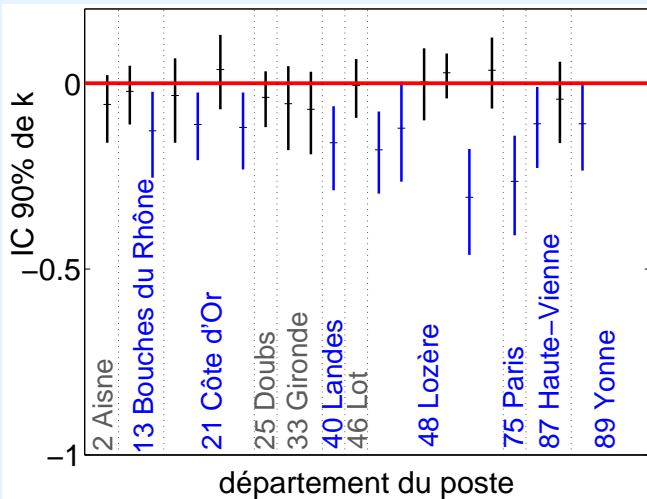
Remise en question de la loi Gumbel ?

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



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 **Conclusions et perspectives**

Plan de l'exposé

- 1 **Maxima annuels de pluie : une analyse multi-durées**
- 2 Conclusions et perspectives

Information multi-durées

Mesures de pluie *Intérêt*

pas de temps 1h

journalière

Information multi-durées

Mesures de pluie

Intérêt

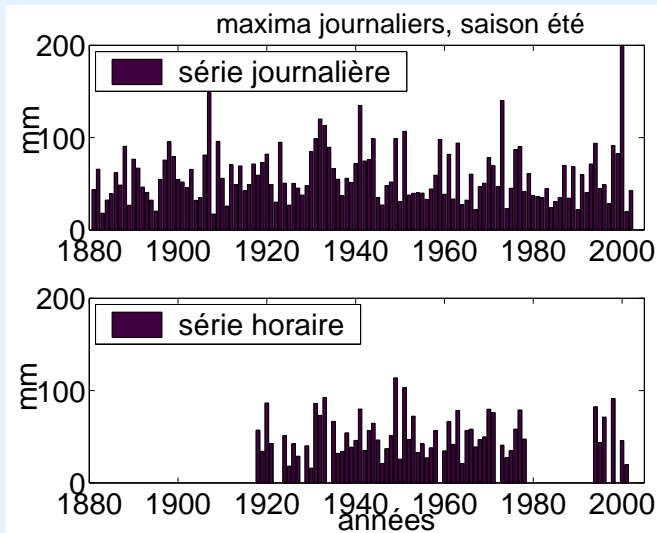
pas de temps 1h

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

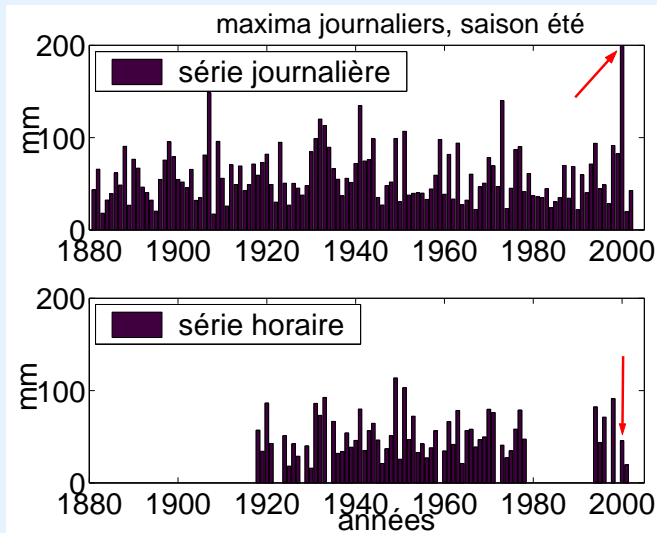
journalière

réseau dense de mesures, données plus
renseignées

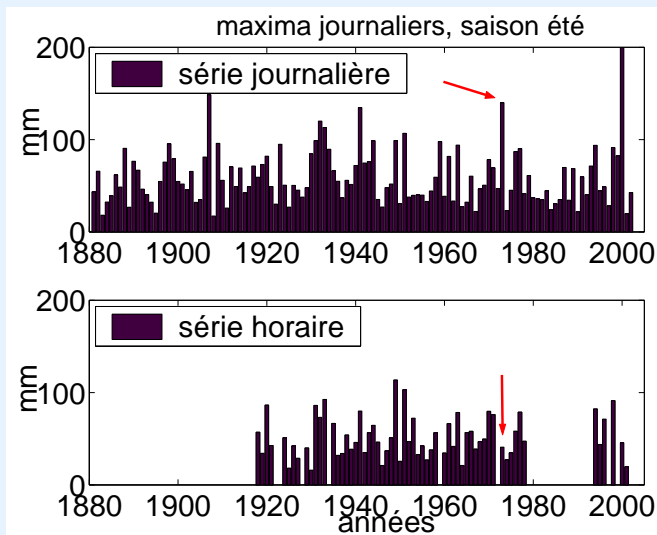
Exemple : information multi-durées à Marseille



