



Université Gaston Berger de Saint-Louis
U.F.R. de Sciences Appliquées et de Technologie

MÉMOIRE DE MASTER
pour l'obtention du diplôme de
MASTER en Mathématiques Appliquées

Soutenu publiquement le ** juillet 2025 par

Kossivi GNOZIGUE

Spécialité : Sciences de Données et Applications

Option : Statistique

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

Encadré par **Pr. Aliou DIOP**

Composition du Jury

Année Scolaire : 2023-2024

Dédicace

À mon père Kédenga GNOZIGUE

À ma mère Hodalou POYODI

À toute ma famille

À la famille du professeur Aliou DIOP

Remerciements

Je rends grâce à Dieu Tout Puissant pour la santé qu'il m'a accordé.

Je tiens ensuite à exprimer ma profonde gratitude à mon directeur de mémoire Professeur Aliou DIOP, à sa famille qui m'a accueilli et soutenu sans condition aucune, au Professeur El Hadji DEME, au docteur Lamine DIOP et aux doctorants du laboratoire d'Études et de Recherches en Statistique et Développement (LERSTAD).

Je souhaite également remercier l'ensemble du corps professoral de l'université Gaston Berger, en particulier les professeurs de l'Unité de Formation et de Recherche en Sciences Appliquées et Technologie (UFR-SAT) pour la qualité de leurs enseignements, ainsi que pour leur disponibilité et leurs précieux conseils qui ont contribué à enrichir ma formation.

Merci également à la communauté togolaise, à mes camarades de promotion et à mon frère et compagnon Richard KONDOZIM.

Résumé

Ce mémoire de recherche se consacre à l'analyse spatio-temporelle des précipitations extrêmes en Afrique de l'ouest, une région particulièrement vulnérable aux aléas climatiques. Notre étude s'appuie sur une double approche méthodologique pour une exploration approfondie de ces phénomènes. D'une part, nous utilisons une méthode classique fondée sur la loi générale des valeurs extrêmes (GEV). Dans ce cadre, la dépendance spatiale des paramètres GEV est capturée par une régression spatiale à coefficients variant dans l'espace. Cette approche permet de cartographier l'influence des facteurs géographiques sur les précipitations extrêmes. D'autre part, nous employons une méthode bayésienne, dont l'inférence repose sur des algorithmes d'apprentissage automatique, notamment le Monte Carlo Hamiltonien (HMC), implémenté avec le langage de programmation Stan. Cette approche bayésienne est cruciale pour quantifier l'incertitude associée aux paramètres du modèle et pour capturer la complexité de la dépendance spatiale entre les différentes stations. Cette dépendance est introduite via une copule gaussienne, justifiée par la relation linéaire observée entre les paramètres GEV et les coordonnées géographiques. Nos analyses sont réalisées sur un vaste ensemble de données de précipitations maximales annuelles, couvrant la période de 2000 à 2023. Ces données ont été recueillies auprès de 1568 stations réparties sur l'ensemble de l'Afrique de l'ouest, délimitée par les coordonnées géographiques $4^{\circ}N - 18^{\circ}N$ et $17.2^{\circ}W - 11.2^{\circ}E$. Enfin, les niveaux de retour des précipitations extrêmes ont été calculés et représentés sous forme de cartes détaillées. Ces visualisations offrent des informations précieuses sur les variations spatiales et temporelles du risque associé aux événements de précipitations extrêmes en Afrique de l'ouest, contribuant ainsi à une meilleure compréhension et à une gestion plus efficace de ces aléas climatiques.

Mots clés : Précipitations extrêmes, distribution générale des valeurs extrêmes (GEV), régression spatiale, méthode bayésienne, Monte Carlo Hamiltonien (HMC) langage de programmation Stan, niveaux de retour.

Abstract

This research thesis focuses on the spatio-temporal analysis of extreme precipitation in West Africa, a region particularly vulnerable to climate hazards. Our study uses a dual methodological approach for an in-depth exploration of these phenomena. On one hand, we employ a classical method based on the Generalized Extreme Value (GEV) distribution. Within this framework, the spatial dependence of GEV parameters is captured by a spatial regression with spatially varying coefficients. This approach allows us to map the influence of geographical factors on extreme precipitation. On the other hand, we use a Bayesian method, whose inference relies on machine learning algorithms, notably Hamiltonian Monte Carlo (HMC), implemented with the Stan programming language. This Bayesian approach is crucial for quantifying the uncertainty associated with the model parameters and for capturing the complex spatial dependence between different stations. This dependence is introduced via a Gaussian copula, justified by the linear relationship observed between GEV parameters and geographical coordinates. Our analyses are performed on a vast dataset of annual maximum precipitation, covering the period from 2000 to 2023. This data was collected from 1568 stations across West Africa, delimited by geographical coordinates $4^{\circ}N - 18^{\circ}N$ et $17.2^{\circ}W - 11.2^{\circ}E$. Finally, the return levels of extreme precipitation have been calculated and represented as detailed maps. These visualizations offer valuable insights into the spatial and temporal variations of the risk associated with extreme precipitation events in West Africa, thereby contributing to a better understanding and more effective management of these climate hazards.

Keywords : Extreme precipitation, Generalized Extreme Value (GEV) distribution, spatial regression, Bayesian method, Hamiltonian Monte Carlo (HMC), Stan programming language, return levels.

Notations et Abréviations

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BHM	Bayesian Hierarchical Models
BMA	Bloc Maxima Analysis
GEV	Generalized Extreme value distribution
GPD	Generalized Pareto Distribution
HMC	Hamiltonian Monte Carlo
iid	indépendantes et identiquement distribuées
LFI	Local Fit Interpolation
MCMC	Monte Carlo Markov Chain
POT	Peaks Over Threshold
RFA	Regional Frequency Analysis
SMLE	Spatial Maximum Likelihood Estimation
TVE	Théorie des Valeurs Extrêmes
v.a	variables aléatoires

Table des matières

Dédicace	i
Remerciement	ii
Résumé	iii
Abstract	iv
Notations et abréviations	v
INTRODUCTION GÉNÉRALE	1
1 LES ÉVÈNEMENTS CLIMATIQUES EXTRÊMES	3
1.1 Les risques climatiques.	3
1.2 Extrêmes pluviométriques	4
1.3 Historique des précipitations en Afrique de l'ouest	5
2 MODÉLISATION DES VALEURS EXTRÊMES	6
2.1 Modélisation GEV par estimation du maximum de vraisemblance	6
2.1.1 Méthodes statistiques classiques de modélisation des extrêmes	6
2.1.2 Inférence avec la GEV	11
2.2 Modélisation hiérarchique bayésienne	15
2.2.1 Introduction à la modélisation hiérarchique bayésienne	15
2.2.2 Inférence bayésienne	15
2.2.3 Structure des modèles hiérarchiques	18
2.2.4 Méthodes d'inférence	23
3 APPLICATIONS ET RÉSULTATS	27
3.1 Données réelles : échantillonnage et prétraitement.	27

3.1.1	Évolution des précipitations	28
3.1.2	Dépendance spatiale et temporelle des paramètres GEV	29
3.2	Modèles	33
3.2.1	Modèles GEV classiques	33
3.2.2	Modèles bayésiens	34
3.3	Choix et diagnostic du meilleur modèle	34
3.3.1	Critères de choix des modèles GEV	34
3.3.2	Diagnostic	35
3.4	Niveaux de retour	35
3.4.1	Modèle GEV classique	35
3.4.2	Modèle bayésien	37
CONCLUSION ET PERSPECTIVES		37
Annexe		40
Bibliographie		42

Table des figures

1.1	Rapport de synthèse du GIEC, 2014 montrant du haut vers le bas, l'évolution moyenne de la température, des précipitations et du niveau de la mer entre 1986-2005 à gauche et 2081-2100 à droite,	4
1.2	Inondation à Dakar (Sénégal) suite à une pluie le 05/08/2022 causant un mort et des dégâts matériels non négligeables	4
1.3	Inondation à Maiduguri (Nigeria) suite à une crue d'un barrage en crue le 10/09/2024. Pire inondation de la ville depuis 30 ans	4
2.1	Simulation des données journalières sur 10 ans selon la loi GEV avec $\xi > 0$. . .	8
2.2	Simulation de 1000 données en considérant $\mu = 0$, $\sigma = 1$ et différentes valeurs de ξ selon le type de distribution	10
3.1	Zone ouest africaine sur laquelle les données ont été échantillonnées.	28
3.2	Visualisation des précipitations maximales annuelles.	29
3.3	Cartographie des paramètres du modèle GEV.	30
3.4	Évolution dans le temps du paramètre de position en fonction de la latitude. . .	30
3.5	Évolution dans le temps du paramètre de position en fonction de la longitude. .	31
3.6	Évolution dans le temps du paramètre d'échelle en fonction de la latitude. . . .	31
3.7	Évolution dans le temps du paramètre d'échelle en fonction de la longitude. . . .	32
3.8	Évolution des paramètres GEV en fonction du temps.	32
3.9	Adéquation entre les quantiles empiriques et ceux théoriques du modèle 3. . . .	35
3.10	Niveaux de retour du modèle GEV classique sur 50 et 100 ans	36
3.11	Convergence des chaînes	37
3.12	Auto corrélation	37
3.13	Carte des niveaux de retour sur 50 et 100 ans du modèle à effets aléatoires par station	38

Liste des tableaux

3.1	Les modèles GEV classiques.	33
3.2	Tableau montrant les valeurs des paramètres estimés.	33
3.3	Valeurs des critères pour chaque modèle	35

INTRODUCTION GÉNÉRALE

Les précipitations jouent un rôle crucial dans les cycles hydrologiques, l'agriculture, la gestion des ressources en eau et la prévention des risques naturels. Cependant, leur variabilité spatio-temporelle, accentuée par les changements climatiques, rend leur modélisation complexe. Une meilleure compréhension et prévision des précipitations sont essentielles pour atténuer les risques climatiques et optimiser les politiques d'adaptation. L'étude vise à analyser les risques climatiques liés aux extrêmes pluviométriques (ici, les inondations) en Afrique de l'ouest afin de développer un modèle prédictif des précipitations intégrant des données géographiques. Les résultats seront présentés sous forme de cartes, offrant une visualisation claire des niveaux de retour, indispensables pour des recommandations opérationnelles.

Ainsi, le document se structure en trois chapitres selon les principaux objectifs précités. Le premier chapitre est consacré à l'étude des risques climatiques en particulier, des conséquences négatives des phénomènes climatiques extrêmes, de l'impact des changements climatiques sur les régimes pluviométriques et de l'historique des précipitations dans la zone ouest africaine. Le deuxième concerne la modélisation : la modélisation classique en utilisant la Generalized Extreme value distribution ([GEV](#)) avec la méthode Spatial Maximum Likelihood Estimation ([SMLE](#)) dans une première partie puis la modélisation bayésienne dans la seconde en utilisant l'algorithme Hamiltonian Monte Carlo ([HMC](#)). Le dernier chapitre présente les différents résultats notamment les valeurs des paramètres estimés, les différents diagnostics et les niveaux de retour sous forme des cartes.

Cette recherche basée sur l'approche Statistique propose un outil de prévision adaptable à différents contextes géographiques. Les résultats pourront éclairer les décideurs publics et les acteurs de la gestion des risques.

LES ÉVÈNEMENTS CLIMATIQUES EXTRÊMES

1.1 Les risques climatiques.

Le risque climatique désigne la probabilité que des événements climatiques extrêmes (inondations, sécheresses, tempêtes, etc.) ou des changements climatiques progressifs (élévation du niveau de la mer, acidification des océans, etc.) aient des conséquences négatives sur les sociétés, les économies et les écosystèmes. À cet effet, il est l'une des préoccupations majeures dans le monde. Les principales causes du risque climatique sont liées aux activités humaines, en particulier l'émission de gaz à effet de serre due à la combustion d'énergies, à la déforestation, . . . et la modification des sols et des écosystèmes, qui réduit leur capacité à absorber le CO_2 . Il existe plusieurs types d'impacts liés à ces risques et ils se font sentir de diverses manières selon les régions. Les principaux aspects sont entre autres :

- **Événements extrêmes** : augmentation de la fréquence et de l'intensité des inondations, des sécheresses, des vagues de chaleur, etc.
- **Élévation du niveau de la mer** : submersion des zones côtières, salinisation des eaux, etc.
- **Sécurité alimentaire** : diminution des rendements agricoles, perturbation des pêcheries, insécurité alimentaire, etc.
- **Santé humaine** : augmentation des maladies liées à la chaleur, problèmes de santé mentale, etc.
- **Migration et déplacement** : déplacement de populations dus aux événements extrêmes et à la dégradation de l'environnement.
- **Économie** : destruction des infrastructures, pertes de production, augmentation des coûts liés aux catastrophes, etc.

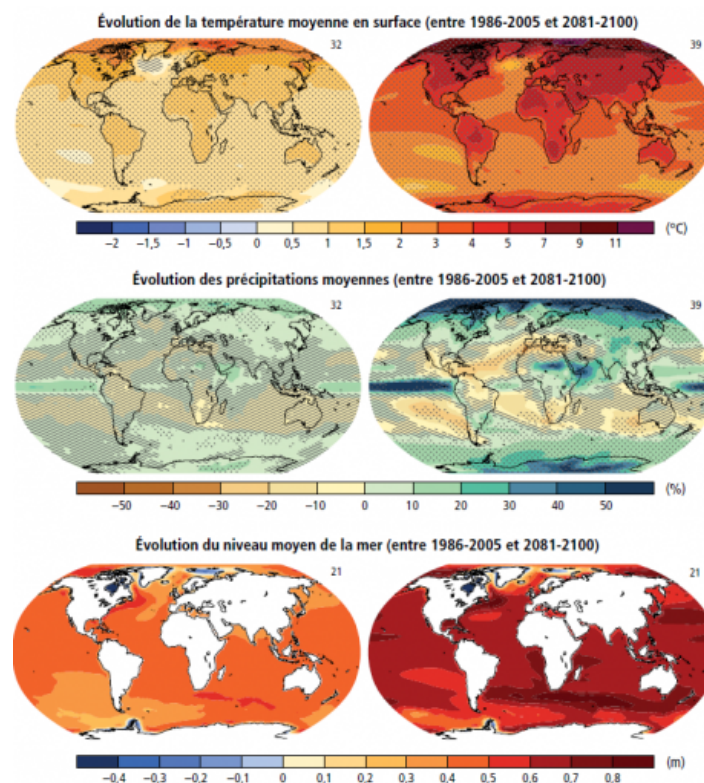


FIGURE 1.1 – Rapport de synthèse du GIEC, 2014 montrant du haut vers le bas, l'évolution moyenne de la température, des précipitations et du niveau de la mer entre 1986-2005 à gauche et 2081-2100 à droite, .

1.2 Extrêmes pluviométriques

l'Afrique de l'ouest connaît des saisons des pluies intenses, et on note un changement climatique à travers une augmentation de la fréquence et de l'intensité des précipitations. Parmi les pays les plus affectés, le Niger, le Nigeria, le Sénégal semblent présenter les pires conséquences. On entend par précipitations extrêmes, des quantités de pluies très rares dépassant un certain



FIGURE 1.2 – Inondation à Dakar (Sénégal) suite à une pluie le 05/08/2022 causant un mort et des dégâts matériels non négligeables



FIGURE 1.3 – Inondation à Maiduguri (Nigeria) suite à une crue d'un barrage en crue le 10/09/2024. Pire inondation de la ville depuis 30 ans

seuil dans une région donnée pendant une certaine période. Au sahel par exemple où une grande

partie de la région est désertique, une observation dépassant 100 mm en 24 heures peut être considérée comme valeur extrême à cause de la rareté de cette quantité et de l'intervalle de temps dans lequel elle a été enregistrée. Dans l'immédiat, très souvent on enregistre malheureusement des cas de décès ou de blessés graves, d'inondations, de glissements de terrain et à la longue des crises alimentaires, des déplacements de population, etc. Il y a aussi des enjeux économiques. Les inondations de 2020 ont coûté près de 1% du PIB au Niger ([Banque Mondiale \(2021\)](#)), un chiffre qui s'élève à 3-5% si l'on inclut les sécheresses récurrentes Programme des Nations Unies pour l'Environnement [PNUE, \(2022\)](#). Les figures 1.2 et 1.3 illustrent des cas d'inondation survenus au Sénégal et au Nigeria avec de lourdes conséquences négatives.

