



الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي  
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

جامعة سعد دحلب البليدة-1-  
Université Saad Dahlab Blida -1-



**Mémoire de fin D'études**

En vue de l'obtention du diplôme Master

**Faculté de sciences**  
**Département : Mathématique**

**Spécialité : Modélisation stochastique et statistique**

**THEME**

**Modélisation statistique des valeurs extrêmes spatiales par des  
processus Max-stable, applications sur précipitations dans le bassin  
de Mitidja**

**Présenté par :**

ATTAR Khadidja

GHANEM Nour El Houda

**Soutenu le : 25/07 / 2019**

**Devant le Jury :**

**Président :** O. TAMI MAA **Université de Blida 1**

**Promoteur :** A. RASSOUL MCA **ENSH de Blida**

**Examineur :** R.FRIHI MAA **Université de Blida 1**

**Promotion : 2018/2019**

## DÉDICACE

Je dédie ce modeste travail :

A MA TRÈS CHÈRE MÈRE : Malika

Tu n'as cessé de me soutenir et de m'encourager durant toutes les années de mes études,  
tu as

toujours été présente à mes côtés pour me consoler quand il fallait. Je t'aime MAMA.

A MON TRÈS CHER PÈRE : Ahmed

Tu as su m'inculquer le sens de la responsabilité, et de la confiance en soi. Tes conseils  
ont

toujours guidé mes pas vers la réussite. Je te dois ce que je suis aujourd'hui et ce que je  
serai

demain et je ferai toujours de mon mieux pour rester ta fierté. Je t'aime PAPA.

A Mes très chères grand-mères

Que ce modeste travail, soit l'expression des vœux que vous n'avez cessé de formuler  
dans

vos prières. Que Dieu vous préserve santé et longue vie.

A la mémoire de mon grand-père Saïd

Qui a été toujours dans mon esprit et dans mon cœur, je vous dédie aujourd'hui ma  
réussite.

Que Dieu, le miséricordieux, vous accueille dans son éternel paradis.

A mes chères amies "Chahinez, Kahina et Fatiha"

qui ont toujours été à mes cotés pour leur soutien et leur amour

A ma très chère sœur Asma

Pour leurs soutiens infinis et leurs aides incessantes, a qui je souhaite un meilleur avenir

A mes cher petits frères Oussama, Fakhro et Islam, que dieu vous protège

A toute la famille ATTAR et la famille AMALOU.

A tous mes proches près et lion de moi

À tous mes collègues de Département mathématique de USDB et spécialement à ma  
promotion de spécialité MSS (2014-2019).

A.khadidja

## DÉDICACE

Je dédie ce mémoire à :

Ma mère, qui a oeuvré pour ma réussite, par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

Mon père, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi.

Mes soeurs Abire , Nihad qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

A Mes très chères grand-mères

A toute la famille **GHANEM** et la famille **YAHYAOU**.

A mes chères amies "**Kiki, Nabila , Soumia , Khouloud**"

Mes professeurs de l'université qui doivent voir dans ce travail la fierté d'un savoir bien acquis.

G.Nour El Houda

## REMERCIEMENT

Nous tenons tout d'abord à remercier Dieu le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce Modeste travail.

En second lieu, nous tenons à remercier notre encadreur Mr : (A.RASSOUL), son précieux conseil et son aide durant toute la période du travail.

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail Et de l'enrichir par leurs propositions.

Enfin, nous tenons également à remercier toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

# TABLE DES MATIÈRES

0.1	Résumé . . . . .	1
0.2	Abstract . . . . .	1
0.3	Notations et abréviations . . . . .	3
<b>1</b>	<b>Modélisation des queues de distributions</b>	<b>7</b>
1.1	Introduction . . . . .	7
1.2	Théorie des valeurs extrêmes dans le cas de variables IID Unidimensionnelles	7
1.2.1	Extrema par blocs . . . . .	8
1.2.2	Estimation . . . . .	11
1.3	Domaine d'attraction . . . . .	12
1.3.1	Le domaine d'attraction de la distribution de Gumbel . . . . .	12
1.3.2	Le domaine d'attraction de la distribution de Weibull . . . . .	13
1.3.3	Le domaine d'attraction de la distribution de Fréchet . . . . .	13
1.4	Loi des dépassements de seuils . . . . .	15
1.4.1	Excès au-delà d'un seuil . . . . .	15
1.4.2	Estimation . . . . .	17
1.4.2.1	Choix du seuil . . . . .	17
1.4.2.2	Estimation des paramètres . . . . .	18
1.5	L'estimation de l'indice de queue . . . . .	19
1.5.1	Estimateur de Pickands . . . . .	19
1.5.2	Estimateur de Hill . . . . .	20
1.5.2.1	Estimateur de Hill Alternatif . . . . .	21
1.6	Exemples de comportements limites dans les queues . . . . .	21
1.7	Conclusion . . . . .	23
<b>2</b>	<b>Processus Max-Stable</b>	<b>24</b>
2.1	Introduction . . . . .	24
2.2	Processus ponctuel . . . . .	25
2.2.1	Lien avec les deux premières approches . . . . .	25
2.2.2	Modélisation avec les excès de seuils . . . . .	26

2.3	Processus max-stables . . . . .	27
2.3.1	Définition et construction . . . . .	27
2.3.2	Quelques modèles max-stables . . . . .	33
2.3.2.1	Modèle de Smith . . . . .	33
2.3.2.2	Modèle de Schlather . . . . .	35
2.3.2.3	Modèle de Brown-Resnick . . . . .	38
2.3.2.4	Modèle de Extremal-t . . . . .	40
2.4	Application aux données spatiales . . . . .	41
2.4.1	Étape 1 : La distribution des valeurs extrêmes généralisées . . . . .	42
2.4.2	Étape 2 : Estimer le coefficient extrême entre chaque paire de sites . . . . .	42
2.4.3	Étape 3 : Estimer l'erreur de chaque type . . . . .	43
2.4.4	Étape 4 : Ajuster un modèle max-stable . . . . .	44
2.4.5	Étape 5. Tester l'ajustement du modèle . . . . .	44
2.5	Mesure de dépendance pour les extrêmes spatiaux . . . . .	45
2.5.1	Madogramme . . . . .	45
2.5.2	Madogramme et coefficient extrême . . . . .	46
2.5.3	F-madogramme . . . . .	47
2.6	Inférence . . . . .	48
2.6.1	Estimation du coefficient extrême . . . . .	50
2.6.2	Le cas simple . . . . .	52
2.6.3	Le cas non simple . . . . .	54
<b>3</b>	<b>Simulation et application des processus max-stable</b> . . . . .	<b>55</b>
3.1	Introduction . . . . .	55
3.2	Qu'est-ce que le package SpatialExtremes? . . . . .	56
3.3	Description des données . . . . .	56
3.3.1	code R . . . . .	56
3.4	Simulation de max-stables . . . . .	60
3.5	Modélisation <i>GEV</i> des précipitations des maxima annuels et sélection de modèle pour la prédiction de quantiles élevés . . . . .	66
3.5.1	Introduction . . . . .	66
3.5.2	Les modèles . . . . .	67
3.5.3	Critère d'information Akaike . . . . .	67
3.5.3.1	Sélection du modèle . . . . .	67
3.5.4	Méthode de validation croisée à en $k$ fold . . . . .	68
3.5.4.1	Sélection du modèle . . . . .	68
3.6	Application . . . . .	69
3.7	Transformation des données . . . . .	70
3.7.1	Les données . . . . .	70
3.8	Analyses des données . . . . .	71
3.8.1	boxplot . . . . .	71

---

3.8.2	Résumé discriptive de la série pluviométrique des quatre stations . . . . .	72
3.8.3	Test d'homogénéité . . . . .	72
3.9	Ajustement de loi GEV . . . . .	73
3.10	Modèle des données . . . . .	75
3.11	Coefficient extrême . . . . .	75
3.12	Estimation des paramètres des modèles . . . . .	77
3.12.1	AIC . . . . .	78
3.13	Conclusion . . . . .	80
3.14	. . . . .	80

## TABLE DES FIGURES

1.1	Représentation graphique de la densité des valeurs extrêmes . . . . .	10
1.2	Fonction de répartition de lois GVE pour différentes valeurs de $\xi$ , avec $\sigma = 1$ et $\mu = 0$ . . . . .	10
1.3	Les graphes représentent les domaines d'attraction des trois lois extrêmes . . . . .	14
1.4	Méthode des excès : $u$ réel suffisamment élevé appelé seuil, $y$ : excès de $x$ au-delà de $u$ . . . . .	15
1.5	Densité des lois de valeurs extrêmes, avec $\xi = 0.5$ pour la loi de Pareto, $\xi = 0$ pour la loi exponentielle et $\xi = -0.5$ pour la loi de Pareto II . . . . .	17
2.1	Illustration de la caractérisation spectrale du Théorème 2.5. Les courbes grises sont les atomes de la forme $\zeta_i Y_i$ et la courbe noire la trajectoire du processus max-stable simple. . . . .	32
2.2	Modèle de Smith . . . . .	33
2.3	Graphe du coefficient extrémal du modèle de Smith avec $\Sigma$ la matrice identité d'ordre 2 . . . . .	34
2.4	Modèle de Schlather. . . . .	35
2.5	Graphes des corrélogrammes de la famille Whittle -Mathérn selon les valeurs de $h \in \mathbb{R}_+$ pour différentes valeurs de $\ell$ , avec $m_1 = m_2 = 1$ : Le noir pour $\ell = 4$ , le rouge pour $\ell = 2$ , le vert pour $\ell = 1$ et le bleu pour $\ell = 0.5$ . . . . .	36
2.6	Graphes des corrélogrammes de la famille de Cauchy selon les valeurs de $h \in \mathbb{R}_+$ pour différentes valeurs de $\ell$ , avec $m_1 = m_2 = 1$ : Le noir pour $\ell = 4$ , le rouge pour $\ell = 2$ , le vert pour $\ell = 1$ et le bleu pour $\ell = 0.75$ . . . . .	36
2.7	Graphes des corrélogrammes de la famille stable selon les valeurs de $h \in \mathbb{R}_+$ pour différentes valeurs de $\ell$ , avec $m_1 = m_2 = 1$ : Le noir pour $\ell = 2$ , le rouge pour $\ell = 1.5$ , le vert pour $\ell = 1$ et le bleu pour $\ell = 0.75$ . . . . .	37
2.8	Graphe du coefficient extrémal du modèle de Schlather où $\rho$ le corrélogramme de Whittle-Mathérn avec $m_1 = m_2 = \ell = 1$ . . . . .	38
2.9	Modèle de Brown-Resnick. . . . .	38
2.10	Modèle de Extremal-t. . . . .	40

2.11 Fonctions des coefficients extrêmes. Gauche : processus Extremal-t avec $\rho(h) = \exp(-h)$ et degrés de liberté $n = 1, 2, 5, 10$ , et À droite : procédé Brown Resnick avec $\gamma(h) = h^a$ et $a = 0.5, 1, 2$ : Pour le modèle extremal-t le cas $n = 1$ correspond à un processus de Schlather. Pour le modèle de Brown Resnick, le cas $a = 2$ correspond à un processus de Smith avec une matrice de covariance égale à la matrice d'identité. . . . .	41
2.12 F-Madogramme empirique (à gauche) et coefficients extrêmes correspondants (à droite) pour les données de précipitations suisses du paquet SpatialExtremes (Ribatet et al., 2013). Les points gris sont des estimations par paires, tandis que les noirs sont des estimations combinées obtenues à partir de 200 cellules. . . . .	49
2.13 Estimation empirique d'un madogramme . . . . .	49
3.1 Plan de la Suisse avec 79 stations pluviométriques . . . . .	57
3.2 Plan de l'Algérie . . . . .	58
3.3 Plan de Algerie avec régions . . . . .	59
3.4 Simulation de processus Schlather . . . . .	62
3.5 Processus de Smith . . . . .	63
3.6 Processus de Schlather . . . . .	64
3.7 Processus de Extremal-t . . . . .	65
3.8 Représentation 3D de Smith . . . . .	66
3.9 Représentation des donn brutesdes pricipitations . . . . .	71
3.10 Données par stations . . . . .	71
3.11 Résumé des données par les boites à moustaches . . . . .	72
3.12 Représentation des distributions GEV . . . . .	74
3.13 . . . . .	75
3.14 Valeurs estimées du coefficient Extrémal par l'approche F-madogramme en fonction de la distance euclidienne( $h$ représente la distance euclidienne entre deux sites d'observation). . . . .	76
3.15 Les points représentent les valeurs estimées du coefficient extrémal par l'approche F-madogramme et la ligne rouge représente le graphe du coefficient extrémal du modèles (schlather,smith,brown,extrémal-t) approprié aux donnés ( $h$ représente la distance euclidienne entre deux sites d'observation) . . . . .	79

