

Rapport de TFE

Processus de Hawkes pour l'aléa sismique



Adel EL HABAIZI

Table des matières

1 Remerciements	2
2 Introduction	2
3 Prérequis mathématiques	3
3.1 Les processus de Hawkes	3
3.1.1 Définition générale	3
3.1.2 Processus de Hawkes spatio-temporel marqué	3
3.2 Simulation par algorithme de branchement multidimensionnel	5
3.2.1 Vraisemblance du processus de Hawkes en dimension 2	6
3.2.2 Inférence des paramètres	9
4 Inférence et incertitudes sur données synthétiques	11
4.1 Bootstrap	11
4.2 Optimisation du noyau naturel / noyau non normalisé	12
5 Le modèle ETAS	13
5.1 Motivations et introduction	13
5.2 Simulation de la magnitude de chaque séisme	14
5.3 Simulation de loi d'Omori généralisée	15
5.4 Tirage du décalage spatial selon le noyau de l'équation (3)	16
6 Calcul détaillé de la vraisemblance pour un processus ETAS	19
6.1 Formule générale de la log-vraisemblance	19
6.2 Intégrale de l'intensité sur $[0, T] \times S$	19
6.3 Somme des densités $\log[\lambda(t_i, x_i, y_i)]$	20
6.4 Expression finale de la log-vraisemblance	21
6.5 Calcul du nombre moyen de répliques n	21
6.6 Log-vraisemblance dans le cas d'intensité de fond inhomogène	22
7 Application sur jeu de données réelles	24
7.1 Présentation des données et choix de modélisation	24
7.2 Estimation de la fonction intensité de fond	25
7.3 Optimisation de la vraisemblance	27
7.4 Log-vraisemblance sur paires et coût quasi linéaire	27
7.5 Intervalles de confiance à partir de la hessienne	28
7.6 Problème mal posé en raison de données incomplètes	29
7.7 Démarches finale du stage	30
7.8 Génération de données synthétiques	34
8 Conclusion et Perspectives	35

1 Remerciements

Je souhaite exprimer ma profonde gratitude à Merlin Keller et Joseph Muré, mes encadrants de stage, pour leur accompagnement constant, la qualité de leurs conseils, leur disponibilité et leur bienveillance, qui ont joué un rôle déterminant dans la réussite de ce travail. Je remercie également Ibrahim Seydi pour son implication, son suivi attentif et les échanges constructifs qui ont enrichi cette étude.

J'adresse aussi mes remerciements à l'ensemble de l'équipe du bâtiment S du département PRISME, pour leur accueil, leur soutien et les discussions scientifiques et humaines stimulantes qui ont marqué mon séjour.

Enfin, je tiens à remercier plus largement toutes les personnes qui, de près ou de loin, ont contribué à la bonne conduite de ce stage et à la progression de mes réflexions.

2 Introduction

Ce stage, réalisé au sein d'EDF R&D , s'inscrit dans un programme de recherche visant à améliorer la modélisation de l'aléa sismique par l'utilisation de processus stochastiques avancés. L'étude s'est focalisée sur les processus de Hawkes et leur variante appliquée à la sismologie, le modèle ETAS (Epidemic-Type Aftershock Sequence), afin de représenter finement la dynamique spatio-temporelle des séismes et de leurs répliques. L'enjeu était double : proposer une modélisation théorique rigoureuse intégrant la structure spatiale et temporelle des événements, et développer des outils numériques performants pour l'estimation des paramètres et la simulation d'événements synthétiques.

Les travaux ont comporté une première phase de formalisation mathématique, portant sur l'intensité conditionnelle, la vraisemblance et les méthodes d'optimisation adaptées, ainsi qu'une étude des choix de noyaux spatio-temporels et de leurs propriétés statistiques. Une seconde phase a été consacrée au développement d'algorithmes de simulation et d'estimation, incluant des techniques de bootstrap pour évaluer la robustesse des estimations, ainsi qu'à l'optimisation de la vraisemblance en présence d'intensités de fond homogènes et inhomogènes. Enfin, les méthodes ont été appliquées à un catalogue sismique couvrant la France métropolitaine et les pays limitrophes sur la période 1965–2016, avec une estimation de l'intensité de fond par lissage à noyau et une validation des modèles par comparaison des statistiques observées et simulées.

L'objectif de cette étude était de proposer une approche alternative pour l'évaluation de l'aléa sismique, en s'écartant des méthodes employées dans les travaux antérieurs. Ces derniers reposaient sur un prétraitement manuel du catalogue sismique, consistant à « déclusteriser » les données, c'est-à-dire à retirer les événements identifiés par les sismologues comme des répliques. Cette étape était nécessaire car une réplique n'est pas indépendante de sa secousse principale, ce qui viole la propriété d'indépendance des accroissements propre au processus de Poisson. Toutefois, cette méthode présente l'inconvénient majeur d'exiger un nettoyage manuel du catalogue, opération coûteuse en temps et susceptible d'introduire des biais. Une fois les données filtrées, l'étape suivante consistait à calibrer un processus de Poisson standard.

La présente étude adopte une approche différente, en mobilisant un processus de Hawkes, c'est-à-dire un processus ponctuel auto-excitant capable de modéliser explicitement l'occurrence des répliques. Ce choix permet donc d'intégrer directement les dépendances entre événements dans le modèle et d'éliminer la nécessité de déclusteriser le catalogue manuellement.

3 Prérequis mathématiques

3.1 Les processus de Hawkes

3.1.1 Définition générale

- **Processus de Hawkes univarié.** Un processus de Hawkes est un processus de comptage $(N_t)_{t \geq 0}$ *auto-excitant*, où chaque événement augmente, au moins temporairement, le taux d'occurrence des événements futurs. C'est donc un processus de comptage avec mémoire, offrant une modélisation *ipso facto* plus riche que celle d'un simple processus de Poisson, même inhomogène.
- **Définition générale de l'intensité conditionnelle.** Pour tout processus de comptage (N_t) défini sur un espace probabilisé filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, l'intensité conditionnelle $\lambda(t)$ dépend de l'historique $\mathcal{F}_t = \sigma\{N_s : s \leq t\}$. Elle se définit par

$$\lambda(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\mathbb{E}[N_{t+\Delta t} - N_t | \mathcal{F}_t]}{\Delta t}.$$

Cette définition formalise le taux instantané d'apparition des événements, conditionné à l'ensemble de l'historique \mathcal{F}_t .

- **Forme spécifique de l'intensité.** Dans le cas d'un processus de Hawkes univarié, l'intensité conditionnelle s'écrit

$$\begin{aligned}\lambda(t) &= \mu + \int_0^t \phi(t-s) dN_s \\ &= \mu + \sum_{t_i < t} \phi(t-t_i),\end{aligned}$$

où $\mu > 0$ est le taux de fond, $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est une fonction de mémoire qui caractérise l'influence des événements passés, et $(t_i)_{i \in \mathbb{N}}$ désigne la suite des instants d'occurrence des événements.

Il est essentiel de souligner que $\lambda(t)$ est une variable aléatoire, puisqu'elle dépend des instants de saut $(t_i)_{i \in \mathbb{N}}$ déjà réalisés. Contrairement à un processus de Poisson inhomogène où l'intensité est déterministe, le processus de Hawkes appartient à la classe des processus de comptage à intensité aléatoire : l'intensité est adaptée à la filtration $(\mathcal{F}_t)_{t \geq 0}$, reflétant ainsi la dépendance explicite au passé des événements.

3.1.2 Processus de Hawkes spatio-temporel marqué

- **Processus spatio-temporel marqué.** En pratique, notre processus qui modélisera les occurrences des séismes doit tenir compte de la position spatiale de ceux-ci ainsi que de l'intensité des séismes précédents. L'intensité stochastique ainsi définie dépend donc des positions spatiales et de la magnitude, il faut donc passer à un processus de Hawkes multi-dimensionnel et généraliser les résultats sans dimension spatiale. On considère un processus de comptage marqué N dont chaque événement est caractérisé par (x_i, y_i, t_i, m_i) dans l'espace \mathbb{R}^2 , le temps \mathbb{R}^+ et l'espace des magnitudes M .
- **Forme spécifique de l'intensité spatio-temporelle marquée.** Pour un processus de Hawkes à deux coordonnées d'espace (x, y) , temps t et magnitude m , l'intensité conditionnelle s'écrit :

$$\begin{aligned}\lambda(x, y, t, m) &= \mu(x, y, m) + \int_0^t \int_{\mathbb{R}^2} \int_M \phi(x-x', y-y', t-s, m-m') dN(x', y', s, m') \\ &= \mu(x, y, m) + \sum_{t_i < t} \phi(x-x_i, y-y_i, t-t_i, m-m_i)\end{aligned}$$

où

- $\mu(x, y, m) > 0$ est le taux de base, potentiellement spatial et dépendant de la magnitude ainsi qu'uniforme en temps (pas de dépendance en t);
- $\phi : \mathbb{R}^2 \times \mathbb{R}^+ \times M \rightarrow \mathbb{R}^+$ est le noyau d'auto-excitation spatio-temporel ;
- M désigne l'espace des magnitudes (typiquement [1 ;9] par exemple)
- Un des buts possibles est de designer les meilleures **familles de fonctions** paramétriques Φ possibles pour la modélisation de notre problème. Puis, le but est de trouver les meilleurs paramètres possibles par maximum de vraisemblance sur données réelles.

3.2 Simulation par algorithme de branchement multidimensionnel

Il nous faut trouver un moyen de simuler un processus de Hawkes, en tenant bien compte des particularités que nous avons énoncées plus tôt. Il existe plusieurs algorithmes qui simulent un processus de Hawkes comme l'algorithme d'Ogata basé sur le principe du Thinning ou l'algorithme d'inversion qui presuppose que la fonction d'intensité cumulée est inversible , mais nous avons choisi l'algorithme de branchement puisqu'il conserve la structure arborescente du processus de Hawkes, ce qui est plus intuitif que les deux autres méthodes de génération.

1. Échantillonnage des événements non-causés en dimension 2.

Tirer le nombre m d'événements "parents" dans $[0, T] \times S$ selon une loi de Poisson de paramètre :

$$m \sim \text{Poisson}(\mu \cdot |S| \cdot T)$$

où $\mu > 0$ est l'intensité de fond, T la durée d'observation, et $|S|$ l'aire de la région spatiale de détection $S \subset \mathbb{R}^2$.

Pour chaque séisme :

- Tirer le temps $t \sim \mathcal{U}([0, T])$,
- Tirer la position $(x, y) \sim \mathcal{U}(S)$, uniformément dans la région spatiale S ,
- (Facultatif selon le modèle) Tirer une magnitude ou tout autre attribut si nécessaire selon la loi considérée pour la magnitude.

2. Initialisation de la génération 0.

Poser $G_0 = \{ \text{tous les parents} \}$.

3. Générations des descendants.

Pour $i = 0, 1, 2, \dots$ tant que $|G_i| > 0$:

- (a) Pour chaque événement $(x_j, y_j, t_j, m_j) \in G_i$:

- i. Tirer le nombre de descendants k selon une loi de Poisson de paramètre

$$C = \int_{t_j}^T \int_{\mathbb{R}^2} \int_M \phi(\cdot) dx' dy' dm' ds.$$

- ii. Pour $r = 1, \dots, k$:

- Tirer un décalage temporel τ_r selon le noyau par rapport au temps **normalisé** $\nu_t(\tau)$ sur $(0, T - t_j)$.
- Tirer un déplacement spatial $(\Delta x_r, \Delta y_r)$ selon le noyau marginal **normalisé** $\nu_s(\Delta x, \Delta y)$.
- Tirer un écart de magnitude δm_r selon le noyau normalisé $\nu_m(\delta m)$.
- Poser l'événement descendant $(x_j + \Delta x_r, y_j + \Delta y_r, t_j + \tau_r, m_j + \delta m_r)$ et l'enregistrer dans $G_i^{(j)}$.

- (b) Poser

$$G_{i+1} = \bigcup_{j=1}^{|G_i|} G_i^{(j)},$$

puis ordonner chronologiquement G_{i+1} .

4. Construction du catalogue complet.

Fusionner toutes les générations :

$$G = \bigcup_{i \geq 0} G_i,$$

puis ordonner G par date et conserver la structure de branchement.

Dans cette version multidimensionnelle, le noyau $\phi(x, y, t, m)$ se décompose souvent en produits de

noyaux marginaux ν_t , ν_s et ν_m pour capturer respectivement la distribution des écarts temporels, des déplacements spatiaux et des écarts de magnitude. À chaque génération, chaque événement parent engendre un nombre de descendants selon l'intensité intégrale $C = \int_{t_j}^T \int_{\mathbb{R}^2} \int_M \phi(\cdot) dx' dy' dm' ds$.

Pour chacun de ces descendants, on effectue trois tirages indépendants :

- un tirage temporel $\tau \sim \nu_t$ pour le décalage en temps ;
- un tirage spatial $(\Delta x, \Delta y) \sim \nu_s$ pour le déplacement dans le plan ;
- un tirage de marque $\delta m \sim \nu_m$ pour l'évolution de la magnitude.

On obtient ainsi une simulation fidèle d'un processus de Hawkes marqué et spatio-temporel, où chaque dimension (espace, temps, magnitude) suit son propre noyau d'auto-excitation.