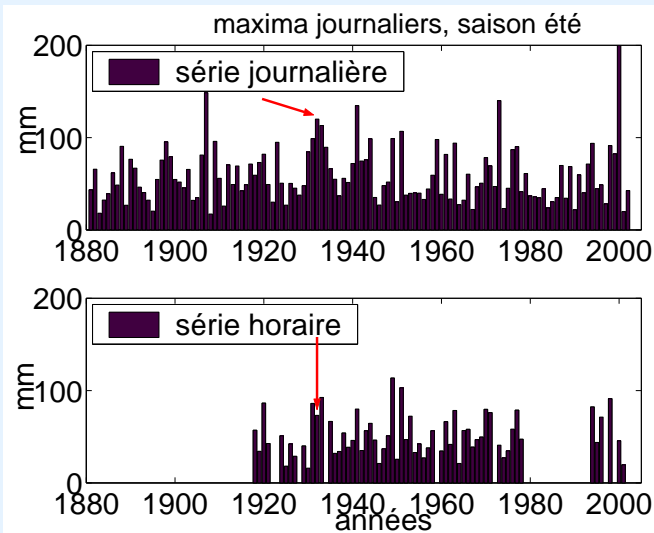
Exemple : information multi-durées à Marseille



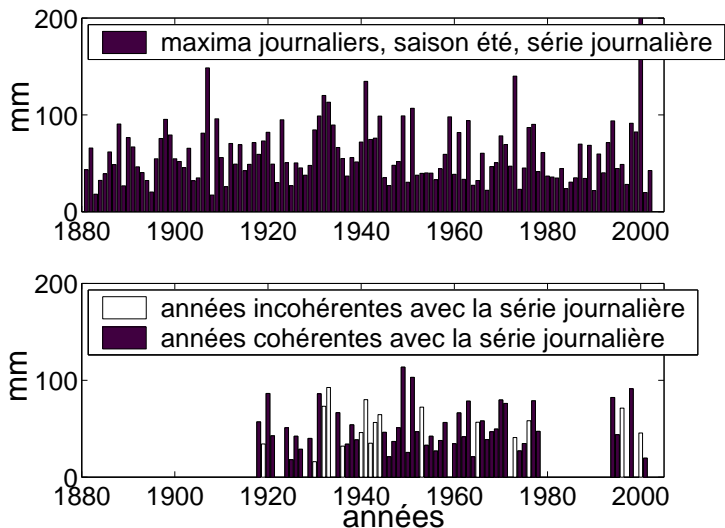
Exemple : information multi-durées à Marseille



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 α_d, β_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) :

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

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) :

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

- 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) :

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

- 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) :

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

- 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) :

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

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

Rappels : valeurs extrêmes en contexte stationnaire

X_1, X_2, \dots suite de v.a. de loi marginale F

stationnaire

$\forall i_1, \dots, i_k, m \in \mathbb{N}, (X_{i_1}, \dots, X_{i_k}) \stackrel{\text{loi}}{=} (X_{i_1+m}, \dots, X_{i_k+m})$.

$\mathcal{D}(u)$: condition de non-dépendance à long terme

$\forall \epsilon > 0, \exists u, l > 0$ tels que $\forall i_1 < \dots < i_p < j_1 < \dots < j_q$ avec $j_1 - i_p > l$ on a

$$\left| P(X_{i_1} < u, \dots, X_{i_p} < u, X_{j_1} < u, \dots, X_{j_q} < u) - P(X_{i_1} < u, \dots, X_{i_p} < u)P(X_{j_1} < u, \dots, X_{j_q} < u) \right| < \epsilon$$

Alors

$$P(\max\{X_1, \dots, X_n\} \leq u) \approx F^{nIE}(u)$$

pour u assez grand, et $IE \in [0, 1]$ l'**indice extrême**l.

$IE = 1 \Rightarrow$ indépendance asymptotique.

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,

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.

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**

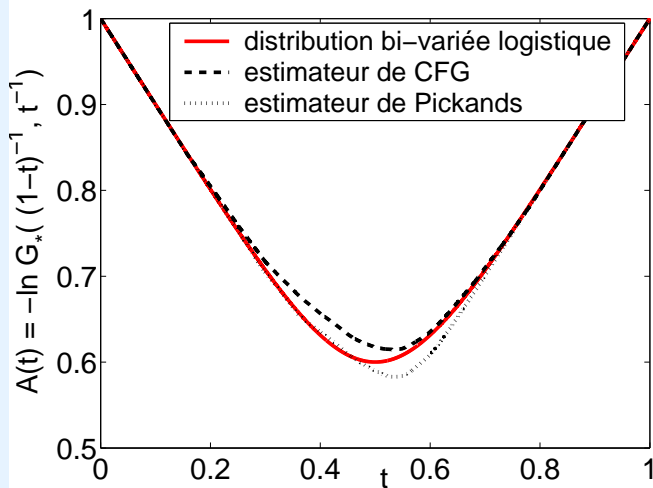
Loi de (P_{24}, P_{72}) : loi bi-variée logistique ?