1.3 Historique des précipitations en Afrique de l'ouest

Comme le montre la figure 1.1, il y a une évolution des précipitations à l'échelle mondiale. En Afrique de l'ouest, l'analyse historique révèle trois périodes clés :

- 1968-1985 : La grande sécheresse sahélienne
- 1990-2010 : Le retour progressif des pluies
- 2010-2024 : L'ère des précipitations explosives

Selon certaines études +30% d'intensité des pluies extrêmes seront enregistrées d'ici 2050 avec des saisons plus courtes mais plus violentes. Dans ce document, nous allons évaluer le risque climatique dû aux événements extrêmes en particulier l'inondation, dans la région ouest-africaine. Parmi les multiples causes de ces inondations, nous nous concentrons sur celle due aux pluies torrentielles. Des études parallèles sont réalisées dans différents domaines afin de trouver des solutions. En Statistique, une modélisation des extrêmes se fait afin de gérer les risques liés aux extrêmes à travers la prévention (définition des seuils d'alerte) et l'anticipation (construction des infrastructures).

MODÉLISATION DES VALEURS EXTRÊMES

La méthode statistique pour analyser et prédire les événements extrêmes comme les inondations est la modélisation. Elle consiste à utiliser des outils mathématiques et statistiques pour représenter le phénomène d'étude en tenant souvent compte de l'incertitude et de la variabilité des données. Nous utilisons dans ce document, deux approches pour parvenir à nos résultats : la première est classique ; elle est basée sur la distribution générale des valeurs extrêmes GEV en raison des données échantillonnées selon des blocs de maxima annuels. La seconde approche est la modélisation bayésienne qui est pratique même avec de petites bases de données et qui permet d'intégrer des connaissances des études antérieures.

2.1 Modélisation GEV par estimation du maximum de vraisemblance

2.1.1 Méthodes statistiques classiques de modélisation des extrêmes

Les événements extrêmes sont difficiles à mesurer à cause de leur rareté, de plus, les séries d'observation sont souvent trop courtes pour garantir une estimation robuste des grands quantiles [Buishand \(1991\)](#). La Théorie des Valeurs Extrêmes (TVE) [Coles \(2001\)](#) ; [Katz et al.\(2002\)](#) ; [Naveau et al. \(2005\)](#) se présente comme la statistique appropriée pour étudier les distributions extrêmes. Elle repose d'une part sur deux méthodes d'échantillonnage : l'échantillonnage Bloc Maxima Analysis (BMA) qui consiste à sélectionner la valeur maximale par période (un an en hydrologie) et l'échantillonnage Peaks Over Threshold (POT) qui sélectionne toutes les valeurs au-dessus d'un seuil. D'autre part, la TVE repose sur des modèles théoriques associés aux méthodes ci-dessus. La distribution GEV pour les séries par bloc maximal et celle Generalized Pareto Distribution (GPD) pour les séries de dépassement d'un seuil.

2.1.1.1 Théorie des valeurs extrêmes

L'étude des comportements probabilistes des valeurs les plus grandes ou les plus petites observées dans un échantillon de données (valeurs extrêmes) est réservée à la TVE. Ainsi, contrairement aux méthodes statistiques classiques qui se concentrent sur la moyenne ou la variance, la TVE s'intéresse aux queues de distribution, là où se situent les événements rares. Le théorème fondamental de la TVE [Fisher et Tippet \(1928\)](#) et [Gnedenko \(1943\)](#) également connu sous le nom de théorème de Fisher-Tippett-Gnedenko, est un résultat central qui décrit les différentes formes des distributions limites des maxima (ou minima) de suites de variables aléatoires (v.a) indépendantes et identiquement distribuées (iid). Par analogie, ce théorème s'articule comme le théorème central limite qui décrit la convergence de la somme de v.a vers une loi normale.

Théorème 2.1. — *Fisher-Tippett-Gnedenko*

Soient X une suite de 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 :

$$\mathbb{P} \left(\frac{M_n - b_n}{a_n} \leq x \right) \xrightarrow{n \rightarrow \infty} G(x), \quad (2.1)$$

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

2.1.1.2 Méthodes de collecte des données

- BMA : Blocs maxima annuels

La méthode de sélection par bloc des maxima est l'une des techniques utilisées en TVE pour analyser les événements extrêmes. Elle consiste à diviser les données par blocs (des années, des mois, etc.) puis à sélectionner la valeur maximale dans chaque bloc.

Soient $(X_i)_{i \in \{1, \dots, M\}}$ une série temporelle de taille M et n blocs de taille $m = 365$ (blocs annuels). Pour chaque bloc j où $j = 1, \dots, n$, le maximum est donné par :

$$M_j = \max(X_{1,j}; X_{2,j}; \dots; X_{m,j}) \quad (2.2)$$

L'échantillon recueilli à la fin de ces étapes représente les données à ajuster selon un modèle des valeurs extrêmes (GEV). Avec ce mode d'échantillonnage, il est indispensable de disposer d'une longue série de données afin d'avoir un échantillon de maxima assez représentatif.

Définition 2.1. La loi du max

Considérons des v.a $X_i, i \in \{1, 2, \dots, n\}$, la loi du maximum est la probabilité telle que le maximum des X_i ne dépasse pas le quantile x (la valeur observée) : $\mathbb{P} \left(\max_{i \in \{1, 2, \dots, n\}} (X_i) < x \right)$

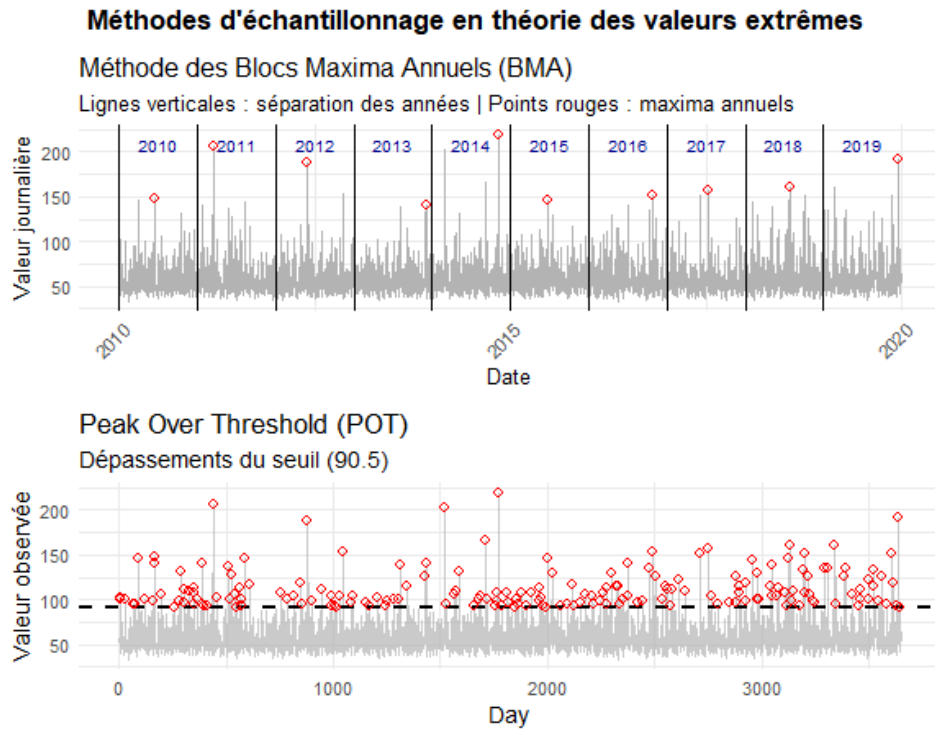


FIGURE 2.1 – Simulation des données journalières sur 10 ans selon la loi GEV avec $\xi > 0$

- POT : Peaks Over Threshold

Définition 2.2. Loi des dépassements

Soit u la valeur au delà de laquelle se fait la sélection des v.a. La loi des dépassements du seuil u est donnée par : $\mathbb{P}(X_i - u < x / X_i > u)$.

Le choix du seuil est très déterminant dans l'étude des valeurs extrêmes. Un seuil trop bas introduit un biais, tandis qu'un seuil trop haut augmente la variance. Le choix de la méthode d'échantillonnage dépend du type d'analyse et de la quantité de données dont on dispose. La méthode BMA basée sur les maxima par bloc est simple et robuste mais fait perdre des informations (une par bloc) tandis que la méthode POT basée sur le dépassement d'un seuil offre une meilleure utilisation des données mais avec le choix du seuil très critique.

2.1.1.3 Modèles théoriques : GEV et GPD

- Loi généralisée des valeurs extrêmes :

Le terme valeurs extrêmes est associé aux distributions introduites par Fisher et Tippet parce

qu'elles peuvent être obtenues comme lois limites (quand la taille de l'échantillon est grande) de la valeur maximale de n v.a iid (2.2). Introduite par JENKINSON (1955), la loi GEV regroupe les trois types de distributions développées par Fisher-Tippett définie par les équations (2.4), (2.5) et (2.6). Sa fonction de répartition (fdr) est donnée par :

$$H_{\xi,\mu,\sigma}(x) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]_+^{-1/\xi} \right\} & \text{si } \xi \neq 0, \\ \exp \left\{ - \exp \left(- \frac{x-\mu}{\sigma} \right) \right\} & \text{si } \xi = 0 \end{cases} \quad (2.3)$$

où : $1 + \xi \left(\frac{x-\mu}{\sigma} \right) > 0$; $a_+ = \max(a, 0)$ et x la valeur de la variable aléatoire.

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

$\xi < 0$: Distribution de Weibull (queue courte)

$$\text{Frechet : } \Phi_\alpha(x) = \begin{cases} 0 & x \leq 0 \\ \exp\{-x^{-\alpha}\} & x > 0 \end{cases}, \quad \alpha > 0 \quad (2.4)$$

$$\text{Weibull : } \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\} & x \leq 0 \\ 1 & x > 0 \end{cases}, \quad \alpha > 0 \quad (2.5)$$

$$\text{Gumbel : } \Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R} \quad (2.6)$$

Les représentations correspondants à ces distributions sont données ci-dessous par la figure 2.2.

- Distribution Pareto généralisée.

La loi standard de la GPD est donnée par :

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta} \right)^{-1/\xi}, & x \geq 0 \text{ si } \xi > 0, \\ 0 \leq x \leq -\beta/\xi \text{ si } \xi < 0, \\ 1 - \exp \left(-\frac{x}{\beta} \right), & x \geq 0 \text{ si } \xi = 0. \end{cases} \quad (2.7)$$

2.1.1.4 Approches d'estimation des paramètres

La méthode flexible et en général la plus utilisée pour estimer les paramètres inconnus θ est le maximum de vraisemblance.

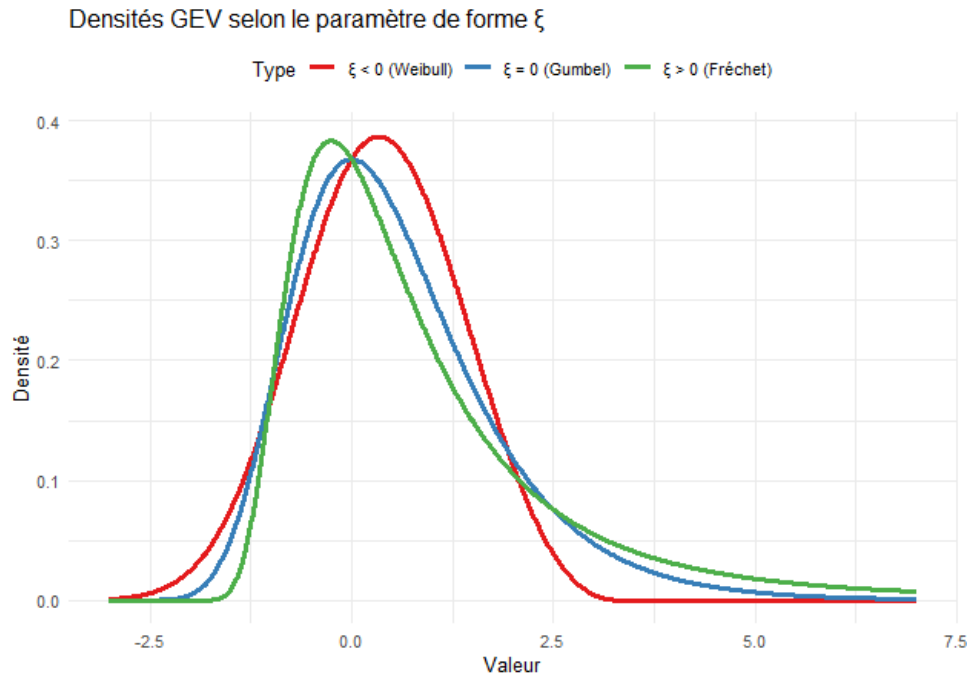


FIGURE 2.2 – Simulation de 1000 données en considérant $\mu = 0$, $\sigma = 1$ et différentes valeurs de ξ selon le type de distribution

Définition 2.3. Fonction de vraisemblance

Soit x_1, \dots, x_n des v.a iid ayant une densité de probabilité $f(x; \theta)$ où θ est le vecteur des paramètres à estimer. La fonction de vraisemblance est la probabilité des données observées en fonction de θ définie par :

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) \quad (2.8)$$

Le principe de l'estimation du maximum de vraisemblance est d'adopter le modèle le plus vraisemblable, car de tous les modèles considérés, c'est celui qui attribue la plus grande probabilité aux données observées. En général, on utilise la fonction vraisemblance logarithmique qui est plus pratique :

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i; \theta) \quad (2.9)$$

Nous pouvons illustrer quatre approches principales pour spatialiser les distributions extrêmes. Le but est de réduire la dépendance entre les v.a.

- i L'analyse fréquentielle régionale (Regional Frequency Analysis (RFA)) : [Hosking et Wallis \(1997\)](#)

Elle consiste à définir des sousrégions considérées homogènes pour lesquelles les paramètres sont assumés constants. La simplicité et la robustesse de cette approche la rendent très populaire en hydrologie.

- ii L'interpolation des distributions extrêmes (Local Fit and Interpolation Local Fit Interpolation (LFI)) : [Blanchet et Lehning \(2010\)](#)

La distribution des pluies extrêmes est estimée localement (à chaque station) puis les niveaux de retour ou les paramètres de la distribution sont interpolés. Différents interpolateurs peuvent être utilisés (régression, plus proche voisin, spline, krigeage, ...).

- iii L'estimation par maximum de vraisemblance d'une distribution spatiale (Spatial Maximum Likelihood Estimation, [SMLE](#))

Cette approche consistant à incorporer des covariables spatiales dans l'estimation par maximum de vraisemblance s'appuie sur les travaux de [Buishand \(1991\)](#). Si de nombreuses études ont utilisé cette approche pour analyser l'évolution temporelle des extrêmes, il y a très peu d'études qui appliquent cette méthode pour spatialiser les distributions extrêmes et proposer des cartes d'aléa [Buishand \(1991\)](#), [Blanchet et Lehning \(2010\)](#).

- iv L'approches bayésienne : [Cooley et al. \(2007\)](#).

Cette approche a récemment été employée pour spatialiser les distributions de précipitations extrêmes. C'est une extension de l'approche SMLÉ dans un cadre bayésien. Elle est recommandable lorsque la répartition spatiale des paramètres de la distribution est trop compliquée pour être correctement représentée par des relations simples.

Nous allons développer dans ce document, les deux dernières approches (l'approche SMLÉ et l'approche bayésienne) en utilisant la GEV

2.1.2 Inférence avec la GEV

2.1.2.1 Structures des modèles

Soit $Z_i, i = 1, \dots, m$ un processus tel que $Z_i \sim GEV(\mu, \sigma, \xi)$ un modèle non stationnaire où $(\mu, \sigma, \xi) = \theta$ représentent respectivement les paramètre de position, d'échelle et de forme. D'après l'équation (2.8), la fonction de vraisemblance s'écrit :

$$L(\theta) = \prod_{i=1}^m GEV(z_i, \mu, \sigma, \xi) \quad (2.10)$$

où $\mu \in \mathbb{R}, \sigma > 0$ et $\xi \in \mathbb{R}$.