Nous avons simulé un tel arbre avec ledit algorithme, que voici :

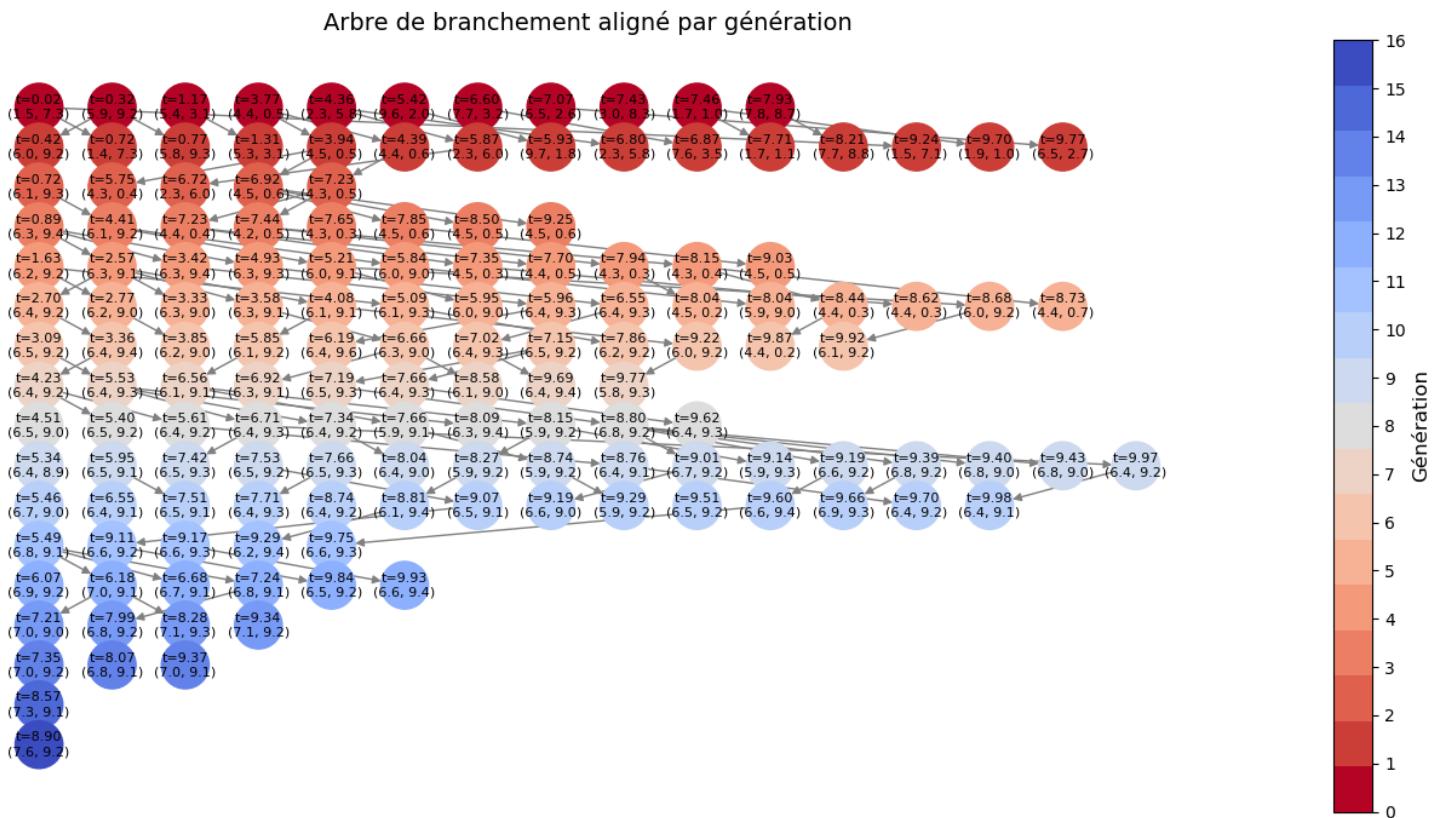


FIGURE 1 – Arbre de Branchement d'un Hawkes avec comme paramètres $\mu : 0.01$, $K : 0.7$, $\beta : 0.5$, $\sigma : 0.1$

3.2.1 Vraisemblance du processus de Hawkes en dimension 2

Soit un processus de Hawkes de dimension 1 défini sur l'intervalle de temps $[0, T]$, et soient $t_1 < t_2 < \dots < t_n$ les instants des n événements observés. On suppose que l'intensité conditionnelle du processus est donnée par :

$$\lambda(t) = \mu(t) + \sum_{t_i < t} \phi(t - t_i)$$

La principale difficulté du processus de Hawkes est que son intensité est une fonction aléatoire, le calcul de sa vraisemblance est plus compliqué que dans le cas d'un Poisson inhomogène. Soit le vecteur (T_1, T_2, \dots, T_n) des n premiers temps de sauts observés du processus de Hawkes. On peut alors écrire la loi de (T_1, T_2, \dots, T_n) comme étant :

$$[T_1, T_2, \dots, T_n] = [T_1][T_2|T_1][T_3|T_2, T_1] \dots [T_n|T_{n-1}, \dots, T_2, T_1]$$

Or, chacune de ces loi conditionnelle est celle du premier instant de saut d'un processus de Poisson inhomogène débutant à T_{i-1} . Ainsi, on a que

$$\begin{aligned}
 [T_1] &= \exp\left(-\int_0^{T_1} \lambda(u)du\right) \cdot \lambda(T_1) \\
 [T_2|T_1] &= \exp\left(\int_{T_1}^{T_2} \lambda(u|T_1)du\right) \cdot \lambda(T_2|T_1) \\
 [T_3|T_2, T_1] &= \exp\left(\int_{T_2}^{T_3} \lambda(u|T_2, T_1)du\right) \cdot \lambda(T_3|T_2, T_1) \\
 &\quad \dots \\
 [T_n|T_{n-1}, \dots, T_1] &= \exp\left(\int_{T_{n-1}}^{T_n} \lambda(u|T_{n-1}, \dots, T_1)du\right) \cdot \lambda(T_n|T_{n-1}, \dots, T_1)
 \end{aligned}$$

d'où , par produit, on a que

$$[T_1, \dots, T_n] = \left(\prod_{i=1}^n \lambda(T_i | \mathcal{F}_{T_i}) \right) \cdot \exp\left(-\int_0^{T_n} \lambda(u | \mathcal{F}_u) du\right) \cdot \mathbf{1}_{\{T_1 < T_2 < \dots < T_n\}}$$

Cette formule est valable en dimension une. En dimension 2 , on a la formule suivante :

$$[(T_1, X_1, Y_1), \dots, (T_n, X_n, Y_n)] = \left(\prod_{i=1}^n \lambda(T_i, X_i, Y_i | \mathcal{F}_{T_i}) \right) \cdot \exp\left(-\int_0^T \int_S \lambda(t, x, y | \mathcal{F}_t) dx dy dt\right) \cdot \mathbf{1}_{\{T_1 < \dots < T_n\}}$$

Calcul de la log-vraisemblance en dimension 2 (spatio-temporel, avec effets de bord).

On considère un processus de Hawkes spatio-temporel défini sur un domaine $[0, T] \times S \subset \mathbb{R} \times \mathbb{R}^2$, avec une intensité conditionnelle donnée par :

$$\lambda(t, x, y) = \mu + \sum_{t_j < t} g(t - t_j, x - x_j, y - y_j)$$

où (t_j, x_j, y_j) sont les événements précédents, et g est un noyau spatio-temporel.

Soient $(t_1, x_1, y_1), \dots, (t_n, x_n, y_n)$ les n événements observés. La log-vraisemblance est alors :

$$\log \mathcal{L}(\theta) = \sum_{i=1}^n \log \left(\mu + \sum_{t_j < t_i} g(t_i - t_j, x_i - x_j, y_i - y_j) \right) - \int_0^T \int_S \lambda(t, x, y) dx dy dt$$

On développe l'intégrale par linéarité :

$$\int_0^T \int_S \lambda(t, x, y) dx dy dt = \mu |S| T + \sum_{j=1}^n \int_{t_j}^T \int_S g(t - t_j, x - x_j, y - y_j) dx dy dt$$

Si le noyau g est de la forme :

$$g(t, x, y) = \alpha e^{-\beta t} \cdot \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

alors, sur un domaine spatial borné S , l'intégrale spatiale n'est plus égale à 1. On note :

$$I(x_j, y_j, \sigma, S) = \int_S \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x - x_j)^2 + (y - y_j)^2}{2\sigma^2}\right) dx dy$$

qui représente la masse de la gaussienne centrée en (x_j, y_j) restreinte au domaine S .

Alors :

$$\int_{t_j}^T \int_S g(t - t_j, x - x_j, y - y_j) dx dy dt = \alpha / \beta \cdot I(x_j, y_j, \sigma, S) \cdot (1 - e^{-\beta(T-t_j)})$$

Finalement, la log-vraisemblance devient :

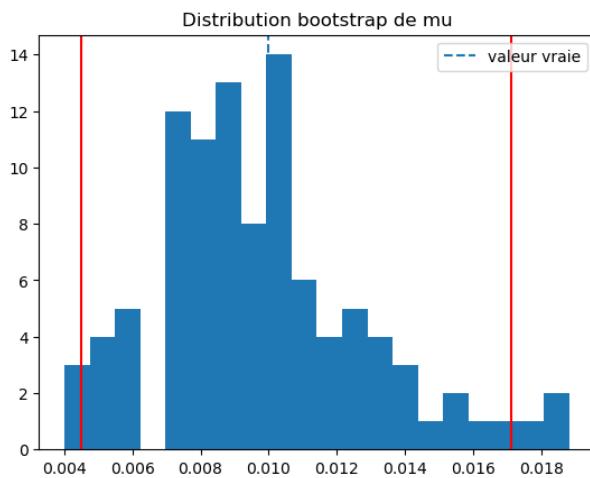
$$\begin{aligned} \log \mathcal{L}(\theta) &= \sum_{i=1}^n \log \left(\mu + \sum_{t_j < t_i} \alpha e^{-\beta(t_i - t_j)} \cdot \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2\sigma^2}\right) \right) \\ &\quad - \mu |S| T - \sum_{j=1}^n \frac{\alpha}{\beta} \cdot I(x_j, y_j, \sigma, S) \cdot (1 - e^{-\beta(T-t_j)}) \end{aligned}$$

Cette expression tient compte des effets de bord induits par le domaine spatial fini S . Elle est utilisée pour l'estimation des paramètres $(\mu, \alpha, \beta, \sigma)$ par maximum de vraisemblance.

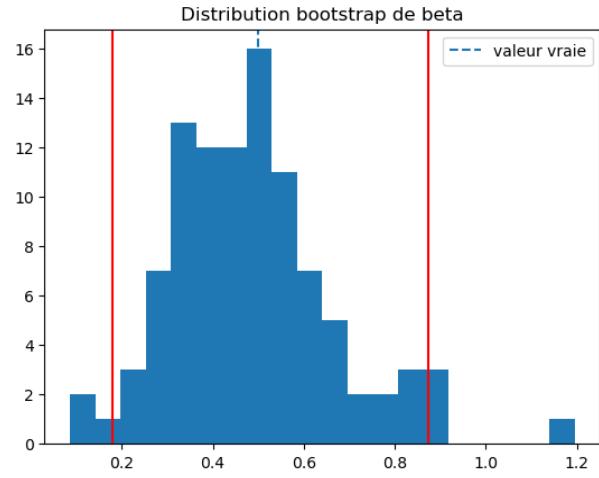
3.2.2 Inférence des paramètres

Une fois la log-vraisemblance écrite, il faut optimiser cette fonction par rapport aux paramètres des familles de fonctions qu'on a choisis, et en pratique on utilise un solveur qui va nous permettre de remonter à ces paramètres.

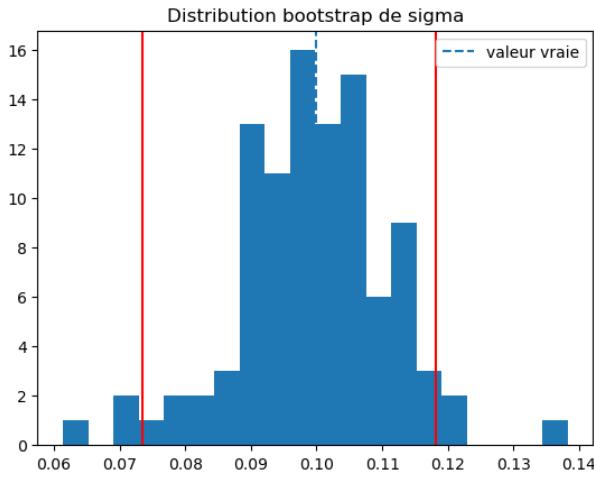
Nous avons décidé de faire du bootstrap et résoudre plusieurs fois le problème d'optimisation pour des arbres différents obtenus avec le même processus de Hawkes avec les mêmes paramètres, l'idée étant d'évaluer un peu la robustesse de notre méthode ainsi que la variation de l'estimateur du Maximum de Vraisemblance (MVE) par rapport à la vraie valeur de chacun des paramètres. Nous avons donc pu faire la simulation pour 100 arbres différents *a priori* issu d'un même processus de Hawkes, puis résolu 100 problèmes d'optimisation et obtenu une distribution sur les 4 paramètres que voici :



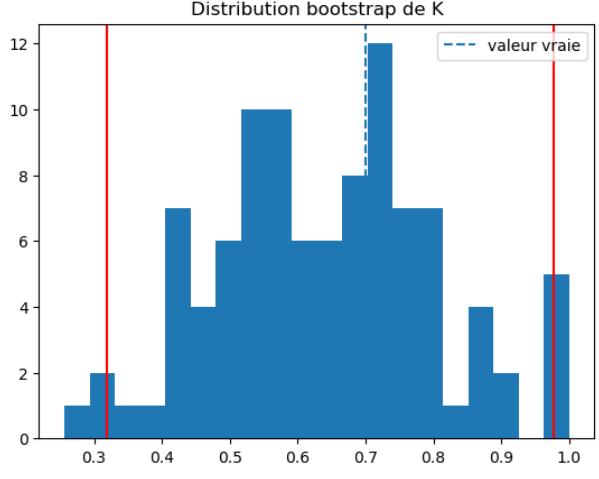
(a) Histogramme des estimations de mu



(b) Histogramme des estimations de beta



(c) Histogramme des estimations de sigma



(d) Histogramme des estimations de K

FIGURE 2 – Comparaison des histogrammes des estimations

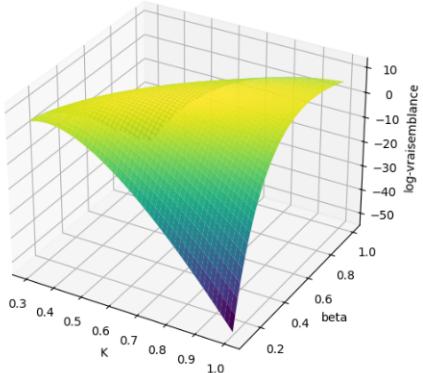
On a également obtenu des intervalles de confiance au seuil de 95 % que voici :

Nous avons ensuite essayé de voir a quoi ressemble la courbe de vraisemblance quand on fixe 2 paramètres parmi les 4 et en faisant varier les deux restants (il y a donc six courbes au total). Il apparaît qu'on a un problème concave dans les 6 cas possibles. On ne peut cependant pas conclure que le problème est concave de manière générale. Il est donc convexe conditionnellement mais pas concave (il faudrait le prouver en calculant la Hessienne en tout point et en montrant qu'elle est semi définie-négative)

Paramètre	Vraie valeur	Moyenne bootstrap	Écart-type bootstrap	CI95%
mu	0.0100	0.0098	0.0031	[0.0045, 0.0171]
K	0.7000	0.6395	0.1620	[0.3185, 0.9781]
beta	0.5000	0.4840	0.1810	[0.1791, 0.8748]
sigma	0.1000	0.0992	0.0114	[0.0734, 0.1182]

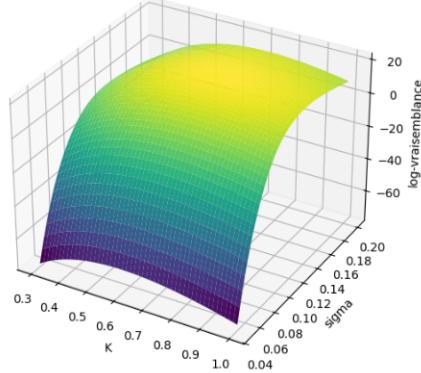
FIGURE 3 – Intervalles de confiance obtenu pour chacun des paramètres.

