

Comportements extrêmes de processus spatio-temporels

Application en océanographie et météorologie

Nicolas Raillard

IRMAR – Université de Rennes 1
CERSAT – IFREMER
École doctorale MATISSE

13 Décembre 2011

Thèse de doctorat réalisée sous la direction de MM. Jian-feng
YAO et Bertrand CHAPRON

Introduction

Contexte

- Motivations pratiques
 - Les extrêmes nécessitent une attention particulière : évènements rares, mais causant le plus de dégats ;
 - Souvent, diverses sources d'informations sont disponibles : mesures in-situ, mesures indirectes, modèle numériques...
- Motivations théoriques
 - Proposer des modèles adaptés aux extrêmes et des estimateurs judicieux ;
 - Mettre en place des procédures de validation de l'ajustement, spécifiques aux extrêmes.

Plan de l'exposé

- 1 Description des données
- 2 Modèles statistiques
 - Théorie probabiliste dans le cas univarié
 - Modélisation de la dépendance
- 3 Inférence statistique
 - Résultats théoriques
 - Résultats numériques
- 4 Application à la hauteur significative des vagues
- 5 Conclusion

Plan de l'exposé

- 1 Description des données
- 2 Modèles statistiques
 - Théorie probabiliste dans le cas univarié
 - Modélisation de la dépendance
- 3 Inférence statistique
 - Résultats théoriques
 - Résultats numériques
- 4 Application à la hauteur significative des vagues
- 5 Conclusion

Description des données

Hauteur significative des vagues : Quatre fois l'écart-type de l'élévation de la surface, ou sous certaines conditions la moyenne du tiers des vagues les plus hautes ; mesure liée à l'énergie des vagues.

Bouées : Données horaires, peu de valeurs manquantes, considérée généralement comme une référence, mais peu de sites de mesure ;

Satellites : Mesures à distance, bonne adéquation aux bouées, mais échantillonnage irrégulier en temps et en espace ;

ERA Interim : Données de réanalyse, sur une grille régulière en temps et en espace, mais lisse les événements extrêmes.

Description des données

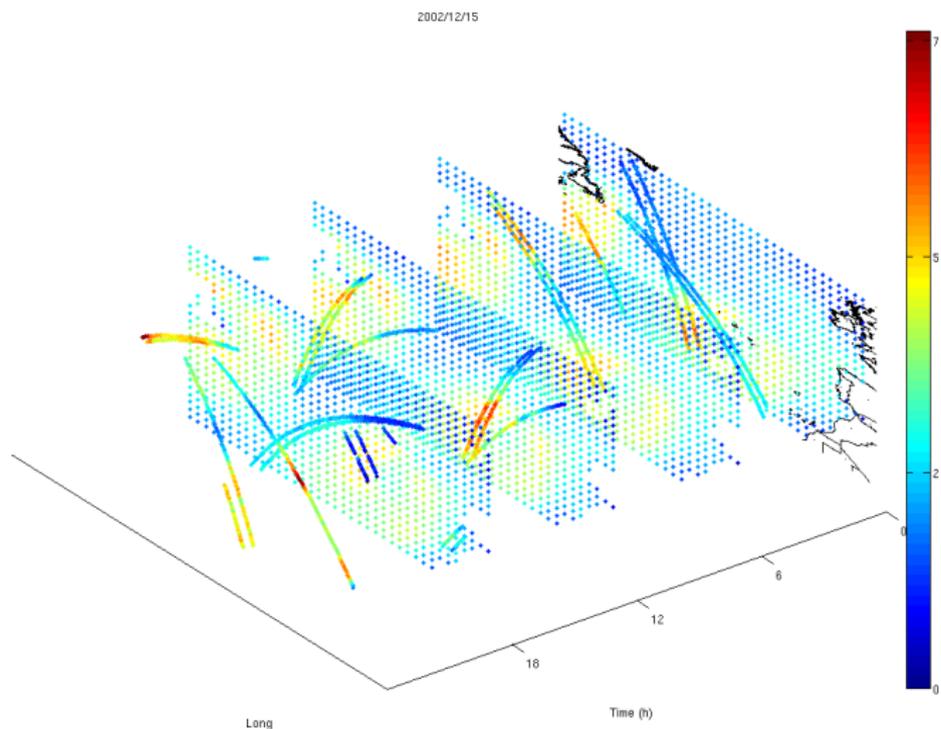


FIGURE: Représentation 3-D des données disponible le 15/12/2002.

Description des données

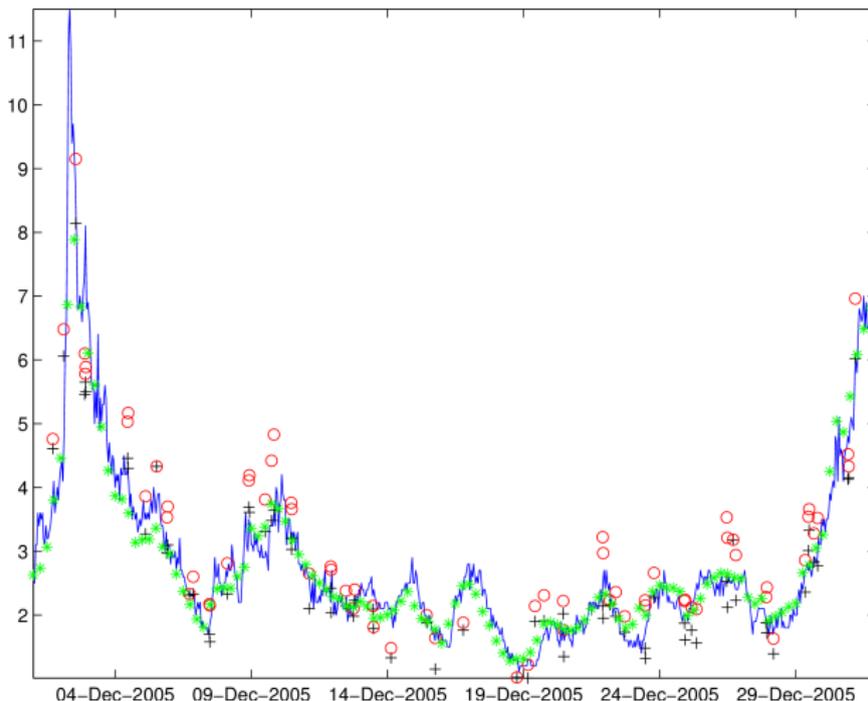


FIGURE: Exemple de série temporelle observée à une bouée (Bleu). Noir et Rouge : satellite ; Vert : ERA-Interim.

Description des données

Objectifs : Calculer par exemple,

- Le niveau de retour à 10 ans : niveau dépassé en moyenne une fois tous les dix ans ;
- Fréquence et durée des évènements extrêmes ;
- Temps inter-clusters.

Deux approches ont été étudiées :

- ① Interpolation après déplacement des traces satellitaires : approche **insuffisante** pour les extrêmes ;
- ② Mise en place d'un modèle spécifique aux extrêmes : approche présentée.

Plan de l'exposé

- 1 Description des données
- 2 Modèles statistiques**
 - Théorie probabiliste dans le cas univarié
 - Modélisation de la dépendance
- 3 Inférence statistique
 - Résultats théoriques
 - Résultats numériques
- 4 Application à la hauteur significative des vagues
- 5 Conclusion

Théorie probabiliste dans le cas univarié

- On s'intéresse à : $M_n = \max_{i=1, \dots, n} X_i$, (X_i) échantillon i.i.d, de f.d.r F ;
- Si $F^n(a_n x + b_n) = \mathbb{P}\left\{\frac{M_n - b_n}{a_n} \leq x\right\} \rightarrow G(x)$, alors G est une **loi max-stable** : $G^n(\alpha_n x + \beta_n) = G(x)$;
- Fisher-Tippett (1928), Gnedenko (1943) :

$$G(x; \mu, \sigma, \xi) = \exp\left[-\left(1 + \xi \frac{x - \mu}{\sigma}\right)_+^{-1/\xi}\right] \in \text{GEV}(\mu, \sigma, \xi) ;$$
- Exemples :
 - Type Weibull : $\xi < 0$ (p.e. Rayleigh...);
 - Type Fréchet (queues lourdes) : $\xi > 0$; Fréchet unitaire $\xi = 1$;
 - $\xi \rightarrow 0$: $G(x) = \exp(-e^{-\frac{x - \mu}{\sigma}})$ (distribution de Gumbel).