## LISTE DES TABLEAUX

1.1	Lois du domaine d'attraction de Gumbel . . . . .	13
1.2	Lois du domaine d'attraction de Weibull. . . . .	13
1.3	Lois du domaine d'attraction de Fréchet . . . . .	14
3.1	Tableau des coordonnées . . . . .	69
3.2	Tableau des données . . . . .	69
3.3	Tableau de summary . . . . .	72
3.4	Test d'homogénéité . . . . .	72
3.5	Estimateurs MLE des paramètre de GEV estimées . . . . .	73
3.6	Valeurs des paramètres estimés de la distribution GEV . . . . .	74
3.7	Tableau des coordonnés . . . . .	76
3.8	Les paramètre estimées de modèle de schlather . . . . .	77
3.9	Tableau de varcov de modèle de Schlather . . . . .	77
3.10	Tableau du paramètres du modèle de Smith . . . . .	77
3.11	Tableau de varcovar de modèle de Smith . . . . .	78
3.12	Tableau estimés de modèle de Brown . . . . .	78
3.13	Tableau de varcovar de modèle de Brown . . . . .	78
3.14	Tableau estimés de modèle de Extrémal-t . . . . .	78
3.15	Tableau de varcovar de modèle de Extrémal-t . . . . .	78
3.16	Le AIC des trois modèle . . . . .	78

## ملخص

يتمثل العمل الحالي في دراسة حدود القيم الإحصائية القصوى للمعطيات بناء على الاحداثيات الفضائية (خطوط الطول, خطوط العرض, الارتفاع) ذات توزيعات القيم القصوى المعممة بعد ان نقدم لمحة عن العمليات المستقرة القصوى ويعرض بعض النماذج مثل (سميث, شلاذر, براون, رزنيك, واكستر يمال تي), نقوم في هذه المذكرة بأمثلة لمحاكات نماذج العمليات القصوى المستقرة بواسطة البرنامج R, كما نقوم بإجراء دراسة عملية على بيانات مناخية لتساقط الامطار في أربعة مواقع لحوض متيجة (احمر العين, فوكة, فارشوفال والبليدة) وذلك بتقدير التوزيع GEV الذي يتوافق مع كل موقع باستخدام البرنامج R ونمذجة عمليات مستقرة لإيجاد تمثيل للمنطقة يتوافق مع الاحداثيات الفضائية .

## 0.1 Résumé

Le travail actuel consiste à étudier les limites des valeurs statistiques maximales des données en fonction des coordonnées spatiales (latitude, longitude et l'altitude) des distributions généralisées des valeurs maximales après avoir présenté un processus max stable et l'affichage de certains modèles (Smith, Schlather, Brown Resnick et Extrémal-t), Dans ce mémoire, nous fournissons des exemples pour simuler des modèles des processus max stables établis par logiciel R, nous menons également une étude pratique sur les données pluviométriques sur quatre sites d'un bassin de Mitidja (FOUKA, AMOUR, FER.A et BLIDA), en estimant la distribution GEV correspondant à chaque site à l'aide du logiciel R et en modélisant des processus max stable pour créer une représentation de surface correspondant aux coordonnées de l'espace.

## 0.2 Abstract

The current work consists in studying the limits of the maximum statistical values of the data as functions of the spatial coordinates (latitude, longitude and altitude) of the generalized distributions of the maximum values after having presented a stable max process and the display some models (Smith, Schlather, Brown Resnick and Extremal-t),

In this thesis, we provide examples to simulate stable max process models established by *R* software, we also conduct a practical study on rainfall data at four sites in a watershed. Mitidja (FOUKA, AMOUR EL AIN, FER.A and BLIDA), estimating the *GEV* distribution corresponding to each site using the *R* software and modeling stable max processes to show a surface representation corresponding to the coordinates of the space.

### 0.3 Notations et abréviations

- $TCL$  : Théorème central limite.
- $TVE$  : Théorie des valeurs extrêmes.
- $i.i.d$  : Indépendant et identiquement distribué.
- $GEV$  : Distribution des Valeurs Extrêmes Généralisée.
- $EMV$  : Méthode du Maximum de Vraisemblance.
- $GPD$  : Distribution de Pareto Généralisée.
- $POT$  : L'approche par dépassements de seuil.
- $BM$  : Block Maxima.
- $AIC$  : Critère d'Information d'AKAIKE
- $\mathbf{1}_{\{A\}}$  : fonction indicatrice de l'événement  $A$ .
- $\bar{F}$  :  $1 - F$ , fonction de survie.
- $\mathcal{N}(\mu, \sigma^2)$  : Loi normale de moyenne  $\mu$  et de variance  $\sigma^2$ .
- $X \stackrel{d}{=} Y$  : Égalité en distribution entre  $X$  et  $Y$ .
- $P(A)$  : La probabilité de l'événement  $A$ .
- $E(X)$  : L'espérance mathématique de la v.a.  $X$ .
- $Var(X)$  : La variance de la variable aléatoire  $X$ .
- $Cov(X, Y)$  : La covariance entre  $X$  et  $Y$ .
- $\mathbb{N}$  : Ensemble des entiers naturels.
- $\mathbb{N}^*$  : Ensemble des entiers naturels non nuls
- $\mathbb{R}$  : Ensemble des réels et  $\mathbb{R}^p = \underbrace{\mathbb{R} \times \mathbb{R} \dots \times \mathbb{R}}_{p \text{ fois}}$ ,  $p \in \mathbb{N}^*$
  
- $v.a.$  : variable aléatoire.
- $p.s.$  : Presque sûrement.

## INTRODUCTION GÉNÉRALE

Au cours des dernières années, nous avons pu observer dans la recherche scientifique, une modélisation des événements rares. Ces derniers sont des événements dont la probabilité d'apparition est trop faible, c'est-à-dire ils se trouvent dans les queues des distributions. Par définition, les événements rares sont des événements ayant une faible probabilité d'apparition. Ils sont dits extrêmes quand il s'agit de valeurs beaucoup plus grandes ou plus petites que celles observées habituellement. Lorsque le comportement de ces événements est dû au hasard on peut étudier leurs lois. La modélisation des événements extrêmes (tels les inondations, les crues, les tempêtes, les pics de pollution etc..) est aujourd'hui un champ de recherche particulièrement actif.

Ces événements extrêmes peuvent causer des dégâts humains et matériels considérables. De telles catastrophes ne peuvent pas toujours être évitées. Cependant, la société peut prendre des actions préventives pour minimiser leurs effets. Pour cela, les spécialistes ont à leur disposition la théorie statistique des valeurs extrêmes. Elle donne un résultat très intéressant parce qu'elle est d'une portée générale.

La théorie des valeurs extrêmes est une branche de la statistique qui essaie d'amener une solution face à des événements rares. Elle repose principalement sur des distributions limites des extrêmes et leurs domaines d'attraction. Cependant, on y retrouve deux modèles : loi généralisée des extrêmes (GEV : " Generalized Extreme Value ") et loi de Paréto généralisée

(GPD : " Generalized Pareto Distribution ").

Tout a commencé avec les auteurs Fisher et Tippett (1928)[26] quand ils étudiaient la résistance des fils de coton puis plus tard Gnedenko (1943)[28] s'est intéressé à ces distributions. Ils ont énoncé un théorème fondamental avec la création de trois domaines d'attraction : domaine d'attraction de Fréchet, Gumbel et Weibull.

Cette théorie est largement utilisée, elle a trouvé des champs d'application en :

- Climatologie : étude des événements climatiques extrêmes (précipitations, températures, chutes de neige), modélisation des grands feux de forêt (voir par exemple Alvarado et al. 1998)[31].
- Hydrologie : crues consécutives à des pluies torrentielles : aux Pays-Bas, digues

menacées par l'effet conjoint des grandes marées et des conditions climatiques en Mer du Nord (Inondation de 1953).

Ces notions ont d'abord été développées pour la théorie univariée des extrêmes et les échantillons indépendants, puis étendues au cadre multivarié et aux échantillons dépendants.

Aujourd'hui, une direction de recherche particulièrement active est celle de la théorie spatiale des extrêmes, où l'on quitte le contexte fini-dimensionnel pour se placer dans un cadre fonctionnel et où les processus occupent une place prépondérante. Cette composante spatiale est présente naturellement dans les applications aux sciences de l'environnement, sciences de la terre et sciences humaines citées plus haut, et un enjeu important est celui de la modélisation de la dépendance spatiale des extrêmes et des problèmes d'inférence relatifs. Dans ce mémoire on essayera de donner une introduction générale à tout ces concepts, en particulier la théorie spatiale des valeurs extrêmes et les processus max-stables.

Ce mémoire est composé d'une introduction générale, de trois chapitres, d'une conclusion générale et d'une bibliographie.

**Le premier chapitre :** Dans ce chapitre, on expose la théorie probabiliste des valeurs extrêmes dans le cas univarié. Au début, dans la première section, on donne Le résultat fondamental de la théorie des valeurs extrêmes, qui est le théorème de Fisher-Tippett (1928)[26] & Gnedenko (1943)[28], est l'analogue du théorème de la limite centrale pour le maximum d'une suite de variables aléatoires indépendantes et de même loi. on présentera le résultat principal de la théorie des valeurs extrêmes qui montre que la loi limite du maximum d'une suite de variables aléatoires indépendantes et de même loi (GEV) ne peut appartenir qu'à l'un des trois types de distributions suivants : Gumbel, Weibull ou Fréchet. Ensuite, dans la deuxième section, on définit la distribution de Paréto généralisée «Generalized Paréto Distribution (GPD)» qui est un outil fondamental pour la méthode des «Excès au-delà d'un seuil»

**Le deuxième chapitre :** Dans ce chapitre, on présentera une introduction en processus max stable, qu'est une généralisation de la théorie des valeurs extrêmes au contexte spatial, qui est le cœur du sujet de ce mémoire, Après. On va voir qu'on peut transformer, sans perte de généralisation, les lois de probabilités marginales d'un processus max-stable en lois de probabilité de Fréchet standards unitaires, ce qui est important pour les applications. On va donner les représentations spectrales des processus max-stables, en particulier la représentation de De Haan & Pickands. Ces représentations spectrales sont à la base de la construction de certains modèles des processus max-stables, Smith, Schlather, Brown Resnick, Extremal-t qui nous permette de mettre en applications les processus max-stables. On définira « le coefficient extrémal » qui est un outil qui caractérise la structure, de dépendance entre les extrêmes spatiaux, et qui mesure le degré de dépendance entre les réalisations d'un processus max-stable. Ensuite, on définira un autre moment d'un processus qui caractérise la dépendance entre ses réalisations, « le F-

madogramme », qui sera utilisé pour l'estimation du coefficient extrémal d'un processus max-stable

**Le troisième chapitre :** dans ce chapitre on a défini des simulations des modèles de processus max stable, en suite on va donner une application de l'ajustement sur les 4 stations de Mitidja pour estimer les paramètres des 4 modèles à l'aide de logiciel R de packages SpatialExtrême.