LL surface: varier K, beta



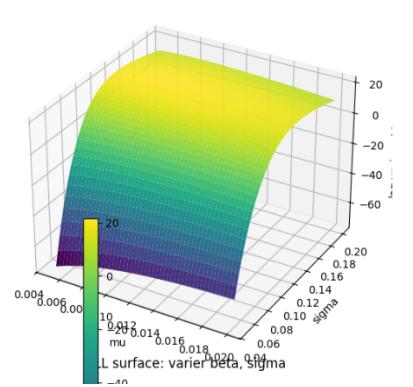
(a) $K-\beta$

LL surface: varier K, sigma



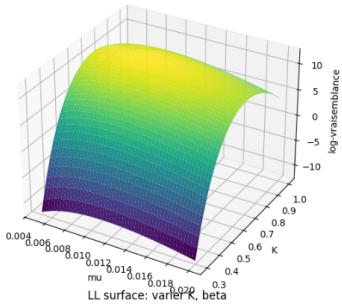
(b) $K-\sigma$

LL surface: varier mu, sigma



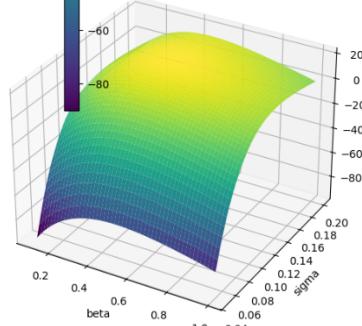
(c) $\mu-\sigma$

LL surface: varier mu, K



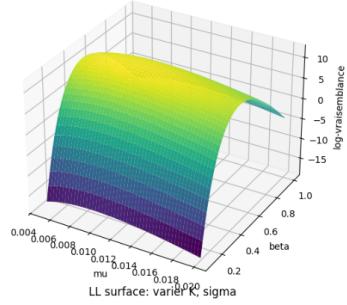
(d) $\mu-K$

LL surface: varier beta, sigma



(e) $\beta-\sigma$

LL surface: varier mu, beta



(f) $\mu-\beta$

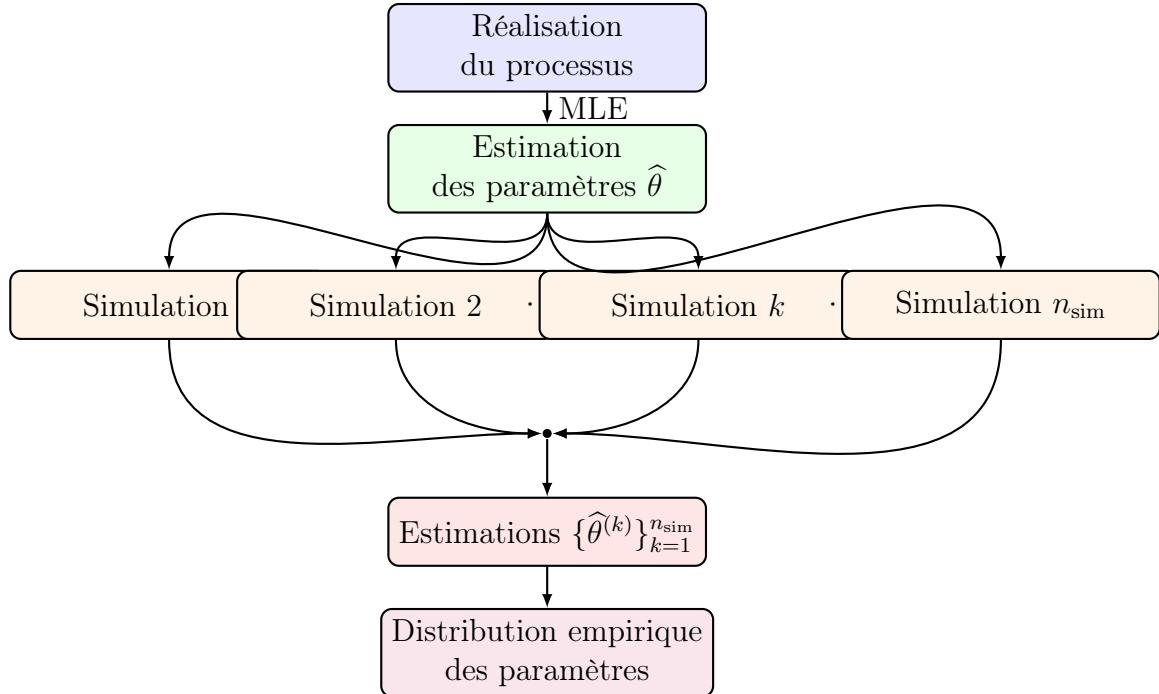
FIGURE 4 – Comparaison des différentes paires de paramètres

4 Inférence et incertitudes sur données synthétiques

4.1 Bootstrap

En réalité, on n'a pas accès à 100 simulations d'un processus mais plutôt à une seule réalisation d'un processus. Pourtant, on veut quand même tester la robustesse de notre estimation.

Ainsi, on va faire du *bootstrap*, c'est-à-dire qu'avec le processus initial, on va obtenir un jeu de paramètres estimés. Puis, avec ces mêmes paramètres estimés, on génère n_{sim} simulations. Pour chacune de ces simulations, on obtient un jeu de paramètres estimés et donc une distribution empirique des estimateurs.



On obtient par exemple les quatre distributions suivantes en suivant le principe énoncé plus haut :

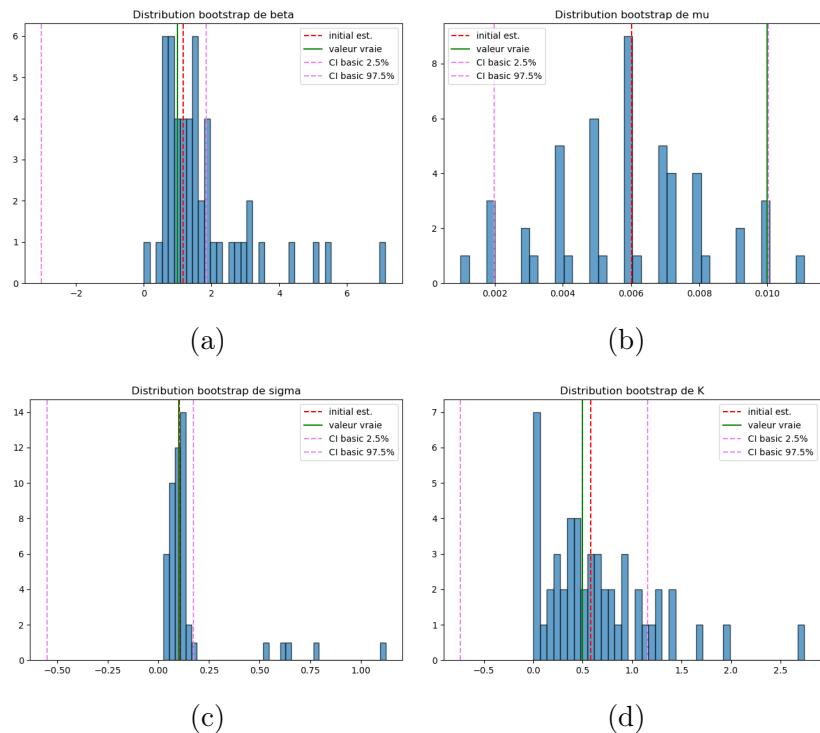


FIGURE 5

4.2 Optimisation du noyau naturel / noyau non normalisé

Il est apparu dans notre modélisation que nous disposions du choix du noyau de mémoire de la fonction de Hawkes à ajuster à nos observations. En effet, la forme du noyau impacte directement la fonction de vraisemblance et donc le problème d'optimisation ainsi que l'ensemble des paramètres estimés. Nous avons comparé deux formulations principales :

- Noyau exponentiel non normalisé :

$$g(t, x, y) = \alpha e^{-\beta t} \cdot \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

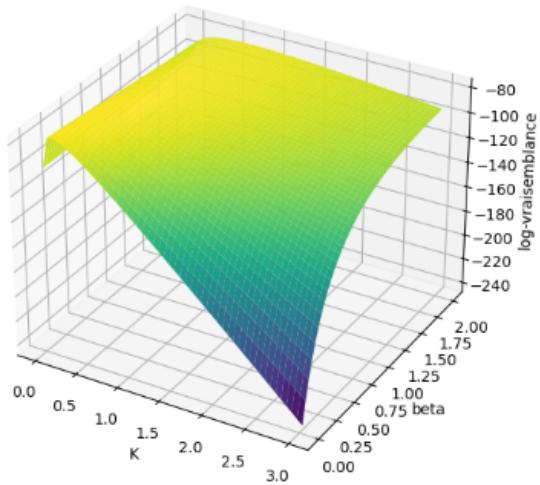
- Noyau naturel (normalisé) :

$$g_{norm}(t, x, y) = \alpha \beta e^{-\beta t} \cdot \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

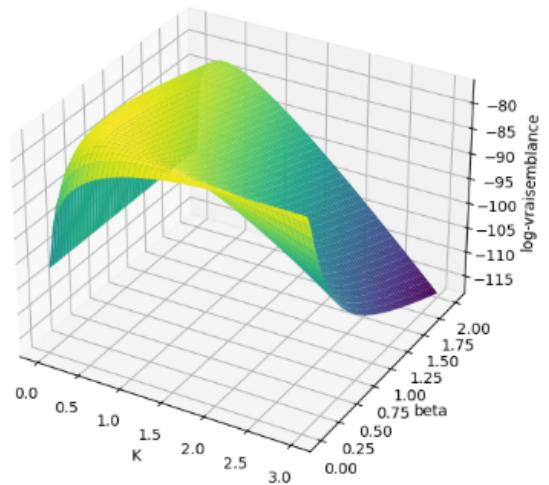
dont la normalisation vérifie : $\int_0^\infty g_{norm}(t) dt = 1$.

L'intérêt du noyau normalisé est que son intégrale en temps et en espace vaut $K \cdot I(x_j, y_j, \sigma, S) \cdot (1 - e^{-\beta(T-t_j)})$, qui est une quantité qui, pour un domaine spatial suffisamment grand et une densité suffisamment concentrée vaut K, c'est-à-dire que le nombre moyen de répliques pour une secousse principale vaut K. Ce paramètre important est seulement piloté par K alors que dans l'autre modélisation, il dépend également de β ce qui ne semble pas naturel comme choix de modélisation.

Cependant, cette modélisation en ne normalisant pas présente un avantage que le problème d'optimisation à résoudre est concave alors que celui normalisé ne l'est pas. Le solveur met en pratique entre 10 à 40 % de temps en moins à trouver les bons paramètres pour faire de l'inférence sur un processus de hawkes quelconque.



(a) vraisemblance non normalisée



(b) Vraisemblance normalisée

FIGURE 6 – Comparaison des vraisemblances pour le même processus

On fait de l'optimisation dans \mathbb{R}^4 ici donc il faut comprendre que l'optimisation est plus compliquée à droite qu'à gauche. On sait aussi qu'on balaie les mêmes familles de fonctions de mémoire pour le processus de Hawkes pour les deux choix de fonctions qu'on a choisis. Il a donc été très intéressant de choisir la fonction non normalisée, résoudre l'optimisation et obtenir les paramètres plus naturels comme le taux de reproduction (le K de g_{norm}) en posant le C^1 -difféomorphisme $\phi(x, y) = (x/y, y)$

5 Le modèle ETAS

5.1 Motivations et introduction

Le modèle ETAS (Epidemic-Type Aftershock Sequence) est un modèle utilisé dans plusieurs domaines comme l'épidémiologie, et qui va nous intéresser ici dans le cadre de la sismologie pour étudier le phénomène de répliques dans les séismes. On peut trouver par exemple une présentation détaillée dans cet article .

Contrairement aux tests précédents, où la fonction de mémoire décroissait de façon exponentielle dans le temps, la fonction de mémoire du processus de Hawkes sous-jacent au modèle ETAS adopte une forme différente. Elle présente en particulier la spécificité de dépendre de la magnitude : l'intensité mémoire est donc modulée par la magnitude du séisme survenu dans le processus de Hawkes. Cette caractéristique apporte un raffinement par rapport au modèle précédent, dans lequel la magnitude n'intervenait pas dans la dynamique de mémoire.

Ainsi, dans le modèle ETAS, l'intensité conditionnelle s'écrit traditionnellement :

$$\lambda(t, x, y \mid \mathcal{F}_t) = \mu + \sum_{i: t_i < t} g(t - t_i, x - x_i, y - y_i, m_i),$$

où :

- μ : taux de sismicité de fond (immigrants), supposé constant dans l'espace et le temps.
- $g(\cdot)$ est le *noyau de déclenchement* (triggering kernel), qui décrit comment chaque séisme (t_i, x_i, y_i, m_i) engendre des répliques dans le temps et dans l'espace.

Dans notre implémentation, nous avons choisi la forme factorisée suivante :

$$g(\tau, \Delta x, \Delta y, m_i) = K e^{\alpha(m_i - m_c)} \times \underbrace{(p-1)c^{p-1}(\tau+c)^{-p}}_{\text{composante temporelle Omori normalisée}} \times \underbrace{f_{\text{spat}}(\Delta x, \Delta y \mid m_i)}_{\text{composante spatiale}}$$

avec :

$$\tau = t - t_i, \quad \Delta x = x - x_i, \quad \Delta y = y - y_i, \quad m_i \text{ la magnitude du "parent".}$$

1. Composante temporelle (loi d'Omori généralisée) On adopte la forme classique :

$$f_t(\tau) = (p-1)c^{p-1}(\tau+c)^{-p}, \quad \tau \geq 0,$$

où

- $p > 1$ est l'exposant d'Omori (décroissance temporelle en puissance).
- $c > 0$ est le délai d'amortissement minimal (évite la divergence à $\tau = 0$).
- f est une densité de probabilité.