De même en se référant à (2.9) on a la fonction de vraisemblance logarithmique suivante :

$$\ell(\mu, \sigma, \xi) = - \sum_{i=1}^m \left\{ \log \sigma + \left(1 + \frac{1}{\xi} \right) \log \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right] + \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (2.11)$$

Avec $1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) > 0$

La structure des modèles dépend du type de processus de Z_i .

1. Estimation de μ

Dans le cadre des valeurs extrêmes, μ représente le paramètre de position ou de localisation (centre) de la distribution.

- Si la série dépend linéairement du temps : $\mu(t) = \beta_0 + \beta_1 t$
- Si la série est sous forme quadratique : $\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2$
- S'il y a un changement de régime : $\mu(t) = \begin{cases} \mu_1 & \text{si } t \leq t_0, \\ \mu_2 & \text{si } t > t_0. \end{cases}$
- Avec effets des covariables $y(t)$, $\mu(t) = \beta_0 + \beta_1 y(t)$

2. Estimation de σ

Le paramètre d'échelle σ permet de quantifier la distribution des extrêmes autour de μ . Il est toujours positif c'est pourquoi on utilise souvent la fonction exponentielle pour assurer sa positivité. $\sigma(t) = \exp(\beta_0 + \beta_1 t)$.

3. Estimation de ξ

Le paramètre de forme permet de contrôler le comportement des extrêmes à travers la queue de distribution. Selon qu'il soit positif, négatif ou nul on a des distributions correspondantes définies dans la section 2.1.1.3 par les équations (2.4), (2.5) et (2.6).

La structure générale de ces modèles peut se présenter sous la forme

$$\theta(t) = h(X^T \beta)$$

où θ est le vecteur des paramètres μ, σ, ξ ; X la matrice des vecteurs du modèle dont la première ligne est unitaire; β le vecteur des paramètres et h une fonction spécifique (le produit scalaire par exemple).

2.1.2.2 Choix du modèle

Après la conception des différents modèles susceptibles d'expliquer et de prédire les données, il est indispensable de choisir le meilleur capable d'y parvenir avec plus de précision, moins d'erreurs et une bonne adéquation de la prédiction par rapport aux données.

Définition 2.4. La log-vraisemblance profilée

Soit θ le vecteur des paramètres à estimer de dimension d , θ_i une composante de θ . La log-vraisemblance profilée de θ peut s'écrire comme $\ell(\theta_i, \theta_{-i})$ où θ_{-i} représente tous les composants de θ sauf θ_i . La vraisemblance profilée de θ_i est définie par : $\ell_p(\theta_i) = \max \ell(\theta_i, \theta_{-i})$.

Cette définition peut se généraliser au cas où θ est divisé en deux sous vecteurs $\theta^{(1)}$ et $\theta^{(2)}$ de

dimensions respectives k et $d - k$. Ainsi, la log-vraisemblance profilée de $\theta^{(1)}$ est donnée par :

$$\ell_p(\theta^{(1)}) = \max_{\theta^{(2)}} \ell(\theta^{(1)}, \theta^{(2)})$$

Théorème 2.2. — Soient x_1, \dots, x_n des réalisations iid d'une famille paramétrique \mathcal{F} et $\hat{\theta}_0$ l'estimateur du maximum de vraisemblance du paramètre $\theta_0 = (\theta^{(1)}, \theta^{(2)})$ de dimension d où $\theta^{(1)}$ est un sous ensemble de dimension k de θ_0 . Alors pour n grand on a formule suivante de la déviance profilée :

$$\mathcal{D}_p = 2\{\ell(\hat{\theta}_0) - \ell_p(\theta^{(1)})\} \sim \chi_{d-k}^2 \quad (2.12)$$

L'une des application du théorème 2.2 est la sélection du modèle.

Soit un modèle \mathcal{M}_1 avec θ (de dimension d) le vecteur des paramètres et soit $\mathcal{M}_0 \subset \mathcal{M}_1$ un modèle dont le vecteur des paramètres est un sous ensemble de θ (dimension $k < d$). Ainsi, $\theta_0 = (\theta^{(1)}, \theta^{(2)})$ avec $\dim(\theta^{(1)}) = k$. Supposons que $\ell_1(\mathcal{M}_1)$ et $\ell_0(\mathcal{M}_0)$ sont respectivement la log-vraisemblance des modèles \mathcal{M}_1 et \mathcal{M}_0 . Alors la déviance s'écrit :

$$\mathcal{D} = 2\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\} \quad (2.13)$$

Pour des modèles imbriqués, le meilleur est donc celui avec une faible déviance et moins de paramètres possibles. Pour des modèles non imbriqués (linéaires et exponentiels par exemple) les critères comme l'Akaike Information Criterion ([AIC](#)) et le Bayesian Information Criterion ([BIC](#)) sont souvent utilisés. Le premier plus adapté pour la prédiction (régression) et le second pour l'inférence. Leur différence se trouve dans la pénalisation qu'ils apportent vis à vis de la déviance. L'AIC pénalise la déviance par le nombre de paramètres $2k$, k étant le nombre de paramètre et le BIC pénalise plus sévèrement pour les grands échantillons $\ln(n)$, n le nombre d'observations. Nous avons les formules correspondantes ci-dessous :

$$AIC = \mathcal{D} + 2k \quad (2.14)$$

$$BIC = \mathcal{D} + k \cdot \ln(n) \quad (2.15)$$

Donc 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.

2.1.2.3 Diagnostic du modèle

Après avoir choisi le meilleur modèle selon les critères définis dans la section précédente, il est important de le diagnostiquer, de vérifier la précision du modèle en terme de sa concordance avec les données qui ont été utilisées pour l'estimer.

Définition 2.5. Probability Plot

Soient $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ des v.a ordonnées de fonction de distribution F et $\tilde{F}(x) = \frac{i}{n+1}$ la fonction de répartition empirique avec $x_{(i)} \leq x \leq x_{(i+1)}$. Le graphe du diagramme des probabilités de la fonction de répartition estimé \hat{F} est donnée par :

$$\left\{ \left(\hat{F}(x_{(i)}), \tilde{F}(x) \right) : i = 1, \dots, n \right\} \quad (2.16)$$

Si \hat{F} est un modèle raisonnable pour la distribution F , les points du diagramme de probabilité doivent être proches de la diagonale unité. Des écarts substantiels par rapport à la linéarité témoignent d'un échec de \hat{F} comme modèle pour les données.

Définition 2.6. Quantile Plot

Soient $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ des v.a dont la fonction de répartition estimée est \hat{F} . Le graphe des quantiles est donnés par :

$$\left\{ \left(\hat{F}^{-1}\left(\frac{i}{n+1}\right), x_{(i)} \right) : i = 1, \dots, n \right\} \quad (2.17)$$

Si \hat{F} est une estimation raisonnable de F , le graphe des quantiles devrait également se composer de points proches de la diagonale unitaire. Le diagramme des probabilités et le diagramme des quantiles contiennent les mêmes informations exprimées sur une échelle différente. Cependant, la perception obtenue sur des échelles différentes peut être importante [Coles \(2001\)](#).

2.1.2.4 Niveaux de retour

Le niveau de retour associé à la période p d'un quantile d'ordre $1 - 1/p$ de H_ξ est le quantile qui est dépassé avec une probabilité $1/p$. Selon le type de loi on a la formule du niveau de retour.

1. Loi du maximum (GEV)

Le quantile z_p au niveau $1 - p$ de la loi GEV s'exprime par :

$$z_p = H_\xi^{-1}(1 - p) = \begin{cases} \mu - \frac{[1 - \{-\log(1 - 1/p)\}^{-\xi}]\sigma}{\xi} & \text{si } \xi \neq 0, \\ \mu - \sigma \log(-\log(1 - 1/p)) & \text{si } \xi = 0. \end{cases} \quad (2.18)$$

2. Loi des dépassements (GPD)

Pour un seuil u fixé, la probabilité de dépassement conditionnelle s'écrit :

$$P(X > x + u) = P(X - u > x \mid X > u) \cdot P(X > u) \quad (2.19)$$

$$P(X > x + u) = \left(1 - G_{\xi, \beta}(x)\right) \cdot \frac{N_u}{N} \quad (2.20)$$

Avec $x = G_{\xi,\beta}^{-1}\left(1 - \frac{p_{x+u} \cdot N_u}{N}\right)$ où N_u est le nombre d'observations dépassant u , N le nombre total d'observations et $G_{\xi,\beta}$ la fonction de répartition GPD.

2.2 Modélisation hiérarchique bayésienne

2.2.1 Introduction à la modélisation hiérarchique bayésienne

La modélisation des précipitations extrêmes nécessite une approche pouvant tenir compte de la complexité des dépendances spatio-temporelles, de l'incertitude sur les paramètres et la rareté des observations. Les modèles Bayesian Hierarchical Models ([BHM](#)) permettent d'intégrer ces contraintes à travers des niveaux de variabilité emboîtés introduits par [Cooley et al. \(2007\)](#). Les méthodes GEV classiques ignorent souvent la complexité des dépendances spatiales, la non stationnarité des phénomènes extrêmes et ont du mal à les ajuster [Lehmann, E. A. \(2012\)](#). L'approche bayésienne permet d'incorporer des connaissances expertes et pré-existantes via les priors. En effet, avec la modélisation classique, il peut exister des structures de dépendance non linéaires et non stationnaires des quantiles extrêmes spatio-temporels. De plus, l'homogénéité spatiale n'est pas toujours garantie ce qui complique la modélisation. Le calcul de la vraisemblance sur une grande région peut s'avérer très compliqué. Ainsi, l'inférence bayésienne va nous permettre de régulariser l'estimation des extrêmes dans ce cadre. Nous allons appliquer les méthodes Monte Carlo Markov Chain ([MCMC](#)) à la GEV pour obtenir une meilleure précision dans le calcul des distributions à posteriori. Nous utiliserons principalement les algorithmes NUTS (No U Turn Sampler) de [Hoffman et Gelman \(2014\)](#) qui sont une version améliorée de [HMC](#).

2.2.2 Inférence bayésienne

La statistique bayésienne est une approche statistique qui repose sur le théorème de Bayes pour la modélisation, l'inférence et la prise de décision à travers des probabilités conditionnelles. Elle cherche dans un premier temps à modéliser à l'aide des distributions de probabilité, les paramètres inconnus puis effectue l'inférence sur ces dernières. Les connaissances préalables (priors) sur les paramètres sont intégrées dans l'analyse. Après, l'incertitude sur le paramètre θ est modélisée par une loi de probabilité $\pi(\theta)$. On fini par l'inférence bayésienne qui consiste à déterminer la loi a posteriori $\pi(\theta | X = x)$.

2.2.2.1 Théorème de Bayes et distributions a posteriori

Définition 2.7. Soit $X = (x_1, x_2, \dots, x_n)$ une suite de v.a.

- L'information a priori sur le paramètre θ , représente la distribution de probabilité qui tient compte des connaissances ou croyances initiales sur θ en dehors de celles apportées par les observations.
- On modélise cette information a priori au travers d'une loi de probabilité, appelée loi a priori. Sa densité est notée $\pi(\theta)$.
- Le modèle statistique paramétrique bayésien est un cadre probabiliste où les paramètres θ sont traités comme des v.a. Le modèle combine la loi a priori $\pi(\theta)$ et la loi des données ou vraisemblance. On appelle loi des observations, la loi conditionnelle de X sachant θ . Sa densité est notée $f(x | \theta)$.

Théorème 2.3. — *Version continue du théorème de Bayes.* [Jerôme DUPUIS \(2007\)](#)

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

$$g(y|x) = \frac{f(x | y)g(y)}{\int f(x | y)g(y)dy} \quad (2.21)$$

On peut éviter de calculer le terme au dénominateur $g_X(x) = \int f(x | y)g(y)dy$ qui ne dépend pas de y . Il est vu comme une constante de normalisation. On peut raisonner par proportionnalité et on utilisera le symbole \propto pour signifier que deux lois sont proportionnelles. On a donc :

$$g(y | x) \propto f(x | y)g(y) \quad (2.22)$$

Définition 2.8. Loi à posteriori

Soit l'espace de probabilité $(\Theta, \mathcal{F}, \pi)$. La loi a posteriori est la loi conditionnelle de θ sachant X , définie par la densité $\pi(\theta | X = x)$ évaluée en $X = x$. On la note par $\pi(\cdot | X)$. Elle donne les valeurs attendues du paramètre en tenant compte des observations et du choix d'a priori. En se référant à (2.21), on a :

$$\pi(\theta | x) = \frac{f(x | \theta)\pi(\theta)}{\int f(x | \theta)\pi(\theta) d\theta} \quad (2.23)$$

La notation (2.22) est souvent reformulée de la façon suivante : $\pi(\theta | x) \propto L(\theta; x)\pi(\theta)$

où $L(\theta; x) := f(x | \theta)$ est la vraisemblance.

Exemple 2.1. [Jerôme DUPUIS \(2007\)](#)

On dispose d'un vecteur d'observations : $x = (x_1, \dots, x_i, \dots, x_n)$, et on considère le modèle bayésien suivant :

$$X_i \mid \theta \sim \text{Bernoulli}(\theta) \quad \text{et} \quad \theta \sim \text{Beta}(a, b).$$

On a :

$$f(x \mid \theta) = \prod_{i=1}^n P(X = x_i \mid \theta) = \theta^s (1 - \theta)^{n-s}$$

où $s = \sum_{i=1}^n x_i$.

Comme $\theta \sim \text{Beta}(a, b)$, on a :

$$\pi(\theta) = \frac{1}{B(a, b)} \theta^{a-1} (1 - \theta)^{b-1} \mathbf{1}_{[0,1]}(\theta).$$

D'autre part on a :

$$\pi(\theta \mid x) = \frac{f(x \mid \theta) \pi(\theta)}{\int_{\Theta} f(x \mid \theta) \pi(\theta) d\theta}.$$

On peut vérifier que :

$$\int_{\Theta} f(x \mid \theta) \pi(\theta) d\theta = B(a, b)$$

où $\alpha = a + s$ et $\beta = b + n - s$. D'où :

$$\pi(\theta \mid x) = \frac{\theta^{\alpha-1} (1 - \theta)^{\beta-1}}{B(\alpha, \beta)} \mathbf{1}_{[0,1]}(\theta).$$

Par conséquent :

$$\theta \mid x \sim \text{Beta} \left(a + \sum_{i=1}^n x_i, b + n - \sum_{i=1}^n x_i \right).$$

2.2.2.2 Vraisemblance marginale et vraisemblance conditionnelle**Définition 2.9.** Vraisemblance conditionnelle [Pawitan \(2001\)](#)

La vraisemblance conditionnelle mesure la probabilité des données observées conditionnellement à un ensemble de paramètres ou effets aléatoires. Elle est utilisée dans les modèles hiérarchiques pour isoler une partie de la variabilité. Soient Y la donnée observée, θ le paramètre à estimer et X les variables explicatives ou effets fixes. La vraisemblance conditionnelle notée $L(\theta \mid Y, X)$ est donnée par :

$$L(\theta \mid Y, X) = \mathbb{P}(Y \mid \theta, X) \tag{2.24}$$

Cette fonction mesure la probabilité d'observer les données Y si les paramètres θ sont donnés.