Théorie probabiliste dans le cas univarié

- On s'intéresse à : $M_n = \max_{i=1, \dots, n} X_i$, (X_i) échantillon i.i.d, de f.d.r F ;
- Si $F^n(a_n x + b_n) = \mathbb{P}\left\{\frac{M_n - b_n}{a_n} \leq x\right\} \rightarrow G(x)$, alors G est une **loi max-stable** : $G^n(\alpha_n x + \beta_n) = G(x)$;
- Fisher-Tippett (1928), Gnedenko (1943) :

$$G(x; \mu, \sigma, \xi) = \exp\left[-\left(1 + \xi \frac{x - \mu}{\sigma}\right)_+^{-1/\xi}\right] \in \text{GEV}(\mu, \sigma, \xi) ;$$
- Exemples :
 - Type Weibull : $\xi < 0$ (p.e. Rayleigh...);
 - Type Fréchet (queues lourdes) : $\xi > 0$; Fréchet unitaire $\xi = 1$;
 - $\xi \rightarrow 0$: $G(x) = \exp(-e^{-\frac{x - \mu}{\sigma}})$ (distribution de Gumbel).

Théorie probabiliste dans le cas univarié

- On s'intéresse à : $M_n = \max_{i=1, \dots, n} X_i$, (X_i) échantillon i.i.d, de f.d.r F ;
- Si $F^n(a_n x + b_n) = \mathbb{P}\left\{\frac{M_n - b_n}{a_n} \leq x\right\} \rightarrow G(x)$, alors G est une **loi max-stable** : $G^n(\alpha_n x + \beta_n) = G(x)$;
- Fisher-Tippett (1928), Gnedenko (1943) :

$$G(x; \mu, \sigma, \xi) = \exp\left[-\left(1 + \xi \frac{x - \mu}{\sigma}\right)_+^{-1/\xi}\right] \in \text{GEV}(\mu, \sigma, \xi) ;$$
- Exemples :
 - Type Weibull : $\xi < 0$ (p.e. Rayleigh...);
 - Type Fréchet (queues lourdes) : $\xi > 0$; Fréchet unitaire $\xi = 1$;
 - $\xi \rightarrow 0$: $G(x) = \exp(-e^{-\frac{x-\mu}{\sigma}})$ (distribution de Gumbel).

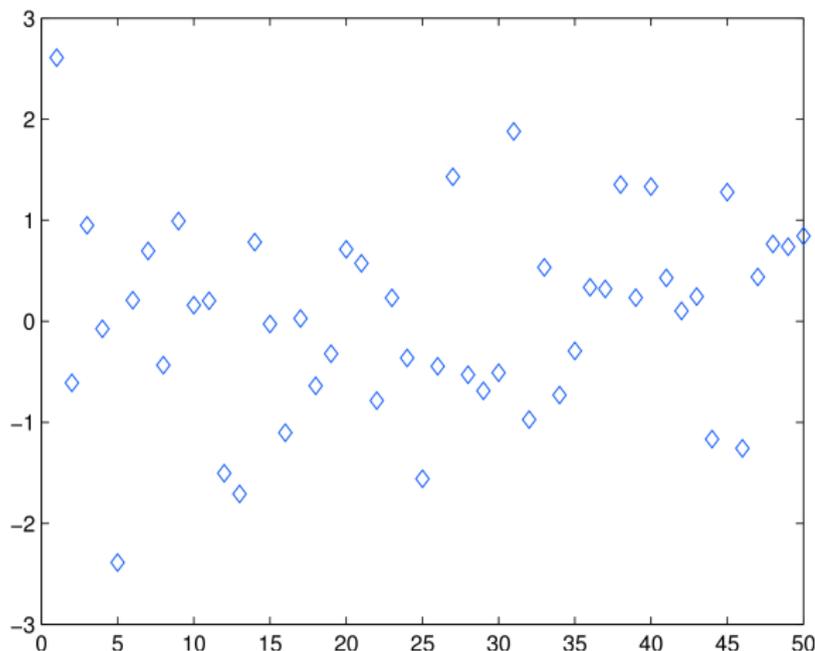
Théorie probabiliste dans le cas univarié

- On s'intéresse à : $M_n = \max_{i=1, \dots, n} X_i$, (X_i) échantillon i.i.d, de f.d.r F ;
- Si $F^n(a_n x + b_n) = \mathbb{P}\left\{\frac{M_n - b_n}{a_n} \leq x\right\} \rightarrow G(x)$, alors G est une **loi max-stable** : $G^n(\alpha_n x + \beta_n) = G(x)$;
- Fisher-Tippett (1928), Gnedenko (1943) :

$$G(x; \mu, \sigma, \xi) = \exp\left[-\left(1 + \xi \frac{x - \mu}{\sigma}\right)_+^{-1/\xi}\right] \in \text{GEV}(\mu, \sigma, \xi) ;$$
- Exemples :
 - Type Weibull : $\xi < 0$ (p.e. Rayleigh...);
 - Type Fréchet (queues lourdes) : $\xi > 0$; Fréchet unitaire $\xi = 1$;
 - $\xi \rightarrow 0$: $G(x) = \exp(-e^{-\frac{x-\mu}{\sigma}})$ (distribution de Gumbel).

Modèles statistiques pour des observations univariées

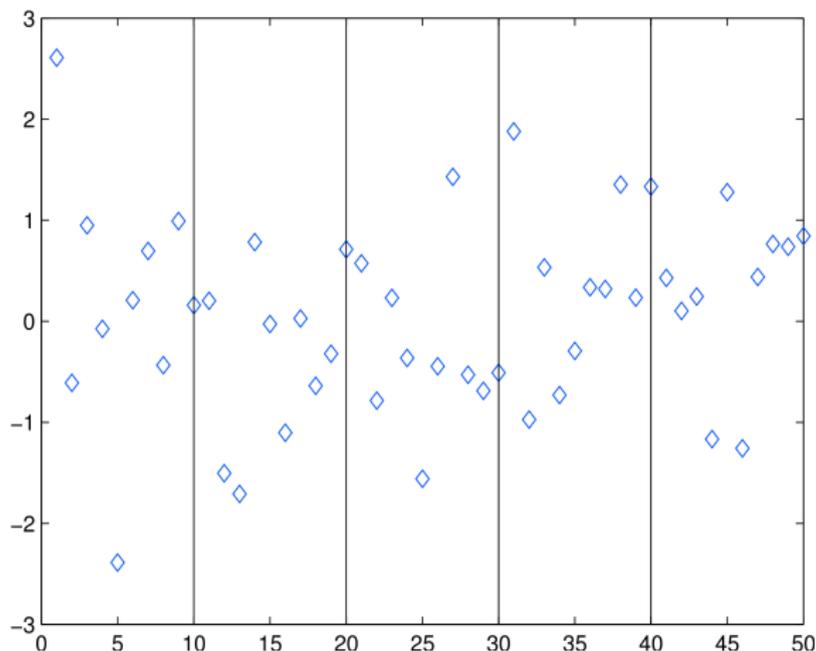
- **Maxima par bloc** $M_n \sim \text{GEV}(\mu, \sigma, \xi)$: perte de données.



