

UFR SAT

Département de Mathématiques Appliquées

SOUTENANCE DE MEMOIRE

MODÉLISATION SPATIO-TEMPORELLE DES PRÉCIPITATIONS EN AFRIQUE DE L'OUEST : ANALYSE DES RISQUES CLIMATIQUES

Présenté par **Kossivi GNOZIGUE**
Sous la direction de **Pr. Aliou DIOP**

15 juillet 2025

- 1 Introduction
- 2 Théorie des valeurs extrêmes (TVE)
- 3 Modélisation hiérarchique bayésienne (BHM)
- 4 Conclusion et perspectives

- ❶ INTRODUCTION
- ❷ THÉORIE DES VALEURS EXTRÊMES
- ❸ MODÉLISATION HIÉRARCHIQUE BAYÉSIENNE
- ❹ CONCLUSION ET PERSPECTIVES

Introduction

- Le risque climatique désigne l'ensemble des conséquences négatives que des phénomènes climatiques, qu'ils soient soudains ou progressifs, peuvent avoir sur nos sociétés, nos économies et nos écosystèmes. Parmi ces phénomènes, on compte :
 - ▷ Événements extrêmes : inondations, sécheresses, tempêtes, etc.
 - ▷ Changements progressifs : élévation du niveau de la mer, acidification des océans, etc.

Notre monde est directement menacé par ces risques. En Afrique de l'Ouest, les inondations sont particulièrement préoccupantes, touchant principalement le Niger, le Nigeria et le Sénégal.



Figure 1 – Inondation à Dakar (Sénégal)
suite à une pluie le 05/08/2022



Figure 2 – Inondation à Maiduguri
(Nigeria) suite à une crue d'un barrage
le 10/09/2024

- Objectif principal : Comprendre et prédire les régimes de précipitations extrêmes qui causent les inondations en Afrique de l'ouest en utilisant la modélisation spatio-temporelle et en se basant sur des données collectées entre 2000 et 2023.
- Méthodologie :
 - ▷ **Approche classique** : Ajuster la loi générale des valeurs extrêmes (GEV) aux données en utilisant la méthode du maximum de vraisemblance.
 - ▷ **Approche bayésienne** : Gérer l'incertitude des paramètres du modèle, capturer les effets aléatoires et la dépendance entre les stations en utilisant la méthode Hamiltonian Monte Carlo (HMC), une version améliorée des Chaînes de Markov Monte Carlo (MCMC).

① INTRODUCTION

② THÉORIE DES VALEURS EXTRÊMES

③ MODÉLISATION HIÉRARCHIQUE BAYÉSIENNE

④ CONCLUSION ET PERSPECTIVES

Théorie des valeurs extrêmes (TVE)

- La TVE fournit des outils pour modéliser la probabilité et l'ampleur des événements rares et extrêmes et repose sur l'idée qu'ils suivent des lois de probabilité différentes des événements courants.
- Elle est basée sur un principal théorème qui est un résultat central décrivant les différentes formes de distributions limites des maxima.

Théorème : Fisher-Tippett-Gnedenko

Soient X une suite de variables aléatoires (v.a) continues et iid $\{X_1, \dots, X_n\}$ de fonction de répartition F et $M_n = \max(X_1, \dots, X_n), \forall n \in \mathbb{N}$ le maximum de ces n v.a. S'il existe des suites de constantes $\{a_n\} > 0$ et $\{b_n\} \in \mathbb{R}$ et une fonction de répartition G non dégénérée telles que :

$$\mathbb{P} \left(\frac{M_n - b_n}{a_n} \leq x \right) \xrightarrow{n \rightarrow \infty} G(x), \quad (1)$$

alors G appartient à la famille des distributions $\text{GEV}(\mu, \sigma, \xi)$, définie sur $x \in \mathbb{R} : 1 + \xi \frac{(x-\mu)}{\sigma} > 0$.