Définition 2.10. Vraisemblance marginale

La vraisemblance marginale intègre la vraisemblance conditionnelle sur tous les effets aléatoires ou paramètres latents. Soient $x = x_1, \dots, x_n$ des paramètres latents et θ des hyperparamètres. D'après [Gelman et al. \(2013\)](#), la vraisemblance marginale est donnée par

$$L(\theta | Y) = \int \mathbb{P}(Y | \theta, x) \mathbb{P}(x | \theta) dx \quad (2.25)$$

Pour un modèle hiérarchique, la vraisemblance jointe se décompose en vraisemblance conditionnelle et distribution a priori : $\mathbb{P}(Y, x) = \mathbb{P}(Y | x) \cdot \mathbb{P}(x)$

2.2.3 Structure des modèles hiérarchiques

Les modèles hiérarchiques bayésiens sont en général structurés en 3 niveaux ou couches. Dans la première couche appelée modèle des données (vraisemblance) où $Y(s) \sim \mathbb{P}(Y(s) | X(s), \theta)$, on identifie la distribution conjointe des données à priori dépendantes grâce à une copule. Au niveau de la deuxième couche hiérarchique ou modèle du processus (régression spatiale latente), on utilise les paramètres de régression spatiale pour modéliser les dépendances spatiales des paramètres de distribution marginales $(\mu(s), \sigma(s), \xi(s))$. Nous capturons la dépendance spatiale résiduelle grâce à un processus latent GP (Gaussian Process) à cause de la dépendance linéaire entre les paramètres et les coordonnées géographiques (figures [3.4](#), [3.5](#), [3.6](#) et [3.7](#)). La dernière couche concerne les hyper paramètres qui contrôlent la distribution des paramètres latents (ou d'autres paramètres). [C. Bracken et al. \(2016\)](#). Le modèle complet s'écrit :

$$\mathbb{P}(Y, X, \theta) = \mathbb{P}(Y | X, \theta) \cdot \mathbb{P}(X | \theta) \cdot \mathbb{P}(\theta) \quad (2.26)$$

Et on cherche la distribution a posteriori : $\mathbb{P}(X, \theta | Y) \propto \mathbb{P}(Y | X, \theta) \mathbb{P}(X | \theta) \mathbb{P}(\theta)$

Notions sur variables latentes, processus latents et paramètres latents.

1. Une variable latente est une quantité inobservable qui influence les données observées. Généralement c'est une variable aléatoire avec une distribution (altitude, proximité de la mer, etc.).
2. Un processus latent est une évolution temporelle ou spatiale d'une variable latente. Il est souvent modélisé par des équations différentielles ou des chaînes de Markov.
3. Un paramètre latent est un paramètre inconnu du modèle qui n'est pas directement observable, mais qui peut être estimé à partir des données. Comme exemple, dans un modèle bayésien hiérarchique, la moyenne μ d'un groupe peut être un paramètre latent

si elle est elle-même tirée d'une distribution a priori. Ou dans la GEV, si le paramètre de localisation μ suit une loi normale, il devient un paramètre latent.

2.2.3.1 Modèle des données : Vraisemblance

Dans la couche de données, les observations extrêmes $Y(s, t)$, s (station) et t (année d'observation) sont modélisées comme dépendantes d'un processus latent $x(s, t)$ et aux effets fixes ou aléatoires θ .

$$Y(s, t) \mid x(s, t), \theta \sim \mathbb{P}(Y(s, t) \mid x(s, t), \theta) \quad (2.27)$$

Exemple 2.2. Si Y suit une GEV spatiale, alors

$$Y(s, t) \mid \mu(s, t), \sigma(s, t), \xi(s, t) \sim GEV(\mu(s, t), \sigma(s, t), \xi(s, t)) \quad (2.28)$$

Ce processus capte la dépendance spatiale et temporelle sur $\mu(s, t)$, $\sigma(s, t)$ et $\xi(s, t)$.

Définition 2.11. : Copules elliptiques [Renard \(2011\)](#); [Sang et Gelfand \(2010\)](#).

Les copules elliptiques sont des fonctions de dépendance utilisées pour modéliser les relations entre v.a multiples tout en gardant une structure de corrélation spécifique.

Exemple 2.3. Soient X et Y deux v.a dépendantes de distributions marginales F_X et F_Y dans cet ordre. La distribution conjointe $F_{X,Y}(x, y)$ s'écrit comme suit :

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$$

où C est la copule qui capture la dépendance entre X et Y .

La copule peut représenter des données spatiales avec n'importe quelle distribution marginale. Un cas particulier des copules est la copule gaussienne basée sur la distribution normale multivariée, généralement utilisée pour les dépendances linéaires.

Soit m le nombre de stations, la distribution jointe de ces m stations peut être modélisée comme une réalisation d'une copule elliptique gaussienne avec la distribution GEV marginale. Cette copule est caractérisée par une matrice de dépendance par paire notée Σ .

Exemple 2.4. Soit Y_1, Y_2, \dots, Y_m un vecteur aléatoire. La fonction de répartition conjointe F_C se définit à partir de la copule gaussienne, qui relie les marginales F_i (via leurs quantiles gaussiens $u_i = \phi^{-1}(F_i(y_i))$) à une loi normale multivariée Φ_Σ . On a :

$$F_C(y_1, \dots, y_m) = \Phi_\Sigma(u_1, \dots, u_m) \quad (2.29)$$

où $\Phi_{\Sigma}(u_1, \dots, u_m)$ est la fonction de répartition conjointe de la distribution normale multivariée avec Σ comme matrice de dépendance. ϕ étant la fonction de répartition d'une distribution normale standard. La densité de probabilité jointe correspondante est un produit entre la densité jointe gaussienne $\Psi_{\Sigma}(u_1, \dots, u_m)$ et le terme de jacobien $\prod_{i=1}^m \frac{du_i}{dy_i} = \prod_{i=1}^m \frac{f_i(y_i)}{\psi(u_i)}$:

$$f_C(y_1, \dots, y_m) = \frac{\prod_{i=1}^m f_i(y_i)}{\prod_{i=1}^m \psi(u_i)} \times \Psi_{\Sigma}(u_1, \dots, u_m) \quad (2.30)$$

Où f_i est la densité marginale GEV au site i , ψ la densité de la distribution normale standard qui est un terme d'ajustement de la transformation vers l'espace gaussien et Ψ_{Σ} la densité normale multivariée. La dépendance entre deux site i et j peut être capturée grâce à un modèle exponentiel :

$$\Sigma(i, j) = \exp \left\{ -\frac{\|s_i - s_j\|}{a} \right\} \quad (2.31)$$

Le paramètre a contrôle la portée spatiale de la dépendance. Plus a est grand, plus la dépendance persiste sur de longues distances. Le choix d'une décroissance exponentielle pour $\Sigma(i, j)$ est motivé par le fait que la décroissance exponentielle est adaptée à des phénomènes où la corrélation diminue rapidement avec la distance (ex. précipitations, températures). De plus, Pour que Σ soit une matrice de covariance définie positive, le noyau exponentiel doit être utilisé avec des distances Euclidiennes (ou une norme appropriée). C'est le cas ici.

On suppose que les paramètres GEV peuvent être décrit à travers une régression spatiale latente où les résidus $\omega_{\mu, \sigma, \xi}(s) \sim GP(0, C_{\mu, \sigma, \xi}(s, s'))$ (un processus stationnaire gaussien de moyenne 0 et de fonction de covariance notée $C_{\gamma}(s, s')$ avec $\gamma = (\mu, \sigma, \xi)^T$. Ainsi, la matrice de covariance correspondante est : $C_{\gamma}(\theta_{\gamma}) = [C_{\gamma}(s_i, s_j; \theta_{\gamma})]_{i,j=1}^m$ avec θ_{γ} les paramètres de la covariance. Donc pour une copule elliptique gaussienne C_m de dimension m , on a : [C. Bracken et al. \(2016\)](#).

$$(Y(s_1, t), Y(s_2, t), \dots, Y(s_m, t)) \sim C_m \left(\Sigma; \left\{ \mu(s), \sigma(s), \xi(s) \right\} \right) \quad (2.32)$$

La vraisemblance de la couche de données est donnée par l'équation suivante :

$$\text{GEV}(Y(s, t) | \mu(s, t), \sigma(s, t), \xi(s, t)) = \prod_{i=1}^m \text{GEV}(Y(s_i, t); \mu(s, t), \sigma(s, t), \xi(s, t)) \quad (2.33)$$

La densité conjointe :

$$f_{\mathbf{Y}}(\mathbf{y} | \mu, \sigma, \xi) = \left[\prod_{i=1}^m f_{\text{GEV}}(y_i | \mu_i, \sigma_i, \xi_i) \right] \times C_{\Sigma}^{\text{Gauss}}(F_1(y_1), \dots, F_m(y_m)) \quad (2.34)$$

Pour chaque observation y_i on a :

$$\log f_{\text{GEV}}(y_i \mid \mu_i, \sigma_i, \xi_i) = -\log \sigma_i - \left(1 + \frac{1}{\xi_i}\right) \log \left[1 + \xi_i \left(\frac{y_i - \mu_i}{\sigma_i}\right)\right] - \left[1 + \xi_i \left(\frac{y_i - \mu_i}{\sigma_i}\right)\right]^{-1/\xi_i}$$

valable pour $1 + \xi \left(\frac{y-\mu}{\sigma}\right) > 0$ et $\sigma > 0$. et la densité de la copule est donnée par :

$$C_{\Sigma}^{\text{Gauss}}(u) = \frac{1}{|\Sigma|^{1/2}} \exp \left(-\frac{1}{2} Q^{\top} (\Sigma^{-1} - I_m) Q \right)$$

Où $C_{\Sigma}^{\text{Gauss}}$ est la copule gaussienne, Σ la matrice de corrélation, $u_i = F_i(y_i)$ la cdf marginale GEV et $Q = \Phi^{-1}(u_i)$ est la distribution normale standard.

DÉMONSTRATION 1 : Dérivation de la densité de la copule gaussienne

Soit $U = (U_1, \dots, U_d)$ avec $U_i \sim \mathcal{U}(0, 1)$ liées par une copule gaussienne de matrice de corrélation Σ . La transformation en variables gaussiennes donne : $Q = (Q_1, \dots, Q_d)$, $Q_i = \Phi^{-1}(U_i)$, avec $Q \sim \mathcal{N}_d(0, \Sigma)$ La densité jointe gaussienne correspondante s'écrit comme suit :

$$f_Q(q) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} q^{\top} \Sigma^{-1} q \right) \quad (2.35)$$

En se référant à 2.30 et en procédant par un changement de variable $U = \Phi(Q)$ on a :

$$c_R^{\text{Gauss}}(u) = f_Q(\Phi^{-1}(u)) \cdot \prod_{i=1}^d \left| \frac{d}{du_i} \Phi^{-1}(u_i) \right| = \frac{1}{|\Sigma|^{1/2}} \exp \left(-\frac{1}{2} Q^{\top} \Sigma^{-1} Q \right) \cdot \prod_{i=1}^d \frac{1}{\phi(Q_i)}$$

On simplifie en utilisant $\phi(Q_i) = \frac{1}{\sqrt{2\pi}} \exp(-Q_i^2/2)$, et on obtient :

$$c_{\Sigma}^{\text{Gauss}}(u) = \frac{1}{|\Sigma|^{1/2}} \exp \left(-\frac{1}{2} Q^{\top} \Sigma^{-1} Q + \frac{1}{2} Q^{\top} I_d Q \right) = \frac{1}{|\Sigma|^{1/2}} \exp \left(-\frac{1}{2} Q^{\top} (\Sigma^{-1} - I_d) Q \right)$$

D'où la formule finale :

$$\boxed{c_{\Sigma}^{\text{Gauss}}(u) = \frac{1}{|\Sigma|^{1/2}} \exp \left(-\frac{1}{2} Q^{\top} (\Sigma^{-1} - I_d) Q \right)} \quad (2.36)$$

où Σ est la matrice de corrélation ($d \times d$), $Q = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$ et I_d est la matrice identité de taille d . Cette dérivation montre comment la dépendance entre les uniformes est capturée par la structure gaussienne sous-jacente.

On obtient la log-vraisemblance complète en combinant la vraisemblance univariée et la

copule ce qui nous donne la formule ci-dessous.

$$\begin{aligned} \mathcal{L}(\boldsymbol{\gamma} \mid \mathbf{y}) = \sum_{i=1}^m & \left[-\log \sigma_i - \left(1 + \frac{1}{\xi_i} \right) \log \left(1 + \xi_i \frac{y_i - \mu_i}{\sigma_i} \right) - \left(1 + \xi_i \frac{y_i - \mu_i}{\sigma_i} \right)^{-1/\xi_i} \right] \\ & - \frac{1}{2} \log |\Sigma| - \frac{1}{2} \mathbf{Q}^\top (\Sigma^{-1} - I_m) \mathbf{Q} \end{aligned} \quad (2.37)$$

2.2.3.2 Modèle des paramètres : effets fixes et aléatoires

Il est question ici de la modélisation de la structure spatiale et temporelle des paramètres latents $x(s, t)$, généralement à travers des processus aléatoires ou gaussiens. Soit θ l'hyperparamètre qui contrôle la dépendance spatiale et temporelle. on a ;

$$x(s, t) \mid \theta \sim \mathbb{P}(x(s, t) \mid \theta) \quad (2.38)$$

Exemple 2.5. Processus gaussien

Les paramètres de régression spatio-temporelle sont modélisés par :

$$\mu(s, t) = \beta_0^\mu + X_\mu^T(s, t)\beta^\mu(s, t) + \omega^\mu(s, t) \quad (2.39)$$

$$\sigma(s, t) = \beta_0^\sigma + X_\sigma^T(s, t)\beta^\sigma(s, t) + \omega^\sigma(s, t) \quad (2.40)$$

$$\xi(s, t) = \beta_0^\xi + X_\xi^T(s, t)\beta^\xi(s, t) + \omega^\xi(s, t) \quad (2.41)$$

β_0^γ des termes d'interception constant ; $X_\gamma^T(s, t)$ des prédicteurs de taille p ;

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

$\omega^\gamma(s, t) \sim GP(0, C(s, s'))$

En général, les covariables utilisées sont la latitude et la longitude. Un coefficient de régression par covariable n'est pas approprié lorsqu'on est dans le cadre spatial avec une vaste région. Il est indispensable de tenir compte de la variation spatiale dans la régression pour chaque paramètre GEV d'où $\beta^\gamma(s)$. D'après [C. Bracket et al. \(2016\)](#), on peut représenter chaque coefficient de régression comme une somme pondérée de fonctions de bases radiales (noyaux gaussiens).

$$\beta_i^\gamma(s) = \sum_{j=1}^k c_{ij}^\gamma \eta_{ij}^\gamma(s; a_{ij}^\gamma) \quad i = 1, \dots, p. \quad (2.42)$$

Les fonctions de bases radiales distribuée sur l'ensemble du domaine sont sous la forme : $\eta_{ij}^\gamma(s; a_{ij}^\gamma) = \exp \left\{ -\frac{\|s - s_i\|^2}{(a_{ij}^\gamma)^2} \right\}$ où $(a_{ij}^\gamma)^2$ est un paramètre d'étendue déterminant l'étendue spatiale de la fonction de base. $\beta_i^\gamma(s)$ est le i ème coefficient de régression (variant spatialement) et γ représente les paramètres GEV. Les fonctions de base radiales, sont placées en des points du domaine appelés noeuds. La somme de ces noyaux gaussiens crée une surface de variation

régulière pour chaque coefficient de régression.

2.2.3.3 Modèle des hyper paramètres : priors

La méthode bayésienne requiert de disposer des informations préalables pour les paramètres de covariance (θ_γ). En considérant la copule gaussienne, la vraisemblance des résidus $\mathbb{P}(\omega \mid \theta_\gamma)$ s'obtient à partir de la densité normale multivariée. $\omega_\gamma \mid \theta_\gamma \sim \text{MNV}(0, \Sigma_\gamma)$ où $\Sigma_\gamma = C(\theta_\gamma)$. La forme paramétrique de la fonction de covariance définie en (2.31) devient :

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

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

2.2.4 Méthodes d'inférence

La modélisation spatiale des extrêmes pluviométriques peut capturer la dépendance spatiale entre les stations et réduire l'incertitude globale des estimations des niveaux de retour en empruntant de la force à travers les emplacements spatiaux [Cooley et al. \(2007\)](#).

Définition 2.12. Un processus spatial latent est un champ aléatoire (souvent Gaussien) sous-jacent qui structure la dépendance spatiale des observations. Soit $Y(s)$ un processus observé en un lieu $s \in S$, où $S \in \mathbb{R}^d$ est un domaine spatial. On suppose que $Y(s)$ dépend d'un processus latent $X(s)$ souvent non observable (l'altitude, la proximité de la mer, etc.). Le modèle général s'écrit :

$$Y(s) = \mu(s) + X(s) + \epsilon(s) \quad (2.43)$$

où $\mu(s)$ est une tendance (fixe ou dépendante de covariables) ; $X(s)$ est un processus spatial latent et $\epsilon(s)$ est un bruit blanc (erreur de mesure).