2. Composante spatiale (noyau de Papadopoulos–Zhuang) Le noyau spatial conditionnel à la magnitude m_i du parent se définit en coordonnées polaires par :

$$f_r(r \mid m_i) = \frac{2r}{d_i} (q-1) \left(1 + \frac{r^2}{d_i}\right)^{-q}, \quad r \geq 0,$$

où :

$$d_i = d \exp[\gamma(m_i - m_c)].$$

En coordonnées cartésiennes, si (x, y) est un point dans le plan, alors

$$r = \sqrt{(x - x_i)^2 + (y - y_i)^2}, \quad \varphi \sim \mathcal{U}(0, 2\pi),$$

et

$$f_{\text{spat}}(\Delta x, \Delta y \mid m_i) = \frac{q-1}{\pi d_i} \left[1 + \frac{\Delta x^2 + \Delta y^2}{d_i}\right]^{-q}.$$

Les constantes et paramètres sont :

- $d > 0$: échelle spatiale de référence (pour $m = m_c$).
- $\gamma \geq 0$: coefficient ajustant l'étendue spatiale en fonction de la magnitude du parent.
- $q > 1$: exposant contrôlant la décroissance spatiale (queue en puissance).
- m_c : magnitude de coupure inférieure (seuil de complétude).

Ainsi, la forme générale du noyau ETAS retenu s'écrit :

$$g(\tau, \Delta x, \Delta y, m_i) = K \exp[\alpha(m_i - m_c)] \times (p-1)c^{p-1}(\tau + c)^{-p} \times \frac{q-1}{\pi d_i} \left[1 + \frac{\Delta x^2 + \Delta y^2}{d_i} \right]^{-q},$$

avec $d_i = d e^{\gamma(m_i - m_c)}$.

Il s'agit ensuite de voir comment simuler des répliques avec une densité mémoire de cette forme.

5.2 Simulation de la magnitude de chaque séisme

Pour affecter une magnitude à chaque événement, il va falloir effectuer un tirage selon la loi de Gutenberg-Richter, qui est une référence dans le domaine de la sismologie. L'affectation de la magnitude est indépendante de la Soit m_c la magnitude de coupure inférieure, m_{\max} la magnitude de coupure supérieure, et β le paramètre de la loi de Gutenberg–Richter. On souhaite générer aléatoirement une magnitude m dans l'intervalle $[m_c, m_{\max}]$ selon la densité tronquée

$$s(m) = \begin{cases} \frac{\beta \exp[-\beta(m - m_c)]}{1 - \exp[-\beta(m_{\max} - m_c)]} & \text{si } m_c \leq m \leq m_{\max}, \\ 0, & \text{sinon.} \end{cases}$$

Nous allons détailler la méthode d'inversion de la CDF pour tirer m selon $s(m)$.

1. Calcul de la constante de normalisation La densité non tronquée sur $[m_c, +\infty)$ est

$$f_{\exp}(m) = \beta e^{-\beta(m - m_c)}, \quad m \geq m_c.$$

Son intégrale de $m = m_c$ à $m = +\infty$ vaut 1. Pour tronquer à m_{\max} , on définit :

$$C = \int_{m_c}^{m_{\max}} \beta e^{-\beta(u - m_c)} du.$$

Calcul de cette intégrale :

$$C = \left[-e^{-\beta(u - m_c)} \right]_{u=m_c}^{u=m_{\max}} = (-e^{-\beta(m_{\max} - m_c)}) - (-e^{-\beta(m_c - m_c)}) = 1 - e^{-\beta(m_{\max} - m_c)}.$$

La densité tronquée $s(m)$ s'écrit alors pour $m \in [m_c, m_{\max}]$:

$$s(m) = \frac{\beta e^{-\beta(m - m_c)}}{C} = \frac{\beta e^{-\beta(m - m_c)}}{1 - e^{-\beta(m_{\max} - m_c)}}.$$

2. Expression de la fonction de répartition (CDF) tronquée La fonction de répartition $F(m)$ associée à $s(m)$ est

$$F(m) = \int_{m_c}^m s(u) du = \frac{1}{C} \int_{m_c}^m \beta e^{-\beta(u - m_c)} du.$$

Or, comme précédemment :

$$\int_{m_c}^m \beta e^{-\beta(u - m_c)} du = \left[-e^{-\beta(u - m_c)} \right]_{u=m_c}^{u=m} = 1 - e^{-\beta(m - m_c)}.$$

Ainsi, pour $m_c \leq m \leq m_{\max}$,

$$F(m) = \frac{1 - e^{-\beta(m-m_c)}}{C} = \frac{1 - e^{-\beta(m-m_c)}}{1 - e^{-\beta(m_{\max}-m_c)}}.$$

On vérifie que $F(m_c) = 0$ et $F(m_{\max}) = 1$.

3. Inversion de la CDF Pour réaliser un tirage par inversion, on suit ces étapes :

1. On génère un nombre uniforme $U \sim \mathcal{U}(0, 1)$.
2. On pose $U = F(m)$. Autrement dit,

$$U = \frac{1 - e^{-\beta(m-m_c)}}{1 - e^{-\beta(m_{\max}-m_c)}}.$$

3. On résout cette équation pour m . En notant à nouveau $C = 1 - e^{-\beta(m_{\max}-m_c)}$, on a :

$$U = \frac{1 - e^{-\beta(m-m_c)}}{C} \implies U C = 1 - e^{-\beta(m-m_c)} \implies e^{-\beta(m-m_c)} = 1 - U C.$$

Puis, en prenant le logarithme népérien :

$$-\beta(m - m_c) = \ln[1 - U C] \implies m - m_c = -\frac{1}{\beta} \ln[1 - U C].$$

Finalement,

$$m = m_c - \frac{1}{\beta} \ln(1 - U [1 - e^{-\beta(m_{\max}-m_c)}]).$$

5.3 Simulation de loi d'Omori généralisée

Nous souhaitons générer un délai temporel $\Delta t \geq 0$ suivant la densité d'Omori généralisée, définie par :

$$f_t(\Delta t) = (p-1) c^{p-1} (\Delta t + c)^{-p}, \quad \Delta t \geq 0,$$

où $p > 1$ est l'exposant et $c > 0$ est un décalage minimal. Nous allons détailler la méthode d'inversion permettant de tirer Δt selon cette densité.

1. Calcul de la fonction de répartition (CDF) On définit la fonction de répartition

$$F(\Delta t) = \int_0^{\Delta t} f_t(u) du = \int_0^{\Delta t} (p-1) c^{p-1} (u+c)^{-p} du.$$

Pour calculer cette intégrale, posons la substitution $v = u + c$. Lorsque $u = 0$, $v = c$. Lorsque $u = \Delta t$, $v = \Delta t + c$. De plus, $du = dv$. On obtient :

$$F(\Delta t) = (p-1) c^{p-1} \int_{v=c}^{v=\Delta t+c} v^{-p} dv.$$

Or, pour $p \neq 1$,

$$\int v^{-p} dv = \frac{v^{-p+1}}{-p+1} = -\frac{v^{-p+1}}{p-1}.$$

Ainsi,

$$\begin{aligned} \int_c^{\Delta t+c} v^{-p} dv &= \left[-\frac{v^{-p+1}}{p-1} \right]_{v=c}^{v=\Delta t+c} = -\frac{(\Delta t + c)^{-p+1}}{p-1} - \left(-\frac{c^{-p+1}}{p-1} \right) \\ &= \frac{c^{-p+1} - (\Delta t + c)^{-p+1}}{p-1}. \end{aligned}$$

En réinjectant dans $F(\Delta t)$:

$$F(\Delta t) = (p-1) c^{p-1} \times \frac{c^{-p+1} - (\Delta t + c)^{-p+1}}{p-1} = c^{p-1} [c^{-p+1} - (\Delta t + c)^{-p+1}].$$

On remarque que $c^{p-1} c^{-p+1} = 1$. Par conséquent,

$$F(\Delta t) = 1 - (\Delta t + c)^{-p+1} c^{p-1}.$$

Or $(\Delta t + c)^{-p+1} c^{p-1} = \left(\frac{c}{\Delta t + c} \right)^{p-1}$. Finalement :

$$F(\Delta t) = 1 - \left(\frac{c}{\Delta t + c} \right)^{p-1}, \quad \Delta t \geq 0.$$

On vérifie que $F(0) = 0$ et $\lim_{\Delta t \rightarrow +\infty} F(\Delta t) = 1$.

2. Tirage par inversion Pour tirer Δt selon f_t , on procède en deux étapes :

1. Tirer $U \sim \mathcal{U}(0, 1)$.
2. Poser $U = F(\Delta t)$, c'est-à-dire

$$U = 1 - \left(\frac{c}{\Delta t + c} \right)^{p-1}.$$

On en déduit

$$1 - U = \left(\frac{c}{\Delta t + c} \right)^{p-1} \implies \frac{c}{\Delta t + c} = (1 - U)^{\frac{1}{p-1}} \implies \Delta t + c = c (1 - U)^{-\frac{1}{p-1}}.$$

Finalement,

$$\Delta t = c \left[(1 - U)^{-\frac{1}{p-1}} - 1 \right], \quad U \sim \mathcal{U}(0, 1).$$

5.4 Tirage du décalage spatial selon le noyau de l'équation (3)

Nous voulons simuler un décalage (dx, dy) autour d'un événement-parent de coordonnées (x_j, y_j) et de magnitude m_j , en respectant exactement la forme du noyau spatial donnée par l'équation (3) de l'article :

$$f(x - x_j, y - y_j, m_j) = \frac{q-1}{\pi d e^{\gamma(m_j-m_c)}} \left[1 + \frac{(x - x_j)^2 + (y - y_j)^2}{d e^{\gamma(m_j-m_c)}} \right]^{-q}, \quad (x, y) \in \mathbb{R}^2,$$

avec les paramètres suivants :

m_c : magnitude de coupure inférieure,

$d > 0$: échelle spatiale de référence (pour $m_j = m_c$),

$\gamma \geq 0$: coefficient qui étend la portée spatiale selon m_j ,

$q > 1$: exposant qui contrôle la décroissance en queue de distribution.

1. Réécriture en coordonnées polaires

Soit

$$\Delta x = x - x_j, \quad \Delta y = y - y_j, \quad r = \sqrt{\Delta x^2 + \Delta y^2}.$$

On pose également

$$d_j := d \exp[\gamma(m_j - m_c)].$$

Alors le noyau s'écrit

$$f(\Delta x, \Delta y | m_j) = \frac{q-1}{\pi d_j} \left[1 + \frac{r^2}{d_j} \right]^{-q}, \quad (\Delta x, \Delta y) \in \mathbb{R}^2.$$

On vérifie que cette densité est normalisée dans le plan : $\iint_{\mathbb{R}^2} f(\Delta x, \Delta y | m_j) d\Delta x d\Delta y = 1$.

2. Passage à la densité radiale $f_r(r)$

En coordonnées polaires, l'élément d'aire s'écrit $d\Delta x d\Delta y = r dr d\varphi$. La densité marginale de r s'obtient donc par :

$$f_r(r | m_j) = \int_0^{2\pi} f(\Delta x = r \cos \varphi, \Delta y = r \sin \varphi | m_j) r d\varphi.$$

Puisque $f(\Delta x, \Delta y | m_j)$ est constant en φ pour un r donné, on a

$$f_r(r | m_j) = (2\pi) \frac{q-1}{\pi d_j} \left[1 + \frac{r^2}{d_j} \right]^{-q} \times r = \frac{2(q-1)}{d_j} r \left[1 + \frac{r^2}{d_j} \right]^{-q}, \quad r \geq 0.$$

On vérifie facilement que $\int_0^\infty f_r(r | m_j) dr = 1$.

3. Calcul de la fonction de répartition $F_r(r)$

Nous calculons la CDF :

$$F_r(r) = \int_0^r \frac{2(q-1)}{d_j} u \left[1 + \frac{u^2}{d_j} \right]^{-q} du.$$

Posons le changement de variable

$$t = 1 + \frac{u^2}{d_j}, \quad dt = \frac{2u}{d_j} du, \quad \text{quand } u=0, t=1, \quad \text{quand } u=r, t=1+\frac{r^2}{d_j}.$$

Alors

$$F_r(r) = (q-1) \int_{t=1}^{t=1+\frac{r^2}{d_j}} t^{-q} dt = (q-1) \left[\frac{t^{-q+1}}{-q+1} \right]_{t=1}^{t=1+\frac{r^2}{d_j}} = 1 - \left[1 + \frac{r^2}{d_j} \right]^{-(q-1)}.$$

Ainsi, pour $r \geq 0$,

$$F_r(r) = 1 - \left(1 + \frac{r^2}{d_j} \right)^{-(q-1)}.$$

4. Tirage par inversion

Pour obtenir un échantillon r selon $f_r(r)$, on procède comme suit :

1. On tire $U \sim \mathcal{U}(0, 1)$.
2. On pose

$$U = F_r(r) = 1 - \left(1 + \frac{r^2}{d_j} \right)^{-(q-1)} \implies 1 - U = \left(1 + \frac{r^2}{d_j} \right)^{-(q-1)}.$$

3. On en déduit

$$1 + \frac{r^2}{d_j} = (1 - U)^{-\frac{1}{q-1}} \implies r^2 = d_j \left[(1 - U)^{-\frac{1}{q-1}} - 1 \right],$$

donc

$$r = \sqrt{d_j \left[(1 - U)^{-1/(q-1)} - 1 \right]}, \quad U \sim \mathcal{U}(0, 1).$$

5. Passage en cartésien Une fois r simulé, on tire un angle $\varphi \sim \mathcal{U}(0, 2\pi)$ et l'on pose

$$dx = r \cos \varphi, \quad dy = r \sin \varphi.$$

Le point enfant se positionne alors en $(x_j + dx, y_j + dy)$.

6 Calcul détaillé de la vraisemblance pour un processus ETAS

Soit $\{(t_i, x_i, y_i, m_i)\}_{i=1}^N$ la réalisation observée d'un processus spatio-temporel ETAS sur la fenêtre $[0, T] \times S$, où

$$S = [x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}], \quad |S| = (x_{\max} - x_{\min})(y_{\max} - y_{\min}).$$

Les paramètres du modèle sont

$$\theta = (\mu, K, \alpha, p, c, q, d, \gamma),$$

avec :

- $\mu > 0$: taux de fond, constant (par unité d'aire et de temps),
- $K > 0, \alpha \geq 0$: productivité dépendant de la magnitude,
- $p > 1, c > 0$: paramètres de la loi d'Omori généralisée,
- $q > 1, d > 0, \gamma \geq 0$: paramètres du noyau spatial de Papadopoulos–Zhuang,
- $m_c < m_i \leq m_{\max}$: magnitudes tronquées (loi de Gutenberg–Richter).

