

## Estimation de la Magnitude Maximale Annuelle par la Théorie des Valeurs Extrêmes

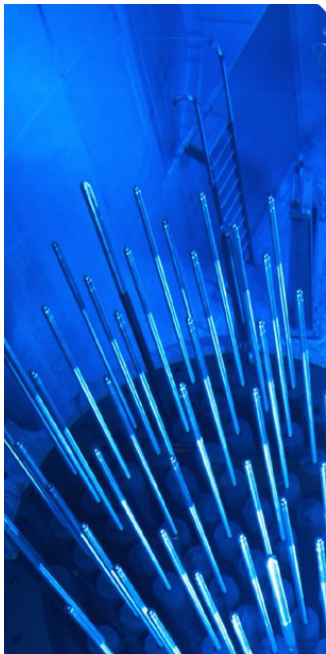
Anne Dutfoy

EDF R&D PERICLES  
[anne.dutfoy@edf.fr](mailto:anne.dutfoy@edf.fr)

Journée Utilisateurs OpenTURNS #11  
15 Juin 2018



CHANGER L'ÉNERGIE ENSEMBLE



# Sommaire

- ① Modélisations probabilistes
  - Séismes
  - Magnitude maximale annuelle
  - Magnitudes
- ② Estimation du modèle
  - Complétude des mesures de séismes
  - Processus Ponctuel de Poisson
  - Estimation du modèle
  - Validation et Sensibilité aux données
- ③ Application à la zone sismique Alpes - Pyrénées
  - Cas de référence
  - Validation
  - Sensibilité aux données de complétude
  - Sensibilité aux incertitudes statistiques
- ④ OpenTURNS
- ⑤ Conclusion

# Estimation de la Magnitude Maximale Annuelle par la Théorie des Valeurs Extrêmes

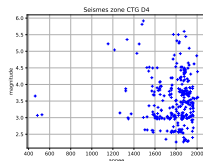
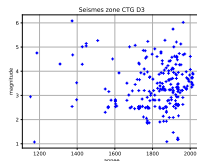
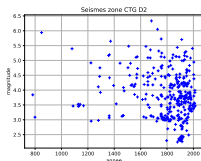
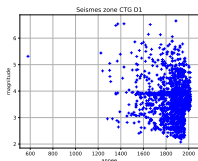
- ① Modélisations probabilistes
  - Séismes
  - Magnitude maximale annuelle
  - Magnitudes
- ② Estimation du modèle
  - Complétude des mesures de séismes
  - Processus Ponctuel de Poisson
  - Estimation du modèle
  - Validation et Sensibilité aux données
- ③ Application à la zone sismique Alpes - Pyrénées
  - Cas de référence
  - Validation
  - Sensibilité aux données de complétude
  - Sensibilité aux incertitudes statistiques
- ④ OpenTURNS
- ⑤ Conclusion

## Séismes : Processus Ponctuel de Poisson

Pour modéliser les arrivées de séismes ainsi que les magnitudes mesurées à chaque séisme, nous considérons le processus  $X(\omega) : \Omega \rightarrow E = \mathbb{R}^+ \times \mathbb{R}_*^+$  défini par :

$$X(\omega) = (T(\omega), M^*(\omega)) \quad (1)$$

où les variables  $T$  et  $M^*$  sont supposées **indépendantes** et caractérisent respectivement les instants aléatoires de séismes et leur magnitude.



### Hypothèse fondamentale

$X(\omega)$  forme un **Processus Ponctuel de Poisson** : la variable aléatoire de comptage  $N(t)$  associée est à **accroissements indépendants et stationnaires** :

- ① les nombres de séismes sur des périodes disjointes sont indépendants
- ② le nombre de séismes sur une période de longueur  $T$  ne dépend que de  $T$  et non de son positionnement dans le temps

$\Rightarrow N(t) - N(s) \sim \mathcal{P}(\mu(t-s))$ ,  $\mu$  (intensité) est le nombre moyen de séismes par unité de temps

$\Rightarrow$  durée entre 2 séismes est une variable aléatoire  $\sim \mathcal{E}(\mu)$

# Séismes : Processus Ponctuel de Poisson

## Mesure Intensité du Processus de Poisson

On peut définir la **mesure intensité**  $\Lambda$  de  $X$ , agissant sur tous les boréliens de  $E = \mathbb{R}^+ \times \mathbb{R}_*^+$ , par son action sur les éléments générateurs :

$$\forall (t_1, t_2, m_1, m_2) \in (\mathbb{R}_*^+)^4, \quad \Lambda([t_1, t_2] \times [m_1, m_2]) = \mu(t_2 - t_1) \mathbb{P}(M^* \in ]m_1, m_2])$$

soit

$$\forall (t_1, t_2, m_1, m_2) \in (\mathbb{R}_*^+)^4, \quad \Lambda([t_1, t_2] \times [m_1, m_2]) = \mu(t_2 - t_1)[\bar{F}^*(m_1) - \bar{F}^*(m_2)] \quad (2)$$

où  $F^*$  est la fonction de répartition de la loi de  $M^*$ .

### Remarques :

La loi produit du processus est justifiée par l'indépendance entre les variables  $T$  et  $M^*$ .