# CHAPITRE 1

## MODÉLISATION DES QUEUES DE DISTRIBUTIONS

### 1.1 Introduction

La théorie des valeurs extrêmes communément appelée « Extreme Value Theory » (*EVT*) en anglais, est une vaste théorie dont le but est d'étudier les événements rares c'est à dire les événements dont la probabilité d'apparition est faible. Autrement dit elle essaie d'amener des éléments de réponse aux intempéries, aux inondations, aux catastrophes naturelles, aux problèmes financiers,... etc. En prédisant leurs occurrences dans les années venir. En d'autres termes on veut estimer des petites probabilités ou des quantités dont la probabilité d'observation est très faible c'est-dire proche de zéro. Ces quantités sont appelées quantiles extrêmes car l'ordre de ces quantiles tend vers zéro lorsque la taille de l'échantillon,  $n$ , tend vers l'infini.

### 1.2 Théorie des valeurs extrêmes dans le cas de variables IID Unidimensionnelles

Le résultat de base de la *TVE* consiste à décrire la loi asymptotique du maximum de  $n$  variables indépendantes et identiquement distribuées (*i.i.d*). La forme de cette loi (notée *GEV*) dépend d'un seul paramètre qui permet de spécifier le comportement de la queue de la loi considérée. Ce paramètre est l'indice des valeurs extrêmes (noté  $\xi$ ). En d'autres termes, le résultat de base de la *TVE* assure que la loi du maximum de  $n$  variables (*i.i.d*) est toujours dans le domaine d'attraction d'une loi *GEV*

### 1.2.1 Extrema par blocs

Si l'on se donne un échantillon de  $n$  variables aléatoires  $(X_i)_{i=1,\dots,n}$  indépendantes et identiquement distribuées (*i.i.d*), la statistique classique s'intéresse à la loi de

$$S_n = \sum_{i=1}^n X_i .$$

convenablement centrée et normalisée. On sait grâce aux travaux de Paul Lévy que  $S_n$  (centré et normalisé) converge en loi vers un élément de la classe des lois stables pour l'addition, et même le plus souvent vers son élément le plus connu. Or, quand on s'intéresse aux queues de distributions, ce n'est plus la moyenne que l'on va chercher à décrire, mais le comportement de  $M_n = \max_{i=1,\dots,n} X_i$ , à nouveau correctement centré et normalisé. De même que dans le cas de  $S_n$ , la famille de lois possibles pour  $M_n$  est connue, cette fois-ci grâce aux travaux de Fisher et Tippett [26]. Comme dans le cas du comportement asymptotique de la moyenne, la famille de lois limites sera stable, non par l'application de la somme, mais par l'application du maximum, d'où appellation de loi max-stables. Cette famille est caractérisée par une équation fonctionnelle portant sur sa fonction de répartition, et dont les solutions sont les membres de la famille *GEV* (Generalized Extreme Value Distribution). Plus précisément, on a le théorème suivant :

**Théorème 1.1 (Fisher-Typett, 1928)** [26] *S'il existe des suites  $(a_n)_{n \in \mathbb{N}} > 0$  et  $(b_n)_{n \in \mathbb{N}} \in \mathbb{R}$  telles que*

$$\lim_{n \rightarrow \infty} P \left\{ a_n^{-1} (M_n - b_n) \leq x \right\} = (F(a_n x + b_n))^n \longrightarrow H(x) \quad (1.1)$$

*alors  $H$  est un membre de la famille *GEV*, c'est-à-dire qu'il existe des paramètres  $\mu, \sigma > 0$  et  $\xi$  tels que*

$$H(x) = \begin{cases} \exp \left[ - \left( 1 - \xi \frac{x - \mu}{\sigma} \right)^{\frac{1}{\xi}} \right], & x / (1 - \xi(x - \mu) / \sigma) > 0, \quad \text{si } \xi \neq 0 \\ \exp(-e^{-x}), & \text{si } \xi = 0 \end{cases} \quad (1.2)$$

*où  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  et  $\xi \in \mathbb{R}$  sont trois paramètres dépendants de la loi initiale des observations  $(X_i)_{i=1,\dots,n}$ .*

Le théorème de Fisher-Tippett fournit, en quelque sorte, la contrepartie du Théorème Central Limite (*TCL*) dans le cas d'événements extrêmes. Cependant, contrairement au *TCL*, où la loi normale est la seule loi limite possible, dans le cas des extrêmes, trois types de loi limite sont possibles

Les trois paramètres de  $H$ ,  $\mu$ ,  $\sigma$  et  $\xi$ , sont respectivement ses paramètres de position, d'échelle, et de queue. Les comportements diffèrent largement suivant ce dernier paramètre : en effet, pour  $\xi > 0$  on a une loi à queue lourde (loi de type Fréchet). pour  $\xi < 0$  on a une loi à support borné (loi de type Weibull), alors que le cas  $\xi = 0$  est intermédiaire, correspondant à une loi de type Gumbel.

La loi  $H$  est l'une des trois lois limites :

### Distribution de Gumbel

$$\mathring{A}(x) = \exp\{-\exp[-x]\}, x \in \mathbb{R}$$

(1.3)

### Distribution de Fréchet

$$\varphi(x) = \begin{cases} \exp\{-x^{-\alpha}\}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

(1.4)

### Distribution de Weibull

$$\psi(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

(1.5)

Sachant que  $\alpha = \frac{1}{\xi}$  c'est l'indice de la queue de la distribution.

La figure 1.1 montre la fonction de répartition de certains membres de cette famille, montrant bien les différents comportements de la queue de la distribution. Il est intéressant à la vue de ce théorème de se demander s'il est possible de prévoir la valeur des paramètres de la loi limite du maximum quand on connaît la loi dont sont issues les observations : c'est la question du domaine d'attraction, défini de la manière suivante :

**Définition 1.1** Une variable aléatoire  $X$  appartient au domaine d'attraction de la distribution des valeurs extrêmes  $F$  s'il existe des suites déterministes  $(a_n)_{n \in \mathbb{N}} > 0$  et  $(b_n)_{n \in \mathbb{N}}$  telles que I.1 est vérifiée.

On peut alors chercher à déterminer dans quel domaine d'attraction appartiennent les lois usuelles. Cette approche n'a pas été retenue dans notre étude car dans le cas général, la loi des observations est inconnue, et on ne peut donc pas étudier le domaine d'attraction de cette façon. Un apprentissage important néanmoins de cette approche est la classification des lois usuelles suivant leur appartenance à un domaine d'attraction de distributions aux queues plus ou moins lourdes . On pourra retenir que la loi normale,

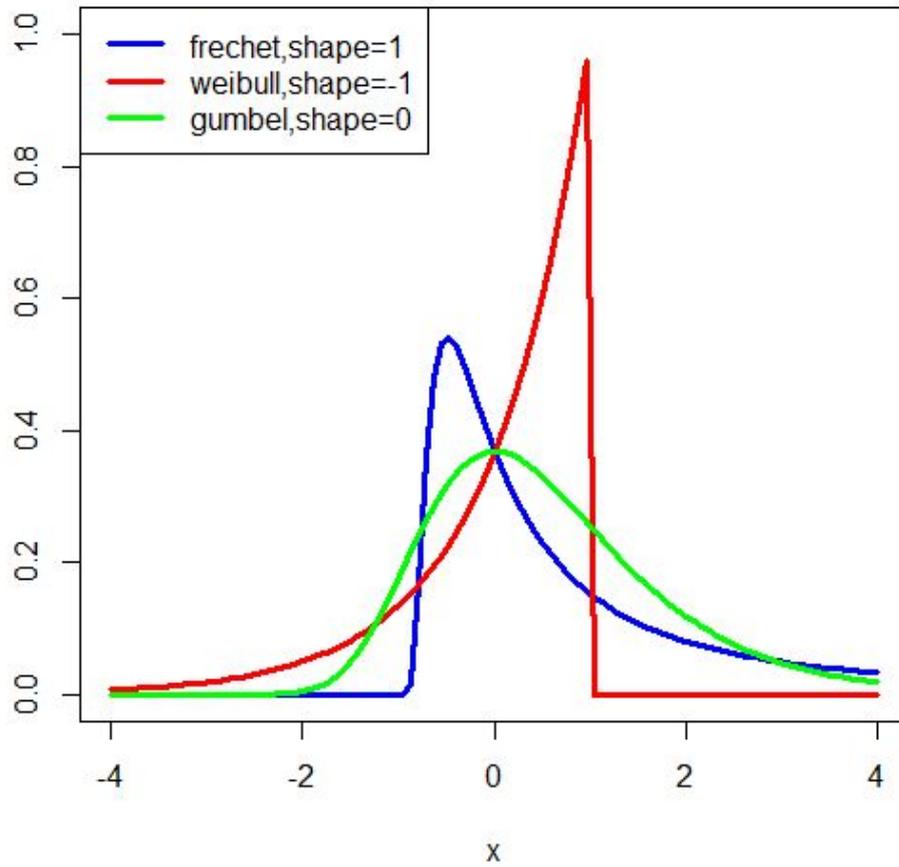


FIGURE 1.1 – Représentation graphique de la densité des valeurs extrêmes

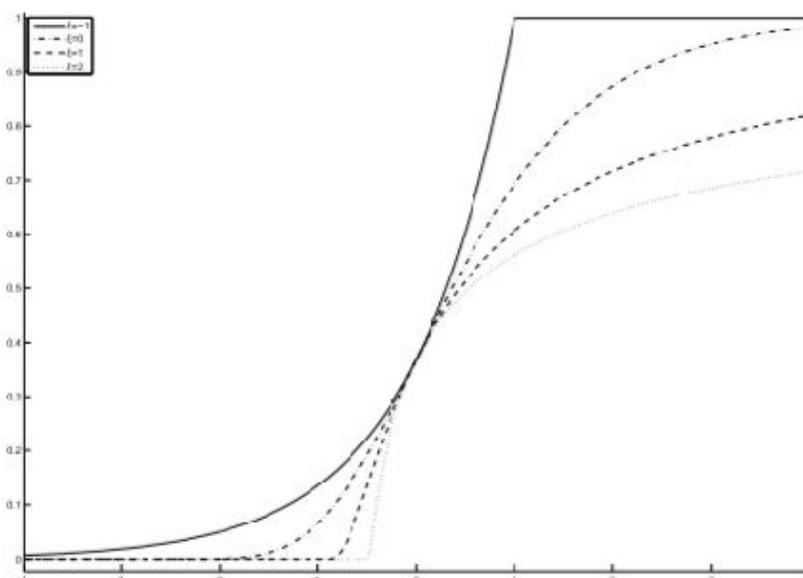


FIGURE 1.2 – Fonction de répartition de lois GVE pour différentes valeurs de  $\xi$ , avec  $\sigma = 1$  et  $\mu = 0$

très utilisée, est dans le domaine d'attraction de la loi de Gumbel, correspondant, on le rappelle, à une queue de distribution légère, quoique non bornée supérieurement.

### 1.2.2 Estimation

D'un point de vue statistique, l'approche couramment employée consiste à utiliser le théorème I.1 pour estimer le comportement du maximum : les données sont séparées en groupes de tailles égales (par exemple des blocs d'une année), sur lesquels on calcule les maxima, et on suppose alors que l'on dispose d'un échantillon *i.i.d* de loi *GEV*, ce qui serait réaliste pour des blocs dont la taille tend vers l'infini. De nombreux travaux se sont attachés à l'estimation des paramètres en partant de cette modélisation : maximum de vraisemblance, la méthode des L-moments Hosking [35], la méthode des moments pondérés pour l'estimation des trois paramètres ; estimateur de Hill 1975 [34], de Pickands 1974 [50], de Dekkers-Einmahl-De Hann 1989 [19] pour l'estimation du paramètre de queue ; chaque méthode citée étant valable dans un domaine particulier du paramètre de queue.

Par exemple, il est connu depuis les travaux de Smith 1987 [63] que l'estimateur du maximum de vraisemblance fonctionne sur un domaine restreint en ce qui concerne les valeurs de  $\xi$ , en partie à cause du fait que le support de la loi dépend de ce paramètre, contrairement aux cas usuels, comme par exemple la famille exponentielle. On a plus particulièrement le résultat suivante :

- Si  $\xi > -0.5$ , l'estimateur du maximum de vraisemblance est régulier, dans le sens où il vérifie les propriétés habituelles d'efficacité et de normalité asymptotique ;
- Si  $-1 < \xi < -0.5$ , l'estimateur du maximum de vraisemblance peut être généralement calculé, mais ne possède pas les propriétés habituelles, en particulier la normalité asymptotique ;
- Si  $\xi < -1$ , l'estimateur du maximum de vraisemblance n'est généralement pas calculable du fait de problèmes numériques.