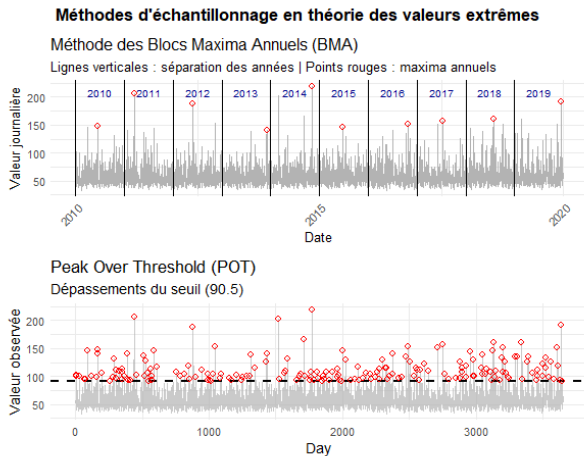


Figure 3 – Simulation d'une série journalières sur 10 ans selon une distribution GEV avec un paramètre de forme positif

Les distributions GEV

- La distribution Pareto généralisée (GPD) pour le POT.
- La distribution généralisée des valeurs extrêmes (GEV) pour les BMA

La fonction de répartition cumulative de la distribution GEV est :

$$H_{\mu,\sigma,\xi}(x) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]_+^{-1/\xi} \right\} & \text{si } \xi \neq 0, \\ \exp \left\{ - \exp \left(- \frac{x-\mu}{\sigma} \right) \right\} & \text{si } \xi = 0 \end{cases} \quad (2)$$

où : $1 + \xi \left(\frac{x-\mu}{\sigma} \right) > 0$

$a_+ = \max(a, 0)$, x la valeur de la variable aléatoire,

$\mu \in \mathbb{R}$ est le paramètre de localisation. $\sigma > 0$ est le paramètre d'échelle et

$\xi \in \mathbb{R}$ est le paramètre de forme, qui détermine le type de queue de distribution :

$\xi > 0$: Distribution de Fréchet (queue lourde)

$\xi = 0$: Distribution de Gumbel (queue légère)

$\xi < 0$: Distribution de Weibull (borne supérieure finie)

La méthode utilisée pour mesurer à quel point un modèle statistique explique des observations est le maximum de vraisemblance. Elle est plus facile à manipuler sous sa forme logarithmique et son application à la pdf de la GEV pour un échantillon de taille n donne :

$$\ell(\mu, \sigma, \xi) = - \sum_{i=1}^n \left\{ \log \sigma + \left(1 + \frac{1}{\xi} \right) \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] + \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (3)$$

Avec $1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) > 0$ pour toutes les observations i

Critère de choix et niveaux de retour

La déviance \mathcal{D} , l'AIC(Akaike Information Criterion) et le BIC(Bayesian Information Criterion) sont des critères qui permettent de sélectionner le meilleur modèle. Celui avec la plus petite valeur et le moins de paramètres.

$$\mathcal{D} = 2\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\} \quad (4)$$

$$AIC = \mathcal{D} + 2k \quad (5)$$

$$BIC = \mathcal{D} + k \cdot \ln(n) \quad (6)$$

La déviance permet de comparer deux modèles pour désigner le meilleur et les critères AIC et BIC permettent d'éviter le sur ajustement à travers des pénalités (nombre de paramètres $2k$ et les grands échantillons $\ln(n)$).