Pour tout  $A \in E$ ,  $\Lambda(A)$  caractérise la variable aléatoire de comptage des séismes dans  $A$  :  $N(A) \sim \mathcal{P}(\Lambda(A))$ .

En considérant  $\Lambda([0, t] \times [0, m]) = \mu t F^*(m)$ , on calcule la **densité**  $\lambda$  est :

$$\lambda(t, m) = \frac{d^2 \Lambda}{\partial m \partial t}([0, t] \times [0, m]) = \mu F^{*'}(m) \quad (3)$$

# Magnitude maximale annuelle I

Les études sur le séisme s'intéressent à la **magnitude maximale annuelle**  $M_{max}$ .

Magnitude maximale annuelle  $M_{max}$

$M_{max}$  est une variable aléatoire définie par :

$$M_{max} = \begin{cases} 0 & \text{si } N_a(\omega) = 0 \\ \max(M_1^*(\omega), \dots, M_{N_a(\omega)}^*(\omega)) & \text{si } N_a(\omega) \geq 1 \end{cases} \quad (4)$$

où  $N_a$  est le nombre annuel (aléatoire) de séismes :  $N_a \sim \mathcal{P}(\mu)$ ,  $M_i^*$  la magnitude du séisme  $i$ , les  $(M_i^*)_i$  sont iid, de loi  $F^*$  et  $(N_a, M^*)$  indépendantes.

De sa fonction de répartition :

$$G(m) = \mathbb{P}(M_{max} \leq m)$$

on détermine les **niveaux de période de retour**  $T$  années : la magnitude  $m_T$  telle que

$$G(m_T) = 1 - 1/T$$

On montre que la durée moyenne entre 2 séismes de période de retour  $T$  années est  $T$ .  
Ex : séisme centenal, millénal.

# Magnitude maximale annuelle

## Loi de $M_{max}$

L'hypothèse de Poisson permet de **calculer exactement** la loi de  $M_{max}$ , de fonction de répartition  $G$  :

$$G(m) = \mathbb{P}(M_{max} \leq m) = e^{-\mu[1-F^*(m)]} \quad \forall m \geq 0 \quad (5)$$

## Objectif

La **connaissance de la loi des magnitudes  $M^*$**  permet donc de connaître la loi de la magnitude maximale annuelle.

## Modélisations de $M^*$

- Loi de Gutenberg-Richter
- **Approche Dépassement de seuil POT : Loi de Pareto**

# $M^*(\omega)$ : Loi de Gutenberg-Richter

## Loi de Gutenberg-Richter

La **loi de Gutenberg-Richter** repose sur la relation empirique entre le nombre moyen annuel  $\mathbb{E}(N_m)$  de séismes de magnitude supérieure à  $m$  et la magnitude  $m$ .

$$\log \mathbb{E}(N_m) = -am + b \quad (6)$$

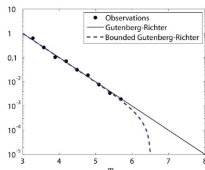
avec  $(a, b)$  des constantes positives.

**Conséquence** : Sous l'hypothèse de processus de Poisson,  $N_a \sim \mathcal{P}(\mu)$  et donc  $N_m \sim \mathcal{P}(\mu \mathbb{P}(M^* > m))$ , d'où :

$$\mathbb{E}(N_m) = \mu \mathbb{P}(M^* > m) \quad (7)$$

De (6) et (7), on déduit :  $\log \mathbb{P}(M^* > m) = -am + d$  où  $d = b - \log \mu$  et donc que  $\mathbb{P}(M^* > m) = 10^{-am+d}$  :

$$M^* - d/a \sim \text{Exp}(a \ln 10)$$



## Limitations de la Loi de Gutenberg-Richter :

- La relation empirique n'est pas valable pour les fortes magnitudes
- Des considérations physiques mènent à borner les magnitudes : on tronque la loi Exponentielle

Beaucoup de variantes ont été proposées pour la loi de  $M^*$ .



Loi  $M^*(\omega)$  : Loi GPD

Revenons à l'expression de la loi de  $M_{\max}$  à partir de celle de  $M^*(\omega)$  :

$$G(m) = \mathbb{P}(M_{\max} \leq m) = e^{-\mu[1-F^*(m)]} \quad \forall m \geq 0$$

**Idée** : pour calculer la loi de  $M_{\max}$ , il faut donc connaître la loi  $F^*$  des magnitudes  $M^*$  ... mais si seules les grandes valeurs de  $M_{\max}$  nous intéressent, il suffit de **connaître  $F^*$  dans sa queue de distribution** !

Pickands, Balkema, de Haan

$F \in \mathcal{D}(G_\xi)$  si et seulement s'il existe une fonction  $\sigma(u) > 0$  telle que :

$$\limsup |\bar{F}_u(y) - \bar{H}_{\sigma_u, \xi}(y + u)| = 0 \quad \text{lorsque } u \rightarrow x_F^- \quad \text{pour } 0 \leq y \leq x_F - u$$

où  $x_F$  est la borne supérieure du support de  $F$ ,  $\bar{F}_u(y) = \mathbb{P}(X > y + u | X > u)$  et  $H_{\sigma_u, \xi}$  la cdf d'une loi de **Pareto Généralisée**.