Ces limitations sont souvent peu contraignantes en pratique, car le cas où  $\xi < -0.5$  correspond à des queues bornées très courtes, qui est un cas assez peu fréquent lorsque l'on s'intéresse à la modélisation des valeurs extrêmes, en particulier lorsque que l'on s'intéresse aux extrêmes de données environnementales ou financières, pour lesquelles on constate des paramètres de queue positifs, ou nuls.

Un des problèmes de cette approche est que toute l'information collectée n'est pas retenue pour l'estimation, étant donné que l'on ne conserve que les maxima sur un bloc, par exemple une année. Il y a alors un compromis à établir entre la taille des blocs, pour justifier de l'utilisation de la loi limite pour le maximum, mais aussi sur le nombre de blocs résultants pour avoir assez d'observations afin de mener à bien l'estimation des paramètres.

Lorsque l'on regroupe les observations pour n'en conserver que le maximum par bloc, on perd non seulement l'information temporelle sur les moments où ces maxima sont

observés et sur leur étendue temporelle comme sur leur tendance à l'agrégation ou à la répulsion, mais on perd aussi une information sur le comportement de la distribution aux alentours du maxima, comme la forme d'un évènement extrême.

De plus, les conditions d'application du théorème de Fisher-Tippett ne sont que rarement vérifiées, étant donné que les observations initiales ne sont pas systématiquement *i.i.d.*, et il faut donc développer des modèles dans ce cas-là : ce point fera l'objet de la dernière partie de ce chapitre, les sections intermédiaires étant dédiées à l'extension de la description des lois max-stables au cas multivarié et à celui des processus stochastiques. Auparavant, nous intéresserons à la description d'une méthode se passant du regroupement par blocs.

## 1.3 Domaine d'attraction

Il faut bien noter que le paramètre conditionne le type de la loi des valeurs extrêmes. Nous présentons dans ce qui suit, les domaines d'attractions dans les trois cas correspondant au paramètre.

On commence par définir les fonctions à variation régulière.

**Définition 1.2** Une fonction  $L$  mesurable et positive sur  $]0, \infty[$  est à variations lentes, si, pour tout  $x > 0$ ,

$$\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1.$$

### 1.3.1 Le domaine d'attraction de la distribution de Gumbel

Cas où  $\xi = 0$  : La loi présente dans la queue une décroissance de type exponentielle, ce qui permet de caractériser dans ce cas, le domaine d'attraction de Gumbel  $D(\Lambda)$ . Ce dernier est délicatement traitable, car il n'y a pas de lien direct entre la queue de la loi et les fonctions variations lentes définies par (Delmas et Jourdain (2006))[20] :

Von Mises (1936), [43] a donné une caractérisation simple pour le domaine d'attraction de Gumbel, formulée par le biais du théorème suivant :

**Théorème 1.2** S'il existe une fonction mesurable  $R$ ; appelée fonction auxiliaire telle que :

$$\lim_{x \rightarrow \omega(F)} \frac{1 - F(t + xR(t))}{1 - F(t)} = \exp(-x)$$

avec

$$\omega(F) = \inf\{x \in \mathbb{R} : F(x) \geq t\}.$$

est le point terminal de  $F$ , alors

$$F \in D(\Lambda)$$

— Le tableau suivant fournit quelques exemples de lois qui appartiennent au domaine d'attraction de Gumbel.

TABLE 1.1 – Lois du domaine d’attraction de Gumbel

Loi	$1 - F(x)$
Benktender II , $\alpha, \beta > 0$	$x^{-(1-\beta)} \exp\left(-\frac{\alpha}{\beta} x^\beta\right)$
Logistic	$\frac{1}{1+\exp(x)}$

### 1.3.2 Le domaine d’attraction de la distribution de Weibull

Cas où  $\xi < 0$  : Le domaine d’attraction dans ce cas est celui de Weibull  $D(\Psi_\xi)$ . Les lois de ce domaine sont bornées à droite, et par conséquent, le point terminal  $\omega(F)$  est fini. Une caractérisation d’appartenance à ce domaine d’attraction est donnée par le théorème suivant (pour la démonstration, voir Gnedenko (1943,)) [28] :

**Théorème 1.3** Une fonction de répartition  $F$  appartient au  $D(\Psi_\xi)$  si et seulement si  $\omega(F) < +\infty$  et

$$\bar{F}\left(\omega(F) - \frac{1}{x}\right) = x^{-\frac{1}{\xi}} L(x)$$

avec :  $\bar{F}$  est la fonction de survie donnée par  $\bar{F}(x) = 1 - F(x)$ ;  $L$  est une fonction variation lente.

- Le tableau suivant présente quelques lois qui appartiennent au domaine d’attraction de Weibull.

TABLE 1.2 – Lois du domaine d’attraction de Weibull.

Loi	$1 - F\left(\omega(F) - \frac{1}{x}\right)$	$\xi$	$L(x)$
Uniforme	$\frac{1}{x}, x > 1$	-1	1
Weibull	$1 - \exp(-x^{-\alpha})$	$-\frac{1}{\alpha}$	$1 - \frac{x^{-\alpha}}{2} + o(x^{-\alpha})$
Reverse Burr ( $\lambda, \beta, \tau$ )	$\left(\frac{\beta}{\beta+x^\tau}\right)^\lambda$ avec $x > 0$	$-\frac{1}{\lambda\tau}$	$\beta^\lambda(1 - \lambda\beta x^{-\tau}) + o(x^{-\tau})$

### 1.3.3 Le domaine d’attraction de la distribution de Fréchet

Cas ou  $\xi > 0$  : Ce cas correspond au domaine d’attraction de Fréchet. Les lois appartenant à ce domaine d’attraction sont caractérisées par une queue à décroissance lente (polynomiale) à l’infini, et un point terminal  $\omega(F) = +\infty$ . Elles sont dites aussi lois à queues lourdes. Une caractérisation de ce domaine d’attraction noté  $D(\Phi_\xi)$  est donnée par le théorème suivant (pour la démonstration, voir Gnedenko (1943)) [28] :

**Théorème 1.4** Une fonction de répartition  $F$  appartient au  $D(\Phi_\xi)$  si et seulement si sa fonction de survie est donnée par :

$$\bar{F}(x) = x^{-\frac{1}{\xi}} L(x)$$

ou  $L$  est une fonction à variation lente.

- Le tableau suivant présente quelques lois qui appartiennent au domaine d’attraction de Fréchet .

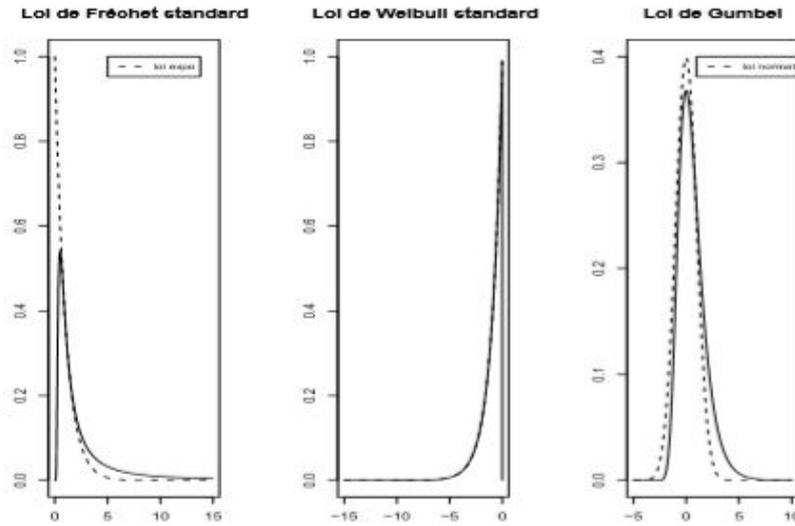


FIGURE 1.3 – Les graphes représentent les domaines d’attraction des trois lois extrêmes

TABLE 1.3 – Lois du domaine d’attraction de Fréchet

Loi	$1 - F(x)$	$\xi$	$l(x)$
Pareto( $\alpha$ ) avec $\alpha > 0$	$x^{-\alpha}$	$\frac{1}{\alpha}$	1
Burr( $\eta, \tau, \lambda$ ) avec $\eta, \tau, \lambda > 0$	$\left(\frac{\eta}{\eta + x^\tau}\right)^\lambda, x > 0$	$\frac{1}{\lambda\tau}$	$\left(\frac{\eta}{1 + \frac{\eta}{x^\tau}}\right)^\lambda$
Pareto Généralisé( $\sigma, \xi$ ) avec $\sigma, \xi > 0$	$\left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}}, x > 0$	$\xi$	$\left(\frac{\sigma}{\xi}\right)^{\frac{1}{\xi}} \left(1 + \frac{\sigma}{\xi x}\right)^{-\frac{1}{\xi}}$

1. Une fonction de répartition  $F$  peut appartenir à deux domaines d’attraction différents  $D_1$  et  $D_2$  (pour plus de détails, voir Resnick (1987))[54].
2. Les variables aléatoires dans les trois domaines d’attraction de Gumbel, de Fréchet et de Weibull sont liées par la relation suivante (voir Embrecht et al. (1997))[24] :

$$X \in D(\Phi_\xi) \iff \log(X^{-\frac{1}{\xi}}) \in D(\Lambda) \iff -X^{-1} \in D(\Psi_\xi)$$

La *GEV* est une approche basée sur la série de maxima de blocs, donc cette méthode a l’inconvénient qu’une seule observation entre dans la modélisation ; cela conduit à une perte d’information contenue dans les autres grandes valeurs de l’échantillon. Contrairement à la *GEV*, la *GPD* est apparue pour remédier les défauts de la *GEV*, cette méthode a l’avantage que plusieurs observations qui dépassent un certain seuil seront présentes dans la modélisation, pour cette raison nous les introduisons dans la section suivante

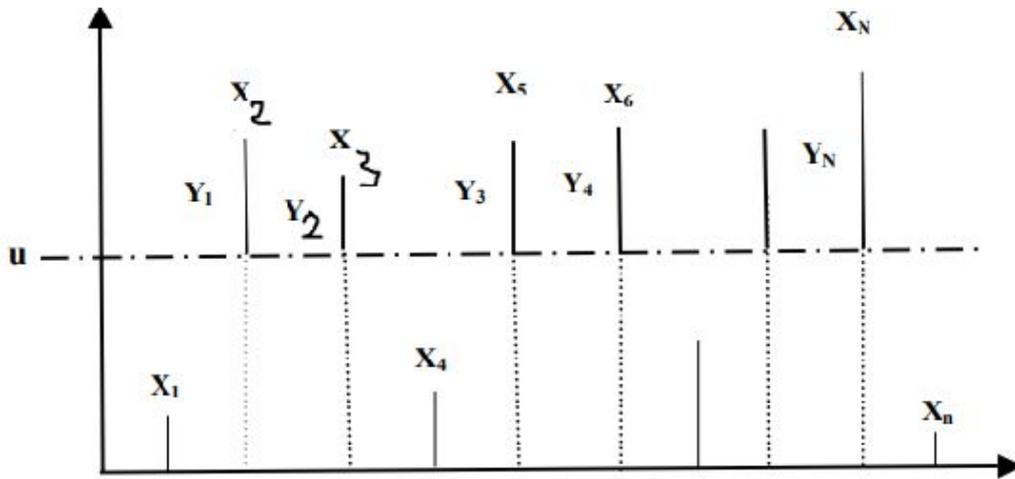


FIGURE 1.4 – Méthode des excès :  $u$  réel suffisamment élevé appelé seuil,  $y$  : excès de  $x$  au-delà de  $u$

## 1.4 Loi des dépassements de seuils

### 1.4.1 Excès au-delà d'un seuil

La méthode des excès au-delà d'un seuil (ou Peak Over Threshold, POT) repose sur le comportement des valeurs observées au-delà d'un seuil donné. En d'autres termes, elle consiste à observer non pas le maximum ou les plus grandes valeurs mais toutes les valeurs des réalisations qui excèdent un certain seuil élevé. L'idée de base de cette approche consiste à choisir un seuil suffisamment élevé et à étudier les excès au-delà de ce seuil. Cette méthode initialement développée par Pickands [1975][46] et abondamment étudiée par divers auteurs tels que de Smith [1987][63], Davison et Smith [1990][65], ou Reiss et Thomas (2001) [53] pour de plus amples références.