Les niveaux de retour d'une période p est le quantile d'ordre $1 - 1/p$ qui est dépassé avec une probabilité $1/p$. Le quantile z_p au niveau $1 - p$ de la loi GEV s'exprime par :

$$z_p = H_\xi^{-1}(1 - p) = \begin{cases} \mu - \frac{[1 - \{-\log(1 - 1/p)\}]^{-\xi} \sigma}{\xi} & \text{si } \xi \neq 0, \\ \mu - \sigma \log(-\log(1 - 1/p)) & \text{si } \xi = 0. \end{cases}$$

Critère de choix et niveaux de retour

La déviance \mathcal{D} , l'AIC(Akaike Information Criterion) et le BIC(Bayesian Information Criterion) sont des critères qui permettent de sélectionner le meilleur modèle. Celui avec la plus petite valeur et le moins de paramètres.

$$\mathcal{D} = 2\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\} \quad (4)$$

$$AIC = \mathcal{D} + 2k \quad (5)$$

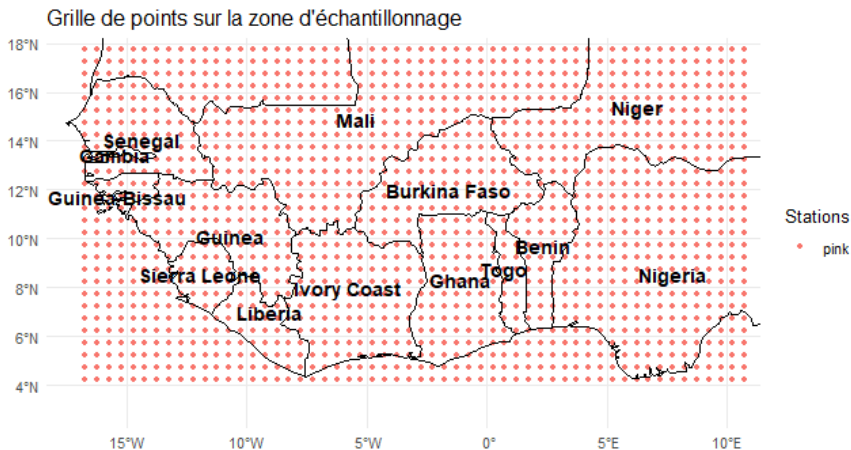
$$BIC = \mathcal{D} + k \cdot \ln(n) \quad (6)$$

La déviance permet de comparer deux modèles pour désigner le meilleur et les critères AIC et BIC permettent d'éviter le sur ajustement à travers des pénalités (nombre de paramètres $2k$ et les grands échantillons $\ln(n)$).

Les niveaux de retour d'une période p est le quantile d'ordre $1 - 1/p$ qui est dépassé avec une probabilité $1/p$. Le quantile z_p au niveau $1 - p$ de la loi GEV s'exprime par :

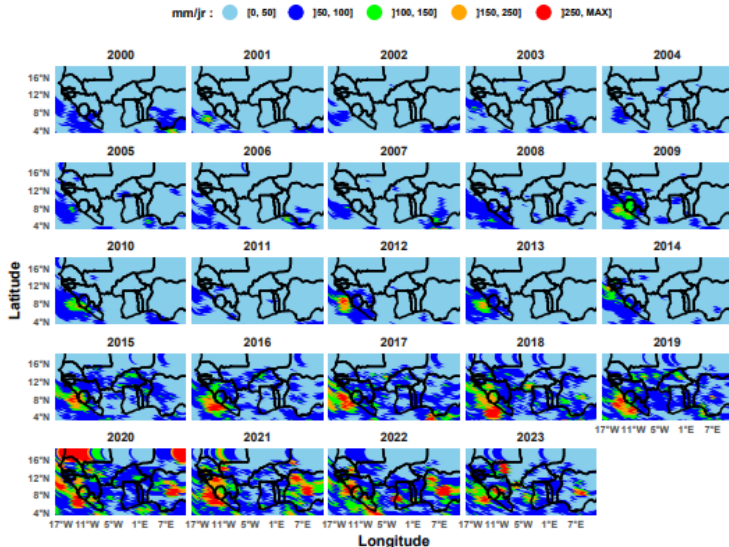
$$z_p = H_{\xi}^{-1}(1 - p) = \begin{cases} \mu - \frac{[1 - \{-\log(1 - 1/p)\}]^{-\xi} \sigma}{\xi} & \text{si } \xi \neq 0, \\ \mu - \sigma \log(-\log(1 - 1/p)) & \text{si } \xi = 0. \end{cases}$$