2.2.4.1 Algorithme Hamiltonian Monte Carlo (HMC)

[Neal \(2011\)](#) Le HMC est une méthode d'échantillonnage de chaînes de Markov Monte Carlo (MCMC) utilisée en inférence bayésienne, notamment dans Stan. Son objectif est de générer des échantillons à partir de distributions de probabilité complexes (par exemple, la distribution postérieure des paramètres d'un modèle) qui ne peuvent pas être échantillonnées directement. Il se présente comme une amélioration des méthodes MCMC simples comme l'algorithme de Metropolis-Hastings. Les méthodes MCMC simples fonctionnent en construisant une "marche aléatoire" dans l'espace des paramètres. À chaque étape, elles proposent un nouveau point et décident de l'accepter ou de le rejeter en fonction d'un critère qui assure que la chaîne finit par converger vers la distribution cible souhaitée. Les limites des MCMC simples sont :

Exploration lente (random walk) : Surtout dans les espaces de grandes dimensions ou lorsque les paramètres sont fortement corrélés. La chaîne se déplace lentement et met beaucoup de temps à explorer toute la distribution.

Faible taux d'acceptation : Pour éviter un déplacement trop lent, on pourrait vouloir faire de grands pas. Mais de grands pas dans l'espace des paramètres ont souvent une faible probabilité d'être acceptés, ce qui conduit à une stagnation de la chaîne.

Comprendre la procédure des HMC

HMC s'inspire de la physique classique, plus précisément de la mécanique Hamiltonienne, pour résoudre ces problèmes. Imaginons une bille roulant sur une surface vallonnée. La surface représente la distribution de probabilité cible (par exemple, la log-densité postérieure). Les vallées correspondent aux régions de haute probabilité, et les collines aux régions de basse probabilité. La position de la bille sur la surface correspond aux valeurs des paramètres du modèle (par exemple, β_0, β_1, σ).

L'énergie potentielle de la bille est l'inverse de la log-densité de probabilité ($-\log(\text{densité})$). Plus la densité est élevée, plus la vallée est profonde et plus l'énergie potentielle est faible.

En plus de sa position, la bille a aussi une impulsion (moment), qui est comme une vitesse et une direction. Il représente une variable auxiliaire introduite dans le modèle. L'énergie cinétique de la bille est liée à son impulsion.

L'Hamiltonien est l'énergie totale du système : $H(\text{position}, \text{impulsion}) = \text{Énergie Potentielle} + \text{Énergie Cinétique}$. Le principe clé de HMC est de simuler la trajectoire de cette bille sur la surface d'énergie potentielle.

Les étapes de l'algorithme HMC

À chaque itération MCMC, HMC effectue les étapes suivantes :

a. Initialiser l'impulsion : À partir de la position actuelle des paramètres q , on tire aléatoirement une nouvelle impulsion p pour la bille, généralement à partir d'une distribution normale indépendante (souvent centrée en 0). C'est la seule étape aléatoire par itération qui permet d'explorer de nouvelles directions.

b. Simuler la dynamique Hamiltonienne (Leapfrog Integrator) : Maintenant que la bille a une position q et une impulsion p , on simule son mouvement sur la surface d'énergie potentielle pendant une durée fixe (nombre de pas L et taille de pas ϵ). Ce mouvement est déterministe et est guidé par les équations de Hamilton. La simulation se fait généralement à l'aide d'un algorithme numérique appelé "Leapfrog Integrator" (intégrateur de saute-mouton). Cet intégrateur préserve l'énergie du système sur le long terme, ce qui est important pour la validité de l'algorithme. Pendant ce mouvement, la bille explore l'espace des paramètres. Lors-

qu'elle monte une colline (diminution de la densité de probabilité, augmentation de l'énergie potentielle), son énergie cinétique diminue (elle ralentit). Lorsqu'elle descend dans une vallée (augmentation de la densité, diminution de l'énergie potentielle), son énergie cinétique augmente (elle accélère). Cela permet à la bille de se déplacer rapidement dans les régions de haute probabilité et de traverser efficacement les régions de basse probabilité pour atteindre d'autres régions de haute probabilité.

c. Critère d'Acceptation-Rejet (Metropolis) : Après avoir simulé la trajectoire et atteint une nouvelle position et impulsion (q^*, p^*) , on applique une étape d'acceptation-rejet de Metropolis-Hastings [Metropolis et al. \(1953\)](#). Cependant, grâce aux propriétés de l'intégrateur Leapfrog (préservation de l'énergie), le taux d'acceptation dans HMC est généralement très élevé (souvent proche de 1). Cela signifie que la plupart des propositions sont acceptées, évitant la stagnation des chaînes. Si la proposition est acceptée, la nouvelle position q^* devient le prochain échantillon. Sinon, la chaîne reste à l'ancienne position q .

Remarque 2.1. Limites et Réglages

Nécessite le gradient : HMC exige que le modèle soit différentiable afin de calculer le gradient de la log-densité de probabilité. Stan utilise la différenciation automatique pour cela, ce qui simplifie grandement son utilisation.

Paramètres de réglage : HMC a des hyperparamètres qui doivent être bien choisis comme la taille de pas (ϵ) qui contrôle la précision de la simulation. Trop petit, la simulation est lente. Trop grand, la simulation est imprécise et le taux d'acceptation baisse ; le nombre de pas (L) qui détermine la longueur de la trajectoire simulée. Trop court, la chaîne ne s'éloigne pas assez. Trop long, cela peut être inefficace ou même faire revenir la chaîne sur ses pas ("U-turn"). C'est là que l'algorithme No-U-Turn Sampler (NUTS), qui est l'implémentation par défaut de HMC dans Stan, intervient. NUTS est une extension intelligente de HMC qui s'adapte automatiquement à la taille de pas optimale (ϵ) et détermine de manière dynamique la longueur optimale de la trajectoire (L), évitant ainsi le problème du "demi-tour" (U-turn).

2.2.4.2 Application avec l'algorithme Hamiltonian Monte Carlo (HMC) sur RStudio

Il n'y a pas de fonction HMC directement intégrée dans les packages R de base comme c'est le cas pour d'autres algorithmes MCMC (par exemple, la fonction `MCMCmetrop1R` du package `MCMCpack` pour Metropolis-Hastings). La raison est que HMC est plus complexe à implémenter, car il nécessite le calcul des gradients de la fonction de vraisemblance (ou plutôt de la log-densité de la distribution cible) par rapport aux paramètres.

Cependant, des packages R spécialisés ont été développés pour faciliter l'utilisation de HMC, notamment en s'appuyant sur des langages de modélisation probabiliste qui gèrent les gradients

automatiquement.

Le package le plus accessible pour faire du HMC en R est `rstan`. Il s'agit d'une interface R vers le langage de modélisation Stan, qui est conçu pour l'échantillonnage bayésien efficace, et dont l'algorithme par défaut est une variante très optimisée de HMC (NUTS - No-U-Turn Sampler).

`rstan` est la solution simple pour HMC en R car il procure :

- Une gestion automatique des gradients : Stan s'occupe de la dérivation des gradients, ce qui est le point le plus complexe de l'implémentation manuelle de HMC.
- Une implémentation optimisée de HMC/NUTS : L'algorithme NUTS de Stan est une version avancée de HMC qui ajuste automatiquement certains paramètres (comme la longueur de la trajectoire), rendant l'échantillonnage encore plus robuste et efficace.
- Un langage de modélisation intuitif : On écrit le modèle bayésien dans un langage simple et déclaratif (le langage Stan), et `rstan` compile ce modèle et exécute l'échantillonneur.

2.2.4.3 Diagnostic

- Évaluer la statistique de Gelman-Rubin \hat{R} , qui est un critère diagnostique clé pour évaluer la convergence des algorithmes MCMC. Il mesure si les chaînes Markoviennes ont suffisamment exploré la distribution a posteriori en comparant la variabilité intra-chaîne et inter-chaînes.

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

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 $\bar{\theta} = \frac{1}{m} \sum_{j=1}^m \bar{\theta}_j$.

- Les graphes des chaînes doivent être stationnaires (pas de tendance visible)

L'algorithme HMC est un outil fondamental en inférence bayésienne, permettant de simuler des distributions complexes sans normalisation explicite. Pour des modèles comme la GEV, une attention particulière doit être portée aux contraintes des paramètres et à l'initialisation.

APPLICATIONS ET RÉSULTATS

3.1 Données réelles : échantillonnage et prétraitement.

Les données proviennent du site « **NASA Prediction Of Worldwide Energy Resources** » sur lequel le projet POWER fournit des ensembles de données solaires et météorologiques provenant de la recherche de la NASA (National Aeronautics and Space Administration) qu'on peut retrouver via le lien suivant : <https://power.larc.nasa.gov> Ces ensembles de données concernent les secteurs de l'énergie renouvelable, de bâtiment durable et d'agro-climatologie. Les paramètres solaires et météorologiques de l'agro-climatologie auxquels nous nous sommes intéressés sont disponibles sous forme de séries chronologiques journalières, mensuelles et annuelles. Tous les paramètres étaient fournis sur la grille de résolution originale (zone maximale de 10×10 degrés sur une grille commune de 0.5×0.5) qui dépend du paramètre. Les paramètres solaires et météorologiques de base sont : la précipitation, la température, la pression de surface, l'humidité, le vent, etc. Les données sont disponibles en format CSV, ASCII et NetCDF et sont mises à jour quotidiennement. Notre étude porte sur la région ouest-africaine délimitée par les coordonnées géographiques $4^{\circ}N - 18^{\circ}N$ en latitude et $17.2^{\circ}W - 11.2^{\circ}E$ en longitude (Figure 3.1). Nous avons analysé l'évolution spatiale des précipitations extrêmes sur une période de 24 ans (2000-2023) à partir d'un réseau de 1568 stations météorologiques réparties sur cette zone. Les données journalières collectées comprennent : les précipitations (variable cible), les covariables météorologiques (température à 2 m, humidité relative à 2 m, pression atmosphérique de surface et vitesse du vent à 2 m). Le traitement à consister à l'extraction des valeurs maximales annuelles de précipitation pour chaque station et l'association des valeurs correspondantes des covariables. Pour les cas particuliers, en occurrence lorsqu'une station enregistre plusieurs événements pluvieux extrêmes de même intensité mais avec des covariables différentes, nous avons retenu la moyenne des valeurs des covariables. La base de données obtenue est une matrice par variable météorologique dont les lignes sont des années (2000 à 2023)

et les colonnes, les stations (1568 au total).

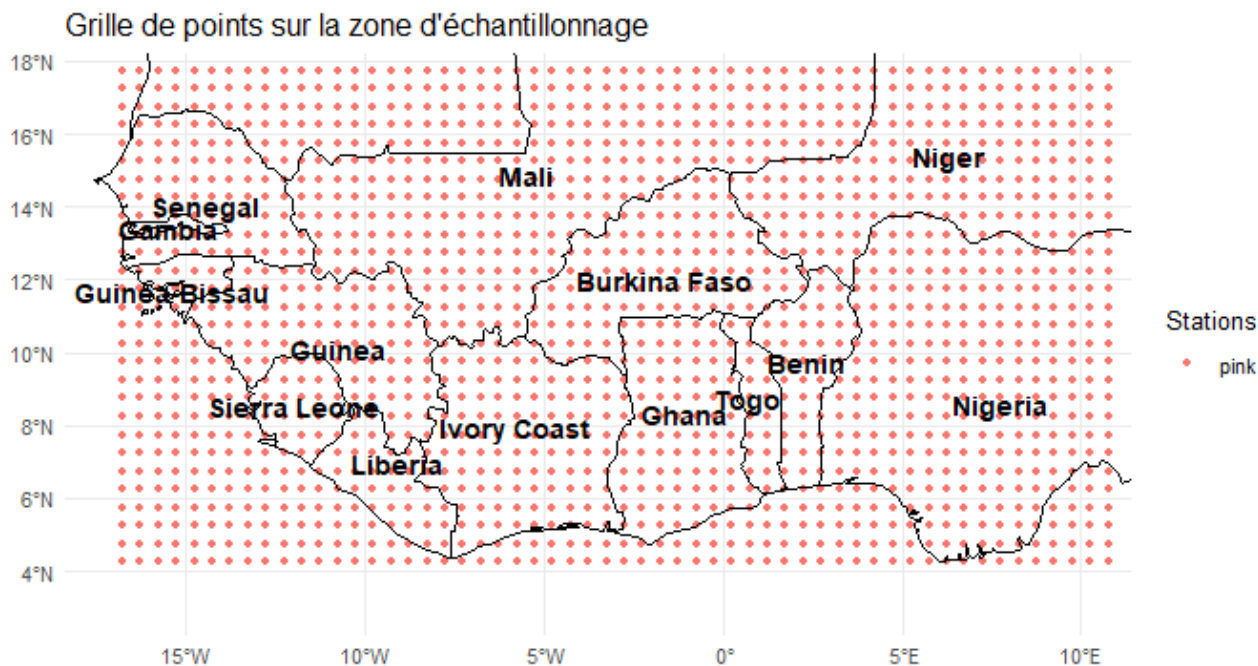


FIGURE 3.1 – Zone ouest africaine sur laquelle les données ont été échantillonnées.

3.1.1 Évolution des précipitations

La figure 3.2 met en évidence une augmentation progressive des précipitations maximales annuelles entre les stations météorologiques, suivant un gradient ouest-est (de la zone côtière vers l'intérieur continental). Alors que les cumuls pluviométriques extrêmes dépassaient rarement 50 mm/jour dans les zones désertiques avant 2014, plusieurs stations ont enregistré des épisodes pluvieux atteignant 250 mm/jour à partir de 2020. La plus grande valeur a été enregistrée en 2022, 2071 mm au cours d'une journée. Cette intensification concerne particulièrement les régions côtières, mais montre également une extension notable vers les zones arides, révélant une modification des régimes pluviométriques à l'échelle régionale.

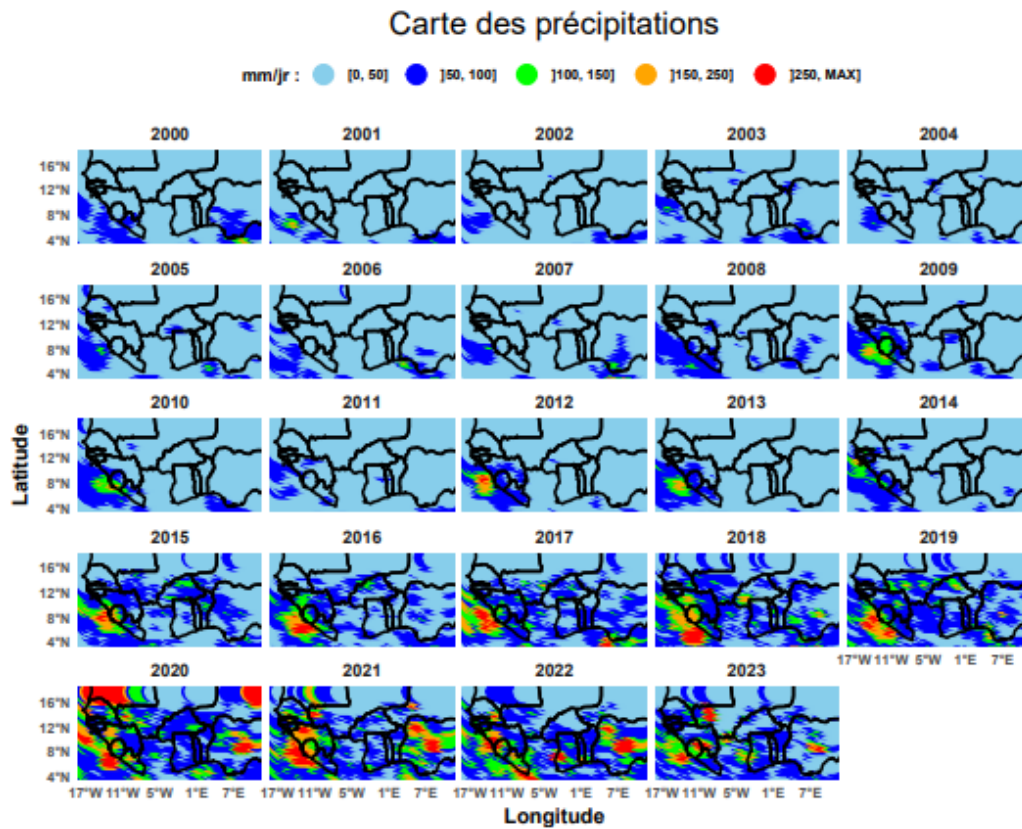


FIGURE 3.2 – Visualisation des précipitations maximales annuelles.