On définit un seuil  $u \in \mathbb{R}$

$$N_u = \text{card} \{i, i = 1, \dots, n, : X_i > u\}$$

et

$$Y_i = X_i - u > 0$$

pour  $1 \leq j \leq N_u$  ou  $N_u$  est le nombre de dépassements du seuil  $u$  par les  $(X_i)_{1 \leq i \leq n}$  et  $Y_1, \dots, Y_{N_u}$  les excès correspondants (Figure 1.4).

On cherche à partir de la loi  $F$  de  $X$  à définir une loi conditionnelle  $F_u$  par rapport au seuil  $u$  pour les variables aléatoires dépassant ce seuil.

On définit alors la loi conditionnelle des excès  $F_u$  par :

$$F_u(y) = \Pr(X - u \leq y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}, y \geq 0 \quad (1.6)$$

Le théorème de Pickands-Balkema-de Haan ci-après donne la forme de la loi limite

pour les valeurs extrêmes : sous certaines conditions de convergence, la loi limite est une loi de Pareto généralisée que l'on notera *GPD*.

**Théorème 1.5 (de Pickands-Balkema-de Haan)** Une fonction de répartition  $F$  appartient au domaine d'attraction maximale de  $H_\xi$ , si et seulement si, il existe une fonction positive  $\beta(u)$  telle que :

$$\lim_{u \rightarrow x_F} \sup_{0 \leq y \leq x_F - u} |F_u(y) - G_{\xi, \beta(u)}(y)| = 0 \quad (1.7)$$

où  $F_u(y)$  est la fonction de répartition conditionnelle excès pour  $u$  élevé

$x_F$  est le point terminal de  $F$

$$x_F = \{x \in \mathbb{R} : F(x) < 1\}$$

et  $G_{\xi, \beta(u)}(y)$  est la *GPD* donnée par :

$$G_{\xi, \beta(u)}(y) := \begin{cases} 1 - \left(1 + \xi \frac{y}{\beta(u)}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\beta(u)}\right), & \xi = 0 \end{cases} \quad (1.8)$$

où :  $y \geq 0$  pour  $\xi \geq 0$  et  $0 \leq y \leq -\frac{\beta(u)}{\xi}$  pour  $\xi < 0$

Ce théorème montre l'existence d'une relation étroite entre la *GPD* et la *GEV* : Pickands [1975][46] a montré que pour n'importe quelle loi  $F$ , l'approximation *GPD* définie par (1.7) n'est vérifiée que s'il existe des constantes de normalisation et une loi non dégénérée telle que le résultat donné par (1.1) est vérifié. Dans ce cas, si  $H$  est écrite sous la forme d'une *GEV* équation (1.2), alors l'indice de queue  $\xi$  est le même que celui de la *GPD* donnée dans (1.8). De même pour la *GPD*, le cas où  $\xi > 0$  correspond aux lois à queues épaisses, pour le quel  $1 - G$  se comporte comme une puissance.  $x^{-\frac{1}{\xi}}$  pour  $x$  assez élevé. Si  $\xi = 0$  on trouve  $1 - \exp\left(-\frac{y}{\beta(u)}\right)$ , c'est la loi exponentielle de paramètre  $\beta$  et enfin pour  $\xi > 0$  c'est la loi pareto de type II à support bornée Figure 1.5 La *GPD* a les propriétés suivantes :

$$E(Y) = \frac{\beta}{1 - \xi}, (\xi < 1) \text{ et } V(Y) = \frac{\beta^2}{(1 - \xi)^2 (1 - 2\xi)}, \left(\xi < \frac{1}{2}\right)$$

En pratique le choix du seuil constitue une difficulté. En fait,  $u$  doit être assez grand pour que l'approximation *GPD* définie par (1.7) soit valide, mais pas trop élevé pour garder un nombre suffisant de dépassements pour estimer les paramètres du modèle. Le seuil doit être choisi de façon à faire un arbitrage, traditionnel en statistiques, entre le biais et la variance. Généralement,  $u$  est déterminé graphiquement en exploitant la linéarité de la fonction d'excès moyenne  $e(u)$  pour la *GPD* (cf. [Embrechts et al., 1997])[24]. Cette technique fournit une aide précieuse, cependant, il ne faut pas attendre d'elle la bonne valeur de  $u$ .

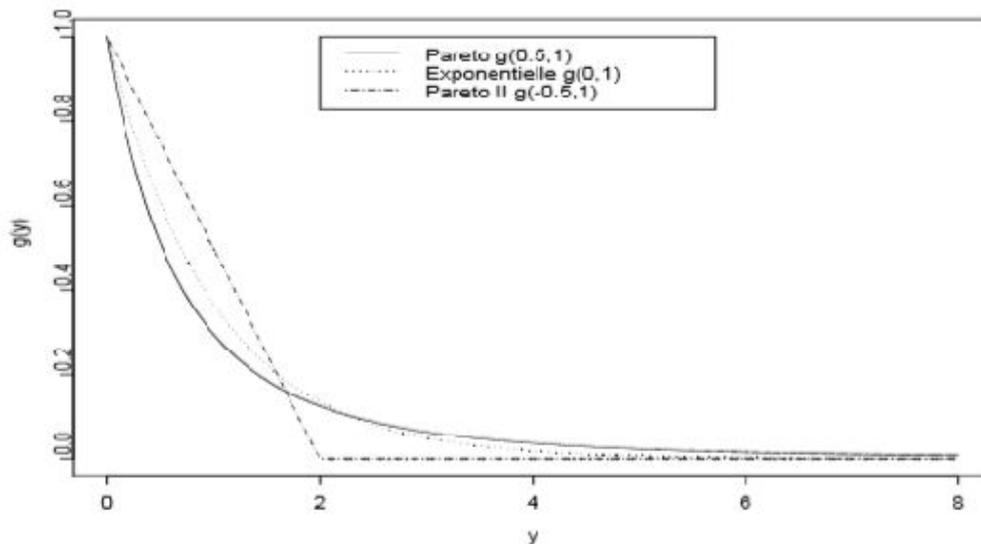


FIGURE 1.5 – Densité des lois de valeurs extrêmes, avec  $\xi = 0.5$  pour la loi de Pareto,  $\xi = 0$  pour la loi exponentielle et  $\xi = -0.5$  pour la loi de Pareto II

En pratique, plusieurs valeurs de  $u$  doivent être testées. Ce problème du choix a suscité de nombreux travaux dans la littérature. Beirlant et al. [1996][2] ont suggéré de choisir le seuil  $u$  qui minimise l’erreur quadratique moyenne asymptotique de l’estimateur de Hill de l’indice de queue, tout en supposant que  $F$  appartient au domaine d’attraction maximale de Fréchet. Parmi d’autres, Hall [1990] [33]; Danielsson et al. [2001][12], ont introduit des approches Bootstrap pour trouver le seuil optimal.

Une autre méthode utilisée par plusieurs auteurs (cf. par exemple, [McNeil et Frey, 2000; Raggad, 2007; Marimoutou et al., 2009])[41] [55][39] consiste à choisir un seuil aléatoire. En fait, ils prennent  $k$  observations excédentaires et le  $k$  peut être déterminé par des méthodes Monte-Carlo. Pour une discussion des principales approches, récemment développées, permettant le choix du seuil optimal, nous renvoyons à Matthys et Beirlant [2000][40].

## 1.4.2 Estimation

### 1.4.2.1 Choix du seuil

Le résultat théorique énoncé précédemment fait apparaître un comportement quand on observe les excès au-delà d’un seuil qui devient arbitrairement grand. Or, en pratique, il faut le laisser fini sous peine de ne plus avoir d’observations le dépassant. On voit alors apparaître une nécessité de compromis entre la justesse de l’approximation, engendrant un biais, et le nombre d’observations conservées, dont la diminution quand le seuil augmente génère une variance d’estimation plus importante. Deux méthodes principales sont utilisées par les praticiens, une exploratoire et une basée sur l’estimation des paramètres :

- Supposons qu’au-delà d’un seuil  $u_0$ ,  $X$  suive effectivement une loi GPD de paramètres  $\sigma_{u_0}$  et  $\xi < 1$ , ce dernier paramètre étant invariant du seuil. On peut montrer

le résultat suivant pour un autre seuil  $u > u_0$  :

$$E(X - u | X > u) = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_{u_0} + \xi u}{1 - \xi}$$

et donc l'espérance des dépassements de seuil  $u$  est une fonction linéaire de  $u$ . Puisque cette quantité est facilement estimable par sa contre partie empirique, nous avons une première approche permettant de choisir un seuil convenable, en cherchant le seuil au-delà du quel la fonction précitée est linéaire.

- Une approche alternative consiste à estimer les paramètres de la loi *GPD* (voir ci-après les méthodes d'estimation) et de chercher une valeur du seuil au-delà de laquelle ces estimations sont stables. Pour obtenir des paramètres indépendants du seuil, Coles (2001)[7] propose de tracer le comportement de  $\sigma^* = \widehat{\sigma}_u - \widehat{\xi}u$  et de  $\widehat{\xi}$

#### 1.4.2.2 Estimation des paramètres

Une fois le seuil fixé, plusieurs approches existent pour l'estimation des paramètres : on peut calculer la vraisemblance associée aux dépassements de seuils, car la loi des excès suit — asymptotiquement— une loi de Pareto généralisée. Il est ainsi possible de construire l'estimateur au maximum de vraisemblance pour estimer les paramètres  $\sigma$  et  $\xi$ , ce dernier étant le même que lorsque l'on s'intéresse au maxima par bloc.

Plus précisément, le modèle retenu est le suivant :

$$P\{X > x | X > u\} = \left[1 + \xi \left(\frac{x - u}{\sigma}\right)\right]^{-\frac{1}{\xi}}$$

Ce que l'on peut aussi écrire :

$$P\{X < x\} = 1 - \lambda_u \left[1 + \xi \left(\frac{x - u}{\sigma}\right)\right]^{-\frac{1}{\xi}}$$

avec  $\lambda_u = P\{X > u\}$ . On voit ainsi qu'il est nécessaire d'estimer trois paramètres, et bien que l'estimation de  $\lambda_u$  puisse se faire trivialement à l'aide du nombre de dépassement, l'erreur d'estimation devrait être prise en compte lors du calcul des intervalles de confiance sur les quantiles extrêmes, ce qui semble rarement fait en pratique. Nous proposerons dans la suite une méthode alternative qui permet de s'affranchir de ce problème, tout en retirant le fait que le paramètre  $\sigma$  estimé dépende du seuil  $u$ , grâce à une autre approche sur les dépassements de seuils.

Le lecteur peut aussi se référer aux articles mentionnés dans la partie précédente et qui s'intéressent à l'estimation de  $\xi$ , vu que les méthodes d'estimation de ce paramètre sont transposables au cas présent.

## 1.5 L'estimation de l'indice de queue

On présente ici trois estimateurs différents, tous basés sur les statistiques d'ordre  $X_{k,n} \leq \dots \leq X_{1,n}$ , obtenue à partir de la série initiale en considérant les  $k$  valeurs les plus grandes (ou les plus petites).  $k$  dépend a priori de  $n$ , même si on ne le mentionnera pas dans la notation : l'idée est d'avoir  $k \rightarrow \infty$  lorsque  $n \rightarrow \infty$ , mais sans prendre « trop » de valeurs de l'échantillon, ce qui conduit à imposer  $\frac{k}{n} \rightarrow 0$ .