Cartographie de la zone d'échantillonnage

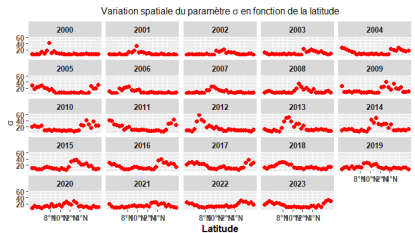
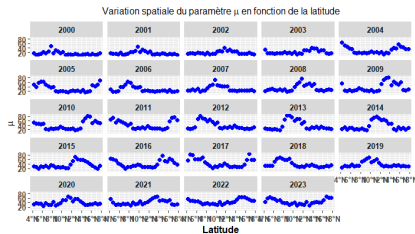
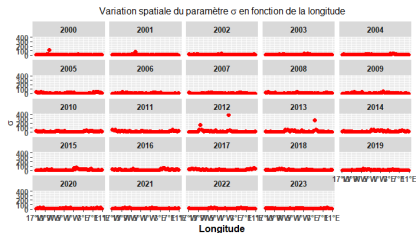
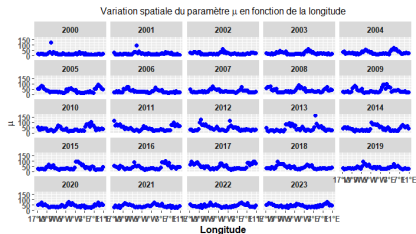


Précipitations maximales annuelles

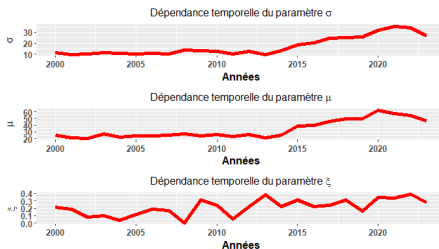
Carte des précipitations



Analyse exploratoire : Variation spatiale des paramètres GEV



Analyse exploratoire : Variation temporelle des paramètres GEV



Suite à ces différentes observations, nous avons développé deux modèles spatiaux se distinguant par la variation du paramètre d'échelle en fonction de la longitude et quatre modèles intégrant une dépendance temporelle, généralement linéaire pour tous les paramètres.

Définition et évaluation des modèles

Table 1 – Les modèles GEV classiques.

Modèles	Localisation	Echelle	Forme
1	$\mu_0 + \mu_1 lat + \mu_2 lon$	$\sigma_0 + \sigma_1 lat + \sigma_2 lon$	ξ_0
2	$\mu_0 + \mu_1 lat + \mu_2 lon$	$\sigma_0 + \sigma_1 lat$	ξ_0
3	$\mu_0 + \mu_1 lat + \mu_2 lon + \mu_3 t$	$\sigma_0 + \sigma_1 lat + \sigma_2 lon + \sigma_3 t$	ξ_0
4	$\mu_0 + \mu_1 lat + \mu_2 lon + \mu_3 t$	$\sigma_0 + \sigma_1 lat + \sigma_2 t$	ξ_0
5	$\mu_0 + \mu_1 lat + \mu_2 lon + \mu_3 t$	$\sigma_0 + \sigma_1 lat + \sigma_2 lon + \sigma_3 t$	$\xi_0 + \xi_1 t$
6	$\mu_0 + \mu_1 lat + \mu_2 lon + \mu_3 t$	$\sigma_0 + \sigma_1 lat + \sigma_2 t$	$\xi_0 + \xi_1 t$

Table 2 – Valeurs des critères pour chaque modèle

Modèles	Déviance	AIC	BIC
1	339324.5	339338.5	339398.2
2	340677.0	340689.0	340740.2
3	339256.2	339274.2	339351
4	340676.9	340692.9	340761.2
5	339277.4	339297.4	339382.7
6	340676.9	340694.9	340771.7