On en déduit que pour  $u$  suffisamment grand :

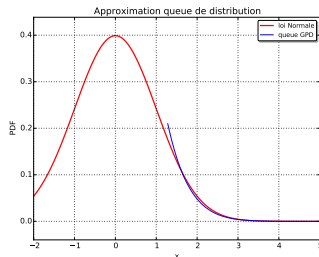
$$\bar{F}_u(y) \simeq \bar{H}_{\sigma_u, \xi}(y + u), \quad \forall y > 0$$

Conséquence : La queue de  $F$  s'écrit, en notant  $\zeta_u = \mathbb{P}(X > U)$  :

$$\bar{F}(x) \simeq \zeta_u \bar{H}_{\sigma_u, \xi}(x - u), \quad \forall x > u$$

Loi  $M^*(\omega)$  : Loi GPDLoi de Pareto Généralisée  $H_{(\sigma,\xi)}$  :

$$H_{(\sigma,\xi)}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-1/\xi} & \forall x > 0, \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma}\right) & \forall x > 0, \xi = 0 \end{cases}$$

Application à la loi de  $M^*$  : pour  $u$  suffisamment grand,  $(M^* - u | M^* > u) \sim H_{\sigma,\xi}$  et donc :

$$\bar{F}^*(m) \simeq \zeta_u^* \bar{H}_{\sigma,\xi}(m - u) \quad \forall m \geq u$$

où  $\zeta_u^* = \mathbb{P}(M^* > u)$ . Cette approximation mène à :

$$G(m) = e^{-\mu \zeta_u^* \bar{H}_{\sigma,\xi}(m - u)} \quad \forall m \geq u$$

Le seuil  $u$  correspond au quantile  $(1 - \zeta_u^*)$  de  $M^*$  et au quantile  $(e^{-\mu \zeta_u^*})$  de  $M_{max}$  :

$$F^*(u) = 1 - \zeta_u^* \quad \text{et} \quad G(u) = e^{-\mu \zeta_u^*}$$

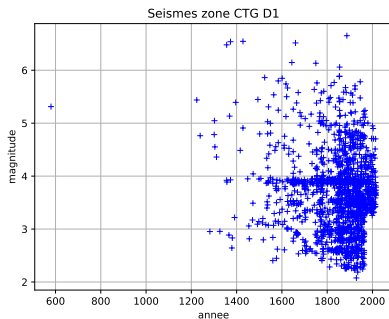
## Objectif

Estimer les paramètres  $(u, \sigma, \xi)$  sur les mesures de séismes.

# Estimation de la Magnitude Maximale Annuelle par la Théorie des Valeurs Extrêmes

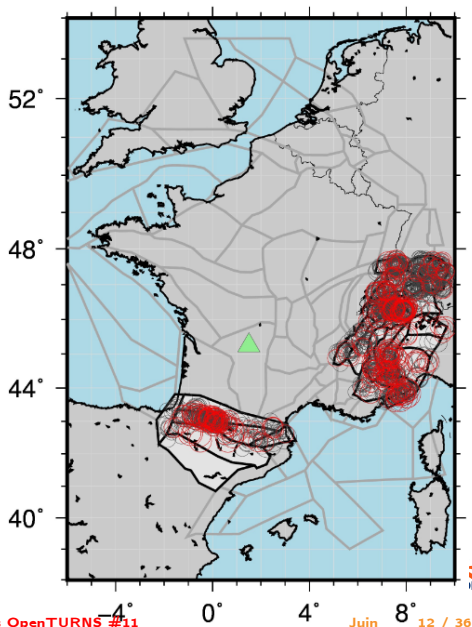
- ① Modélisations probabilistes
  - Séismes
  - Magnitude maximale annuelle
  - Magnitudes
- ② Estimation du modèle
  - Complétude des mesures de séismes
  - Processus Ponctuel de Poisson
  - Estimation du modèle
  - Validation et Sensibilité aux données
- ③ Application à la zone sismique Alpes - Pyrénées
  - Cas de référence
  - Validation
  - Sensibilité aux données de complétude
  - Sensibilité aux incertitudes statistiques
- ④ OpenTURNS
- ⑤ Conclusion

## Exemple de données sismiques



## Zone Alpes-Pyrénées - Mesures de séismes

Nbre de séismes	1806
Dates	[580, 2016]
Magnitude	[2.08, 6.65]
Séisme max	1887



# Complétude des mesures

## Complétude

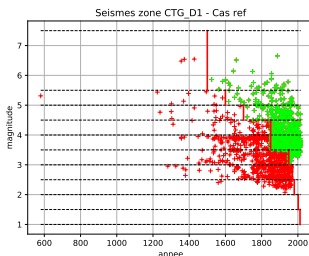
Les séismes de faible magnitude ne sont mesurés que depuis peu. Les mesures de faible magnitude sont d'autant moins *fiables* qu'elles sont anciennes.

**Il est nécessaire de tronquer les mesures aux dates de complétude.**

Les sismologues ont défini un tableau des périodes de complétude pour chaque classe de magnitude : date à partir de laquelle les observations sont conservées.