- Amélioration : conserver tous les dépassements d'un seuil u .
- Deux approches possibles :

Modèles statistiques pour des observations univariées

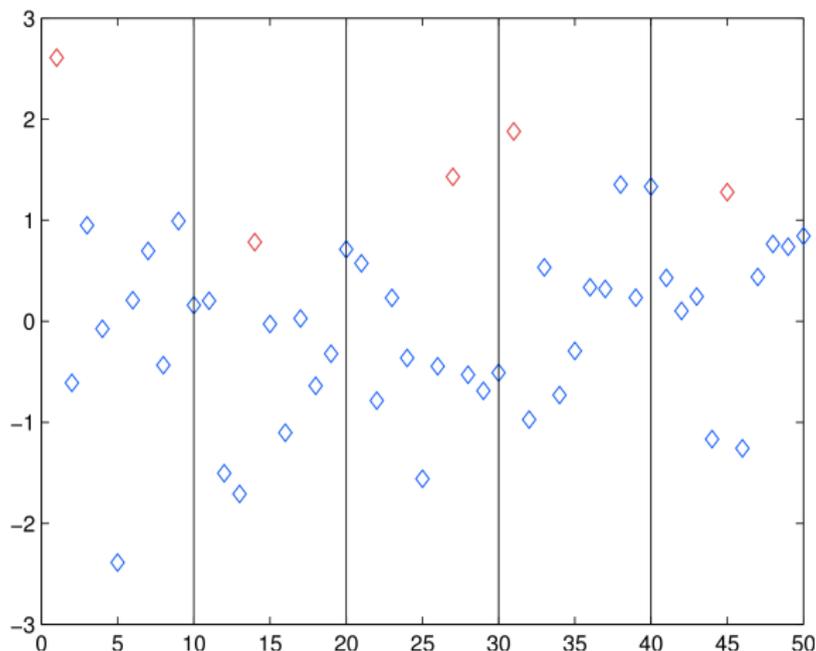
- **Maxima par bloc** $M_n \sim \text{GEV}(\mu, \sigma, \xi)$: perte de données.



- **Amélioration** : conserver tous les dépassements d'un seuil u .
- Deux approches possibles :

Modèles statistiques pour des observations univariées

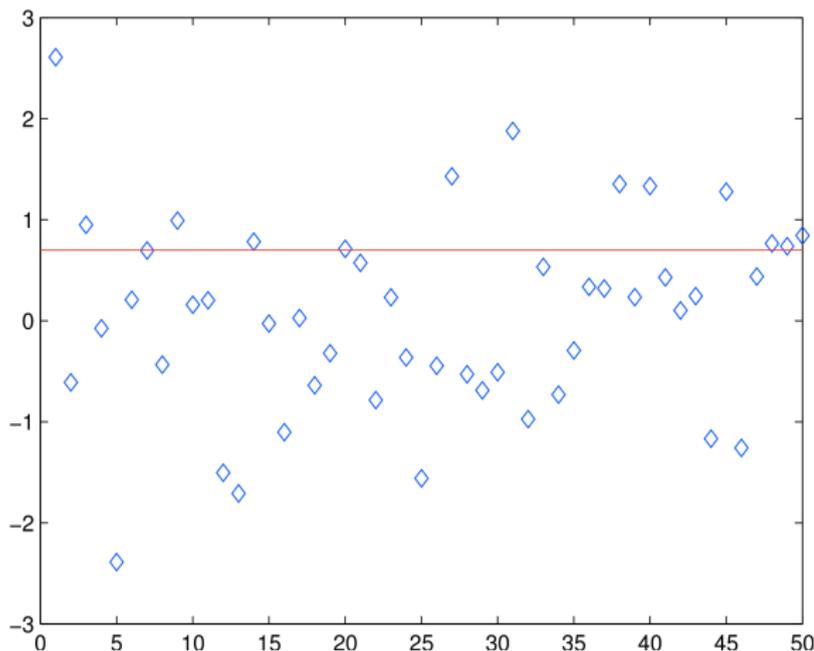
- **Maxima par bloc** $M_n \sim \text{GEV}(\mu, \sigma, \xi)$: perte de données.



- **Amélioration** : conserver tous les dépassements d'un seuil u .
- Deux approches possibles :

Modèles statistiques pour des observations univariées

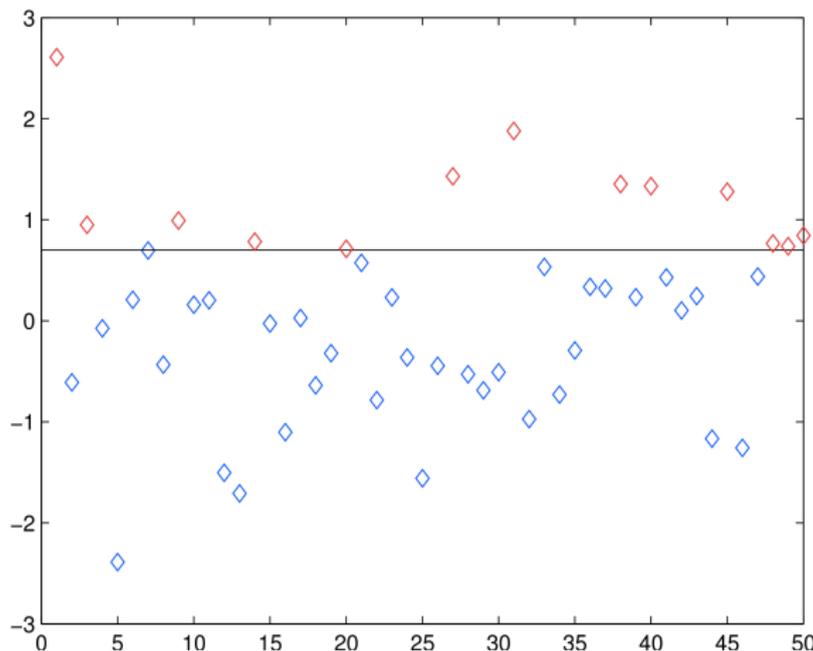
- Maxima par bloc $M_n \sim \text{GEV}(\mu, \sigma, \xi)$: perte de données.
- **Amélioration** : conserver tous les dépassements d'un seuil u .



- Deux approches possibles :

Modèles statistiques pour des observations univariées

- Maxima par bloc $M_n \sim \text{GEV}(\mu, \sigma, \xi)$: perte de données.
- **Amélioration** : conserver tous les dépassements d'un seuil u .



- Deux approches possibles :

Modèles statistiques pour des observations univariées

- Maxima par bloc $M_n \sim \text{GEV}(\mu, \sigma, \xi)$: perte de données.
- **Amélioration** : conserver tous les dépassements d'un seuil u .
- Deux approches possibles :
 - **POT** : $X|X > u \sim \text{GPD}(\lambda, \tilde{\sigma}, \xi)$, avec $\lambda = \mathbb{P}(X > u)$;
 - Censure : si $X_i > u$, $X_i = \tilde{X}_i$ où $\tilde{X}_i \sim \text{GEV}(\mu, \sigma, \xi)$.
- **Problème** : les observations consécutives sont (souvent) dépendantes \Rightarrow nécessité d'un modèle dynamique spécifique aux extrêmes.

Modèles statistiques pour des observations univariées

- Maxima par bloc $M_n \sim \text{GEV}(\mu, \sigma, \xi)$: perte de données.
- **Amélioration** : conserver tous les dépassements d'un seuil u .
- Deux approches possibles :
 - POT : $X|X > u \sim \text{GPD}(\lambda, \bar{\sigma}, \xi)$, avec $\lambda = \mathbb{P}(X > u)$;
 - **Censure** : si $X_i > u$, $X_i = \tilde{X}_i$ où $\tilde{X}_i \sim \text{GEV}(\mu, \sigma, \xi)$.
- **Problème** : les observations consécutives sont (souvent) dépendantes \Rightarrow nécessité d'un modèle dynamique spécifique aux extrêmes.

Modèles statistiques pour des observations univariées

- Maxima par bloc $M_n \sim \text{GEV}(\mu, \sigma, \xi)$: perte de données.
- **Amélioration** : conserver tous les dépassements d'un seuil u .
- Deux approches possibles :
 - POT : $X|X > u \sim \text{GPD}(\lambda, \tilde{\sigma}, \xi)$, avec $\lambda = \mathbb{P}(X > u)$;
 - **Censure** : si $X_i > u$, $X_i = \tilde{X}_i$ où $\tilde{X}_i \sim \text{GEV}(\mu, \sigma, \xi)$.
- **Problème** : les observations consécutives sont (souvent) dépendantes \Rightarrow nécessité d'un modèle dynamique spécifique aux extrêmes.