Définition et évaluation des modèles

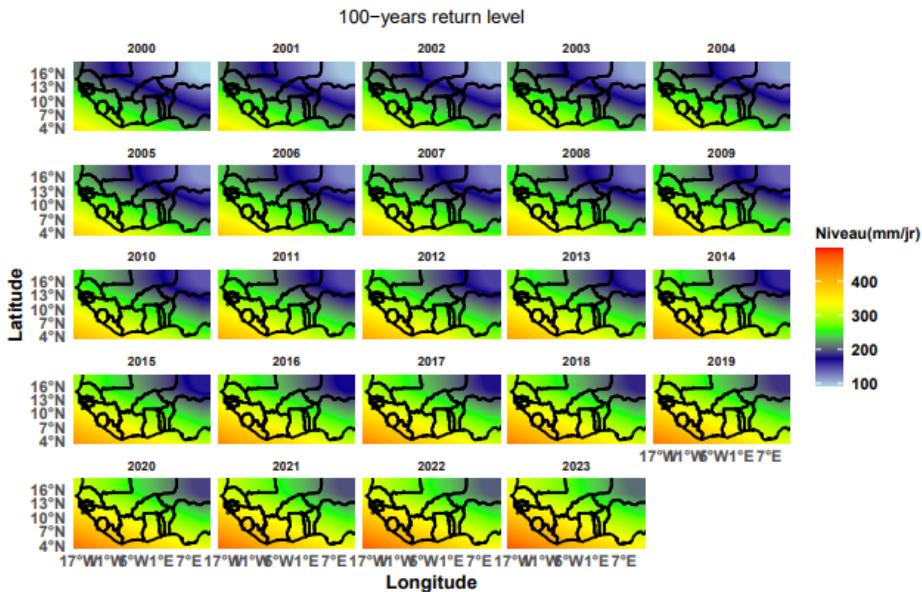
Table 1 – Les modèles GEV classiques.

Modèles	Localisation	Echelle	Forme
1	$\mu_0 + \mu_1 lat + \mu_2 lon$	$\sigma_0 + \sigma_1 lat + \sigma_2 lon$	ξ_0
2	$\mu_0 + \mu_1 lat + \mu_2 lon$	$\sigma_0 + \sigma_1 lat$	ξ_0
3	$\mu_0 + \mu_1 lat + \mu_2 lon + \mu_3 t$	$\sigma_0 + \sigma_1 lat + \sigma_2 lon + \sigma_3 t$	ξ_0
4	$\mu_0 + \mu_1 lat + \mu_2 lon + \mu_3 t$	$\sigma_0 + \sigma_1 lat + \sigma_2 t$	ξ_0
5	$\mu_0 + \mu_1 lat + \mu_2 lon + \mu_3 t$	$\sigma_0 + \sigma_1 lat + \sigma_2 lon + \sigma_3 t$	$\xi_0 + \xi_1 t$
6	$\mu_0 + \mu_1 lat + \mu_2 lon + \mu_3 t$	$\sigma_0 + \sigma_1 lat + \sigma_2 t$	$\xi_0 + \xi_1 t$

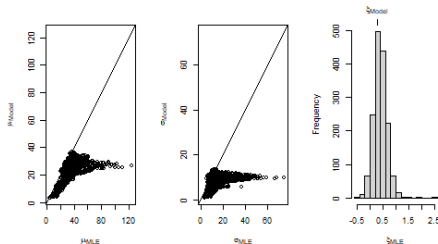
Table 2 – Valeurs des critères pour chaque modèle

Modèles	Déviance	AIC	BIC
1	339324.5	339338.5	339398.2
2	340677.0	340689.0	340740.2
3	339256.2	339274.2	339351
4	340676.9	340692.9	340761.2
5	339277.4	339297.4	339382.7
6	340676.9	340694.9	340771.7