### Classe | Date

[1, 1.5]	2012
[1.5, 2]	2002
[2, 2.5]	1980
[2.5, 3]	1970
[3, 3.5]	1950
[3.5, 4]	1850
[4, 4.5]	1850
[4.5, 5]	1700
[5, 5.5]	1600
$\geq 5.5$	1500



### Zone CTD - D1 - Après troncature

Nbre de séismes	851
Dates	[1524, 2016]
Magnitude	[3.04, 6.65]
Séisme max	1887
Classes non représentées	[1, 3]

## Difficulté de modélisation

Comment prendre en compte les périodes d'observation de chaque classe de magnitude dans l'estimation de  $(u, \sigma, \xi)$  ? On a besoin d'inférer le modèle simultanément sur les instants de séismes **ET** les magnitudes.

# Domaine d'intérêt



## Ensemble des données

On s'intéresse à l'ensemble :

$$A_0 = \cup A_i, \quad A_i = [T_f - D_i, T_f] \times [m_i, m_{i+1}]$$

Comme les  $A_i$  sont disjoints,  $\Lambda(A) = \sum_i \Lambda(A_i)$  :

$$\Lambda(A_0) = \sum_i \mu D_i [\bar{F}^*(m_i) - \bar{F}^*(m_{i+1})]$$

## Lien avec les valeurs extrêmes

L'approximation de la distribution de  $F^*$  n'est valide que dans sa queue :

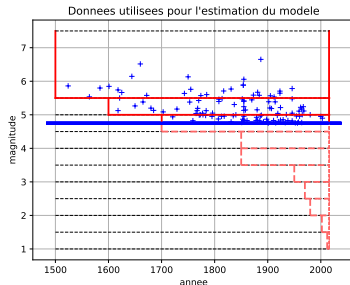
$$\bar{F}^*(m) \simeq \zeta_u^* \bar{H}_{\sigma, \xi}(m - u) \quad \forall m > u$$

où  $\bar{F}^*(u) \simeq 1$ .

⇒ On seuille  $A_0$  à la valeur élevée  $u$  :

$$A_u = A_0 \cap \mathbb{R} \times [u, +\infty[$$

Le nombre de séismes dans  $A_u$  suit  $\mathcal{P}(\Lambda(A_u))$ .

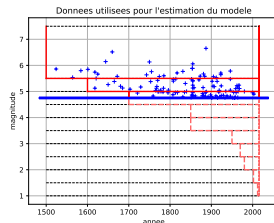


# Vraisemblance d'un processus Ponctuel de Poisson

## Théorème

On considère un processus ponctuel de Poisson sur  $\mathbb{R}^p$ , d'intensité  $\Lambda$  de densité  $\lambda$ , ie telle que  $\Lambda(\mathcal{R}) = \int_{\mathcal{R}} \lambda(\underline{s}) d\underline{s}$ , pour tout borélien  $\mathcal{R}$  de  $\mathbb{R}^p$ . On note  $\mathcal{E}_n = (n, \{\underline{x}_1, \dots, \underline{x}_n\})$  une réalisation du processus sur  $\mathcal{R}$ . La vraisemblance de  $\mathcal{E}_n$  s'écrit :

$$\mathcal{L}(\mathcal{E}_n) = \exp(-\Lambda(\mathcal{R})) \prod_{i=1}^n \lambda(\underline{x}_i) \quad (8)$$



## Estimation du modèle sur les données par Max de Vraisemblance

On considère les données :  $(t_k, m_k)_{1 \leq k \leq n_u}$  contenues dans le Borélien  $A_u$  :

$$\mathcal{L} = e^{-\sum_{i \in I} \lambda(A_i)} (\mu \zeta_u^*)^{n_u} \prod_{k=1}^{n_u} H'_{\sigma, \xi}(m_k - u), \quad I = \{i | C_i \cap A_u \neq \emptyset\}$$

Si  $m_i > u$ ,  $\lambda(A_i) = \mu D_i \zeta_u^* [\bar{H}^*(m_i - u) - \bar{H}^*(m_{i+1} - u)]$ .

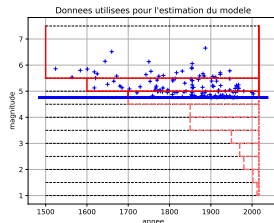
Pour la classe tronquée  $A_{i_u}^+$  :  $\lambda(A_{i_u}^+) = \mu D_{i_u} \zeta_u^* [1 - \bar{H}^*(m_{i_u+1} - u)]$ .

# Optimisation sous contraintes

L'optimisation se fait sur le log de la vraisemblance.