Modèles statistiques pour des observations univariées

- Maxima par bloc $M_n \sim \text{GEV}(\mu, \sigma, \xi)$: perte de données.
- **Amélioration** : conserver tous les dépassements d'un seuil u .
- Deux approches possibles :
 - POT : $X|X > u \sim \text{GPD}(\lambda, \bar{\sigma}, \xi)$, avec $\lambda = \mathbb{P}(X > u)$;
 - Censure : si $X_i > u$, $X_i = \tilde{X}_i$ où $\tilde{X}_i \sim \text{GEV}(\mu, \sigma, \xi)$.
- **Problème** : les observations consécutives sont (souvent) dépendantes \Rightarrow nécessité d'un modèle dynamique spécifique aux extrêmes.

Modélisation de la dépendance

Processus extrémal gaussien

Extension du concept de **loi max-stable aux processus stochastiques** (de Haan, 1986 ; Smith, 1990) :

Définition

Soit Z tel que, pour (Z_i) des copies indépendantes de Z et des suites a_n et b_n :

$$\frac{1}{a_n} \left(\max_{i=1, \dots, n} (Z_1, \dots, Z_n) - b_n \right) \stackrel{\mathcal{L}}{=} Z.$$

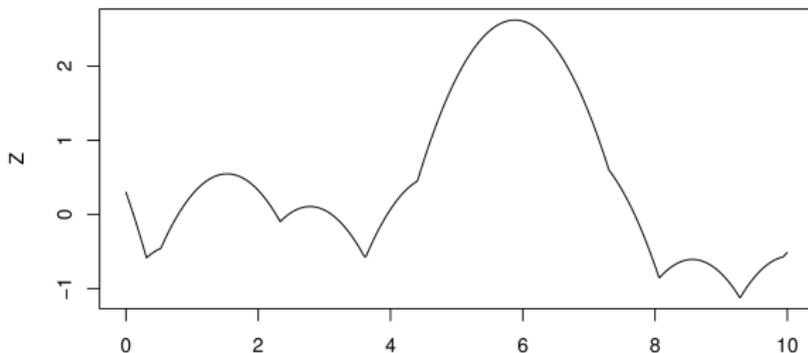
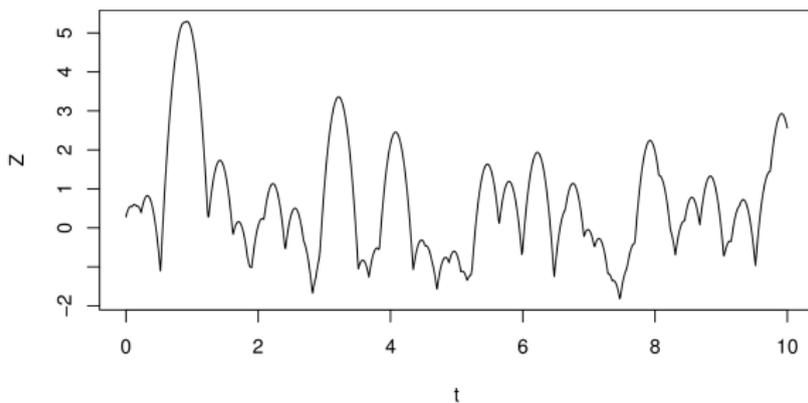
Alors Z est appelé processus max-stable.

Cas Particulier (Smith, 1990) :

$\forall t \in \mathbb{R}$, $Z(t) = \max_{i=1, \dots} \left\{ \frac{\zeta_i}{\nu \sqrt{2\pi}} \exp \left(-\frac{(s_i - t)^2}{2\nu^2} \right) \right\}$ où (ζ_i, s_i) sont les points d'un processus de Poisson sur $\mathbb{R}_+^* \times \mathbb{R}$, d'intensité $\zeta^{-2} d\zeta \times ds$

Modélisation de la dépendance

Processus extrémal gaussien



Modélisation des dépassements de seuil

Modèle pour les excès

- Observations $\{X_{t_i}\}_{i=1,\dots,n}$
 - Si $X_{t_i} > u$, $X_{t_i} = \tilde{X}_{t_i}$ où \tilde{X}_t est un processus max-stable :
 - $\tilde{X}_i \sim G(\cdot; \mu, \sigma, \xi) \in GEV(\mu, \sigma, \xi)$
 - $\log \tilde{X}_i = -\log \tilde{X}_{t_i} \mu, \sigma, \xi$ (\sim Fréchet)
 - $\{Z_i\}$ est un processus stochastique, appelé la *tail process*, qui est \mathbb{R}^d -valable asymptotiquement via la loi
- $$P\{X_{t_i} \leq x_i, X_{t_j} \leq x_j\} \sim \left[\frac{1}{x_i} + \frac{1}{x_j} - \frac{1}{x_i x_j} \right] \frac{1}{x_i x_j} (x_i x_j - u)$$
- $\tilde{X}_i = \frac{1}{\sigma} \log \tilde{X}_{t_i} \mu, \sigma, \xi$
- $P\{X_{t_i} \leq x_i, X_{t_j} \leq x_j\}$: connue si x_i et $x_j > u$, censurée sinon ;

Modélisation des dépassements de seuil

Modèle pour les excès

- Observations $\{X_{t_i}\}_{i=1,\dots,n}$
- Si $X_{t_i} > u$, $X_{t_i} = \tilde{X}_{t_i}$ où \tilde{X}_t est un processus max-stable :
 - $\tilde{X}_t \sim G(\cdot; \mu, \sigma, \xi) \in GEV(\mu, \sigma, \xi)$
 - Soit $Z_t := -1/\log G(\tilde{X}_t; \mu, \sigma, \xi) \sim \text{Fréchet}(1)$
 - $\{Z_t\}$ est un processus extrêmeal Gaussien, seule la loi des paires est calculable analytiquement via la f.d.r :

$$\mathbb{P}\{Z(s) \leq z_i, Z(t) \leq z_j\} = \exp\left[-\frac{1}{z_i} \Phi\left(\frac{a}{2} + \frac{1}{a} \log \frac{z_j}{z_i}\right) - \frac{1}{z_j} \Phi\left(\frac{a}{2} + \frac{1}{a} \log \frac{z_i}{z_j}\right)\right]$$

où Φ est la f.d.r d'une $\mathcal{N}(0, 1)$, $a^2 = \frac{s-t}{\nu}$. ν : paramètre de dépendance

- $\mathbb{P}\{X_{t_i} \leq x_i, X_{t_j} \leq x_j\}$: connue si x_i et $x_j > u$, censurée sinon ;

Modélisation des dépassements de seuil

Modèle pour les excès

- Observations $\{X_{t_i}\}_{i=1,\dots,n}$
- Si $X_{t_i} > u$, $X_{t_i} = \tilde{X}_{t_i}$ où \tilde{X}_t est un processus max-stable :
 - $\tilde{X}_t \sim G(\cdot; \mu, \sigma, \xi) \in GEV(\mu, \sigma, \xi)$
 - Soit $Z_t := -1/\log G(\tilde{X}_t; \mu, \sigma, \xi) \sim \text{Fréchet}(1)$
 - $\{Z_t\}$ est un processus extrémal Gaussien, seule la loi des paires est calculable analytiquement via la f.d.r :

$$\mathbb{P}\{Z(s) \leq z_i, Z(t) \leq z_j\} = \exp\left[-\frac{1}{z_i} \Phi\left(\frac{a}{2} + \frac{1}{a} \log \frac{z_j}{z_i}\right) - \frac{1}{z_j} \Phi\left(\frac{a}{2} + \frac{1}{a} \log \frac{z_i}{z_j}\right)\right]$$

où Φ est la f.d.r d'une $\mathcal{N}(0, 1)$, $a^2 = \frac{s-t}{\nu}$. ν : paramètre de dépendance