Carte des niveaux de retour



Limites de la GEV classique et Approches de solutions



Limites de l'approche GEV avec la méthode MLE :

- Possible existence des effets aléatoires dus aux covariables latentes comme l'altitude, la proximité de la mère.
- Faible quantité d'observation par station

Avantage de l'approche bayésienne :

- Capture de la dépendance à travers les effets aléatoires gaussiens.
- Pratique avec peu de données par station.
- Intégration des connaissances préalables et des données auxiliaires.
- Prise en compte de l'incertitude associée aux paramètres.

- 1 INTRODUCTION
- 2 THÉORIE DES VALEURS EXTRÊMES
- 3 **MODÉLISATION HIÉRARCHIQUE BAYÉSIENNE**
- 4 CONCLUSION ET PERSPECTIVES

Modélisation hiérarchique bayésienne : Théorème de Bayes

L'approche bayésienne est basée sur le théorème de Bayes. Ce dernier permet de mettre à jour les probabilités a priori en fonction des données observées.

Théorème

Soient X et Y deux variables aléatoires définies sur l'espace de probabilité (Θ, \mathcal{F}, P) . X et Y de loi jointe de densité $\phi(x, y)$, de densité conditionnelle $f(x | y)$ et de densité marginale $g(y) = \int \phi(x, y) dx$, alors la formule de Bayes est donnée par

$$g(y|x) = \frac{f(x | y)g(y)}{\int f(x | y)g(y)dy} \quad (7)$$

Le terme au dénominateur est une constante de normalisation. Donc l'équation (7) peut s'écrire sous la forme :

$$g(y | x) \propto f(x | y)g(y) \quad (8)$$

1 Modèle des données : Vraisemblance

$$Y(s, t) \mid \mu(s, t), \sigma(s, t), \xi(s, t) \sim GEV(\mu(s, t), \sigma(s, t), \xi(s, t)) \quad (9)$$

Il capte la dépendance spatio-temporelle sur $\mu(s, t)$, $\sigma(s, t)$ et $\xi(s, t)$.

2 Modèle des paramètres : régression spatio-temporelle

$$\gamma(s, t) = \beta_0^\gamma + X_\gamma^T(s, t)\beta^\gamma(s, t) + \omega^\gamma(s, t) \quad (10)$$

β_0^γ intercepts constants; $X_\gamma^T(s, t)$ matrice des prédicteurs;

$\beta^\gamma(s, t) = [\beta_1^\gamma(s, t), \dots, \beta_p^\gamma(s, t)]^T$ vecteur des coefficients de régression;

$\omega^\gamma(s, t) \sim GP(0, C(s, s'))$ et $\gamma = (\mu, \sigma, \xi)^T$

3 Modèle des hyper paramètres : Priors

En considérant la copule gaussienne à cause de la dépendance linéaire, on a : $\omega_\gamma \mid \theta_\gamma \sim MNV(0, \Sigma_\gamma)$ où $\Sigma_\gamma = C(\theta_\gamma)$ la fonction de covariance :

$$C(s_i, s_j; \theta_\gamma) = \begin{cases} \delta_\gamma^2 \exp \left\{ -\frac{\|s_i - s_j\|}{a_\gamma} \right\} & \text{pour } i \neq j \\ \delta_\gamma^2 & \text{pour } i = j \end{cases} \quad (11)$$

Avec $\theta_\gamma = (\delta_\gamma^2, a_\gamma)$ comme priors où $\delta_\gamma^2 \sim \mathcal{N}(a, b)$ est le seuil partiel et $a_\gamma \sim \mathcal{LN}(a, b)$ l'étendue et $a, b \in \mathbb{R}$.