Les contraintes sur les paramètres  $(u, \sigma, \xi)$  sont les suivantes :

- $C_1 : \forall m_k \in A, \quad 1 + \xi \left( \frac{m_k - u}{\sigma} \right) > 0;$
- $C_2 : \sigma > 0;$
- $C_3(a) : \xi \in [-1, 1]$  ou  $C_3(b) : \xi \in [-1, 0[.$



## Informations géologiques supplémentaires

- Des considérations physiques imposent que la magnitude d'un séisme, liée à l'**énergie dégagée par le séisme**, ne peut pas prendre des valeurs inconsiderablement grandes :  $M^*$  **est nécessairement bornée** :  $\xi < 0$  : la loi des magnitudes est dans le domaine d'attraction de Weibul. On peut donc imposer la contrainte  $C_3(b)$  plutôt que  $C_3(a)$ .
- Si les sismologues déterminent une **magnitude maximale**  $m_{sup}$  dans une zone donnée, cette information se traduit par une contrainte sur la borne supérieure de la loi de Pareto du domaine de Weibul, qui se calcule à partir de  $(u, \sigma, \xi)$  :

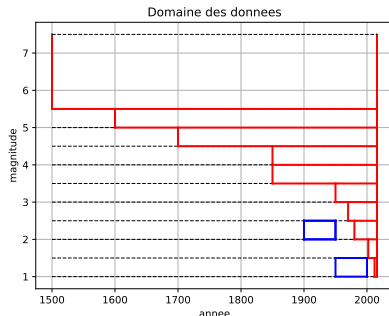
$$C_4 : \quad u - \frac{\sigma}{\xi} \leq m_{sup}$$



# Processus Ponctuel de Poisson

## Intérêt du modèle

L'intérêt de ce modèle est de pouvoir s'adapter aux plages (temps, magnitude) où l'on possède les données.



Exemple : Ce modèle permet aussi de considérer les situations où un capteur n'a pas fourni de mesures pendant une durée limitée :

- panne,
- guerre,
- perte de mesures,
- ...

# Validation et Sensibilité

## Tests de validation des modèles

Pour valider le modèle processus de Poisson, on valide sur chaque classe  $C_k$  :

- **la loi de comptage** : nombre annuel de sésimes. Comparaison entre la loi empirique et celle du modèle de Poisson d'intensité  $\Lambda_a^k = \Lambda([0, 1] \times C_k)$
- **la loi des magnitudes** : comparaison entre la loi empirique et celle du modèle de Pareto (loi conditionnelle).

Remarque : Les séismes de faible magnitude étant mesurés depuis peu de temps relativement à ceux de forte magnitude, le modèle de Poisson empirique est calé sur peu de données et est donc de moindre précision. C'est ce que l'on observe de manière générale sur l'ensemble des graphes de ces classes de faible magnitude.

## Sensibilité aux données

Deux types de sensibilité sont analysées :

- **sensibilité aux données de complétude** : prise en compte des autres valeurs (inf et sup) données par les sismologues,
- **sensibilité aux mesures elles-mêmes** : incertitudes statistiques des estimations : loi des estimateurs des paramètres obtenue par Bootstrap et conséquence sur la loi de la magnitude maximale annuelle.

# Estimation de la Magnitude Maximale Annuelle par la Théorie des Valeurs Extrêmes

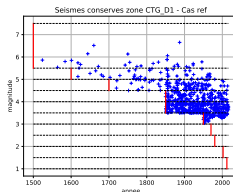
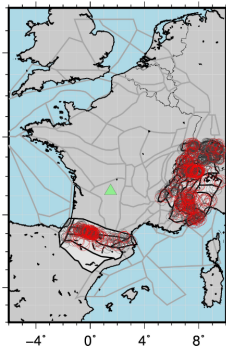
- ① Modélisations probabilistes
  - Séismes
  - Magnitude maximale annuelle
  - Magnitudes
- ② Estimation du modèle
  - Complétude des mesures de séismes
  - Processus Ponctuel de Poisson
  - Estimation du modèle
  - Validation et Sensibilité aux données
- ③ Application à la zone sismique Alpes - Pyrénées
  - Cas de référence
  - Validation
  - Sensibilité aux données de complétude
  - Sensibilité aux incertitudes statistiques
- ④ OpenTURNS
- ⑤ Conclusion

# Cas d'application : France, Alpes et Pyrénées

## Catalogue de sismicité

Les données proviennent du catalogue de sismicité déclusterisée *Catalogue\_Agregats\_GTR* qui découpe la France en 6 macro-zones homogènes.

On étudie la zone 1 qui possède le plus de mesures de séismes.



### Zone CTD - D1 - Après troncature

Nbre de séismes	851
Dates	[1524, 2016]
Magnitude	[3.04, 6.65]
Séisme min	1953
Séisme max	1887
Classes non représentées	[1, 3]

## Cas d'application : France, Alpes et Pyrénées

Loi de  $M^*$ 

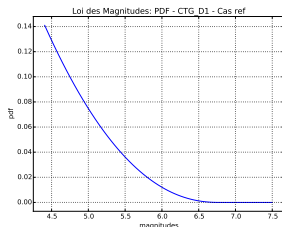
Les paramètres de la loi de Pareto obtenus sont :

$$(u, \sigma, \xi) = (1, 0.740, -0.312) \quad 1 - \hat{\zeta}_u^* = 0.896$$

Borne supérieure des magnitudes : Comme  $\xi < 0$ , le modèle de Pareto (modèle de Weibull) a une borne supérieure :

$$M_{sup} = u - \frac{\sigma}{\xi} = 6.78$$

La loi de la Magnitude est déterminée à partir de son quantile 0.90.