Incidentement, cela implique que se posera la question du choix optimal de  $k$ . En effet, il est indispensable de calculer ces estimateurs sur les queues de distribution. Choisir un  $k$  trop élevé engendre le risque de prendre en compte des valeurs qui ne sont pas extrêmes, à l'inverse, un sous-échantillon trop petit ne permet pas aux estimateurs d'atteindre leur niveau de stabilité. Enfin, on retiendra que l'approche non paramétrique n'est envisageable que si l'on dispose d'un nombre important d'observations : dans le cas où les échantillons sont de petite taille

### 1.5.1 Estimateur de Pickands

L'estimateur de Pickands est construit en utilisant trois statistique d'ordres. Cet estimateur a l'avantage d'être valable quel que soit le domaine d'attraction de la distribution et par conséquent, du domaine de définition de l'indice des valeurs extrêmes. Pickands (1975)[46] démontre la consistance faible de son estimateur. La convergence forte

ainsi que la normalité asymptotique ont été démontrées par Dekkers et de Haan (1989)[19].

**Définition 1.3** *On suppose que  $\{X_i, i = 1, \dots, n\}$  est une suite de variables aléatoires indépendantes de loi  $F$  appartenant à l'un des domaines d'attractions. Soit  $(k_n)_{n \geq 1}$  une suite d'entiers avec  $1 \leq k_n < n$ , l'estimateur de Pickands est défini par*

$$\widehat{\xi}_{k_n}^p = \frac{1}{\ln 2} \ln \left( \frac{X_{n-k_n+1,n} - X_{n-2k_n+1,n}}{X_{n-2k_n+1,n} - X_{n-4k_n+1,n}} \right) \quad (1.9)$$

**Théorème 1.6 (Propriétés de l'estimateur de Pickands)** *Soit  $(k_n)_{n \geq 1}$  une suite d'entiers telle que  $1 \leq k_n < n$ ,  $k_n \rightarrow \infty$  et  $k_n/n \rightarrow 0$  quand  $n \rightarrow \infty$*

- *Alors,  $\widehat{\xi}_{k_n}^p$  converge en probabilité vers  $\xi$ .*
- *Si de plus  $k_n/\ln(\ln(n)) \rightarrow \infty$  quand  $n \rightarrow \infty$ , Alors  $\widehat{\xi}_{k_n}^p$  converge presque sûrement vers  $\xi$ .*
- *Sous des conditions additionnelles sur la suite  $k_n$  et la fonction de répartition  $F$  que l'on pourra consulter dans Dekkers et de Haan (1989)[21],*

$$\sqrt{k_n}(\widehat{\xi}_{k_n}^p - \xi) \xrightarrow{\text{Loi}} N \left( 0, \frac{\xi^2 (2^{2\xi+1} + 1)}{4(\ln 2)^2 (2^\xi - 1)^2} \right) \quad (1.10)$$

### 1.5.2 Estimateur de Hill

Cet estimateur a été introduit par Hill (1975)[34] pour estimer d'une manière non paramétrique le paramètre de queue des lois appartenant au D(Fréchet). Pour construire son estimateur, Hill utilise la méthode du maximum de vraisemblance sur l'ensemble des  $k_n$  plus grandes observations d'un échantillon. Un grand nombre de travaux théoriques ont été consacrés à l'étude des propriétés de l'estimateur de Hill. Mason (1982)[42] a démontré la consistance faible et Deheuvels, Haeusler et Mason ont établi la consistance forte dans Deheuvels et al. (1988)[18]. La normalité asymptotique est due entre autres à Davis et Resnick (1984)[14], Csörgö et Mason (1985)[11], Haeusler et Teugels (1985)[32] et Smith (1987)[63].

**Définition 1.4** Soit  $(k_n)_{n \geq 1}$  une suite d'entiers avec  $1 < k_n \leq n$ , l'estimateur de Hill est défini par

$$\widehat{\xi}_{k_n}^H = \frac{1}{k_n - 1} \sum_{j=1}^{k_n-1} \ln \left( \frac{X_{n-j+1,n}}{X_{n-k_n+1,n}} \right) \quad (1.11)$$

Si on choisit  $k, n \rightarrow +\infty$ , de sorte que  $\frac{k}{n} \rightarrow 0$  alors on peut montrer que  $\lim_{k \rightarrow \infty} \widehat{\xi}_{k,n}^H = \xi$  et l'estimateur de Hill est de plus asymptotiquement normal :

$$\sqrt{k} \frac{\widehat{\xi}_{k,n}^H - \xi}{\xi} \xrightarrow{\text{Loi}} N(0, 1) \quad (1.12)$$

Le théorème suivant présente les propriétés de l'estimateur de Hill :

**Théorème 1.7 (Propriétés de l'estimateur de Hill)** Soit  $(k_n)_{n \geq 1}$  une suite d'entiers avec  $1 < k_n \leq n, k_n \rightarrow \infty$  et  $k_n/n \rightarrow 0$  quand  $n \rightarrow \infty$

- Si les  $X_1, \dots, X_n$  sont i.i.d. Alors l'estimateur de Hill  $\widehat{\xi}^H$  converge en probabilité vers  $\xi$ .
- Si  $\frac{k}{n} \rightarrow 0$ ,  $\frac{k}{\ln(\ln(n))} \rightarrow \infty$  et les  $X_1, \dots, X_n$  sont i.i.d. Alors l'estimateur de Hill converge presque sûrement vers  $\xi$ .
- Si les  $X_1, \dots, X_n$  sont i.i.d et  $\bar{F}$  est à variations régulières. Alors

$$\sqrt{k_n} (\widehat{\xi}_{k_n}^H - \xi) \xrightarrow{\text{Loi}} N(0, \xi^2) \quad (1.13)$$

Pour la preuve de ce théorème, voir Hill (1975)[34].

la convergence étant en loi. Cet estimateur est l'estimateur du maximum de vraisemblance dans le cas particulier du modèle

$$\bar{F}(x) = 1 - F(x) = Cx^{-\frac{1}{\xi}}$$

on reconnaît ici une distribution de Pareto d'indice  $\alpha = \frac{1}{\xi}$ . Dans le cas général du domaine de Fréchet, la fonction de survie est de la forme

$$\bar{F}(x) = 1 - F(x) = x^{-\frac{1}{\xi}} L(x)$$

avec  $L$  une fonction à variation lente. Cela induit un biais important sur l'estimateur de Hill, qui est donc en pratique d'un maniement délicat. Dans le cas général, la fonction  $L$  apparaît comme un paramètre de nuisance de dimension infinie, qui complique l'estimation

### 1.5.2.1 Estimateur de Hill Alternatif

Même si l'estimateur de Hill est bien moins volatil que celui de Pickand (quand on le représente en fonction du nombre d'observations considérées), il manque de précision et sa phase de stabilisation, qui devrait correspondre à la valeur de l'indice de queue, est parfois difficile à identifier. RESNICK et STARICA [1997][57] ont mis au point une méthode très simple pour effectuer un « agrandissement » du démarrage du graphe de l'estimateur de Hill. Pour ce faire, l'idée est de changer d'échelle sur l'axe des abscisses pour passer en échelle logarithmique. Ainsi, la partie de courbe correspondant aux premières valeurs extrémales est agrandie et plus précise, par opposition à la partie de courbe correspondant aux valeurs qui ne font probablement pas partie de la queue de la distribution. En pratique, au lieu de représenter graphiquement  $\{k, \xi_{k,n}^H\}$ , on trace  $\{k, \xi_{n^\theta, n}^H\}$  avec  $0 \leq \theta \leq 1$ .

## 1.6 Exemples de comportements limites dans les queues

Dans cette section, nous proposons trois exemples illustrant comment se manifestent les lois limites de la  $GEV$  et la  $GPD$  en pratique, compte tenu de différentes hypothèses sur la loi  $F$  de la variable parente  $X$  de l'échantillon.

**Exemple 1.1 (Cas de la loi exponentielle)** *Pour la loi exponentielle de paramètre  $\lambda = 0$ , la fonction de répartition de cette loi est*

$$\forall x \in \mathbb{R}, F(x) = \Pr(X \leq x) = \begin{cases} 1 - e^{-x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

En posant  $b_n = \log(n)$  et  $a_n = 1$  alors

$$\forall x \in \mathbb{R}, \forall n \in \mathbb{N}^*, \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = \Pr(M_n \leq a_n x + b_n) = F^n(x + \ln n)$$

Alors :

$$\begin{aligned}
 \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) &= \begin{cases} (1 + e^{-x - \log(n)})^n & \text{si } x + \ln n > 0 \\ 0 & \text{si } x + \ln n \leq 0 \end{cases} & (1.14) \\
 &= \begin{cases} \left(1 - \frac{e^{-x}}{n}\right) \\ \rightarrow \exp\{-e^{-x}\} \\ = A(x) \end{cases}
 \end{aligned}$$

c'est-à-dire que le maximum convenablement normalisé de la loi exponentielle converge vers la loi de Gumbel. Cette loi appartient au domaine d'attraction maximal de Gumbel. Aussi, les lois dans le domaine d'attraction de la loi de Gumbel sont parfois appelées lois de type exponentiel. En ce qui concerne la méthode des excès au-delà d'un seuil  $u$ , en prenant  $\beta_u = 1$  alors

$$\begin{aligned}
 F_u(y) &= \frac{F(u+y) - F(u)}{1 - F(u)}, y \geq 0 & (1.15) \\
 &= \frac{e^{-u} - e^{-u-y}}{e^{-u}} \\
 &= 1 - e^{-y}
 \end{aligned}$$

pour tout  $y > 0$ . Aussi, la limite est la loi GPD de paramètre  $\xi = 0$  avec  $\beta_u = 1$  notons que dans ce cas, la loi GPD n'est pas simplement la loi limite, mais il s'agit de la loi exacte pour tout  $u$ .

**Exemple 1.2 (cas de la loi de Pareto)** pour la loi de Pareto de fonction de répartition  $F(x) = 1 - cx^{-\alpha}$ , ou  $c > 0$  et  $\alpha > 0$ . En posant,  $b_n = 0$  et  $a_n = (nc)^{\frac{1}{\alpha}}$  alors on a pour  $x > 0$

$$\begin{aligned}
 F^n(a_n x - b_n) &= (1 - c(a_n x)^{-\alpha})^n = (1 - ca_n^{-\alpha} x^{-\alpha})^n & (1.16) \\
 &= \left(1 - \frac{x^{-\alpha}}{n}\right)^n \\
 &\rightarrow \exp\{-x^{-\alpha}\} \\
 &= \varphi_\alpha(x)
 \end{aligned}$$

qui est la loi de Fréchet. La loi de Pareto appartient au domaine d'attraction maximal de Fréchet. Aussi, les lois dans le domaine d'attraction de la loi de Fréchet sont parfois appelées lois de type Pareto. En se basant sur la méthode des excès au-delà d'un seuil  $u$  et en considérant  $\beta_u = ub$  pour  $b > 0$ . on a alors

$$\begin{aligned}
 F_u(y) &= \frac{F(u + uby) - F(u)}{1 - F(u)}, y \geq 0 & (1.17) \\
 &= \frac{cu^{-\alpha} - c(u + uby)^{-\alpha}}{cu^{-\alpha}} \\
 &= 1 - (1 - by)^{-\alpha}
 \end{aligned}$$

pour  $\xi = \frac{1}{\alpha}$  et  $b = \xi$ , la limite est alors la loi GPD de paramètre  $\xi$

**Exemple 1.3 (Cas de la loi normale)** Soit  $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$  la fonction de répartition de la loi  $N(0, 1)$ . D'après Feller [1968][27],  $1 - F(x)$  est équivalente à  $\frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  au voisinage de l'infini, donc  $1 - F(x) \sim \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  quant  $x \rightarrow +\infty$  alors :

$$\lim_{u \rightarrow +\infty} \frac{1 - F\left(u + \frac{z}{u}\right)}{1 - F(u)} = \lim_{u \rightarrow +\infty} \left[ \left(1 + \frac{z}{u^2}\right) \exp\left\{-\frac{1}{2}\left(u + \frac{z}{u}\right)^2 + \frac{1}{2}u^2\right\} \right] = e^{-z}$$

Si on suppose dans un premier temps que  $\beta_u = \frac{1}{u}$ , alors

$$1 - \frac{1 - F\left(u + \frac{z}{u}\right)}{1 - F(u)} = \frac{F(u + \beta_u z) - F(u)}{1 - F(u)} \rightarrow 1 - e^{-z} \text{ quant } u \rightarrow +\infty$$

et par la suite la loi limite des excès au-delà d'un seuil est la loi exponentielle. Dans un second temps si on considère  $b_n$  solution de  $F(b_n) = 1 - \frac{1}{n}$  et  $a_n = \frac{1}{b_n}$ , on obtient

$$n\{1 - F(a_n x + b_n)\} = \frac{\{1 - F(a_n x + b_n)\}}{1 - F(b_n)} \rightarrow e^{-x}$$

et donc

$$\lim_{n \rightarrow +\infty} F^n(a_n x + b_n) = \lim_{n \rightarrow +\infty} \left(1 - \frac{e^{-x}}{n}\right)^n = \exp\{-e^{-x}\}$$

converge vers la loi de Gumbel.

comme il a été remarqué par Smith[2003][64], bien que loi de Gumbel et la loi exponentielle soient respectivement les limites exactes des maxima et des excès au-delà d'un seuil, les meilleures approximations sont obtenues en utilisant la loi GEV et la loi GPD. Ces lois sont très générales et permettent d'inclure même le cas où  $\xi \neq 0$ . En fait, il est généralement préférable de considérer ces deux lois même si on pense que la loi de Gumbel ou la loi exponentielle sont les vraies lois limites.