L'intensité conditionnelle spatio-temporelle du processus s'écrit :

$$\lambda(t, x, y | \mathcal{H}_{t^-}) = \mu + \sum_{t_j < t} g(t - t_j, x - x_j, y - y_j, m_j),$$

avec, pour chaque parent j :

$$g(\tau, \Delta x, \Delta y, m_j) = K e^{\alpha(m_j - m_c)} \times (p - 1) c^{p-1} (\tau + c)^{-p} \times \frac{q - 1}{\pi d_j} \left[1 + \frac{\Delta x^2 + \Delta y^2}{d_j} \right]^{-q},$$

où

$$\tau = t - t_j, \quad (\Delta x, \Delta y) = (x - x_j, y - y_j), \quad d_j = d \exp[\gamma(m_j - m_c)].$$

6.1 Formule générale de la log-vraisemblance

La vraisemblance d'un processus à intensité $\lambda(t, x, y)$ observé sur $[0, T] \times S$ est :

$$\mathcal{L}(\theta) = \exp \left(- \iint_{[0, T] \times S} \lambda(u, x, y) dx dy du \right) \times \prod_{i=1}^N \lambda(t_i, x_i, y_i).$$

Nous privilégions la log-vraisemblance :

$$\log \mathcal{L}(\theta) = - \iint_{[0, T] \times S} \lambda(u, x, y) dx dy du + \sum_{i=1}^N \log [\lambda(t_i, x_i, y_i)].$$

Nous détaillons chacun de ces deux termes.

6.2 Intégrale de l'intensité sur $[0, T] \times S$

$$\iint_{[0, T] \times S} \lambda(u, x, y) dx dy du = \underbrace{\iint_{[0, T] \times S} \mu dx dy du}_{\text{(A) terme fond}} + \underbrace{\iint_{[0, T] \times S} \sum_{t_j < u} g(u - t_j, x - x_j, y - y_j, m_j) dx dy du}_{\text{(B) terme auto-excitation}}$$

(A) Terme fond

$$\iint_{[0, T] \times S} \mu dx dy du = \mu \times \left(\int_0^T du \right) \times \left(\int_S dx dy \right) = \mu T |S|.$$

(B) Terme auto-excitation On échange somme et intégrale :

$$\begin{aligned} & \iint_{0 \leq u \leq T, (x,y) \in S} \sum_{t_j < u} g(u - t_j, \Delta x, \Delta y, m_j) dx dy du \\ &= \sum_{j=1}^N \int_{u=t_j}^T \int_{(x,y) \in S} g(u - t_j, x - x_j, y - y_j, m_j) dx dy du. \end{aligned}$$

Pour un j fixé, posons $\tau = u - t_j$ (donc $\tau \in [0, T - t_j]$). On a :

$$I_j = \int_{\tau=0}^{T-t_j} \int_{(x,y) \in S} K e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (\tau + c)^{-p} \times \frac{q-1}{\pi d_j} \left[1 + \frac{(x-x_j)^2 + (y-y_j)^2}{d_j} \right]^{-q} dx dy d\tau.$$

Les dépendances en τ et en (x, y) se factorisent :

$$I_j = K e^{\alpha(m_j - m_c)} \times \underbrace{\int_0^{T-t_j} (p-1) c^{p-1} (\tau + c)^{-p} d\tau}_{A_j} \times \underbrace{\int_{(x,y) \in S} \frac{q-1}{\pi d_j} \left[1 + \frac{(x-x_j)^2 + (y-y_j)^2}{d_j} \right]^{-q} dx dy}_{B_j}.$$

— Calcul de A_j :

$$A_j = \int_0^{T-t_j} (p-1) c^{p-1} (\tau + c)^{-p} d\tau = \left[-c^{p-1} (\tau + c)^{-p+1} \right]_{\tau=0}^{\tau=T-t_j} = 1 - \left(\frac{c}{T - t_j + c} \right)^{p-1}.$$

— Calcul de B_j :

$$B_j = \iint_{(x,y) \in S} \frac{q-1}{\pi d_j} \left[1 + \frac{(x-x_j)^2 + (y-y_j)^2}{d_j} \right]^{-q} dx dy,$$

qui est la probabilité qu'une réplique née au point (x_j, y_j) tombe à l'intérieur de S .

— Si $S = \mathbb{R}^2$, alors $B_j = 1$.

— Sinon, on évalue numériquement B_j (par quadrature 2D ou via une fonction spatial_integral).

En conclusion :

$$I_j = K e^{\alpha(m_j - m_c)} \left[1 - \left(\frac{c}{T - t_j + c} \right)^{p-1} \right] B_j.$$

Donc :

$$\iint_{[0,T] \times S} \sum_{t_j < u} g(\cdot) dx dy du = \sum_{j=1}^N K e^{\alpha(m_j - m_c)} \left[1 - \left(\frac{c}{T - t_j + c} \right)^{p-1} \right] B_j.$$

Ainsi, l'intégrale totale d'intensité se réécrit :

$$\iint_{[0,T] \times S} \lambda(u, x, y) dx dy du = \underbrace{\mu |S| T}_{\text{(fond)}} + \underbrace{\sum_{j=1}^N K e^{\alpha(m_j - m_c)} \left[1 - (c/(T - t_j + c))^{p-1} \right] B_j}_{\text{(auto-excitation)}}.$$

6.3 Somme des densités $\log[\lambda(t_i, x_i, y_i)]$

Pour chaque événement $i = 1, \dots, N$, l'intensité au point (t_i, x_i, y_i) vaut :

$$\lambda(t_i, x_i, y_i) = \mu + \sum_{j < i} K e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (t_i - t_j + c)^{-p} \times \frac{q-1}{\pi d_j} \left[1 + \frac{(x_i - x_j)^2 + (y_i - y_j)^2}{d_j} \right]^{-q}.$$

Posons la contribution du parent j à l'événement i :

$$G_{j \rightarrow i} = K e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (t_i - t_j + c)^{-p} \times \frac{q-1}{\pi d_j} \left[1 + \frac{(x_i - x_j)^2 + (y_i - y_j)^2}{d_j} \right]^{-q}.$$

Alors

$$\lambda(t_i, x_i, y_i) = \mu + \sum_{j < i} G_{j \rightarrow i}.$$

Le deuxième terme de la log-vraisemblance est donc :

$$\sum_{i=1}^N \log[\lambda(t_i, x_i, y_i)] = \sum_{i=1}^N \log \left[\mu + \sum_{j < i} G_{j \rightarrow i} \right].$$

6.4 Expression finale de la log-vraisemblance

En combinant l'intégrale d'intensité calculée au paragraphe précédent et la somme des log-intensités, on obtient :

$$\log \mathcal{L}(\theta) = - \underbrace{\left[\mu |S| T + \sum_{j=1}^N K e^{\alpha(m_j - m_c)} \left[1 - (c/(T - t_j + c))^{p-1} \right] B_j \right]}_{\text{(fond)}} + \underbrace{\sum_{i=1}^N \log \left[\mu + \sum_{j < i} G_{j \rightarrow i} \right]}_{\text{(événements)}}.$$

Plus explicitement :

$$\begin{aligned} \log \mathcal{L}(\theta) = & -\mu |S| T - \sum_{j=1}^N K e^{\alpha(m_j - m_c)} \left[1 - \left(\frac{c}{T - t_j + c} \right)^{p-1} \right] B_j \\ & + \sum_{i=1}^N \log \left[\mu + \sum_{j < i} K e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (t_i - t_j + c)^{-p} \times \frac{q-1}{\pi d_j} \left(1 + \frac{(x_i - x_j)^2 + (y_i - y_j)^2}{d_j} \right)^{-q} \right]. \end{aligned}$$

où :

$$d_j = d e^{\gamma(m_j - m_c)}, \quad B_j = \iint_{(x,y) \in S} \frac{q-1}{\pi d_j} \left[1 + \frac{(x-x_j)^2 + (y-y_j)^2}{d_j} \right]^{-q} dx dy.$$

Cette expression fournit la vraisemblance exacte du processus ETAS observé sur $[0, T] \times S$. Pour l'estimation des paramètres par maximum de vraisemblance, on maximise $\log \mathcal{L}(\theta)$ par rapport à $\theta = (\mu, K, \alpha, p, c, q, d, \gamma)$, en évaluant numériquement les B_j et les sommes $\sum_{j < i} G_{j \rightarrow i}$.

6.5 Calcul du nombre moyen de répliques n

Dans le cadre du modèle ETAS, on sait que la distribution des magnitudes des répliques suit une loi de Gutenberg-Richter tronquée, exprimée ainsi :

$$s(m) = \frac{\beta e^{-\beta(m-m_c)}}{1 - e^{-\beta(m_{\max} - m_c)}} \quad \text{pour } m \in [m_c, m_{\max}]$$

Par ailleurs, le nombre moyen de répliques directement déclenchées par un événement de magnitude m est donné par :

$$k(m) = K e^{\alpha(m - m_c)}$$

On obtient alors le nombre moyen total de répliques (appelé souvent productivité moyenne) par :

$$n = \int_{m_c}^{m_{\max}} k(m) s(m) dm$$

Insérant les expressions de $k(m)$ et $s(m)$, on a :

$$n = \int_{m_c}^{m_{\max}} A e^{\alpha(m-m_c)} \cdot \frac{\beta e^{-\beta(m-m_c)}}{1 - e^{-\beta(m_{\max}-m_c)}} dm$$

Ce qui peut être simplifié en :

$$n = \frac{A\beta}{1 - e^{-\beta(m_{\max}-m_c)}} \int_{m_c}^{m_{\max}} e^{(\alpha-\beta)(m-m_c)} dm$$

On procède au changement de variable $u = m - m_c$, ce qui donne :

$$n = \frac{A\beta}{1 - e^{-\beta(m_{\max}-m_c)}} \int_0^{m_{\max}-m_c} e^{(\alpha-\beta)u} du$$

Deux cas sont à distinguer :

Cas 1 : $\alpha \neq \beta$

Dans ce cas, l'intégrale s'évalue directement :

$$\int_0^{m_{\max}-m_c} e^{(\alpha-\beta)u} du = \frac{e^{(\alpha-\beta)(m_{\max}-m_c)} - 1}{\alpha - \beta}$$

D'où :

$$n = \frac{A\beta}{1 - e^{-\beta(m_{\max}-m_c)}} \cdot \frac{e^{(\alpha-\beta)(m_{\max}-m_c)} - 1}{\alpha - \beta}$$

Cas 2 : $\alpha = \beta$

Dans ce cas limite, on a :

$$\int_0^{m_{\max}-m_c} e^{0 \cdot u} du = m_{\max} - m_c$$

Et donc :

$$n = \frac{A\beta(m_{\max} - m_c)}{1 - e^{-\beta(m_{\max}-m_c)}}$$

La formule du nombre moyen de répliques n s'écrit donc :

$$n = \begin{cases} \frac{A\beta}{1 - e^{-\beta(m_{\max}-m_c)}} \cdot \frac{e^{(\alpha-\beta)(m_{\max}-m_c)} - 1}{\alpha - \beta} & \text{si } \alpha \neq \beta \\ \frac{A\beta(m_{\max} - m_c)}{1 - e^{-\beta(m_{\max}-m_c)}} & \text{si } \alpha = \beta \end{cases}$$

6.6 Log-vraisemblance dans le cas d'intensité de fond inhomogène

Dans le cas où la fonction d'intensité de fond ne dépend que de l'espace, on pose

$$\mu(x, y) = \mu kde(x, y),$$

où $kde(x, y)$ est une densité spatiale sur S normalisée telle que $\iint_S kde(x, y) dx dy = 1$, avec la constante μ estimée de sorte que

$$\int_{[0,T] \times S} \mu(x, y) du dx dy = \mu \times T \hat{=} N,$$

et N le nombre total d'événements observés.

En recalculant la vraisemblance, on obtient :

$$\begin{aligned} \log \mathcal{L}(\theta) \hat{=} & -N - \sum_{j=1}^N K e^{\alpha(m_j - m_c)} \left[1 - \left(\frac{c}{T-t_j+c} \right)^{p-1} \right] B_j \\ & + \sum_{i=1}^N \log \left[\mu(x_i, y_i) + \sum_{j < i} K e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (t_i - t_j + c)^{-p} \times \frac{q-1}{\pi d_j} \left(1 + \frac{(x_i - x_j)^2 + (y_i - y_j)^2}{d_j} \right)^{-q} \right]. \end{aligned}$$

7 Application sur jeu de données réelles

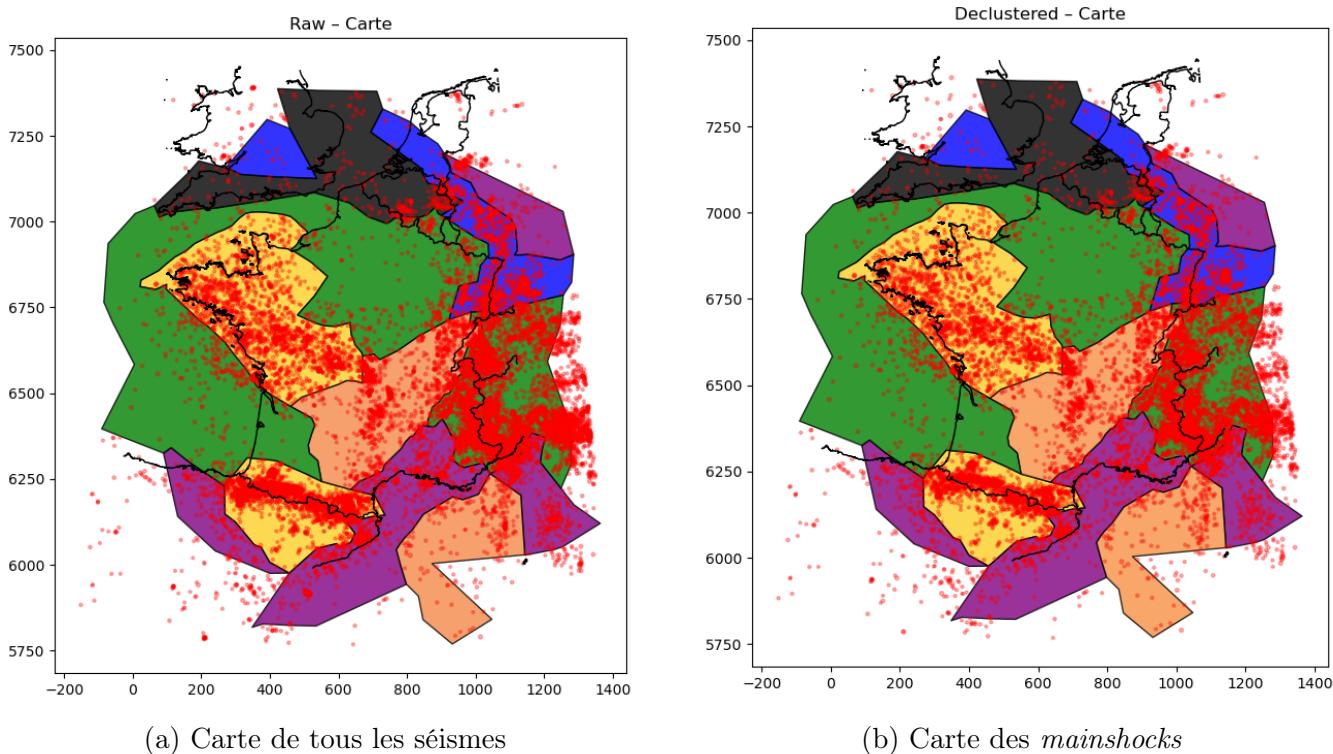
7.1 Présentation des données et choix de modélisation

Les données utilisées pour calibrer par maximum de vraisemblance le modèle ETAS sont décrites dans cette section.