## Loi des magnitudes (Cas Ref)

Quantile $1 - \hat{\zeta}_u^* = 89.6\%$	$u = 4.41$
Quantile 90%	4.44
Quantile 95%	4.89
Quantile 99%	5.64

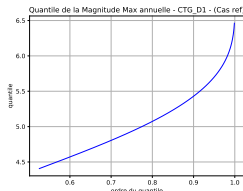
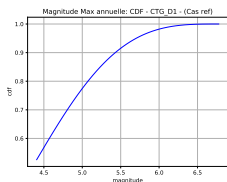
## Cas d'application : France, Alpes et Pyrénées

Loi de  $M_{max}$ 

La fonction de répartition de  $M_{max}$  s'écrit :

$$G(m) = e^{-\mu \zeta_u^* \bar{H}_{\sigma, \xi}(m-u)} \quad \forall m \geq u \quad \hat{\mu} = 5.46 \quad (9)$$

La loi de la Magnitude maximale annuelle est déterminée à partir de son quantile  $e^{-\mu \zeta_u^*} = 0.56$ .

Loi  $M_{max}$  (cas Ref)

Ordre minimal	56.5%
Quantile minimal	$u = 4.41$
Quantile 90%	5.38
Quantile 95%	5.66

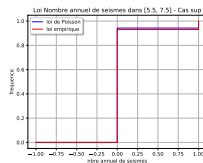
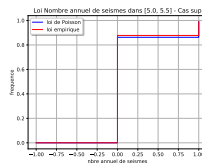
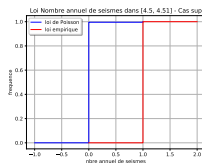
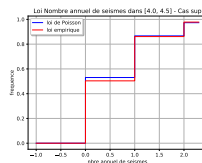
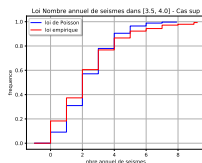
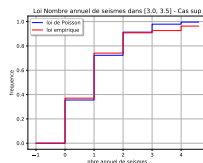
Période de retour centennale :  $m_{100} = 6.10$

Période de retour 200 ans :  $m_{200} = 6.23$

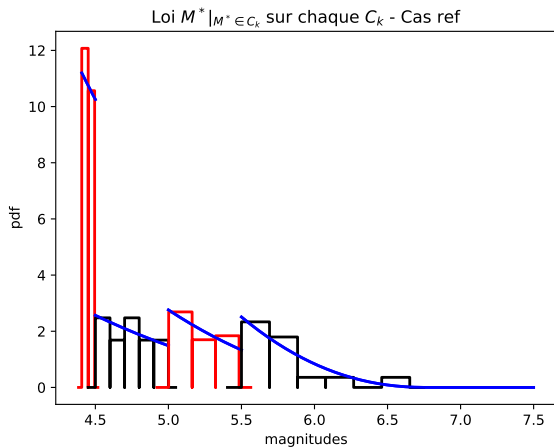
Remarque : On vérifie que les séismes de magnitude  $m \geq m_{100}$  sont apparus en 1644 ( $m = 6.15$ ), 1660 ( $m = 6.52$ ), 1750 ( $m = 6.13$ ), 1887 ( $m = 6.65$ ). Ceux de magnitude  $m \geq m_{200}$  sont apparus en 1660 ( $m = 6.52$ ), 1887 ( $m = 6.65$ ).

Attention : peu de données par rapport aux périodes de retour considérées.

## Validation graphique du modèle de Poisson



## Validation graphique du modèle de Pareto

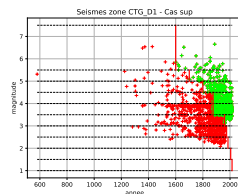
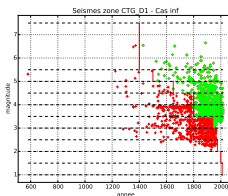
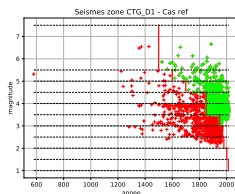




## Sensibilité aux dates de complétudes

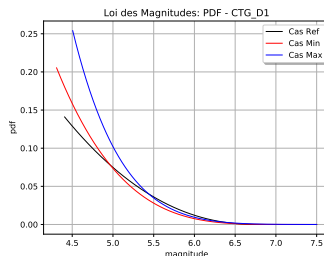
Les experts ne sont pas tous d'accord entre eux !

	Mesures brutes	Cas Ref	Cas Inf	Cas Sup
Nbre de séismes	1806	851	948	644
Dates	[580, 2016]	[1524, 2016]	[1428, 2016]	[1601, 2016]
Magnitude	[2.08, 6.65]	[3.04, 6.65]	[3.04, 6.65]	[3.14, 6.65]
Séisme max	1887	1887	1887	1887



## Loi des Magnitudes :

Loi Magnitude	Cas Ref	Cas Inf	Cas Sup
$1 - \zeta_u^* \%$	89.6%	87.5%	87.2%
$(u, \sigma, \xi)$	(4.41, 0.740, -0.312)	(4.31, 0.608, -0.210)	(4.51, 0.503, -0.136)
Quantile 90%	4.44	4.44	4.63
Quantile 95%	4.89	4.81	4.95
Quantile 99%	5.64	5.50	5.59
$M_{sup}^*$	6.78	7.21	8.19
$n_u$	189	197	154