Définition des modèles

Pour chaque observation y_i (précipitation maximale annuelle)

- ❶ **Modèle GEV avec effets spatiaux linéaires : Capture des gradients spatiaux linéaires.** $y_i \sim \text{GEV}(\mu_i, \sigma_i, \xi)$ Où

$$\mu_i = \beta_0^\mu + \beta_1^\mu \text{lat}_i + \beta_2^\mu \text{lon}_i, \quad \sigma_i = \beta_0^\sigma + \beta_1^\sigma \text{lon}_i \quad \text{et} \quad \xi = \beta_0^\xi$$

Priors : $\beta_0^\mu, \beta_1^\mu, \beta_2^\mu \sim \mathcal{N}(0, 5)$, $\beta_0^\sigma, \beta_1^\sigma \sim \mathcal{LN}(0, 1)$ et $\beta_0^\xi \sim \mathcal{N}(0, 0.5)$

- ❷ **Modèle GEV avec effets aléatoires par station : Capture de la variation inter-stations.** Pour la station j et l'observation i :

$$y_{ij} \sim \text{GEV}(\mu_{ij}, \sigma_{ij}, \xi) \text{ Où}$$

$$\mu_{ij} = \beta_0^\mu + u_j^\mu, u_j^\mu \sim \mathcal{N}(0, \tau_\mu), \quad \sigma_{ij} = \beta_0^\sigma + u_j^\sigma, u_j^\sigma \sim \mathcal{N}(0, \tau_\sigma) \text{ et } \xi = \beta_0^\xi$$

Priors : $\beta_0^\mu \sim \mathcal{N}(0, 10)$, $\beta_0^\sigma \sim \mathcal{LN}(0, 1)$, $\beta_0^\xi \sim \mathcal{N}(0, 0.5)$, $\tau_\mu, \tau_\sigma \sim \mathcal{LN}(0, 1)$

- ❸ **Modèle GEV avec processus gaussien : Capture des motifs spatiaux non linéaires.** $y_i \sim \text{GEV}(\mu_i, \sigma_i, \xi)$ Où

$$\mu_i = \beta_0^\mu + f_1(\mathbf{s}_i), \quad \sigma_i = \beta_0^\sigma + f_2(\mathbf{s}_i) \quad \text{et} \quad \xi = \beta_0^\xi$$

$$f_1 \sim \mathcal{GP}(0, K_1), \quad f_2 \sim \mathcal{GP}(0, K_2), \quad K_k(\mathbf{s}_i, \mathbf{s}_j) = \eta_k^2 \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{2\rho_k^2}\right)$$

avec $k \in \{1, 2\}$

Priors : $\eta_1, \eta_2 \sim \mathcal{N}(0, 3)$, $\rho_1, \rho_2 \sim \text{Inv-Gamma}(2, 1)$, $\beta_0^\mu \sim \mathcal{N}(0, 10)$, $\beta_0^\sigma \sim \mathcal{LN}(0, 1)$ et $\beta_0^\xi \sim \mathcal{N}(0, 0.5)$

Étapes principales de l'algorithme Hamiltonian Monte Carlo :

- ➊ **Initialiser l'impulsion** : On définit l'état actuel des paramètres (position q) et on tire aléatoirement une impulsion p , on définit un pas d'intégration ϵ et un total de pas L .
- ➋ **Simuler la dynamique Hamiltonienne** : On simule une trajectoire dans l'espace (q, p) grâce aux équations de Hamilton qui décrivent l'évolution d'un système pouvant conserver l'énergie totale et les intégrateurs numériques comme l'intégrateur de Leapfrog. Cela permet une évolution rapide des paramètres q et p en prenant de petits pas ϵ sur un total de L pas.
- ➌ **Critère d'acceptation-Rejet** : On applique le critère avec un taux d'acceptation généralement élevé, évitant la stagnation. Si la proposition est acceptée, la nouvelle position est q^* sinon, la chaîne reste à l'ancienne position q .