- $\mathbb{P}\{X_{t_i} \leq x_i, X_{t_j} \leq x_j\}$: connue si x_i et $x_j > u$, censurée sinon ;

Modélisation des dépassements de seuil

Modèle pour les excès

- Observations $\{X_{t_i}\}_{i=1,\dots,n}$
- Si $X_{t_i} > u$, $X_{t_i} = \tilde{X}_{t_i}$ où \tilde{X}_t est un processus max-stable :
 - $\tilde{X}_t \sim G(\cdot; \mu, \sigma, \xi) \in GEV(\mu, \sigma, \xi)$
 - Soit $Z_t := -1/\log G(\tilde{X}_t; \mu, \sigma, \xi) \sim \text{Fréchet}(1)$
 - $\{Z_t\}$ est un processus extrémal Gaussien, seule la loi des paires est calculable analytiquement via la f.d.r :

$$\mathbb{P}\{Z(s) \leq z_i, Z(t) \leq z_j\} = \exp\left[-\frac{1}{z_i} \Phi\left(\frac{a}{2} + \frac{1}{a} \log \frac{z_j}{z_i}\right) - \frac{1}{z_j} \Phi\left(\frac{a}{2} + \frac{1}{a} \log \frac{z_i}{z_j}\right)\right]$$

où Φ est la f.d.r d'une $\mathcal{N}(0, 1)$, $a^2 = \frac{s-t}{\nu}$. ν : paramètre de dépendance

- $\mathbb{P}\{X_{t_i} \leq x_i, X_{t_j} \leq x_j\}$: connue si x_i et $x_j > u$, censurée sinon ;

Modélisation des dépassements de seuil

Modèle pour les excès

- Observations $\{X_{t_i}\}_{i=1,\dots,n}$
- Si $X_{t_i} > u$, $X_{t_i} = \tilde{X}_{t_i}$ où \tilde{X}_t est un processus max-stable :
 - $\tilde{X}_t \sim G(\cdot; \mu, \sigma, \xi) \in GEV(\mu, \sigma, \xi)$
 - Soit $Z_t := -1/\log G(\tilde{X}_t; \mu, \sigma, \xi) \sim \text{Fréchet}(1)$
 - $\{Z_t\}$ est un processus extrêmeal Gaussien, seule la **loi des paires** est calculable analytiquement via la f.d.r :

$$\mathbb{P}\{Z(s) \leq z_i, Z(t) \leq z_j\} = \exp\left[-\frac{1}{z_i}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_j}{z_i}\right) - \frac{1}{z_j}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_i}{z_j}\right)\right]$$

où Φ est la f.d.r d'une $\mathcal{N}(0, 1)$, $a^2 = \frac{s-t}{\nu}$. ν : paramètre de dépendance

- $\mathbb{P}\{X_{t_i} \leq x_i, X_{t_j} \leq x_j\}$: connue si x_i et $x_j > u$, censurée sinon ;

Modélisation des dépassements de seuil

Modèle pour les excès

- Observations $\{X_{t_i}\}_{i=1,\dots,n}$
- Si $X_{t_i} > u$, $X_{t_i} = \tilde{X}_{t_i}$ où \tilde{X}_t est un processus max-stable :
 - $\tilde{X}_t \sim G(\cdot; \mu, \sigma, \xi) \in GEV(\mu, \sigma, \xi)$
 - Soit $Z_t := -1/\log G(\tilde{X}_t; \mu, \sigma, \xi) \sim \text{Fréchet}(1)$
 - $\{Z_t\}$ est un processus extrêmeal Gaussien, seule la **loi des paires** est calculable analytiquement via la f.d.r :

$$\mathbb{P}\{Z(s) \leq z_i, Z(t) \leq z_j\} = \exp\left[-\frac{1}{z_i}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_j}{z_i}\right) - \frac{1}{z_j}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_i}{z_j}\right)\right]$$

où Φ est la f.d.r d'une $\mathcal{N}(0, 1)$, $a^2 = \frac{s-t}{\nu}$. ν : paramètre de dépendance

- $\mathbb{P}\{X_{t_i} \leq x_i, X_{t_j} \leq x_j\}$: connue si x_i et $x_j > u$, censurée sinon ;

Plan de l'exposé

- 1 Description des données
- 2 Modèles statistiques
 - Théorie probabiliste dans le cas univarié
 - Modélisation de la dépendance
- 3 Inférence statistique**
 - Résultats théoriques
 - Résultats numériques
- 4 Application à la hauteur significative des vagues
- 5 Conclusion

Estimation des paramètres

Définition

La vraisemblance exacte est inaccessible.

Définition

- Vraisemblance indépendante :
 $IL(x_{t_1}, \dots, x_{t_n}; \theta) = \prod_{i=1}^n p(x_{t_i}; \theta)$ estimateur associé : *MILE* ;
- Vraisemblance **composite** :
 $CL_K(x_{t_1}, \dots, x_{t_n}; \theta) = \prod_{i=1}^{n-K} \prod_{j=i+1}^{i+K} p(x_{t_i}, x_{t_j}; \theta)$ estimateur associé : *MPL_KE* ;
- Vraisemblance markovienne :
 $ML^{(1)}(x_{t_1}, \dots, x_{t_n}; \theta) = CL_1(x_{t_1}, \dots, x_{t_n}; \theta) / IL(x_{t_1}, \dots, x_{t_n}; \theta)$
 estimateur associé : *MMLE*

Estimation des paramètres

Propriétés

Théorème

Soit $\hat{\nu} = \arg \max CL_1(x_{t_1}, \dots, x_{t_n}; \theta)$. Si $\nu^* \in \Theta := [\nu_-, \nu_+]$ où $0 < \nu_- < \nu_+$, alors $\hat{\nu}$ est un estimateur convergent de ν^* .

Idées de la preuve : Estimation en deux étapes.

- 1 Estimation des paramètres marginaux ;
- 2 Estimation du paramètre de dépendance.

Estimation des paramètres

Propriétés

Théorème

Soit $\hat{\nu} = \arg \max CL_1(x_{t_1}, \dots, x_{t_n}; \theta)$. Si $\nu^* \in \Theta := [\nu_-, \nu_+]$ où $0 < \nu_- < \nu_+$, alors $\hat{\nu}$ est un estimateur convergent de ν^* .

Idées de la preuve : Estimation en deux étapes.

- 1 Estimation des paramètres marginaux ;
- 2 Estimation du paramètre de dépendance.

Estimation des paramètres

Propriétés

Théorème

Soit $\hat{\nu} = \arg \max CL_1(x_{t_1}, \dots, x_{t_n}; \theta)$. Si $\nu^* \in \Theta := [\nu_-, \nu_+]$ où $0 < \nu_- < \nu_+$, alors $\hat{\nu}$ est un estimateur convergent de ν^* .

Idées de la preuve : Estimation en deux étapes.

- 1 Estimation des paramètres marginaux ;
- 2 Estimation du paramètre de dépendance.

Estimation des paramètres

Propriétés

Théorème

Soit $\hat{\nu} = \arg \max CL_1(x_{t_1}, \dots, x_{t_n}; \theta)$. Si $\nu^* \in \Theta := [\nu_-, \nu_+]$ où $0 < \nu_- < \nu_+$, alors $\hat{\nu}$ est un estimateur convergent de ν^* .

Idées de la preuve : Estimation en deux étapes.

- 1 Estimation des paramètres marginaux ;
- 2 Estimation du paramètre de dépendance.

Estimation des paramètres

Propriétés

Théorème de Pzfangal

Soit $\{Z_i\}_{i=1,\dots,n}$ un processus stationnaire ergodique dont la loi dépend d'un paramètre $\nu^* \in \Theta$ avec Θ un compact de \mathbb{R} et soit Q_n un contraste tel que $Q_n(\nu) = \frac{1}{n} \sum_{i=1}^{n-1} f(Z_i, Z_{i+1}; \nu)$, où f est une fonction mesurable continue en ν . Si

- 1 $\mathbb{E} \inf_{\nu \in \Theta} f(Z_1, Z_2; \nu) > -\infty$;
- 2 $\nu \mapsto \mathbb{E} f(Z_1, Z_2; \nu)$ possède un unique minimum fini en ν^* .

alors l'estimateur au minimum de contraste

$\hat{\nu}_n = \arg \min_{\nu \in \Theta} Q_n(\nu)$ est fortement consistant :

$$\lim_{n \rightarrow \infty} \hat{\nu}_n = \nu^* \text{ p.s.}$$

Estimation des paramètres

Propriétés

Un processus Z extrémal gaussien, est ergodique (Stoev, 2008).