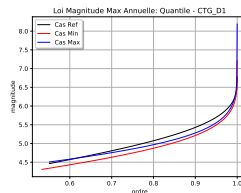
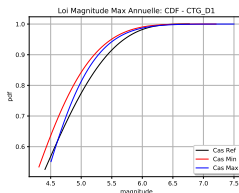
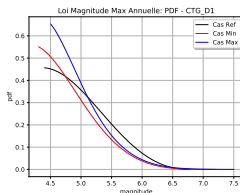


## Conclusion

L'estimation des quantiles extrêmes est relativement peu sensible aux dates de complétude. En revanche, l'estimation de la borne supérieure du support y est très sensible : rien de surprenant ! Il est préférable de baser l'étude des valeurs extrêmes de  $M^*$  sur ses quantiles et non sur  $M_{sup}^*$ .

## Loi de la Magnitude Maximale Annuelle :

Loi Magnitude Maximale Annuelle	Cas Ref	Cas Inf	Cas Sup
$\mu$	5.46	5.02	4.66
Ordre minimal	56.5%	53.4%	55.1%
Quantile min	4.41	4.31	4.51
Quantile 90%	5.38	5.21	5.28
Quantile 95%	5.66	5.49	5.55
<b>Période retour 100 ans (quant. 99%)</b>	<b>6.10</b>	<b>5.99</b>	<b>6.08</b>
<b>Période retour 200 ans (quant. 99.5%)</b>	<b>6.23</b>	<b>6.16</b>	<b>6.27</b>



## Conclusion

L'estimation des quantiles extrêmes est relativement peu sensible aux dates de complétude.

## Incertitudes statistiques

On se limite au **cas de Référence**. Les paramètres du modèle,  $(u, \sigma, \xi)$ ,  $\zeta_u^*$  et  $\mu$ , maximisent la vraisemblance du modèle.

### Technique du Bootstrap

Pour estimer la dispersion dans l'estimation de ces paramètres, on utilise la **technique du Bootstrap** : on génère plusieurs échantillons issus de la loi empirique de l'échantillon initial, à partir desquels on évalue les paramètres. On estime ensuite la dispersion des estimateurs en reconstruisant leur loi de manière non paramétrique (reconstruction à noyaux).

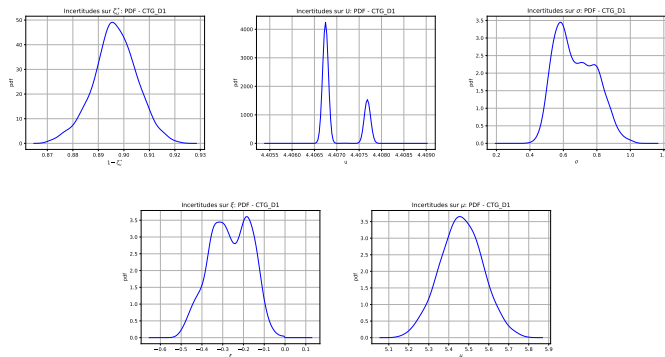
Paramétrages du Bootstrap :  $10^3$  échantillons de même taille que l'échantillon initial.

### Sensibilité aux données

On étudie l'impact de cette variabilité sur :

- l'estimation de  $(u, \sigma, \xi)$ ,  $\zeta_u^*$  et  $\mu$
- la loi de  $M_{max}$ , son quantile 0.9

## Lois marginales 1d des estimateurs

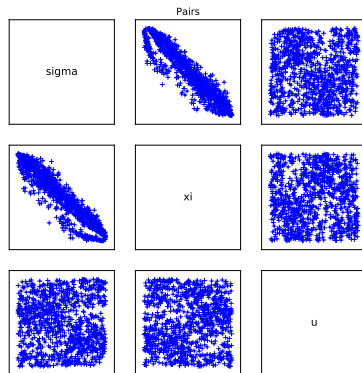
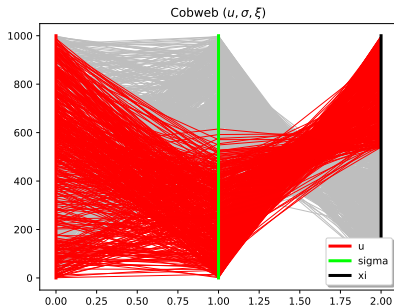


## Conclusions

- $\zeta_u^*$  principalement  $\simeq 0.9$  : on est bien dans le régime *extrême*
- 2 seuils  $u$  très proches
- Une valeur de  $\sigma$  principale  $\simeq 0.6$  et un plateau proche  $\simeq [0.65; 0.8]$
- 2 valeurs de  $\xi$  différentes :  $\simeq -0.33$  et  $\simeq -0.18$  : **peut avoir un gros impact !**
- $\mu$  principalement  $\simeq 5.45$

## Dépendances entre les estimateurs

Diagrammes Cobweb et Pairs (espace des rangs) :



## Conclusions

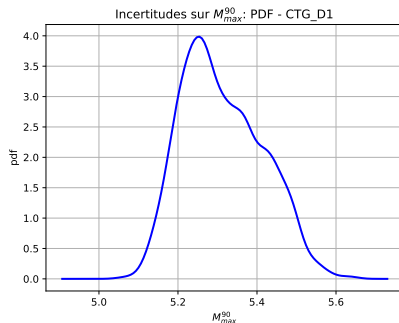
On constate une relation de :

- dépendance  $(\sigma, \xi)$
- indépendance apparente entre  $u$  et  $(\sigma, \xi)$

Quantile 0.9 de  $M_{max}$ Quantile 0.9 de  $M_{max}$ 

Il est sans doute plus stable de s'intéresser au **quantile 0.9 de la loi de  $M_{max}$**  plutôt qu'à sa borne supérieure.