## 1.7 Conclusion

Dans ce chapitre, nous avons fait un rappel sur la théorie des valeurs extrêmes, en mentionnant les différentes caractéristiques et les notions de base qui sont très utiles pour l'estimation des quantiles extrêmes et le niveau de retour que nous aborderons dans le chapitre suivant

## 2.1 Introduction

Les techniques de processus ponctuels sont largement utilisées dans la théorie des valeurs extrêmes parce qu'elles fournissent des outils qui peuvent être utilisés pour prouver plusieurs résultats asymptotiques avec élégance et perspicacité.

Pour donner une idée générale, soit à construire des modèles pour la distribution aléatoire des points dans un espace qu'on prend souvent comme une partie de  $\mathbb{R}_+$ , de  $\mathbb{R}$  ou de  $\mathbb{R}^d$  ( $d \geq 1$ ). Par exemple, les emplacements des arbres dans une forêt, les localisations et les instants où se produit un ouragan ou les répliques à la suite d'un tremblement de terre, les instants d'arrivée des clients dans une file d'attente, les instants de panne d'une machine ou d'un groupe de machines dans une usine, etc. Alors, les processus ponctuels sont des candidats idéaux pour ce genre de modélisation.

Pour être un peu plus précis, soient  $\mathbb{E}$  un espace et  $x = \{x_i, i \in I \subset \mathbb{N}^*\}$  un ensemble de points de  $\mathbb{E}$  considérés comme des réalisations de certaines variables aléatoires aux valeurs dans  $\mathbb{E}$ .

Un processus ponctuel  $N$  est une distribution aléatoire des points de  $x$  dans l'espace  $\mathbb{E}$  qui à chaque ensemble  $A \subseteq \mathbb{E}$ ,  $N(A)$  est le nombre de points de  $x$  dans  $A$ .

Les plus importants des processus ponctuels dans la théorie des valeurs extrêmes sont ceux pour lesquels le nombre de points de  $x$  qui appartiennent à l'ensemble  $A$  suit une loi de probabilité de Poisson. Ces processus qu'on appelle « **Processus Ponctuels de Poisson** », qu'on va voir plus loin, sont une généralisation des processus de Poisson classiques (homogènes) où  $\mathbb{E} = \mathbb{R}_+$  est considéré comme l'espace de temps,  $x$  est l'ensemble des instants d'occurrence d'un certain événement aléatoire,  $A$  est un intervalle de temps et  $N(A)$  est le nombre de réalisations de l'événement pendant la période du temps  $|A|$  ( $|\cdot|$  est la mesure de Lebesgue sur  $\mathbb{R}_+$ )

## 2.2 Processus ponctuel

On appelle processus ponctuel sur un ensemble  $\mathcal{A}$  un processus stochastique dont la réalisation est un ensemble de points dans  $\mathcal{A}$ . La loi de ce processus se définit grâce à une mesure de comptage  $N(\cdot)$  qui, pour un ensemble borélien  $A \subset \mathcal{A}$ , donne le nombre de points dans  $A$  :

$$N(A) = \sum_{i \geq 1} \mathbf{1}_{\{x_i \in A\}} \quad (2.1)$$

où  $\{X_i\}_{i \geq 1}$  sont les points du processus ponctuel et  $\mathbf{1}$  est la fonction indicatrice.

La fonction  $\Lambda : A \rightarrow \mathbb{E}[N(A)]$  est appelée mesure d'intensité du processus ponctuel. Cette mesure permet de donner la définition d'un processus ponctuel de Poisson.

**Définition 2.1 (Processus de Poisson).** *Un processus ponctuel est dit de Poisson d'intensité  $\Lambda$  si pour tout  $k \geq 1$  et pour tous ensembles boréliens  $A, A_1, \dots, A_k \subset \mathcal{A}$  disjoints, on a :*

1.  $N(A) \sim P(\Lambda(A))$ , alors :

$$\Pr(N(A) = k) = \begin{cases} \frac{e^{-\mu(A)} (\mu(A))^k}{k!}, & \text{si } \mu(A) < \infty \\ 0, & \text{sinon} \end{cases} \quad (2.2)$$

2.  $N(A_1), \dots, N(A_k)$  sont mutuellement indépendants.

Le théorème suivant (cf. Théorème de Coles, 2001[7]) permet de définir une troisième représentation des valeurs extrêmes d'une variable aléatoire  $X$  qui utilise le cadre d'un processus ponctuel de Poisson.

**Théorème 2.1 (Convergence vers un processus de Poisson).** *Soit  $X$  une variable aléatoire et  $\{X_1, X_2, \dots\}$  des répliques i.i.d. de  $X$ . Si  $X$  appartient au domaine d'attraction d'une loi GEV( $\mu, \sigma, \xi$ ), alors la suite de processus ponctuels  $\{\mathcal{P}_n\}_n$  définie sur  $A = [0, 1] \times \mathbb{R}$  par :*

$$\mathcal{P}_n = \left\{ \left( \frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right) : i = 1, \dots, n \right\} \quad (2.3)$$

converge vers un processus de Poisson  $P$  sur  $[0, 1] \times C$  de mesure d'intensité

$$\Lambda([a, b] \times [z, \infty]) = (a, b) \left( 1 + \xi \frac{z - \mu}{\sigma} \right)^{-1/\xi} \quad (2.4)$$

où

$$C = \{z \in \mathbb{R} : 1 + \xi(z - \mu)/\sigma > 0\}.$$

### 2.2.1 Lien avec les deux premières approches

**Coles(2001)[7]**

Il est possible de faire le lien entre la représentation des valeurs extrêmes par un processus de Poisson et les deux approches décrites dans la section précédente, à savoir les maxima par blocs et les excès de seuil.

Si on définit l'ensemble  $A_z = (0, 1) \times (z, \infty)$ , le théorème 2.1 permet d'écrire :

$$\begin{aligned} \Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) &= \Pr(N_n(A_z) = 0) \\ &\xrightarrow{n \rightarrow +\infty} \Pr(N(A_z) = 0) \\ &= \exp(-\Lambda(A_z)) \\ &= \exp\left(-\left[1 + \xi \frac{z - \mu}{\sigma}\right]_+^{-1/\xi}\right). \end{aligned} \quad (2.5)$$

On retrouve donc bien la forme de la fonction de répartition de la loi *GEV* de paramètres  $\mu$ ,  $\sigma$  et  $\xi$ .

Pour faire le lien avec la méthode des excès de seuil, Coles (2001)[7] écrit la mesure d'intensité  $\Lambda$  comme le produit  $\Lambda(A_z) = \Lambda_1([t_1, t_2]) \times \Lambda_2([z, \infty])$ , où :

$$\Lambda_1([t_1, t_2]) = (t_2 - t_1) \quad \text{et} \quad \Lambda_2([z, \infty]) = \left[1 + \xi \frac{z - \mu}{\sigma}\right]^{-1/\xi} \quad (2.6)$$

Pour  $Z \geq u$ , on obtient alors, grace au théorème 2.1 :

$$\begin{aligned} \Pr\left(\frac{X_i - b_n}{a_n} > z \mid \frac{X_i - b_n}{a_n} > u\right) &= \frac{\Lambda_1([t_1, t_2])\Lambda_2([z, \infty])}{\Lambda_1([t_1, t_2])\Lambda_2([u, \infty])} \\ &= \left(\frac{1 + \xi \frac{z - \mu}{\sigma}}{1 + \xi \frac{u - \mu}{\sigma}}\right)^{-1/\xi} \\ &= \left(1 + \frac{\xi \frac{z - \mu}{\sigma}}{1 + \xi \frac{u - \mu}{\sigma}}\right)^{-1/\xi} \\ &= \left(1 + \xi \frac{z - u}{\sigma + \xi(u - \mu)}\right)^{-1/\xi} \\ &= \left(1 + \xi \frac{z - u}{\tau}\right)^{-1/\xi} \end{aligned} \quad (2.7)$$

On retrouve la fonction de répartition de la loi *GP* de paramètres  $\tau$  et  $\xi$ .

## 2.2.2 Modélisation avec les excès de seuils

? Les liens faits précédemment avec la méthode des excès de seuil permettent de reformuler le théorème précidant de la façon suivante (**cf. Théorème de Coles, 2001[7]**).

**Théorème 2.2** Soient  $\{X_1, X_2, \dots\}$  des répliques *i.i.d.* d'une variable aléatoire  $X$  et

$$N_n = \left\{ \left( \frac{i}{n+1}, X_i \right) : i = 1, \dots, n \right\} \quad (2.8)$$

Si les maxima renormalisés des  $X_i$  convergent vers une loi non dégénérée  $G$ , alors pour un seuil  $u$  assez grand, sur les régions de la forme  $(0, 1) \times [u, \infty)$ , le processus ponctuel  $N_n$  est approximativement Poisson de mesure

$$\Lambda(A) = (t_2 - t_1) \left[ 1 + \xi \frac{z - \mu}{\sigma} \right]^{-1/\xi} \quad (2.9)$$

sur  $A = [t_2 - t_1] \times [z, \infty)$ .

En pratique, si on observe  $n_{ann}$  années d'observations, le théoème précédent fait correspondre les paramètres  $\mu, \sigma$  et  $\xi$  de la formule 2.9 avec la loi *GEV* associée au maximum sur  $n_{ann}$  années d'observations. Il est donc préférable de travailler sur l'échelle annuelle en appliquant à 2.9 la modification :

$$\Lambda(A) = n_{ann} (t_2 - t_1) \left[ 1 + \xi \frac{z - \mu}{\sigma} \right]^{-1/\xi} \quad (2.10)$$

afin que les paramètres *GEV* correspondent à ceux de la loi ajustée sur les maxima annuels.

## 2.3 Processus max-stables

Le processus  $\{Z(s)_{s \in X}\}$  des maxima standardisés en Fréchet unitaire peut être modélisé par **un processus max-stable** [de Haan, 1984][15]. Les processus max-stables généralisent la théorie des valeurs extrêmes au cas spatial en modélisant la structure de dépendance entre les maxima  $Z(s)$ .

### 2.3.1 Définition et construction

**Définition 2.2** de Haan (1984)[15] :  $Z(\cdot)$  est max-stable si pour  $Z_i, i = 1, \dots, n$ , copies indépendantes de  $Z$ ,

$$\max_{i=1, \dots, n} Z_i \stackrel{d}{=} nZ \quad (2.11)$$

**Définition 2.3** Un processus  $Z$  est dit **max-stable** si disposant pour tout  $n \in \mathbb{N}$  et copies *i.i.d.* de  $Z, Z_1, \dots, Z_n$ , il existe des fonctions (continues)  $a_n(s) > 0$  et  $b_n(s) \in \mathbb{R}$  telles que

$$\max_{i=1, \dots, n} \frac{Z_i(s) - b_n(s)}{a_n(s)} \stackrel{d}{=} Z(s) \quad (2.12)$$

Remarque. L' égalité en loi précédente est à interpréter au sens *fidi*, i.e., toutes les fonctions de répartition  $k$ -variées,  $k \in \mathbb{N}^*$ , sont égales.