3.1.2 Dépendance spatiale et temporelle des paramètres GEV

L'application de la distribution GEV aux données de 24 ans pour chaque station révèle une variation spatiale des paramètres GEV (figure 3.3). Cependant, cette figure ne permet pas de déterminer la nature exacte de cette dépendance. C'est pourquoi nous avons analysé la variation des paramètres GEV en fonction des coordonnées géographiques (longitude et latitude). Pour ce faire, nous avons fixé tour à tour chaque coordonnée géographique. Ainsi, pour chaque année, nous avons obtenu des données en fonction de la latitude (en fixant la longitude) ou en fonction de la longitude (en fixant la latitude). Cette approche nous a permis de mettre en évidence la dépendance des paramètres GEV par rapport à chaque coordonnée géographique pour chaque année, comme l'illustrent les figures 3.4, 3.5, 3.6 et 3.7. En intégrant le facteur temps, la figure 3.8 suggère une dépendance quasi linéaire pour les paramètres μ , σ et ξ .

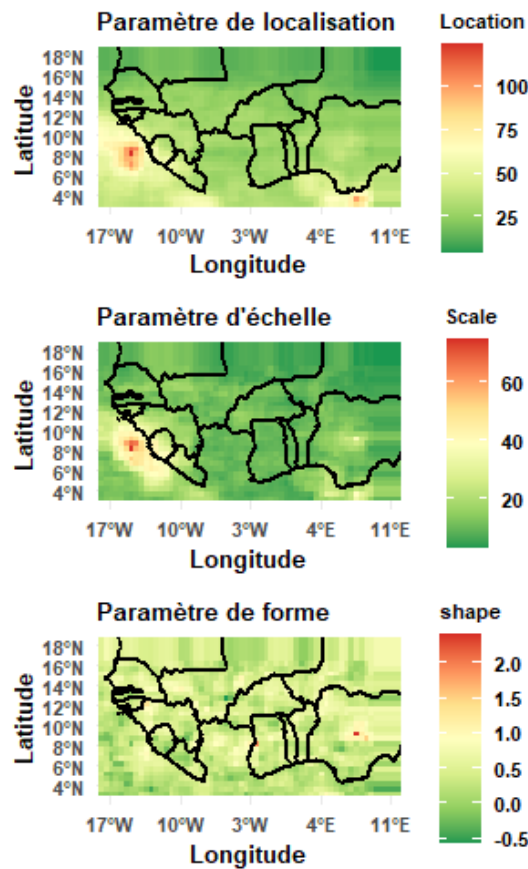


FIGURE 3.3 – Cartographie des paramètres du modèle GEV.

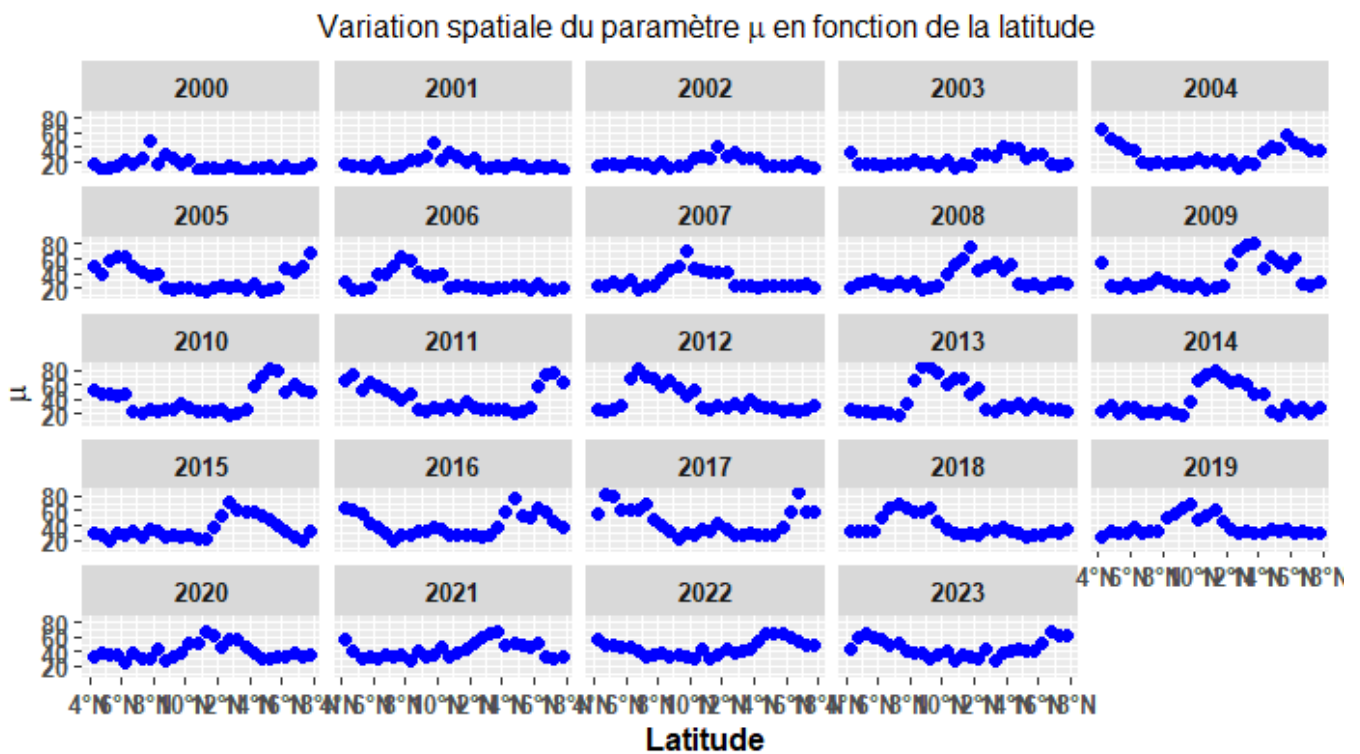


FIGURE 3.4 – Évolution dans le temps du paramètre de position en fonction de la latitude.

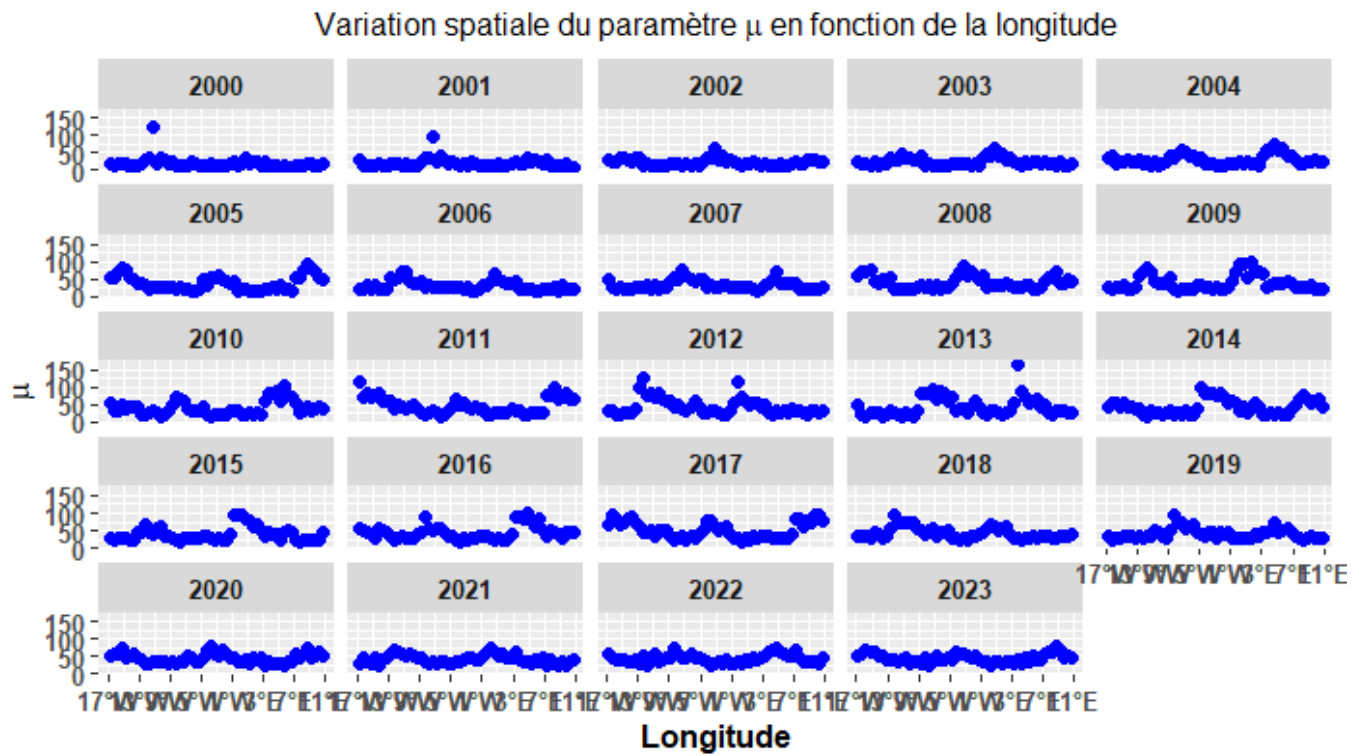


FIGURE 3.5 – Évolution dans le temps du paramètre de position en fonction de la longitude.

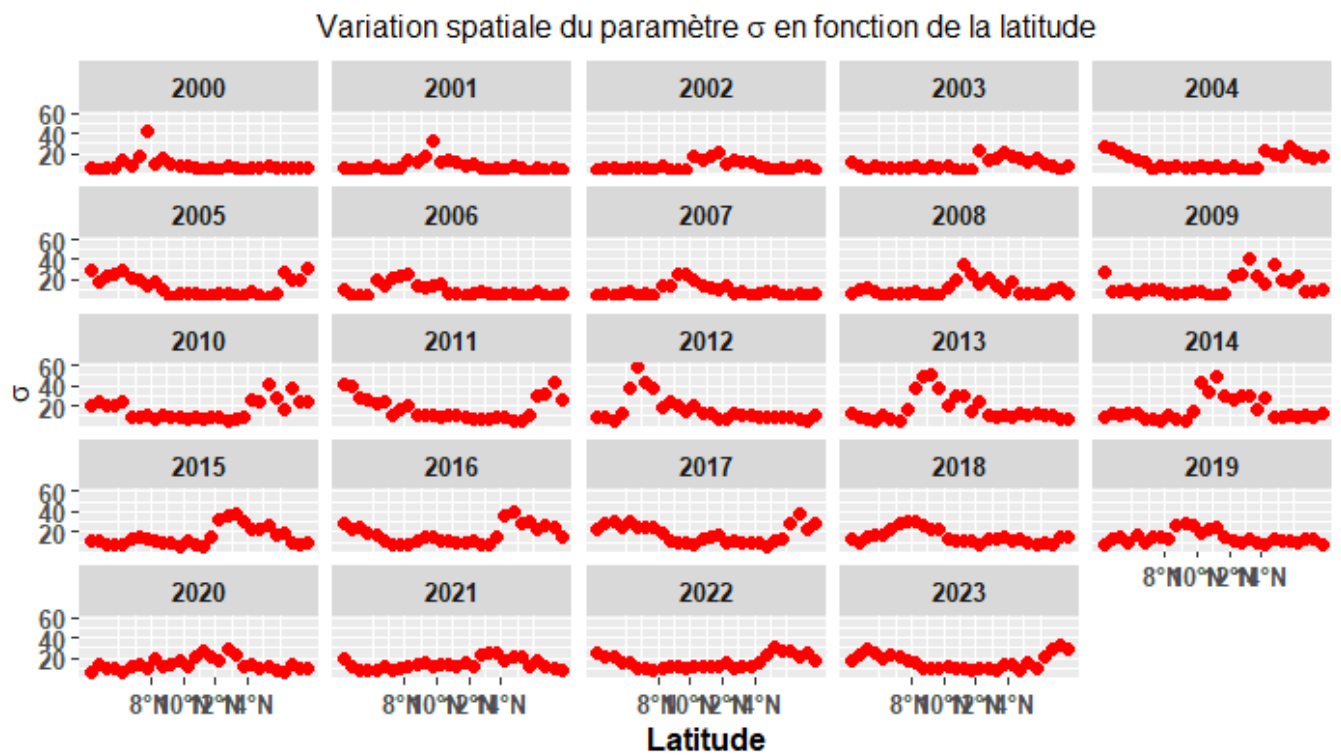


FIGURE 3.6 – Évolution dans le temps du paramètre d'échelle en fonction de la latitude.

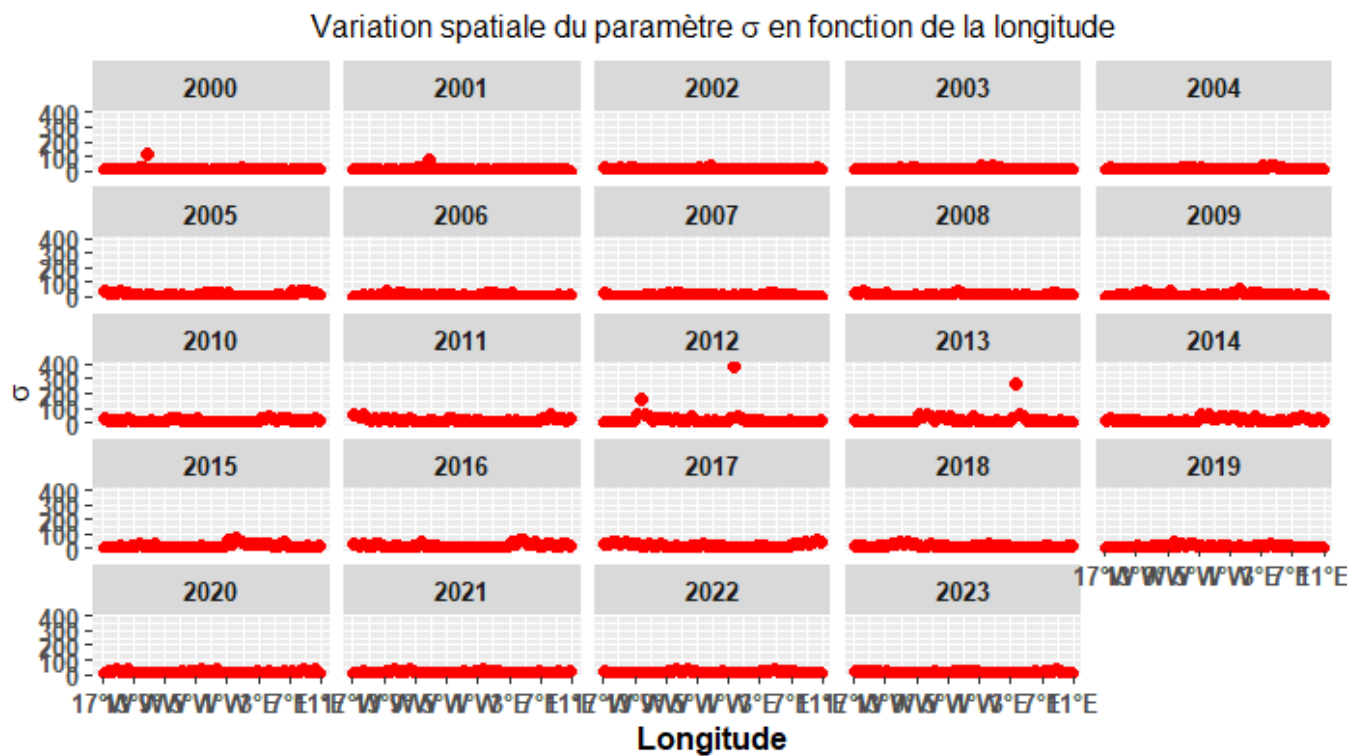


FIGURE 3.7 – Évolution dans le temps du paramètre d'échelle en fonction de la longitude.