Définitions

- $F_Z(z_1, z_2; \nu) := \mathbb{P}(Z_1 < z_1, Z_2 < z_2) = \exp \{-V(z_1, z_2; \nu)\}$
- $V(z_1, z_2; \nu) := \frac{\Phi(a/2+1/a \log \frac{z_2}{z_1})}{z_1} + \frac{\Phi(a/2+1/a \log \frac{z_1}{z_2})}{z_2}$
- $p(Z_1, Z_2; \nu) := \frac{\partial^2}{\partial z_1 \partial z_2} F_Z(z_1, z_2; \nu)$;
- $f(Z_1, Z_2; \nu) := -\log p(Z_1, Z_2; \nu)$;

Expression du contraste

$$f(z_1, z_2; \nu) = V(z_1, z_2; \nu) - \log \left[\frac{\Phi(w)\Phi(v)}{a^2 z_1 z_2} + \frac{\varphi(w)}{a z_1^2 z_2} \right]$$

avec

- $w := w(z_1, z_2; \nu) := a/2 + 1/a \log \frac{z_2}{z_1}$
- $v := v(z_1, z_2; \nu) := a/2 - w$

Estimation des paramètres

Propriétés

Preuve de (1) (Minoration)

On a d'après le slide précédent :

$$f(z_1, z_2; \nu) = V(z_1, z_2; \nu) - \log \left[\frac{\Phi(w)\Phi(v)}{a^2 z_1 z_2} + \frac{\varphi(w)}{a z_1^2 z_2} \right].$$

- On a grâce à la borne de Fréchet (1951) :

$$\inf_{\nu} \{V(z_1, z_2; \nu)\} \geq \min \left(-\frac{1}{z_1}, -\frac{1}{z_2} \right);$$

- Si $\nu \in [\nu_-, \nu_+]$, $\log \left[\frac{\Phi(w)\Phi(v)}{a^2 z_1 z_2} + \frac{\varphi(w)}{a z_1^2 z_2} \right] \geq 1 - \frac{\nu_+^2}{z_1 z_2}$

Si $Z \sim \text{Fréchet}(1)$, $1/Z \sim \mathcal{E}(1)$, donc $\mathbb{E} \frac{1}{Z} = 1 < \infty$;

Donc $\mathbb{E} \inf_{\nu \in \Theta} f(Z_1, Z_2; \nu) > -\infty$

Estimation des paramètres

Propriétés

Preuve de (2) (Identifiabilité)

Distance de Kullback-Leibler :

$$K(P_{\nu^*}, P_{\nu}) := \mathbb{E}_{\nu^*} \left[-\log \frac{p(Z_1, Z_2; \nu)}{p(Z_1, Z_2; \nu^*)} \right]$$

- $K = 0 \Leftrightarrow \forall z_1, z_2 : p(z_1, z_2; \nu^*) = p(z_1, z_2; \nu)$
- En particulier : pour $z_1 = z_2 = z$, on a :

$$\begin{aligned} \exp \left[-\frac{2}{z} \Phi \left(\frac{1}{2\nu} \right) \right] & \left[\frac{\Phi \left(\frac{1}{2\nu} \right)^2}{z^2} \nu^2 + \frac{\varphi \left(\frac{1}{2\nu} \right)}{z^3} \nu \right] \\ & = \exp \left[-\frac{2}{z} \Phi \left(\frac{1}{2\nu^*} \right) \right] \left[\frac{\Phi \left(\frac{1}{2\nu^*} \right)^2}{z^2} \nu^{*2} + \frac{\varphi \left(\frac{1}{2\nu^*} \right)}{z^3} \nu^* \right] \end{aligned}$$

- Quand $z \rightarrow 0 : \Phi \left(\frac{1}{\nu} \right) = \Phi \left(\frac{1}{\nu^*} \right)$

Estimation des paramètres

Etude empirique : Convergence

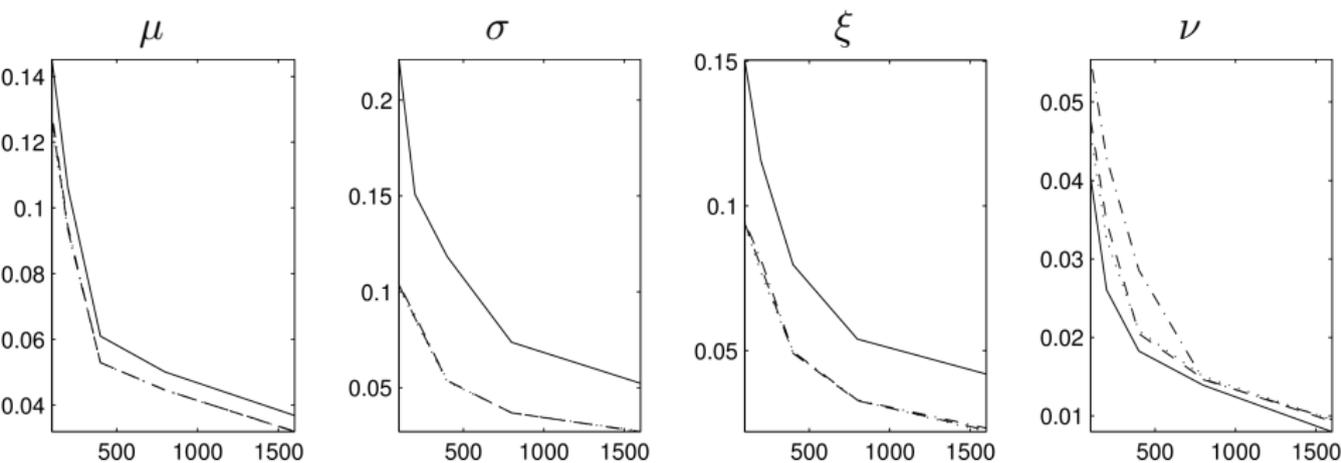


FIGURE: RMSE en fonction de la taille de l'échantillon. Ligne continue : $MILE$; ligne pointillée : $MMLE$; Tirets : MPL_1E ; Trait mixte : MPL_5E .

Estimation des paramètres

Etude empirique : Influence de la censure

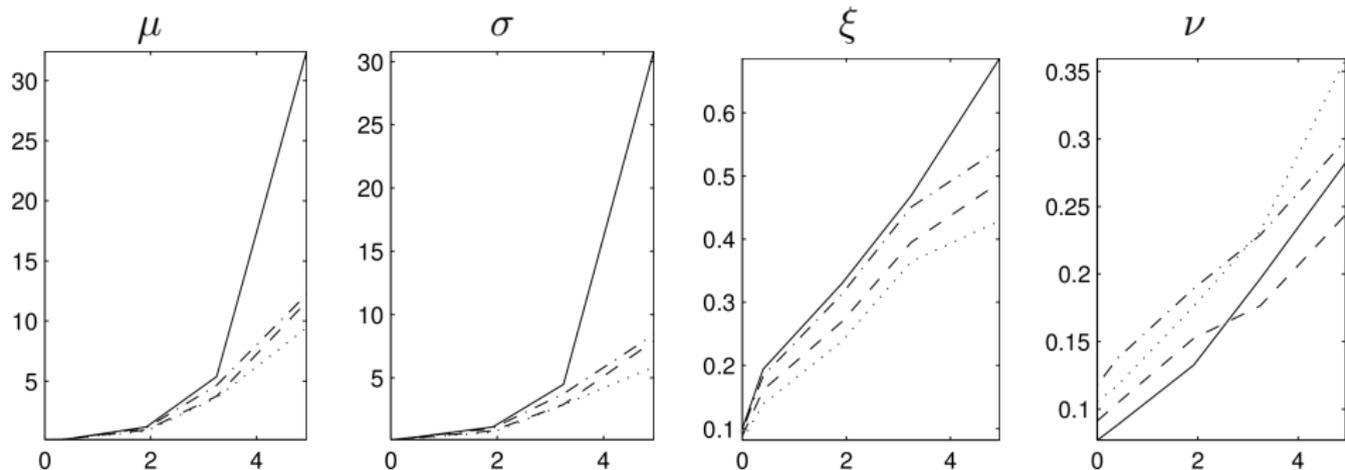


FIGURE: RMSE en fonction du seuil. Ligne continue : $MILE$; ligne pointillée : $MMLE$; Tirets : MPL_1E ; Trait mixte : MPL_5E .