Le choix du catalogue sismique à utiliser pour estimer un modèle ETAS a fait l'objet d'une réflexion approfondie. Dans un premier temps, nous avons envisagé de restreindre notre étude à la France métropolitaine, en supposant une intensité de fond uniforme sur un domaine rectangulaire. Les différentes quantités nécessaires à l'estimation étaient alors calculées sur ce domaine S , de manière à prendre en compte les effets de bord.

Nous avons également exploré une approche fondée sur les zones sismologiques définies par les experts, dans laquelle un processus ETAS indépendant, avec ses propres paramètres (a priori différents), est supposé se réaliser dans chaque zone. Toutefois, cette option a été écartée. En effet, le calcul des intégrales nécessaires sur des surfaces polygonales, ainsi que la gestion des bordures entre zones, s'est révélé trop complexe au regard du gain incertain en précision.

Par ailleurs, une modélisation réaliste de l'aléa sismique en France requiert l'inclusion des séismes survenant dans les régions frontalières. Négliger les événements des Pyrénées espagnoles ou des Alpes italiennes introduirait des effets de bord susceptibles de biaiser les estimations des paramètres du modèle. C'est ce constat qui nous a conduits à retenir un catalogue couvrant à la fois la France métropolitaine et les pays limitrophes, pour la période 1965–2016.



Quatre grandes stratégies de modélisation avaient donc été envisagées :

- Modèle ETAS unique avec intensité de fond homogène ;
- Modèle ETAS par zones avec intensité de fond homogène ;
- Modèle ETAS unique avec intensité de fond inhomogène ;
- Modèle ETAS par zones avec intensité de fond inhomogène.

Chaque approche présente ses avantages et ses limites. Par exemple, le modèle par zones avec intensité de fond inhomogène est potentiellement le plus précis, mais également le plus complexe à mettre en œuvre. Il suppose le calcul d'intégrales sur des polygones, la vérification de l'appartenance d'une réplique à une zone, et ne prend pas en compte les interactions entre zones voisines. Toutefois,

il permet de refléter la réalité selon laquelle les paramètres de réPLICATION peuvent varier d'une zone à l'autre, ce que ne permet pas le modèle ETAS unique.

L'analyse empirique des ratios entre le nombre total de séismes et le nombre de *mainshocks* dans chaque zone montre que ces ratios varient significativement. Cela suggère que certaines régions génèrent davantage de répliques que d'autres — une observation probablement liée à des mécanismes physiques spécifiques, encore mal identifiés. Il est néanmoins peu plausible que ces différences soient très marquées, et les paramètres de réPLICATION restent globalement similaires d'une zone à l'autre.

En revanche, supposer que l'intensité de fond est homogène sur l'ensemble du territoire constitue une hypothèse irréaliste. Elle ignore la structure tectonique du territoire et conduit à une mauvaise modélisation de la distribution spatiale des séismes. C'est pourquoi l'approche utilisant un modèle ETAS unique avec intensité de fond homogène a été écartée.

En conclusion, pour concilier réalisme physique et faisabilité numérique dans un temps raisonnable, tout en limitant les effets de bord, nous avons choisi de calibrer un **modèle ETAS unique avec intensité de fond inhomogène**.

7.2 Estimation de la fonction intensité de fond

Nous allons commencer par trouver la fonction intensité de fond sur tout l'espace considéré.

Nous avons pu faire cela car les sismologues peuvent, à partir d'un catalogue sismique, séparer à la main les séismes qui sont des mainshocks et les séismes qui sont des répliques.

Il aurait en effet été impossible d'estimer cette intensité de fond inhomogène sur le catalogue complet qui contient les répliques. Nous avons donc utilisé le catalogue déclusterisé pour estimer cette fonction .

Pour estimer cette fonction d'intensité de fond, nous avons eu recours à la méthode de lissage par noyau, connue sous le nom de *Kernel Density Estimation* (KDE). Cette méthode permet d'obtenir une estimation non paramétrique de la densité spatiale sous-jacente à une série de points localisés.

Classiquement, la méthode KDE fournit une estimation d'une densité spatiale de probabilité. Or, dans notre cas, nous cherchons à estimer une intensité de fond $\mu(x, y)$, c'est-à-dire une fonction qui donne, en chaque point de l'espace, le taux d'occurrence moyen des mainshocks par unité de surface et par unité de temps.

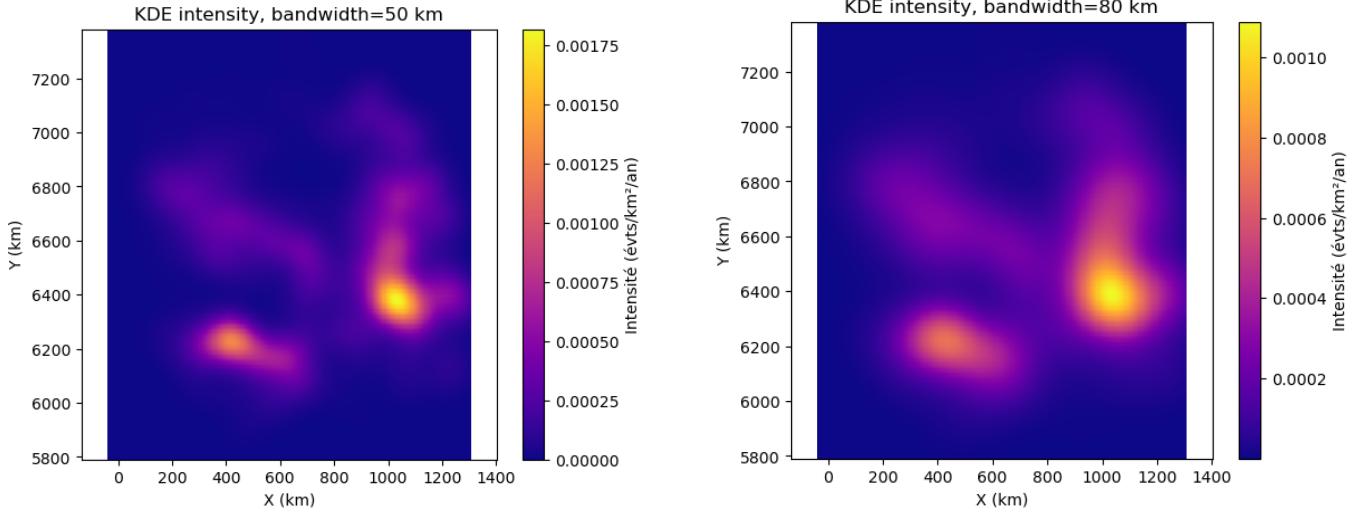
Soient (x_i, y_i) les positions des n mainshocks du catalogue déclusterisé, observés sur une durée totale T . L'estimateur de l'intensité s'écrit alors :

$$\hat{\mu}(x, y) = \frac{1}{Th^2} \sum_{i=1}^n K\left(\frac{x - x_i}{h}, \frac{y - y_i}{h}\right)$$

où :

- T est la durée d'observation (en années ou en jours) ;
- h est la bande passante (en kilomètres) ;
- K est un noyau bidimensionnel normalisé (souvent gaussien) ;
- le facteur $1/h^2$ provient du changement d'échelle dans le cas bidimensionnel.

Cette expression peut se comprendre comme une KDE classique, multipliée par le facteur n/T , qui représente le taux moyen de séismes de fond observés par unité de temps. Ainsi, l'intégrale de $\hat{\mu}(x, y)$ sur tout l'espace donne bien n/T , c'est-à-dire le nombre moyen de séismes principaux par an (ou par jour).



(a) Intensité estimée, bande passante 50 km

(b) Intensité estimée, bande passante 80 km

FIGURE 8 – Comparaison des estimations d'intensité avec deux largeurs de bande différentes.

En ce qui concerne le paramètre h , il contrôle le niveau de lissage de l'estimation :

- Un h trop faible suit de près les points de données, révélant de petites structures spatiales mais amplifiant le bruit (forte variance).
- Un h trop élevé produit une carte très lissée, stable statistiquement (faible variance) mais susceptible d'effacer des zones de concentration sismique (biais important).

Le choix optimal de h se fait donc en équilibrant la réduction du bruit et la préservation des véritables contrastes géologiques. Nous avons observé qu'une valeur de h entre 50 et 100km donnait des cartes d'intensité cohérente avec la sismicité des zones comme les Alpes, les Pyrénées ou le nord de la France.

7.3 Optimisation de la vraisemblance

Une fois la fonction d'intensité de fond $\hat{\mu}(x, y)$ estimée à partir du catalogue déclusterisé, nous disposons de tous les éléments nécessaires pour optimiser les autres paramètres du modèle ETAS par maximum de vraisemblance. Le fait de calibrer μ séparément présente un avantage notable : cela supprime un paramètre de la phase d'optimisation, réduisant ainsi la dimension du problème et la complexité numérique.

Dans sa forme naïve, le calcul de la vraisemblance ETAS implique une double somme sur l'ensemble des paires (i, j) d'événements avec $t_i > t_j$, ce qui entraîne un coût algorithmique en $\mathcal{O}(N^2)$, N étant le nombre d'événements considérés. Pour de grands catalogues, ce coût devient prohibitif.

Un compromis doit donc être trouvé entre :

- un N suffisamment grand pour que le maximum de vraisemblance soit bien identifié et que les paramètres estimés soient stables,
- un N suffisamment limité pour que le calcul et l'optimisation (y compris l'évaluation des gradients) restent rapides et réalisables.

Dans notre étude, ce compromis a conduit à retenir des fenêtres temporelles de 5 à 10 ans, contenant typiquement entre 2000 et 4000 événements. Cette plage permet d'obtenir des estimations robustes tout en conservant des temps de calcul acceptables.

7.4 Log-vraisemblance sur paires et coût quasi linéaire

Nous avons aussi fait le choix de réduire le coût de calcul en ne calculant les fonctions mémoire (celles faisant intervenir les $t_j - t_i$) que sur des paires de points pas trop distants en temps et en espace (typiquement une année en temps et 100km en espace) , les indices de ces paires de points étant calculés au début une bonne fois pour toutes. C'est assez cohérent comme choix puisqu'un point disons j assez loin en temps et ou en espace d'un point aura une influence négligeable ce qui fait qu'on ne sommerra que des zéros dans le calcul de la vraisemblance, c'est donc du temps perdu....

On ne conserve donc que les couples pertinents dans une fenêtre mémoire temporelle

$$\mathcal{P} = \{(i, j) : 0 < t_i - t_j \leq A\},$$

(on peut ajouter un cutoff spatial, mais ce n'est pas nécessaire ici). On **pré-calcule une fois** les vecteurs d'indices $(i_k, j_k)_{k=1}^M$ ainsi que $\Delta t_k = t_{i_k} - t_{j_k}$ et $r_k^2 = (x_{i_k} - x_{j_k})^2 + (y_{i_k} - y_{j_k})^2$, et les valeurs de fond $\mu_i = \mu(x_i, y_i)$.

Le terme parent→enfant est

$$G_{j \rightarrow i}(\theta) = K e^{\alpha(m_j - m_c)} (p - 1) c^{p-1} (t_i - t_j + c)^{-p} \frac{q - 1}{\pi d_j} \left(1 + \frac{r_{ij}^2}{d_j}\right)^{-q}, \quad d_j = d e^{\gamma(m_j - m_c)}.$$

La log-vraisemblance (à constante près en $(K, \alpha, p, c, q, d, \gamma)$) s'écrit

$$\ell(\theta) = - \sum_{j=1}^N K e^{\alpha(m_j - m_c)} \left[1 - \left(\frac{c}{T - t_j + c} \right)^{p-1} \right] + \sum_{i=1}^N \log \left(\mu_i + \sum_{(i,j) \in \mathcal{P}} G_{j \rightarrow i}(\theta) \right).$$

L'avantage de cette écriture est qu'évidemment le cardinal de \mathcal{P} sera beaucoup plus petit que N , ce qui accélère considérablement le temps de calcul de la log-vraisemblance, et accélère considérablement l'optimisation.

7.5 Intervalles de confiance à partir de la hessienne

Dans le cadre du maximum de vraisemblance, on note $\hat{\theta}$ l'estimateur obtenu en maximisant la log-vraisemblance $\ell(\theta)$. Sous des conditions régulières, le théorème de l'information de Fisher stipule que $\hat{\theta}$ est asymptotiquement distribué selon une loi normale :

$$\hat{\theta} \stackrel{\text{approx.}}{\sim} \mathcal{N}(\theta^*, I(\theta^*)^{-1}),$$

où $I(\theta^*)$ est la matrice d'information de Fisher, définie par :

$$I(\theta) = -\mathbb{E} [\nabla_\theta^2 \ell(\theta)].$$

En pratique, $I(\theta^*)$ est inconnu et on l'approche par la **hessienne observée** :

$$\widehat{I}(\hat{\theta}) = -\nabla_\theta^2 \ell(\theta) \Big|_{\theta=\hat{\theta}}.$$

L'estimation de la variance asymptotique de $\hat{\theta}$ est alors donnée par :

$$\widehat{\text{Var}}(\hat{\theta}) = \widehat{I}(\hat{\theta})^{-1}.$$

Pour un paramètre scalaire θ_k , on obtient l'intervalle de confiance asymptotique à $100(1 - \alpha)\%$:

$$\hat{\theta}_k \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\theta}_k)},$$

où $z_{1-\alpha/2}$ est le quantile de la loi normale standard (par exemple $z_{0.975} \approx 1.96$ pour un intervalle à 95%).

Cette approche, appelée *intervalle de Wald*, est valable lorsque N est suffisamment grand et que les hypothèses de régularité du maximum de vraisemblance sont vérifiées. Cela a l'air d'être le cas pour nous car les fonctions

7.6 Problème mal posé en raison de données incomplètes

Dans certains cas, l'estimation de certains paramètres temporels rencontre un problème dans le sens où le solveur nous renvoie $c = 0$ et $p = +\infty$. Cela arrive car nos données ne sont pas assez précises en temps, il arrive très souvent, selon le catalogue que l'on utilise, que des séismes soient confondus. En effet, selon le catalogue, la précision temporelle est au jour ou à la seconde près ce qui peut entraîner que des séismes arrivent au même moment.