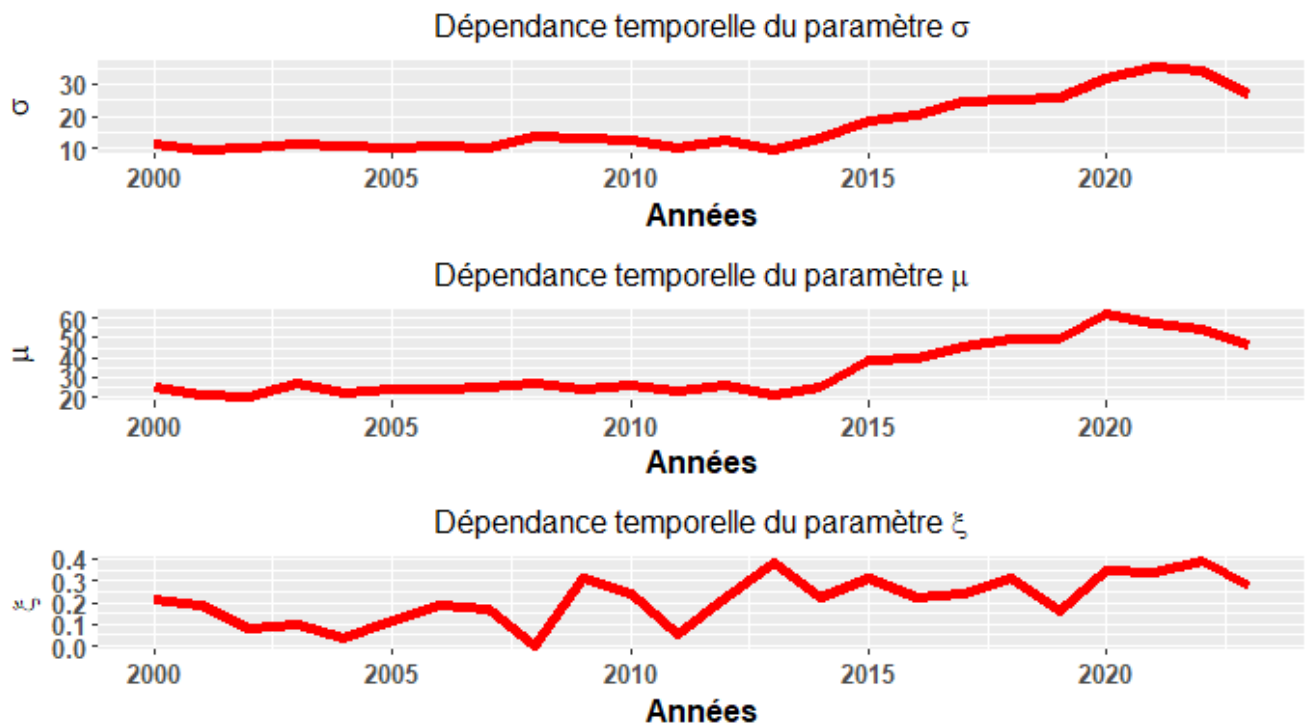


FIGURE 3.8 – Évolution des paramètres GEV en fonction du temps.

Nous avons initialement envisagé une dépendance de type sinusoïdale, bien que son existence sur la période de 24 ans analysée ne soit pas clairement établie. Après avoir testé différentes périodes (24, 12 et 5 ans), nous avons constaté des niveaux de retour sur 100 ans très élevés (au

moins 3×10^4) pour le meilleur modèle où le paramètre ξ dépendait de $\sin(t)$. La seule exception fut la période de 5 ans, où le meilleur modèle présentait une dépendance ξ par rapport au temps. Suite à ces observations, nous avons décidé d'exclure l'hypothèse de dépendance sinusoïdale.

3.2 Modèles

3.2.1 Modèles GEV classiques

Nous avons développé les modèles suivants : 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. Nous avons estimé les

TABLE 3.1 – Les modèles GEV classiques.

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

paramètres de chacun des modèles présentés précédemment, et ces résultats sont récapitulés dans le tableau 3.2. Une observation attentive des coefficients du paramètre de forme montre que leur valeur absolue excède 0.1. Cette observation est cruciale, car elle indique que le cas de Gumbel ($\xi = 0$) n'est pas pertinent pour le calcul des niveaux de retour (2.18). En pratique, nous avons considéré ξ nul lorsque sa valeur absolue est inférieure à 10^{-6} , compte tenu de l'impossibilité d'obtenir une valeur de ξ exactement égale à zéro.

TABLE 3.2 – Tableau montrant les valeurs des paramètres estimés.

Estimations	Modèle1	Modèle2	Modèle3	Modèle4	Modèle5	Modèle6
μ_0	53.25479	51.93073	50.96573	47.23999	52.61253	51.75354
μ_1	-2.13763	-2.003659	-1.969361	-2.005949	-2.065836	-2.002953
μ_2	-0.4103754	-0.1352346	-0.4176337	-0.1367772	-0.4107481	-0.136742
μ_3	*****	*****	2.286912	4.724565	-0.3404369	0.1698026
σ_0	24.63661	24.8414	22.68041	20.023	18.0529	25.03797
σ_1	-0.8893558	-0.857108	-0.7500399	-0.8594044	-0.831334	-0.8563929
σ_2	-0.2979556	*****	-0.3140811	*****	-0.306505	-0.2099486
σ_3	*****	*****	0.1952695	4.837822	5.755051	*****
ξ_0	0.4004958	0.4134662	0.3996152	0.4125777	-0.1027155	0.4561981
ξ_1	*****	*****	*****	*****	0.5020913	-0.04389385

3.2.2 Modèles bayésiens

- **Modèle GEV simple : Capture de la tendance globale.** Pour chaque observation y_i :

$$y_i \sim \text{GEV}(\mu, \sigma, \xi) \text{ où } \mu = \beta_0^\mu, \sigma = \beta_0^\sigma \text{ et } \xi = \beta_0^\xi \quad (3.1)$$

Priors : $\beta_0^\mu \sim \mathcal{N}(0, 10)$, $\beta_0^\sigma \sim \mathcal{LN}(0, 5)$ et $\beta_0^\xi \sim \mathcal{N}(0, 0.5)$.

- **Modèle GEV avec effets spatiaux linéaires : Capture des gradients spatiaux linéaires.**

$$y_i \sim \text{GEV}(\mu_i, \sigma_i, \xi) \text{ où } \mu_i = \beta_0^\mu + \beta_1^\mu \cdot \text{lat}_i + \beta_2^\mu \cdot \text{lon}_i, \sigma_i = \beta_0^\sigma + \beta_1^\sigma \cdot \text{lon}_i \text{ et } \xi = \beta_0^\xi \quad (3.2)$$

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 $\beta_0^\xi \sim \mathcal{N}(0, 0.5)$

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

$$y_{ij} \sim \text{GEV}(\mu_{ij}, \sigma_{ij}, \xi) \quad (3.3)$$

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

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

- **Modèle GEV avec processus gaussien : Capture des motifs spatiaux non linéaires complexes.**

$$y_i \sim \text{GEV}(\mu_i, \sigma_i, \xi) \quad (3.4)$$

Où $\mu_i = \beta_0^\mu + f_1(\mathbf{s}_i)$, $f_1 \sim \mathcal{GP}(0, K_1)$, $\sigma_i = \beta_0^\sigma + f_2(\mathbf{s}_i)$, $f_2 \sim \mathcal{GP}(0, K_2)$,

$K_k(\mathbf{s}_i, \mathbf{s}_j) = \eta_k^2 \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{2\rho_k^2}\right)$, $k \in \{1, 2\}$, et $\xi = \beta_0^\xi$

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 $\beta_0^\xi \sim \mathcal{N}(0, 0.5)$

3.3 Choix et diagnostique du meilleur modèle

3.3.1 Critères de choix des modèles GEV

Le choix du meilleur modèle s'est basé sur la déviance (2.13), l'AIC (2.14) et le BIC (2.15), le modèle optimal étant celui qui minimise ces critères. Par conséquent, le modèle 3 a été sélectionné pour sa performance supérieure en termes de prédiction et d'explication.

TABLE 3.3 – Valeurs des critères pour chaque modèle

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

3.3.2 Diagnostic

En Statistique, le diagnostic d'un modèle consiste à évaluer sa qualité d'ajustement, ses performances prédictives et la validité de ses hypothèses. Nous avons étudié l'adéquation du modèle par rapport aux données figure 3.9.

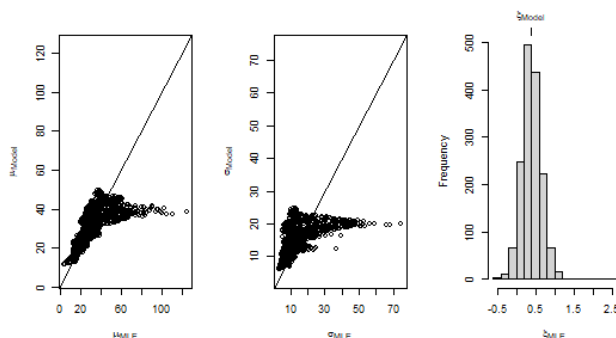


FIGURE 3.9 – Adéquation entre les quantiles empiriques et ceux théoriques du modèle 3.

Malgré une bonne adéquation générale avec la diagonale unité, quelques déviations des points sont observées. Cela suggère soit l'existence d'un modèle supérieur non exploré, soit des limites de la méthode d'inférence utilisée. Ces limites peuvent être attribuées au fait que la gestion totale de la dépendance entre les stations n'est pas prise en compte. En effet, la dépendance spatiale est structurée par les covariables (latitude et longitude). On peut noter l'existence des effets aléatoires sur les stations dus aux variables latentes (altitude, proximité de la mère, etc.). De plus, la GEV n'est pas optimale lorsqu'on dispose de peu d'observations par stations.

3.4 Niveaux de retour

3.4.1 Modèle GEV classique

Les figures définies en 3.10 illustrent une augmentation générale des quantiles de précipitations à mesure que la période de retour s'allonge. On observe clairement une disparition

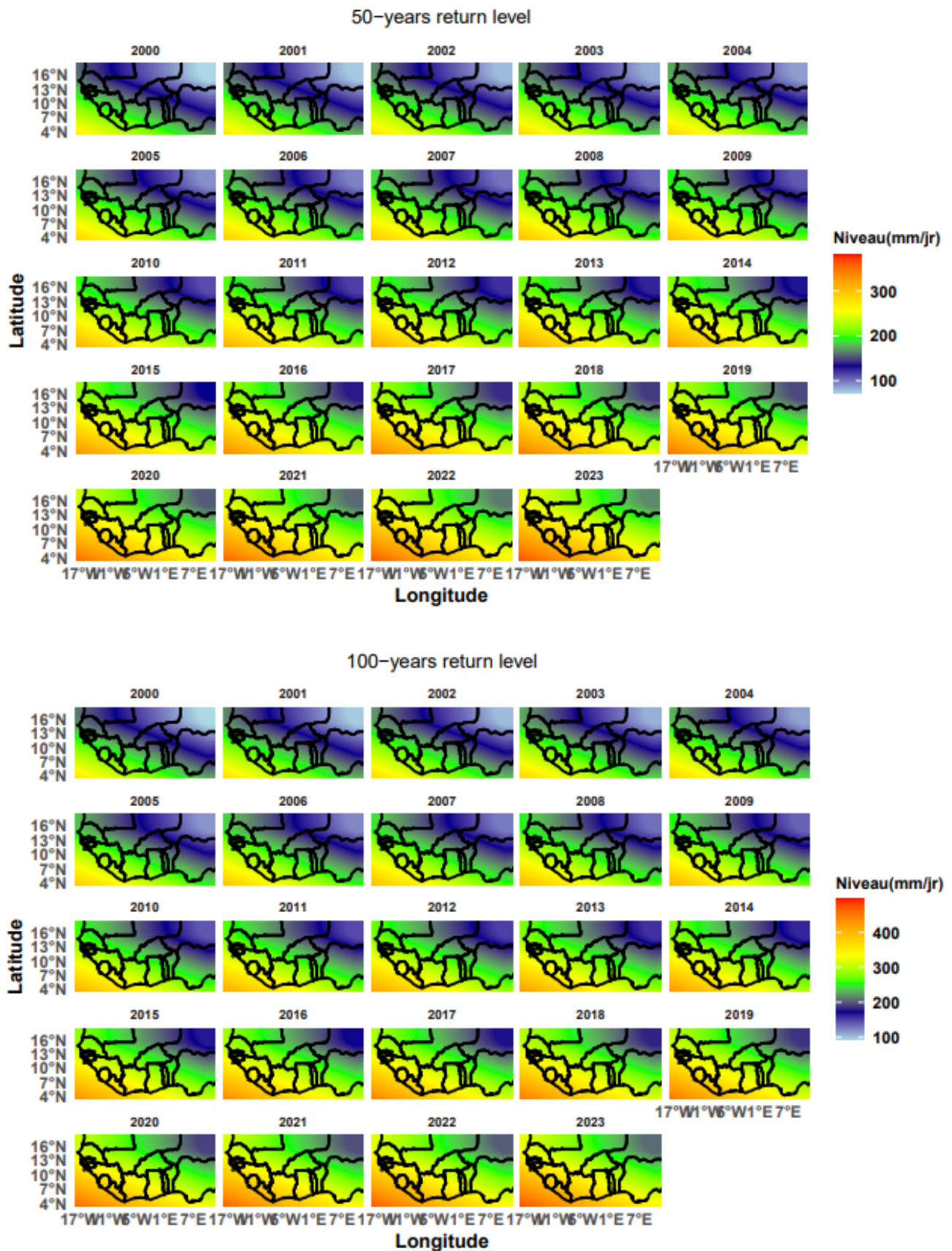


FIGURE 3.10 – Niveaux de retour du modèle GEV classique sur 50 et 100 ans

progressive des faibles quantiles (indiqués en bleu, correspondant aux zones désertiques) au profit de l'apparition du rouge, signalant ainsi des valeurs de précipitations plus élevées au fil

du temps.

3.4.2 Modèle bayésien

Parmi les modèles présentés dans la section 3.2.2, celui avec processus gaussien (3.2.2) n'a pas fait l'objet d'une application sur nos données. Le meilleur parmi les trois autres est celui qui capture la variation inter-stations, défini par l'équation (3.3). Ces modèles ont été exécutés en utilisant 4 chaînes, 2000 itérations, et une période de "warmup" (phase d'initialisation) de 1000 itérations. L'exécution s'est déroulée sur un ordinateur équipé de 32 Go de RAM et 16 coeurs pendant 4 heures de temps. Il est important de noter que toutes les chaînes ont parfaitement convergé, et nous avons observé une bonne autocorrélation des "intercepts" pour chaque chaîne.

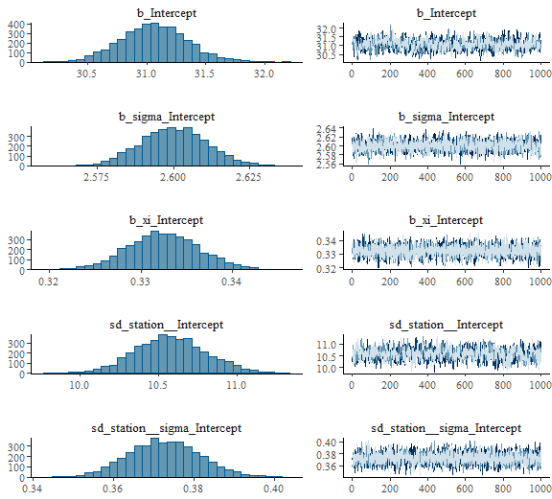


FIGURE 3.11 – Convergence des chaînes

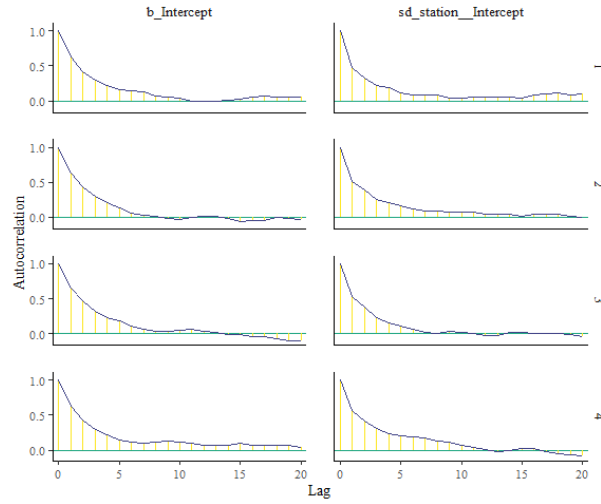


FIGURE 3.12 – Auto corrélation

Aucun des paramètres n'a présenté une statistique de Gelman-Rubin (\hat{R}) dépassant 1.05. On a plutôt des valeurs sensiblement égal à 1 (1.0045, 1.0000, 0.9997) pour les paramètres (μ, σ, ξ) dans cet ordre. Ce qui témoigne d'une bonne convergence.