Robustesse

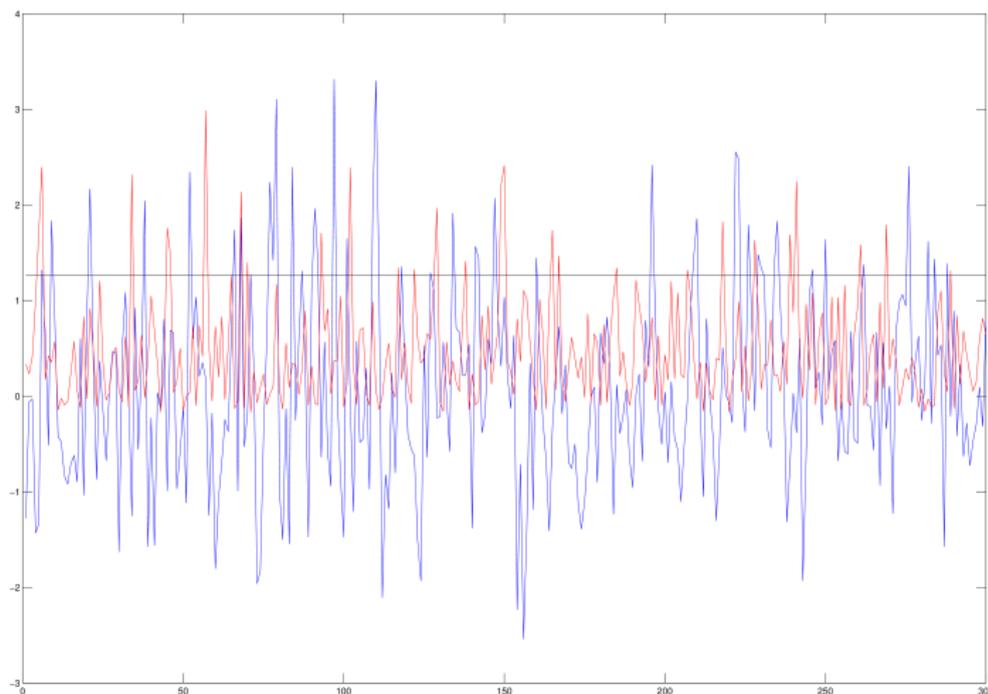
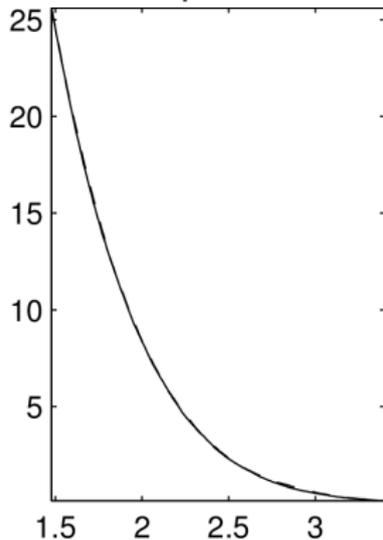


FIGURE: Exemple de données simulées (Bleu) contre le modèle ajusté (Rouge) et le seuil retenu pour l'ajustement (Noir).

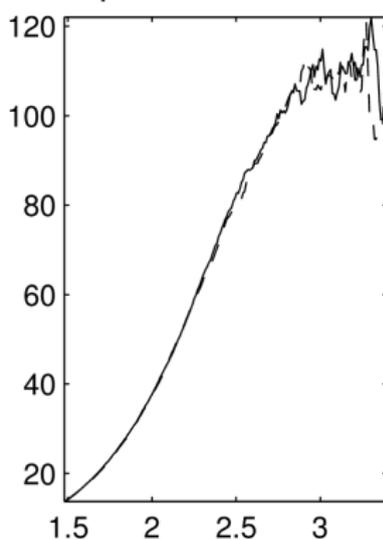
Robustesse

Adéquation à un échantillon i.i.d

Nb. de dépassements



Temps inter-clusters



Longueur des clusters

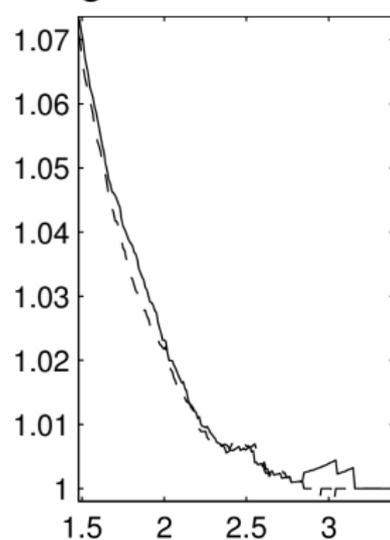
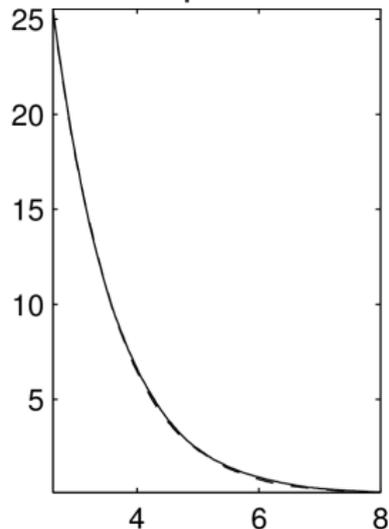


FIGURE: Comparaison du comportement extrême d'un échantillon i.i.d de loi normale contre le modèle ajusté.

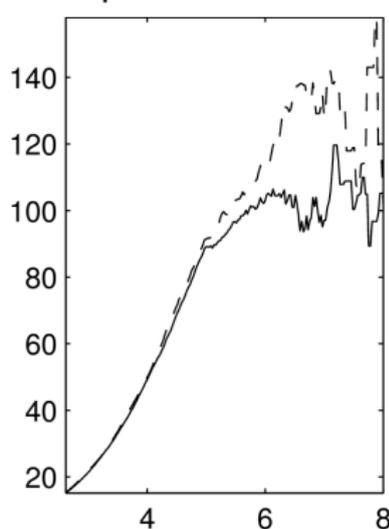
Robustesse

Adéquation à un logARMAX

Nb. de dépassements



Temps inter-clusters



Longueur des clusters

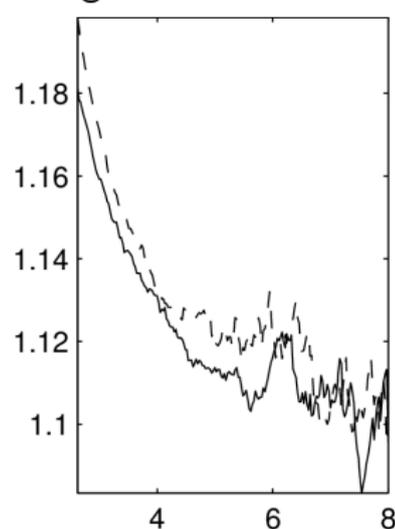
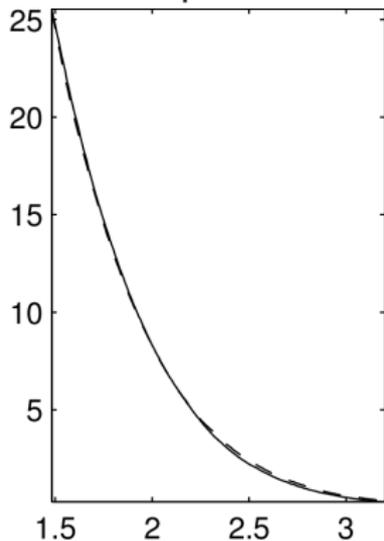


FIGURE: Comparaison du comportement extrême d'un logARMAX de paramètre $\varphi = 0.9$ contre le modèle ajusté.