Rappelons l'expression du terme de contribution :

$$G_{j \rightarrow i}(\theta) = K e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (t_i - t_j + c)^{-p} \frac{q-1}{\pi d_j} \left(1 + \frac{r_{ij}^2}{d_j}\right)^{-q}, \quad d_j = d e^{\gamma(m_j - m_c)}.$$

Ainsi, si pour un couple (i, j) on a $t_i = t_j$, on obtient

$$G_{j \rightarrow i}(\theta) = K e^{\alpha(m_j - m_c)} (p-1) c^{p-1} c^{-p} \frac{q-1}{\pi d_j} \left(1 + \frac{r_{ij}^2}{d_j}\right)^{-q}.$$

Ce qui se simplifie en

$$G_{j \rightarrow i}(\theta) = K e^{\alpha(m_j - m_c)} \frac{p-1}{c} \frac{q-1}{\pi d_j} \left(1 + \frac{r_{ij}^2}{d_j}\right)^{-q}.$$

On voit alors que :

1. Si $c = 0$, le terme devient *singulier*, c'est-à-dire $G_{j \rightarrow i}(\theta) \rightarrow +\infty$ dès qu'un couple (t_i, t_j) coïncide.
2. Même lorsque $c > 0$ est fixé, on remarque que $G_{j \rightarrow i}(\theta)$ croît linéairement en p (via le facteur $(p-1)/c$). Dans la log-vraisemblance :

$$\ell(\theta) = - \sum_{j=1}^N K e^{\alpha(m_j - m_c)} \left[1 - \left(\frac{c}{T-t_j+c} \right)^{p-1} \right] + \sum_{i=1}^N \log \left(\mu_i + \sum_{(i,j) \in \mathcal{P}} G_{j \rightarrow i}(\theta) \right),$$

chaque terme $\log(\mu_i + G_{j \rightarrow i}(\theta))$ devient alors de plus en plus grand lorsque $p \rightarrow +\infty$.

3. Par conséquent, la maximisation de $\ell(\theta)$ pousse toujours vers $p \rightarrow +\infty$ pour exploiter cette divergence, et ce même si c est fixé à une valeur strictement positive.

La présence de temps égaux $t_i = t_j$ entraîne une singularité structurelle du modèle ETAS. Fixer $c > 0$ empêche seulement la divergence immédiate, mais ne supprime pas la tendance de l'optimisation à envoyer p vers $+\infty$ à cause du comportement de $G_{j \rightarrow i}(\theta)$. Ceci est fondamental puisque par construction d'un processus de Saut, il est presque sûrement impossible d'avoir deux sauts en même temps. Le calcul de la vraisemblance repose implicitement sur ce principe, puisqu'on calcule la vraisemblance d'un processus qui ne peut pas avoir plus d'un saut au même moment. Ce principe n'étant pas respecté, on calcule donc la vraisemblance pour un processus qui ne peut pas être un modèle ETAS, il aboutit donc à ce que l'estimation par maximum de vraisemblance est faussée. On pourrait penser qu'une précision à la seconde suffit, mais en vérifiant, il y avait 5 séismes qui arrivaient à la seconde près au même moment donc il a été nécessaire dans notre catalogue sismique d'être précis à la dixième de seconde près.

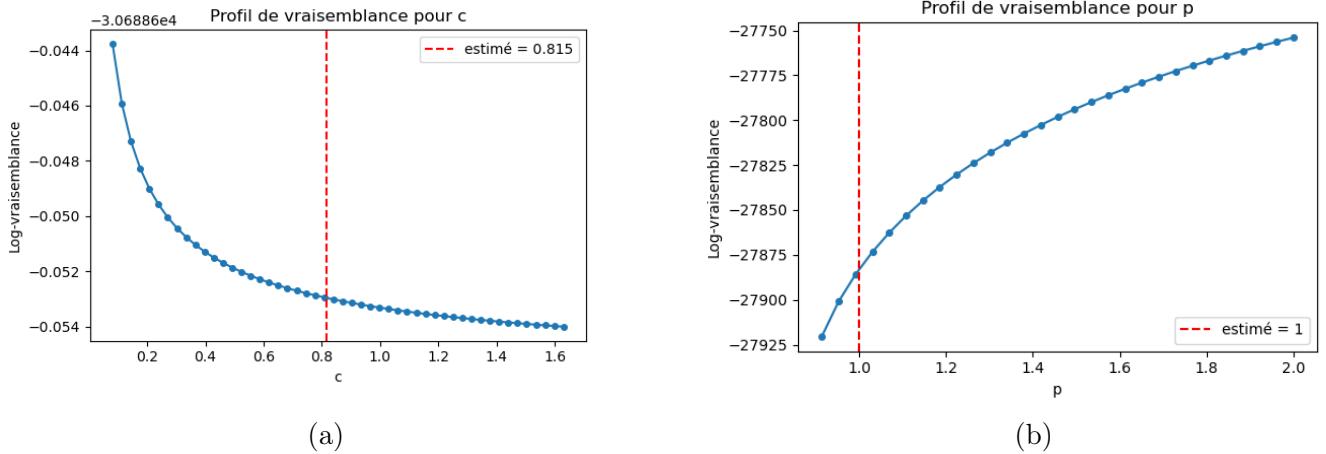


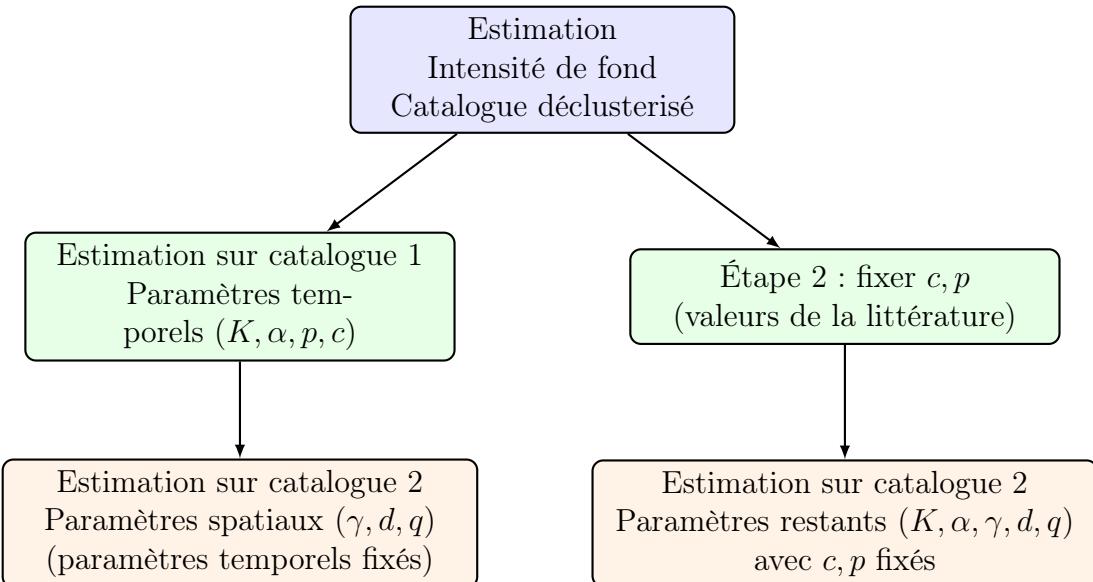
FIGURE 9 – Vraisemblance maximale pour $c = 0$ et $p = +\infty$

7.7 Démarches finales du stage

Afin de surmonter les difficultés évoquées dans la section précédente, nous avons retenu une démarche en trois étapes, qui constitue l'aboutissement de ce stage. L'idée générale est de décomposer l'estimation des paramètres du modèle ETAS en blocs cohérents, en exploitant des jeux de données adaptés à chaque niveau de précision.

- **Étape 1 : Intensité de fond.** Nous commençons par estimer séparément l'intensité de fond, à partir d'un catalogue déclusterisé et défini sur une surface géographique plus vaste que la seule France. Cette étape permet de réduire la complexité de l'optimisation ultérieure.
- **Étape 2 : Paramètres temporels.** Sur un catalogue sismique centré sur la France métropolitaine et disposant d'une précision temporelle à la dixième de seconde, nous estimons les paramètres temporels (K, α, p, c). À ce stade, nous considérons le processus uniquement comme fonction du temps et de la magnitude, et nous adaptons la vraisemblance en conséquence.
- **Étape 3 : Paramètres spatiaux.** Enfin, nous utilisons un catalogue élargi spatialement (mais moins précis temporellement). Nous réintroduisons la composante spatiale dans le modèle et estimons les paramètres spatiaux restants, tout en fixant les paramètres temporels obtenus à l'étape précédente.

Cette approche progressive permet de réduire le nombre de paramètres libres à chaque étape et d'améliorer la stabilité numérique de l'optimisation.



Pour mieux visualiser les catalogues qu'on a utilisé, on peut afficher en bleu le catalogue qu'on a utilisé pour l'estimation des paramètres temporels, et en rouge celui utilisé pour l'estimation des paramètres spatiaux.

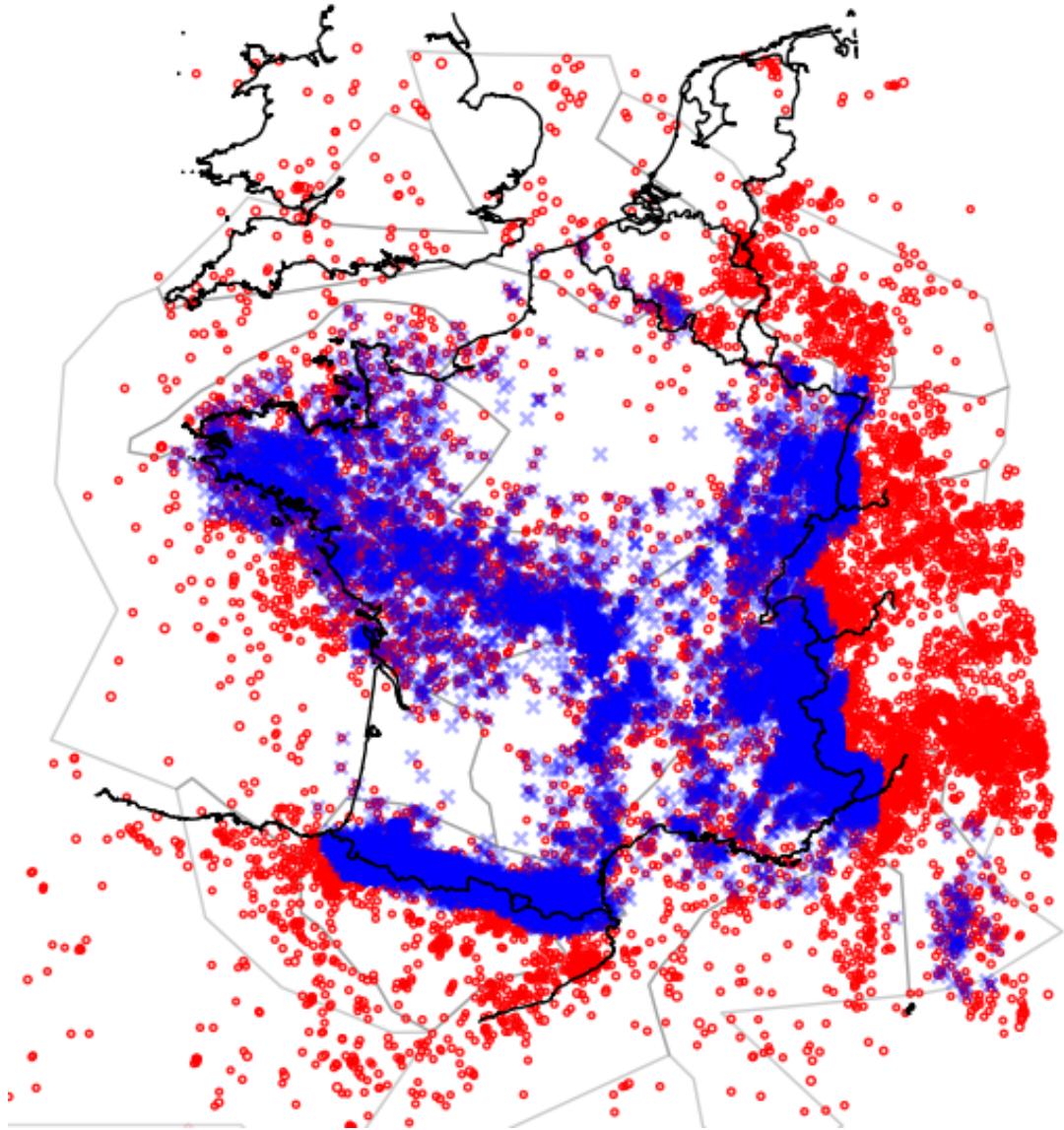


FIGURE 10 – Catalogue 1 (en bleu) et 2 en (rouge)

Cette approche n'est certes pas idéale, mais elle s'impose compte tenu de la qualité et de la structure des données disponibles. Optimiser successivement sur les marginales temporelles puis spatiales n'équivaut pas à optimiser directement la vraisemblance complète, dans notre cas du catalogue 1 et nous allons voir pourquoi. Toutefois, cette stratégie présente un intérêt computationnel non négligeable, puisqu'elle réduit la dimension de l'espace de paramètres à explorer lors de chaque étape d'optimisation. Pourquoi dans notre cas , optimiser les marginales n'est pas équivalent à optimiser par rapport à tous les paramètres? Nous aurions pu nous dire qu'en prenant comme fonction de mémoire un noyau séparable en temps et en espace, cela aurait pu marcher ...