Quantile 0.9 de $M_{max}$	Cas Ref
Moyenne	5.31
Ecart-type	0.10
Quantile 0.9	5.46
Valeur initiale	5.38
Quantile associé	0.73



## Conclusion

On constate une **relative stabilité du quantile 0.9 de  $M_{max}$** .

Remarque : Le quantile 0.9 de  $M_{max}$  a été estimé à 5.38 et a une probabilité 0.9 d'être inférieur à 5.46.

# Estimation de la Magnitude Maximale Annuelle par la Théorie des Valeurs Extrêmes

- ① Modélisations probabilistes
  - Séismes
  - Magnitude maximale annuelle
  - Magnitudes
- ② Estimation du modèle
  - Complétude des mesures de séismes
  - Processus Ponctuel de Poisson
  - Estimation du modèle
  - Validation et Sensibilité aux données
- ③ Application à la zone sismique Alpes - Pyrénées
  - Cas de référence
  - Validation
  - Sensibilité aux données de complétude
  - Sensibilité aux incertitudes statistiques
- ④ OpenTURNS
- ⑤ Conclusion



## OpenTURNPythonFunction

Utilisation de la *classe* *OpenTURNPythonFunction* :

- une **classe** : capitaliser des informations internes d'un appel à l'autre de la fonction.
- la classe *OpenTURNPythonFunction* : créer dans le monde python une classe qui peut être utilisée par la librairie C++ en tant que *Function*.

## Classe *logVraisemblanceGPD\_Completude*

- les données *fiabiles* sont stockées à l'initialisation de la classe
- les fonctions contraintes du problème d'optimisation sont créés à l'initialisation de la classe
- la fonction *vraisemblance* s'exprime à partir du triplet  $(u, \sigma, \xi)$

```
class logVraisemblanceGPD_Completude(OpenTURNPythonFunction):
    def __init__(self, sample, completudeDurees):
        # sample : date, mag
        self.sample_ = sample

        -----
        # creation de la fonction contrainte
        self.constraint_ = contrainteGPD_Completude(sample)
        -----

    def _exec(self, x):
        # (x[0], x[1], x[2]) = (sigma, xi, u)

# Corps de l etude
logVraisPP = logVraisemblanceGPD_Completude(donnees, completudeDurees)
logVraisFct = Function(logVraisPP)
```

## OptimizationProblem

Utilisation de la **classe OptimizationProblem** : définition du problème d'optimisation

- fonction à optimiser
- fonctions contraintes d'inégalités
- bornes

## BootstrapExperiment

Utilisation de la **classe BootstrapExperiment** : génération d'un échantillon de même taille

**Problème d'optimisation :**

```
optimPb = OptimizationProblem(logVraisFct)

optimPb.setInequalityConstraint(contFct)
optimPb.setBounds(bounds)
optimPb.setMinimization(False)

solver = CobyLa(optimPb)
solver.setStartingPoint(startPoint)
solver.setVerbose(verbose)

solver.run()
result = solver.getResult()
optimalPoint = result.getOptimalPoint()
```

**Bootstrap :**

```
# tirage avec remise d un echantillon
# de meme taille
experiment = BootstrapExperiment(donnees)
myExperiment = experiment.generate()
```

# Estimation de la Magnitude Maximale Annuelle par la Théorie des Valeurs Extrêmes

- ① Modélisations probabilistes
  - Séismes
  - Magnitude maximale annuelle
  - Magnitudes
- ② Estimation du modèle
  - Complétude des mesures de séismes
  - Processus Ponctuel de Poisson
  - Estimation du modèle
  - Validation et Sensibilité aux données
- ③ Application à la zone sismique Alpes - Pyrénées
  - Cas de référence
  - Validation
  - Sensibilité aux données de complétude
  - Sensibilité aux incertitudes statistiques
- ④ OpenTURNS
- ⑤ Conclusion

# Conclusion

## Situation complexe ...

- Point de vue **mathématique** : données et modèle probabiliste
- Point de vue **outil informatique** : pas de package ni dans R ni dans python (y compris OpenTURNS ...)

## mais pas désespérée ...

**OpenTURNS fournit des éléments haut niveau qui permettent de créer rapidement et facilement toute l'algorithmie nécessaire :**

- création du modèle probabiliste : fonction de vraisemblance
- inférence du modèle : optimisation sous contrainte
- analyse de la dispersion : technique du Bootstrap
- exploitation graphique
- niveau de performance suffisant pour mener à bien des explorations numériques