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 :

α_J, α'	loi lognormale (0,100)
β_J	loi normale (0,100)
k_J	loi uniforme sur [-1,1]
Θ	loi uniforme sur [1,24]
ϑ, ϑ'	loi normale (0,100), tronquée en 0
ε	loi lognormale (0,100)
ϕ (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

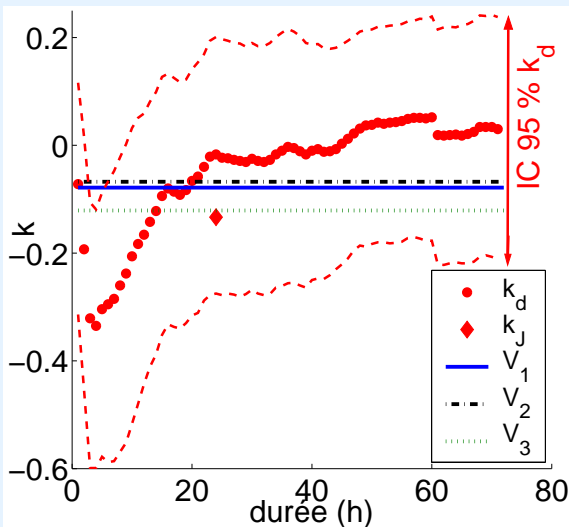
- 1 V_1 : indépendance, données de 1h,6h,12h,24h,72h et journalières
- 2 V_2 : indépendance, données de 1h,24h,72h, journalières sur période disjointe des données horaires
- 3 V_3 : idem V_2 avec modélisation extrême bi-variée de (24h,72h)

Modélisation bi-variée extrême de V_3

IC à 90% du paramètre ϕ : [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.

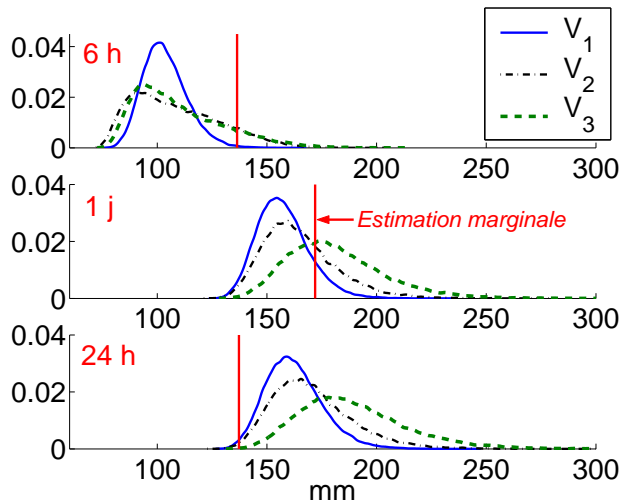
Queues de distribution de type **GEV** : IC à 90% :

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

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

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

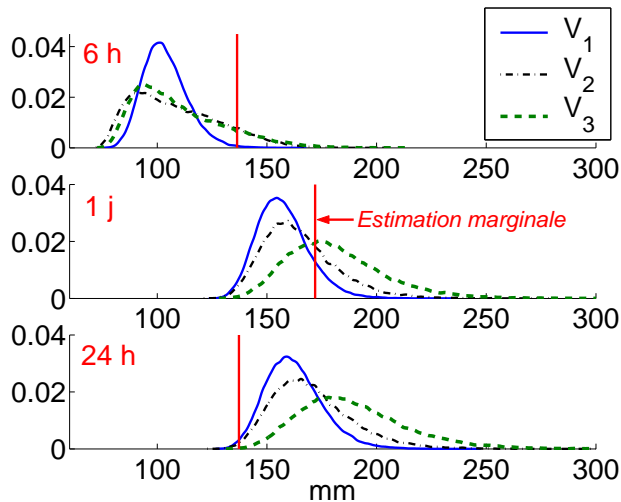
Distributions a posteriori des quantiles de période de retour 100 ans à Marseille



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

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 **Conclusions et perspectives**

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.

- 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.
- **Caractériser l'effet de la persistance des averses sur les événements extrêmes.**

BAYESIAN COMPARISON OF DIFFERENT RAINFALL DEPTH-DURATION-FREQUENCY

Aurélie MULLER (ENS Cachan),
Jean-Noël BACRO (Univ. Montpellier 2),
Michel LANG (Cemagref Lyon)

Sous presse dans Stochastic Environmental Research and Risk Assessment

mai 2007, groupe de travail SAMA