On considère donc la vraisemblance :

$$\begin{aligned} \mathcal{L}(\theta) = & \exp \left(-N - \sum_{j=1}^N K e^{\alpha(m_j - m_c)} \left[1 - \left(\frac{c}{T-t_j+c} \right)^{p-1} \right] B_j \right) \\ & \times \prod_{i=1}^N \left[\mu(x_i, y_i) + \sum_{j < i} K e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (t_i - t_j + c)^{-p \frac{q-1}{\pi d_j}} \left(1 + \frac{(x_i - x_j)^2 + (y_i - y_j)^2}{d_j} \right)^{-q} \right]. \end{aligned}$$

En intégrant cette vraisemblance sur S^N , on peut procéder par récurrence descendante sur l'indice i : à chaque étape, l'intégrale $\int_S \cdot dx_i dy_i$ se fait entrer sous le produit, de sorte que l'on réduit successivement la dimension de l'intégrale. On obtient ainsi

$$\begin{aligned} \int_{S^N} \mathcal{L}(\theta) \prod_{i=1}^N dx_i dy_i &= \exp \left(-N - \sum_{j=1}^N K e^{\alpha(m_j - m_c)} \left[1 - \left(\frac{c}{T-t_j+c} \right)^{p-1} \right] B_j \right) \\ &\quad \times \prod_{i=1}^N \int_S \left[\mu(x_i, y_i) + \sum_{j < i} K e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (t_i - t_j + c)^{-p} \right. \\ &\quad \left. \times \frac{q-1}{\pi d_j} \left(1 + \frac{(x_i - x_j)^2 + (y_i - y_j)^2}{d_j} \right)^{-q} \right] dx_i dy_i. \end{aligned}$$

Or, par construction du terme de fond, on a

$$\int_S \mu(x_i, y_i) dx_i dy_i = \frac{N}{T}.$$

En revanche, les intégrales spatiales issues des noyaux de réplique ne se simplifient pas par 1 ou une constante : elles dépendent des frontières du domaine S ! Ces effets de bords, très important dans le catalogue 1, empêchent de retrouver exactement la forme de la vraisemblance d'un processus de Hawkes purement temporel-magnitude. Dit simplement, les effets de bords du catalogue 1 ne permettent pas d'optimiser sur les marginales plutôt que sur la vraisemblance totale. Si l'on rajoute que l'on n'a pas les mêmes catalogues et que donc la fonction vraisemblance n'est pas la même, l'estimation des paramètres temporels c et p est biaisée. C'est pourquoi nous avons essayé de fixer ces paramètres à des valeurs trouvées dans la littérature, en l'occurrence issues d'une estimation du modèle ETAS sur le contexte italien [5], puis de maximiser la vraisemblance des autres paramètres. Nous obtenons les résultats suivant :

Intervalles de confiance (95%) au point optimal:				
paramètre	estimate	std_err	IC95_low	IC95_high
K	0.078970	0.002288	0.074486	0.083454
alpha	1.251018	0.033692	1.184982	1.317053
p	1.750000	NaN	NaN	NaN
c	0.020000	NaN	NaN	NaN
q	1.724541	0.031862	1.662092	1.786991
d	1.769446	0.169596	1.437045	2.101848
gamma	0.925741	0.073435	0.781811	1.069671
m_c	2.000000	NaN	NaN	NaN

FIGURE 11 – Estimation des paramètres et intervalles de confiance

Cependant, cette approche n'est valide que si les paramètres temporels sont les mêmes en France et en Italie, ce que rien ne garantit.

En complément, nous avons appliqué la démarche en trois étapes, décrite ci-dessus, à une sous partie du catalogue, restreinte à un intervalle de 10 ans dans lequel aucune paire de séisme n'a été constaté au cours d'une même journée. Cela nous a permis d'exploiter le catalogue complet, sans la troncature spatiale du catalogue 1 (cf. Figure 10), et donc sous l'hypothèse que les intégrales spatiales B_{ij} sont toutes égales à 1. On supprime ainsi le risque de biais du transfert des valeurs de p et de c . En revanche, il y a un autre risque de biais du à la troncature temporelle.

Profils de vraisemblance

Après estimation conjointe des paramètres ETAS, nous traçons les *profils de vraisemblance* : pour chaque paramètre θ_k , on fixe les $N - 1$ autres à leur EMV $\hat{\theta}_{-k}$ et on recalcule la log-vraisemblance $\ell(\theta_k; \hat{\theta}_{-k})$ sur une grille. Sur les figures, le maximum du profil (ligne rouge pointillée) coïncide avec l'EMV, ce qui confirme la cohérence de l'optimum. La courbure autour du sommet indique l'identifiabilité : pics étroits \Rightarrow faible incertitude (p. ex. K, α, γ), profils plus plats \Rightarrow incertitude plus large (p. ex. c, d). Globalement, les profils valident le caractère vraisemblable des paramètres estimés. Ces profils de vraisemblance sont obtenus par estimation des paramètres temporels, puis spatiaux.

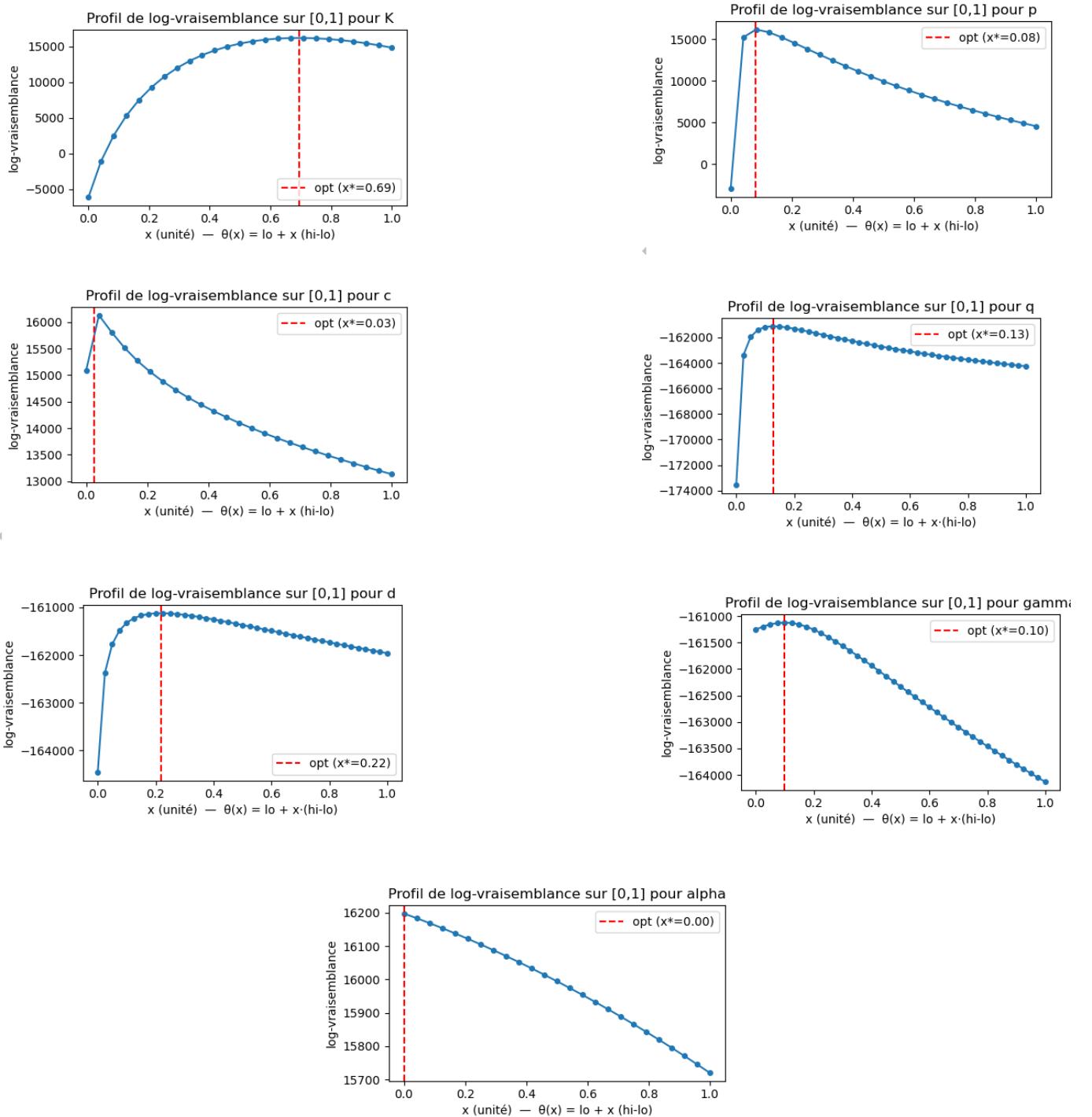


FIGURE 12 – Ensemble des 7 profils de vraisemblances.

7.8 Génération de données synthétiques

Une fois les paramètres du modèle ETAS inhomogène estimés comme indiqué ci-dessus, il est possible de régénérer des catalogues sismiques artificiels afin d'évaluer la capacité du modèle à reproduire les caractéristiques observées. Ces données synthétiques sont obtenues en simulant le processus ETAS calibré, sur une période d'une année, en utilisant la fonction d'intensité de fond estimée et les paramètres de réPLICATION optimisés.

On a donc deux réalisations indépendantes de ce processus simulé. Sur chaque carte, les mainshocks sont représentés par des cercles rouges, tandis que les répliques générées par le mécanisme de déclenchement sont indiquées par des croix bleues.

On observe que la répartition spatiale des événements est cohérente avec les intensités de fond du domaine étudié : les zones historiquement plus actives dans le catalogue réel (par exemple le sud-est et la région pyrénéenne) apparaissent également plus denses dans les simulations. De plus, le nombre de répliques simulées par mainshock reste réaliste, confirmant que la calibration du modèle capture correctement la dynamique de déclenchement.

Ces résultats montrent que le modèle ETAS inhomogène ajusté est capable de reproduire des patrons spatio-temporels plausibles, renforçant ainsi la confiance dans son utilisation pour la modélisation de l'aléa sismique.

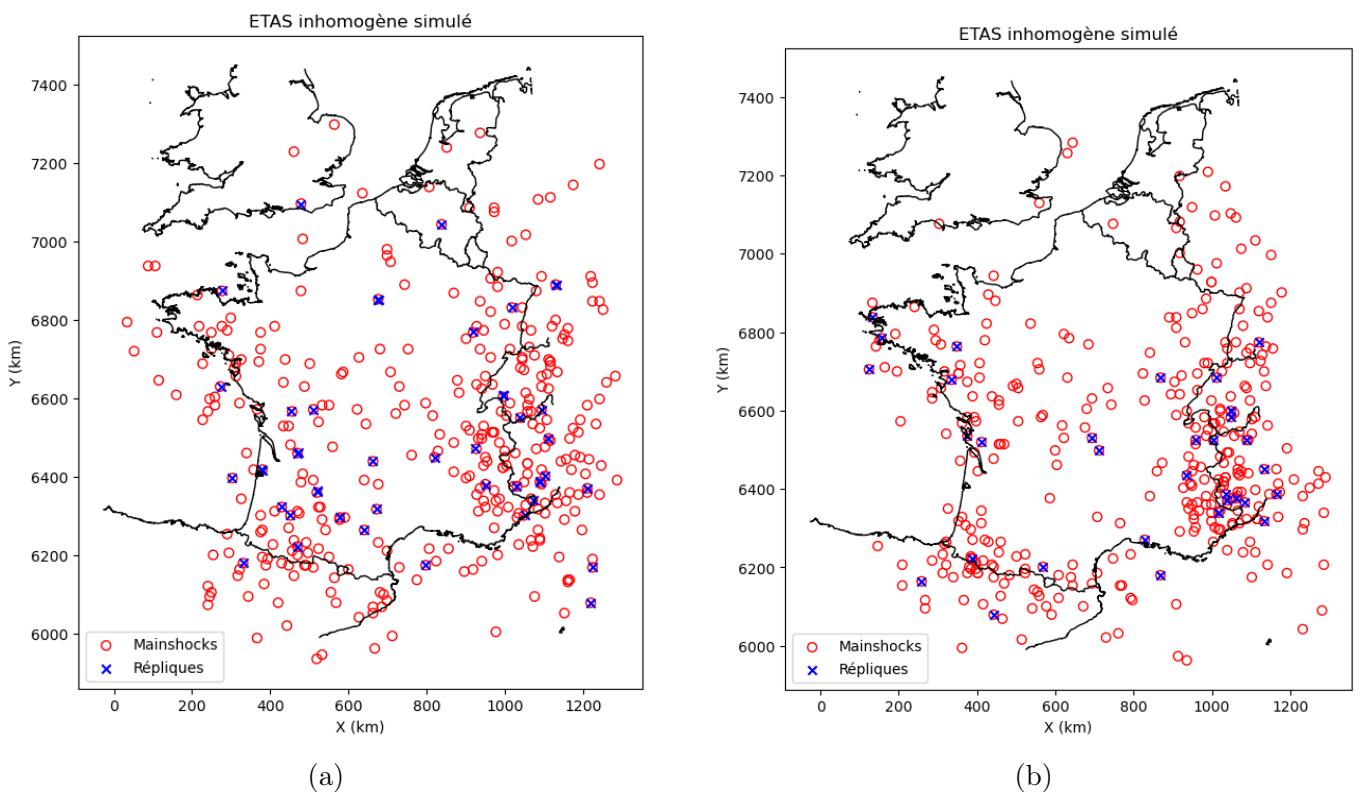


FIGURE 13 – Génération de données sur 1 an avec des ETAS de paramètres différents .

8 Conclusion et Perspectives

En conclusion, ce stage a permis de mettre en évidence la pertinence du modèle ETAS pour décrire la dynamique spatio-temporelle des séismes en France et dans les régions limitrophes. Malgré certaines limites liées à la qualité et à l'exhaustivité des données disponibles, les résultats obtenus se sont révélés cohérents et prometteurs, confortant ainsi le choix de ce cadre de modélisation pour l'étude de l'aléa sismique.

Cette expérience a permis de mettre en pratique des outils variés des mathématiques appliquées, tout en apportant des réponses concrètes aux difficultés rencontrées. Elle ouvre également des perspectives pour améliorer le modèle ETAS et l'adapter à des applications pratiques en évaluation du risque sismique, ainsi que pour explorer de nouvelles approches d'inférence de ses paramètres.

En effet, nous n'avons pas eu l'occasion d'explorer des méthodes bayésiennes comme les méthodes ABC (Approximate Bayesian Computation) pour l'estimation des paramètres du modèle ETAS comme le propose Deutsch et al. (2021), ou des méthodes à base de Machine Learning voire de Deep Learning. Pour ces dernières, il aurait été très intéressant d'élaborer une démarche de génération d'un grand volume de données synthétiques vraisemblables sur lesquelles réaliser un entraînement, et en utilisant comme algorithmes prédicteurs une structure de ML/DL capable de capturer les corrélations temporelles et spatiales comme les LSTM ou les Transformers, idée que propose Zuo et al. (2020) dans son article paru en 2020 ...

Références

- [1] Junhyeon Kwon, Yingcai Zheng, Mikyoung Jun, *Flexible spatio-temporal Hawkes process models for earthquake occurrences*, Spatial Statistics, Volume 54, 2023.
- [2] A. Bernabeu, J. Zhuang, J. Mateu, *Spatio-Temporal Hawkes Point Processes : A Review*, JABES, 2024. <https://doi.org/10.1007/s13253-024-00653-7>
- [3] J. Derek Tucker, Lyndsay Shand, John R. Lewis, *Handling missing data in self-exciting point process models*, Spatial Statistics, Volume 29, 2019.
- [4] Jiancang Zhuang, Ting Wang, Koji Kirosugi, *Detection and replenishment of missing data in marked point processes*, Statistica Sinica, 2020. 10.5705/ss.202017.0403
- [5] Nevena Sipcic, Mohsen Kohrangi, Athanasios N. Papadopoulos, Warner Marzocchi, Paolo Bazzurro, *The Effect of Seismic Sequences in Probabilistic Seismic Hazard Analysis* Bulletin of the Seismological Society of America (2022) 112 (3) : 1694–1709.