① Vérification et adéquation du modèle

- Convergence des chaînes.
- La statistique de Gelman-Rubin \hat{R} :

$$\hat{R} = \sqrt{\frac{\widehat{VAR}(\theta)}{W}} \text{ avec } \widehat{VAR}(\theta) = \left(1 - \frac{1}{n}\right)W + \frac{1}{n}B \quad (12)$$

où W et B sont respectivement les variances intra-chaînes (variance moyenne dans chaque chaîne) et inter-chaînes (variance entre les moyennes des chaînes). Donc pour m chaînes, $W = \frac{1}{m} \sum_{j=1}^m s_j^2$ où $s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\theta_{ij} - \bar{\theta}_j)^2$ et $B = \frac{n}{m-1} \sum_{j=1}^m (\bar{\theta}_j - \bar{\theta})^2$ où $\bar{\theta} = \frac{1}{m} \sum_{j=1}^m \bar{\theta}_j$.

② Performances prédictives : WAIC(Watanabe-Akaike Information Criterion) et LOO-CV (Leave-One-Out Cross-Validation)

Table 3 – Valeurs des critères WAIC et LOO-CV

Modèles	WAIC	LOO-CV
1	329597.4	329585.8
2	289681.0	289675.5

Convergence et carte des niveaux de retour sur 100 ans

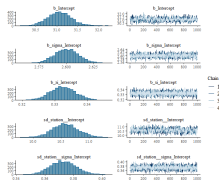


Figure 4 – Convergence des chaînes

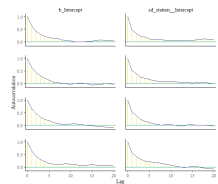
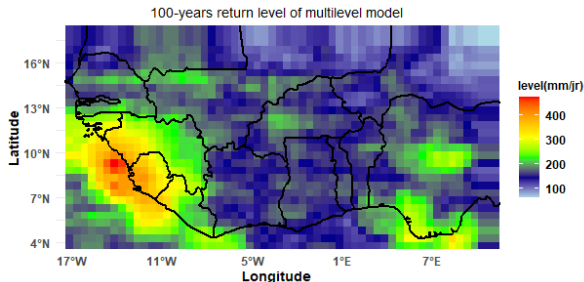


Figure 5 – Auto corrélation



- 1 INTRODUCTION
- 2 THÉORIE DES VALEURS EXTRÊMES (TVE)
- 3 MODÉLISATION HIÉRARCHIQUE BAYÉSIENNE (BHM)
- 4 CONCLUSION ET PERSPECTIVES

Ce travail a souligné l'importance cruciale de la modélisation spatio-temporelle pour analyser les précipitations extrêmes en Afrique de l'ouest. Nous avons vu que si l'approche GEV classique, basée sur la méthode du maximum de vraisemblance, a ses limites, notamment face à la complexité de la variabilité spatiale et au volume des données, l'approche bayésienne se révèle être une alternative puissante.

Grâce au cadre bayésien et à l'algorithme HMC implémenté via Stan, nous avons pu quantifier l'incertitude des paramètres et intégrer la gestion de la dépendance spatiale complexe entre les stations, offrant ainsi des estimations plus fiables au prix d'un coût computationnel très élevé.

Les cartes des niveaux de retour sont un apport majeur de cette étude. Elles fournissent une visualisation claire des risques de précipitations extrêmes, des informations essentielles pour la planification et la gestion des ressources en eau, l'agriculture et les stratégies d'adaptation au changement climatique dans la région.

Application de :

- Un modèle avec processus gaussiens afin de capturer des motifs spatiaux non linéaires complexes.
- Des approches par apprentissage machine pour combiner robustesse et flexibilité.
- Des variables climatologiques telles que la température, la pression de surface, l'humidité relative, etc. reste également une piste exploitable.
- Des approches numériques comme INLA(Integrated Nested Laplace Approximation) plus rapide que les méthodes approchées (HMC) en terme de calcul.

MERCI POUR VOTRE
AIMABLE ATTENTION !