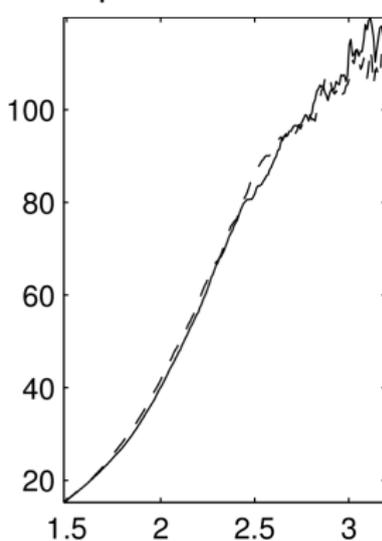
Robustesse

Adéquation à un AR(1)

Nb. de dépassements



Temps inter-clusters



Longueur des clusters

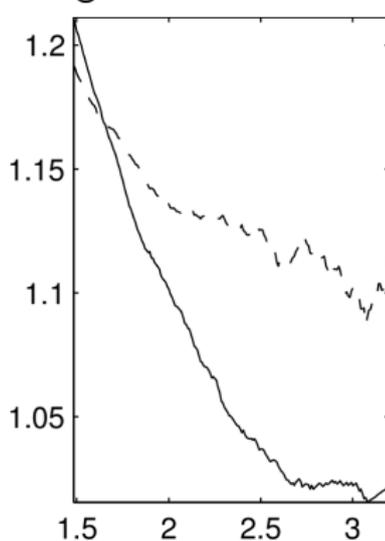


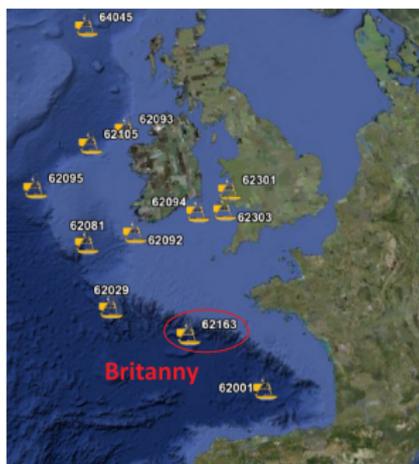
FIGURE: Comparaison du comportement extrême d'un AR(1) de paramètre $\varphi = 0.2$ contre le modèle ajusté.

Plan de l'exposé

- 1 Description des données
- 2 Modèles statistiques
 - Théorie probabiliste dans le cas univarié
 - Modélisation de la dépendance
- 3 Inférence statistique
 - Résultats théoriques
 - Résultats numériques
- 4 Application à la hauteur significative des vagues**
- 5 Conclusion

Application la hauteur significative des vagues

- Bouée Brittany ;
- Mois de **Décembre** (stationnarité) ;
- **8 ans** de données ;
- Seuil : quantile à 93%.



Jeu de données	ξ	σ	μ	θ	$q_{100 \text{ years}}$
ERA-Interim (POT)	-0.13	0.40	7.31	0.25	8.65
ERA-Interim (CPL)	0.14	0.68	4.21	0.06	14.42
Satellite(CPL)	0.10	0.95	3.22	0.60	15.77
Buoy (CPL)	0.05	1.03	4.55	0.25	14.60

TABLE: Ajustement des modèles sur les différents jeux de données.

Application la hauteur significative des vagues

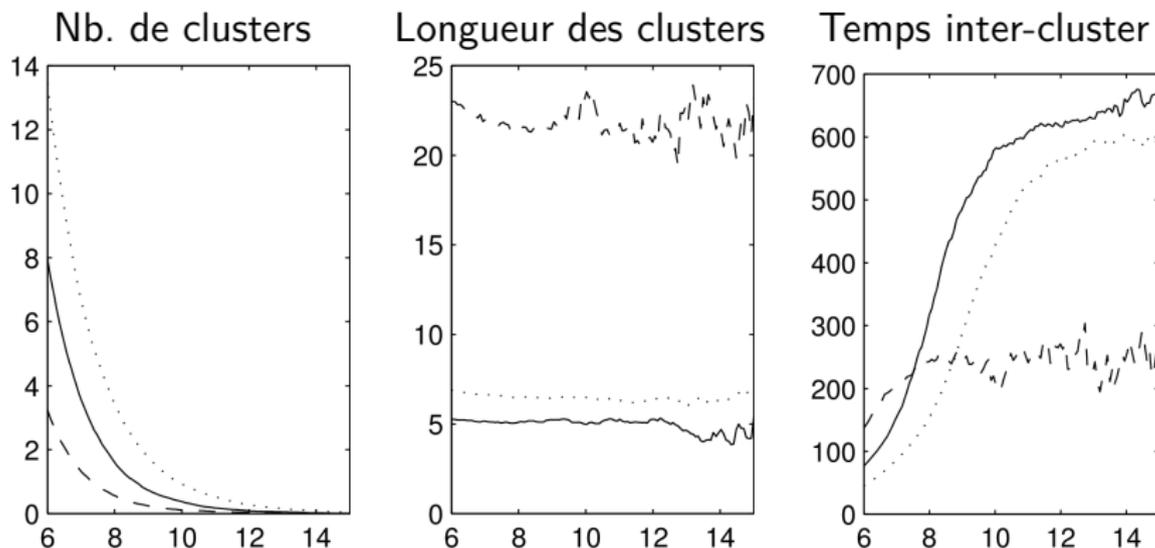


FIGURE: Ligne continue : modèle ajusté sur la bouée ; Ligne pointillée : modèle ajusté sur le satellite ; Tirets : modèle ajusté sur ERA-Interim.

Plan de l'exposé

- 1 Description des données
- 2 Modèles statistiques
 - Théorie probabiliste dans le cas univarié
 - Modélisation de la dépendance
- 3 Inférence statistique
 - Résultats théoriques
 - Résultats numériques
- 4 Application à la hauteur significative des vagues
- 5 Conclusion

Conclusion

... et perspectives

- Le modèle proposé est flexible et permet de prendre en compte des données réparties de façon irrégulière ou avec des valeurs manquantes ;
- Des outils de diagnostics ont été proposés et appliqués à la vérification de l'ajustement des modèles ;
- Perspectives :
 - Résultats théoriques : normalité asymptotique, choix de K , influence de la censure, choix de modèle...
 - Extensions : modélisation spatio-temporelle, autres processus max-stables...



D. R. Cox.

A note on pseudolikelihood constructed from marginal densities.

Biometrika, 92 :729–737, 2004.



R. A. Davis.

Stable limits for partial sums of dependent random variables.

The Annals of Probability, 11(2) :262–269, 1983.



L. de Haan.

A spectral representation for max-stable processes.

Annals of Statistics, 12(4) :1194–1204, 1984.



S. A. Padoan, M. Ribatet, and S. A. Sisson.

Likelihood-based inference for max-stable processes.

Journal of the American Statistical Association,
105(489) :263–277, 2010.



M. Ribatet, T. B. M. J. Ouarda, E. Sauquet, and J. M. Gresillon.

Modeling all exceedances above a threshold using an extremal dependence structure : Inferences on several flood characteristics.

Water Resources Research, 45(3), March 2009.



R. L. Smith.

Max-stable processes and spatial extremes.
Unpublished, 1990.



R. L. Smith, J. A. Tawn, and S. G. Coles.

Markov chain models for threshold exceedances.

Biometrika, 84(2) :249–268, 1997.



C. Varin.

On composite marginal likelihoods.

Advances in Statistical Analysis, 92(1) :1–28, 2008.