Ce modèle est un modèle à effets mixtes où chaque station est traitée comme ayant une influence aléatoire sur l'observation. Il décompose la variabilité en effets fixes β_0^γ (effets constants à travers toutes les stations) où γ est le vecteur des paramètres GEV et des effets aléatoires u_j (variations spécifiques à chaque station). Les effets aléatoires utilisés ont pour but, la capture de la corrélation entre les observations d'une même station.

Les niveaux de retour sur une période de 50 et 100 ans ont été obtenus en appliquant l'équation 2.18. Ainsi nous avons les figures suivantes qui cartographie ces niveaux de retour.

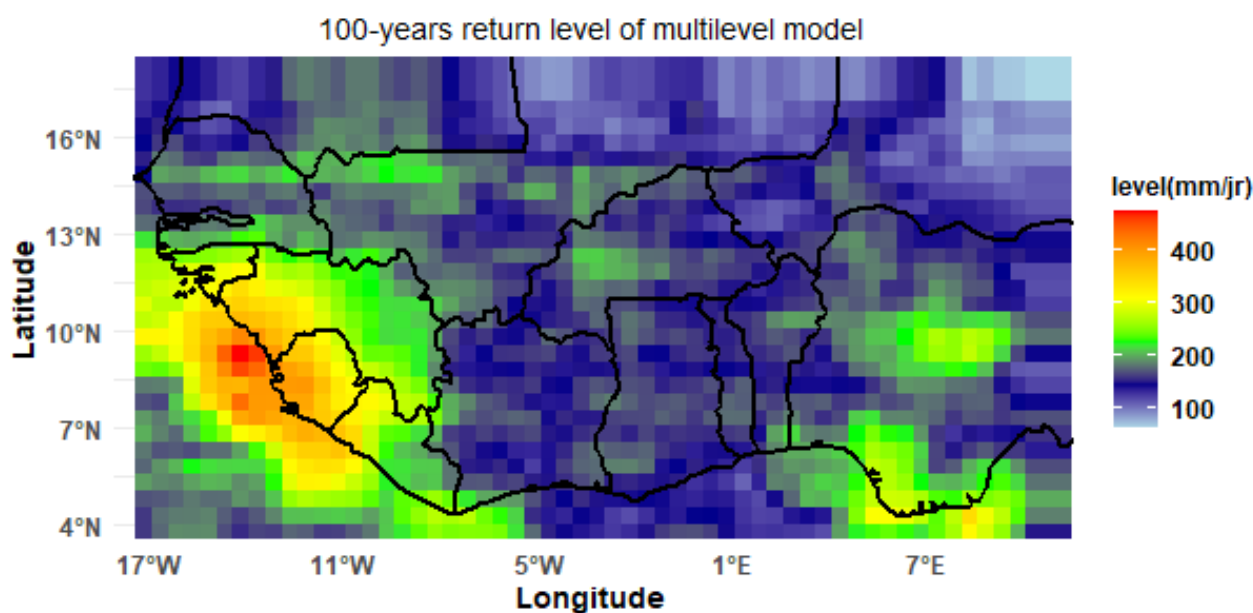
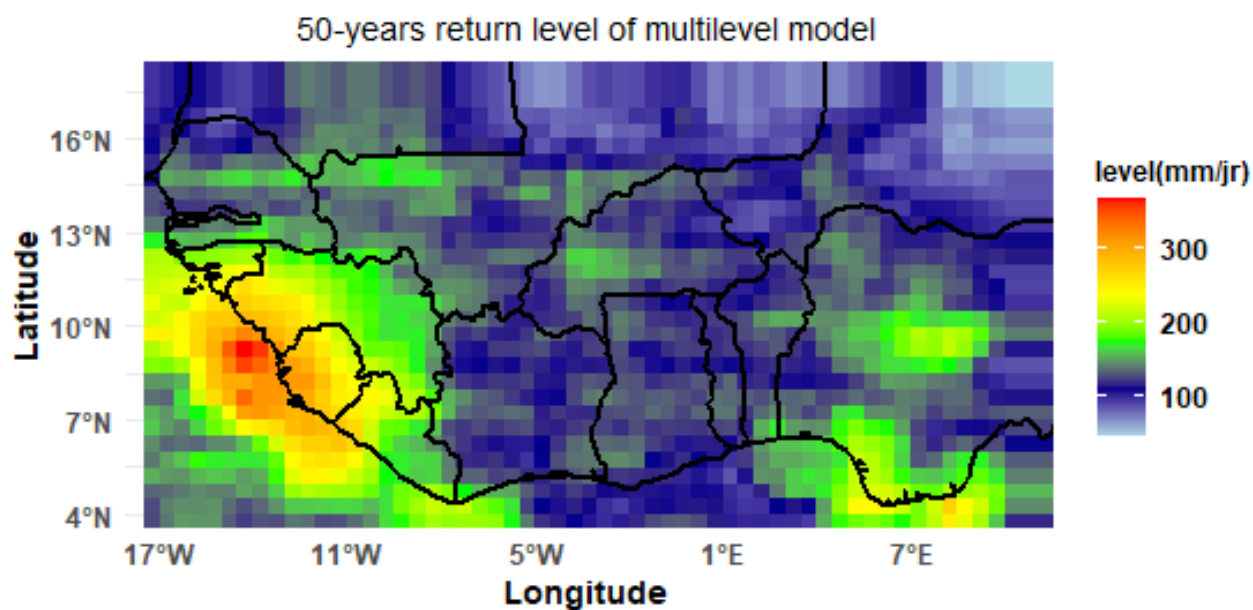


FIGURE 3.13 – Carte des niveaux de retour sur 50 et 100 ans du modèle à effets aléatoires par station

CONCLUSION ET PERSPECTIVES

Ce travail a démontré l'importance cruciale d'une approche de modélisation spatio-temporelle robuste pour l'analyse des précipitations extrêmes en Afrique de l'Ouest. En utilisant l'approche GEV intégrant la méthode du maximum de vraisemblance, nous avons remarqué les limites de cette approche particulièrement lorsqu'il s'agit de gérer la complexité des variabilités spatiales et lorsqu'il y a peu de données. L'adoption du cadre bayésien a été utilisée comme alternative dans le but de surmonter ces défis en offrant une capacité supérieure à quantifier l'incertitude des paramètres et à intégrer la dépendance spatiale entre les stations de manière cohérente moyennant un coût computationnel très élevé. L'application de l'algorithme HMC via Stan a prouvé son efficacité pour l'inférence dans des modèles complexes, fournissant des estimations fiables des paramètres. Les cartes illustrant les niveaux de retour représentent un apport majeur de cette étude. Elles offrent une visualisation claire des risques de précipitations extrêmes, informations essentielles pour la planification et la gestion des ressources en eau, l'agriculture, ainsi que pour la mise en place de stratégies d'adaptation face aux changements climatiques dans la région.

Ce travail ouvre la voie à des recherches futures axées sur l'intégration de prédicteurs environnementaux supplémentaires ou l'exploration de structures de dépendance plus complexes et des techniques d'apprentissage machine pour améliorer davantage la prévision des événements extrêmes. Ci-dessous les principaux points :

- Modèles à processus gaussiens : Pour les modèles bayésiens, ils permettent de capturer des motifs spatiaux non linéaires complexes.
- Approches par apprentissage automatique : On peut envisager l'utilisation de Random Forests quantiles ou de Gradient Boosting (comme XGBoost avec une fonction de perte quantile) pour modéliser des relations non linéaires, ou des réseaux de neurones pour les extrêmes (par exemple, Extreme Value Learning) pour apprendre les queues de distribution. Les modèles hybrides, combinant statistiques des valeurs extrêmes et apprentissage automatique, offrent également un bon compromis entre robustesse et flexibilité.

- Enfin, l'intégration de variables climatologiques telles que la température, la pression de surface ou l'humidité relative représente également une voie prometteuse.

Annexe

Exemple de code HMC avec rstan (régression linéaire bayésienne)

1. Installation et chargement des packages

```
“R
install.packages("rstan")
install.packages("loo") # Pour l'analyse des résultats
library(rstan)
rstan_options(auto_write = TRUE) # Sauvegarde les modèles compilés
options(mc.cores = parallel::detectCores()) # Utilise les processeurs pour le calcul pa-
rallèle
““
```

2. Création de données

```
“R
set.seed(123)
N <- 100 # Nombre d'observations
x <- rnorm(N, mean = 0, sd = 1)
beta_0_vrai <- 2
beta_1_vrai <- 0.5
sigma_vrai <- 1
y <- beta_0_vrai + beta_1_vrai * x + rnorm(N, mean = 0, sd = sigma_vrai)
data_stan <- list(N = N, x = x, y = y) # Préparation les données pour Stan
““
```

3. Définition du modèle Stan (dans une chaîne de caractères)

```
“R
model_code <- "
data
int<lower=0> N; // Nombre d'observations
```

```

vector[N] x; // Variable explicative
vector[N] y; // Variable réponse
parameters
real beta_0; // Intercepte
real beta_1; // Pente
real<lower=1e-6> sigma; // Écart-type des erreurs (doit être positif)
model
// A priori (distrib. de probabilité avant de voir les données)
beta_0 ~ normal(0, 10); // A priori non informatif pour beta_0
beta_1 ~ normal(0, 10); // A priori non informatif pour beta_1
sigma ~ lognormal(0, 2); // A priori pour sigma
// Vraisemblance (comment les données sont générées, étant donné les paramètres)
y ~ normal(beta_0 + beta_1 * x, sigma);
// Optionnel : Génération de quantités dérivées après l'échantillonnage
generated quantities
vector[N] y_rep; // Prédiction pour la validation du modèle
for (n in 1:N)
y_rep[n] = normal_rng(beta_0 + beta_1 * x[n], sigma);

" ""

```

4. Exécution de l'échantillonneur HMC (avec `stan_model` et `sampling`)

```

""R
# Compilation du modèle Stan
stan_reg_model <- stan_model(model_code = model_code) # Exécution de l'échan-
tillonneur HMC (NUTS par défaut) : c'est la partie HMC de l'algorithme
fit_stan <- sampling(stan_reg_model,
data = data_stan,
chains = 4, # Nombre de chaînes MCMC indépendantes
iter = 2000, # Nombre total d'itérations par chaîne (warmup + sampling)
warmup = 1000, # Nombre d'itérations pour la phase d'entraînement
thin = 1, # Stocke chaque 1er échantillon (pas de sous-échantillonnage ici)
seed = 42) # Pour la reproductibilité
print(fit_stan, pars = c("beta_0", "beta_1", "sigma")) # Résumé des résultats
# Visualiser les chaînes et les distributions a posteriori
traceplot(fit_stan, pars = c("beta_0", "beta_1", "sigma"))
plot(fit_stan, pars = c("beta_0", "beta_1", "sigma")) ""

```

Bibliographie

- [1] Buishand, T. A. (1991). *Extreme rainfall estimation by combining data from several sites*. Hydrological Sciences Journal, *36*(4), 345-365. DOI : 10.1080/02626669109492519
- [2] Bracken, C. et al. (2016). *Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain*, Water Resour. Res., 52, 66436655, doi :10.1002/
- [3] Lehmann, E. A. (2012). *Bayesian hierarchical modelling of rainfall extremes*. Extremes, 15(3), 283310. <https://doi.org/10.1007/s10687-012-0153-2>
- [4] Caragea, P. C., et Smith, R. L. (2017). *Asymptotic Properties of Bayes Estimators for Latent Nonstationary Spatial Processes*. Bayesian Analysis, 12(1), 128. <https://doi.org/10.1214/16-BA10022016WR018768>.
- [5] Castruccio, S., Huser, R., et Genton, M. G. (2014). *High-Order Composite Likelihood Inference for Max-Stable Distributions and Processes*. Journal of Computational and Graphical Statistics, 23(3), 606630. <https://doi.org/10.1080/10618600.2013.842158>
- [6] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. London : Springer-Verlag. DOI : 10.1007/978-1-4471-3675-0
- [7] Cooley, D., Nychka, D., et Naveau, P. (2007). *Bayesian Spatial Modeling of Extreme Precipitation Return Levels*. Journal of the American Statistical Association, 102(479), 824840. <https://doi.org/10.1198/016214507000000780>
- [8] DUPUIS, J., (2007). *Statistique bayésienne et algorithmes MCMC*, (LSP-UPS)
- [9] Fisher, R. A., et Tippett, L. H. C. (1928). *Limiting forms of the frequency distribution of the largest or smallest member of a sample*. Proceedings of the Cambridge Philosophical Society, *24*, 180-190. DOI : 10.1017/S0305004100015681
- [10] Gelman et al. (2013). *Bayesian Data Analysis*, (Chapman et Hall), Chap. 7.
- [11] Gnedenko, B. (1943). *Sur la distribution limite du terme maximum d'une série aléatoire*. Annals of Mathematics, *44*(3), 423-453. DOI : 10.2307/1968974

- [12] Hastings, W. K. (1970). *Monte Carlo sampling methods using Markov chains and their applications*. Biometrika, 57(1), 97109. <https://doi.org/10.1093/biomet/57.1.97>
- [13] JENKINSON, A.F. (1955). The frequency distribution of the annual maximum (or minimum) of meteorological elements. Quart. J. Roy. Meteor. Soc., 81, 158-171.
- [14] Katz, R. W., Parlange, M. B., et Naveau, P. (2002). *Statistics of extremes in hydrology*. Advances in Water Resources, *25*(8-12), 1287-1304. DOI : 10.1016/S0309-1708(02)00056-8
- [15] Metropolis, N. et al. (1953). *Equation of state calculations by fast computing machines*. The Journal of Chemical Physics, 21(6), 10871092. <https://doi.org/10.1063/1.1699114>
- [16] Naveau, P., Nogaï, M., Ammann, C., Yiou, P., Cooley, D., et Jomelli, V. (2005). *Statistical methods for the analysis of climate extremes*. Comptes Rendus Geoscience, *337*(10-11), 1013-1022. DOI : 10.1016/j.crte.2005.05.004
- [17] Neal, R. M. (2011). *MCMC using Hamiltonian dynamics*, In S. Brooks, A. Gelman, G. Jones, and X. L. Meng (Eds.), Handbook of Markov Chain Monte Carlo (pp. 113-162). Chapman and Hall/CRC.
- [18] Panthou G., Théo Vischel, Thierry Lebel, Juliette Blanchet, Guillaume Quantin et Abdou Ali (2015). *Estimation de cartes d'aléa pluviométrique en Afrique de l'Ouest : comparaison de différentes approches*, La Houille Blanche, 101 :6, 42-48, DOI : 10.1051/lhb/20150067
- [19] Pawitan, Y. (2001). *In All Likelihood : Statistical Modelling and Inference Using Likelihood*. Oxford University Press.
- [20] Renard, B. (2011). *A Bayesian Hierarchical Approach to Regional Frequency Analysis*. Water Resources Research, 47(11), W11513. <https://doi.org/10.1029/2010WR010089>
- [21] Rue, H., Martino, S., et Chopin, N. (2009). *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations*. Journal of the Royal Statistical Society : Series B (Statistical Methodology), 71(2), 319392.
- [22] Sang, H., et Gelfand, A. E. (2010). *Continuous Spatial Process Models for Spatial Extreme Values*. Journal of Agricultural, Biological, and Environmental Statistics, 15(1), 4965. <https://doi.org/10.1007/s13253-009-0007-9>