Sans aucune perte de généralité et afin de se concentrer uniquement sur la structure de dépendance du processus, on supposera souvent que les lois univariées sont fixées. Ce choix simplifiera la théorie mais il faudra bien entendu relâcher cette hypothèse plus tard.

**Définition 2.4** La variable aléatoire, non dégénérée,  $X$  ou la loi de probabilité de  $X$  ou, encore, la fonction de distribution  $F$  de  $X$  est dite **max-stable**, s'il existe des constantes  $a_n \in \mathbb{R}_+^*$  et  $b_n \in \mathbb{R}$

telles que :

$$M_n \stackrel{d}{=} a_n X + b_n$$

Ou, ce qui est équivalent, s'il existe des constantes  $a_n \in \mathbb{R}_+^*$  et  $b_n \in \mathbb{R}$  telles que :

$$F^n(a_n x + b_n) = F(x) \text{ pour tout } n \in \mathbb{N}^* \text{ et } x \in \mathbb{R}. \quad (2.14)$$

**Exemple 2.1 (Coles (2001)) [7]** : Supposons que  $X$  est une variable aléatoire suivant une loi de probabilité de Fréchet standard, i.e :

$$\forall x \in \mathbb{R}, F(x) = \Pr(X \leq x) = \begin{cases} \exp(-x^{-1}) & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} \quad (2.15)$$

Posons  $a_n = n$  et  $b_n = 0$ ; on a alors :

$$\begin{aligned} \forall x \in \mathbb{R}, \forall n \in \mathbb{N}^*, F^n(a_n x + b_n) &= F^n(nx) \\ &= \begin{cases} \left[ \exp(-nx^{-1}) \right]^n & \text{si } nx > 0 \\ 0 & \text{si } nx \leq 0 \end{cases} \\ &= \begin{cases} \exp(-x^{-1}) & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} \\ &= F(x) \end{aligned} \quad (2.16)$$

Donc :

$$\forall x \in \mathbb{R}, \text{ for all } n \in \mathbb{N}^*, \exists a_n = n \text{ et } b_n = 0 \text{ telque } F^n(a_n x + b_n) = F(x) \quad (2.17)$$

On déduit que la loi de probabilité de Fréchet standard est une **loi max-stable**.

**Théorème 2.3** La classe des lois max-stables coïncide avec la classe de toutes les lois limites possibles, non dégénérées, de la suite de maximums, normalisés, d'une suite de variables aléatoires indépendantes et identiquement distribuées.

**Remarque 2.1** Le théorème ci-dessus nous donne une propriété importante des lois **max-stables**, dite «propriété limite des lois max-stables», qui nous montre que toute loi max-stable est une loi limite pour la suite des maximums, normalisés, d'une suite de variables aléatoires indépendantes et identiquement distribuées et, en plus, les lois max-stables sont les seules et uniques lois limites possibles.

**Définition 2.5 ([De Haan and Ferreira, 2006]) [17]** Un processus max-stable est dit **simple** si dans 2.12  $b_n \equiv 0$  et  $a_n \equiv n$ . Cela implique donc, cf. EVT univariée, que

$$\Pr\{\eta(s) \leq z\} = \exp\left(-\frac{1}{z}\right), \quad z > 0 \quad . \quad s \in \chi. \quad (2.18)$$

Autrement dit les marges sont **Fréchet unitaires**.

Le théorème suivant justifie pourquoi les processus max-stables jouent un rôle central pour les extrêmes spatiaux.

**Théorème 2.4** Soit  $\{Y(s) : s \in \chi\}$  un processus stochastique. Supposons qu'il existe deux suites de fonctions de normalisation  $\{c_n > 0 : n \geq 1\}$  et  $\{d_n \in \mathbb{R} : n \geq 1\}$  telles que pour tout  $n \geq 1$  et  $n$  copies *i.i.d.* de  $Y$ ,  $Y_1, \dots, Y_n$ , on ait

$$\max_{i=1, \dots, n} \frac{Y_i(s) - d_n(s)}{c_n(s)} \xrightarrow{D} Z(s), \quad n \rightarrow \infty, \quad (2.19)$$

où  $\{Z(s) : s \in \chi\}$  est un processus stochastique non dégénéré. Alors nécessairement  $Z$  est un processus max-stable.

Démonstration.

On utilise les arguments standards pour les maxima par bloc cf. *GEV* pour *EVT* univariée et loi max-stable pour *MEVT*. D'ailleurs la preuve pour la *MEVT* est suffisante mais on donne une preuve heuristique tout de même.

Posons  $n = m \times \ell$  et faisons tendre  $m \rightarrow \infty$ . Sous les conditions du théorème, on a donc

$$\max_{i=1, \dots, n} \frac{Y_i - d_n}{c_n} = \max_{j=1, \dots, \ell} \max_{i=1, \dots, m} \frac{Y_{i,j} - d_n}{c_n}, \quad (2.20)$$

où  $Y_{i,j} = Y_{(j-1) \times \ell + i}$ .

puisque

$$\frac{Y_{i,j} - d_n}{c_n} = \frac{c_m}{c_n} \frac{Y_{i,j} - d_m}{c_m} + \frac{c_m}{c_n} \frac{d_m - d_n}{c_m} \rightarrow \frac{1}{\tilde{c}_\ell} Z_j - \frac{\tilde{d}_\ell}{\tilde{c}_\ell}, \quad (2.21)$$

$Z_1, \dots, Z_\ell$  sont des copies *i.i.d.* de  $Z$ . On a donc d'une part :

$$\max_{i=1, \dots, n} \frac{Y_i - d_n}{c_n} \xrightarrow{D} Z, \quad n \rightarrow \infty, \quad (2.22)$$

et d'autre part

$$\max_{i=1, \dots, n} \frac{Y_i - d_n}{c_n} \xrightarrow{D} \max_{i=1, \dots, \ell} \frac{Z_j - \tilde{d}_\ell}{\tilde{c}_\ell}, \quad n \rightarrow \infty, \quad (2.23)$$

montrant ainsi que  $Z$  doit être max-stable.

Remarque. Ce théorème nous donne une justification pour utiliser les processus max-stables quand on s'attache à modéliser par exemple les maxima annuels ponctuels.

Les processus max-stables sont donc un peu nos processus Gaussiens pour les extrêmes spatiaux. La définition [de Haan and Ferreira, 2006][17] n'est pas très parlante au niveau de la structure du processus. Est-il possible d'avoir une représentation de cette famille des processus.

En se restreignant aux processus max-stable simples, on sait déjà via la *MEVT* que pour tout  $k \in \mathbb{N}_*$  et  $s = (s_1, \dots, s_k) \in \chi^k$ , on a

$$\Pr \{\eta(s_1) \leq z_1, \dots, \eta(s_k) \leq z_k\} = \exp \{-V_s(z_1, \dots, z_k)\}, \quad (2.24)$$

où  $V_s$  est une fonction positive homogène d'ordre  $-1$ , i.e. ,

$$V_s(cz) = c^{-1} V_s(z). \quad (2.25)$$

C'est un premier pas mais pas vraiment plus utile, le théorème suivant fait bien mieux.

**Théorème 2.5** ([Penrose, 1992; De Haan, 1984]) [49][15] Soit  $\{\eta(s), s \in \chi\}$  un processus max-stable simple. Alors nous avons

$$\eta \stackrel{d}{=} \max_{i \geq 1} \zeta_i Y_i, \quad (2.26)$$

où  $\{\zeta_i : i \geq 1\}$  sont les points d'un processus de Poisson sur  $(0, \infty)$  d'intensité

$$d\Lambda(\zeta) = \zeta^{-2} d\zeta$$

et  $Y_1, Y_2, \dots$ , des copies i.i.d d'un processus stochastique positif  $Y$  tel que  $\mathbb{E}\{Y(s)\} = 1$  pour tout  $s \in \chi$ .

Démonstration. Nous allons seulement montrer que le processus défini est bien max-stable simple. On va le faire en 2 temps uniquement a titre pédagogique.

Commençons par montrer que les marges sont Fréchet unitaires. Soit  $s \in \chi$ , on a donc

$$\begin{aligned} \Pr\{\eta(s) \leq z\} &= \Pr\{\zeta_i Y_i(s) \leq z, i \geq 1\} \\ &= \Pr\left\{ \begin{array}{c} \text{aucun point du processus de Poisson dans} \\ \text{l'ensemble } \{(\zeta, y) : \zeta y > z\} \end{array} \right\} \\ &= \exp[-\Lambda\{0, z\}^c] \\ &= \exp\left\{-\int_{\{\zeta y > z\}} dP^Y(y) \zeta^{-2} d\zeta\right\} \\ &= \exp\left\{-\int \frac{y}{x} dP^Y(y)\right\} \\ &= \exp\left(-\frac{1}{z}\right). \end{aligned} \quad (2.27)$$

Montrons que c'est max-stable, i.e.,  $n^{-1} \max(\eta_1, \dots, \eta_n) \stackrel{d}{=} \eta$

Soient  $k \geq 1, s = (s_1, \dots, s_k) \in \chi^k$  et  $z = (z_1, \dots, z_k) \in (0, \infty)^k$ . On a

$$\begin{aligned}
 \Pr\{\eta(s) \leq z\} &= \Pr\{\text{aucun point dans } (0, z)^c\} & (2.28) \\
 &= \exp[-\Lambda\{(0, z)^c\}] \\
 &= \exp\left[-1_{\{\zeta > \min_j z/y\}} \zeta^{-2} d\zeta dP^Y(y)\right] \\
 &= \exp\left[-\int \frac{1}{\min_j \frac{z_j}{y_j}} dP^Y(y)\right] \\
 &= \exp\left[-\mathbb{E}\left\{\max_{j=1, \dots, k} \frac{Y(s_j)}{z_j}\right\}\right].
 \end{aligned}$$

D'où

$$\begin{aligned}
 \Pr\{n^{-1}\eta(s) \leq z\}^n &= \exp\left[-\mathbb{E}\left\{\max_{j=1, \dots, k} \frac{Y(s_j)}{nz_j}\right\}\right]^n & (2.29) \\
 &= P\{\eta(s) \leq z\}.
 \end{aligned}$$

**Proposition 2.1** *En passant dans la preuve on a montré que les f.d.r. fini-dimensionnelles sont de la forme*

$$\Pr\{\eta(s_1) \leq z_1, \dots, \eta(s_k) \leq z_k\} = \exp\left[-\mathbb{E}\left\{\max_{j=1, \dots, k} \frac{Y(s_j)}{z_j}\right\}\right]. \quad (2.30)$$

qui, en accord avec la MEVT, est bien de la forme  $\exp\{-V(z_1, \dots, z_k)\}$  avec  $V$  homogène d'ordre  $-1$ .

La représentation (2.26) s'appelle la **représentation spectrale** ou **caractérisation spectrale** du processus max-stable et la **figure 2.1** illustre cette dernière. Il est important de garder en mémoire que cette représentation n'est **pas unique** dans le sens où deux processus  $Y_1$  et  $Y_2$  de lois différentes peuvent conduire au même processus max-stable

**Exemple 2.2** *Posons  $Y_1 \sim Y$  et  $Y_2 \sim XY$  où  $X$  est une v.a. positive ind. du processus  $Y$  et telle que  $E(X) = 1$  (on est donc bien dans les conditions de la représentation spectrale).*

Soit  $\eta_2 = \max_{i \geq 1} \zeta_i Y_{2,i}$ . Clairement on a

$$P\{\eta_2(s) \leq z\} = \exp\left[-E\left\{\max_{i=1, \dots, k} \frac{XY(s_j)}{z_j}\right\}\right] = \exp\left[-E\left\{\max_{i=1, \dots, k} \frac{Y(s_j)}{z_j}\right\}\right] = P\{\eta(s) \leq z\}. \quad (2.31)$$

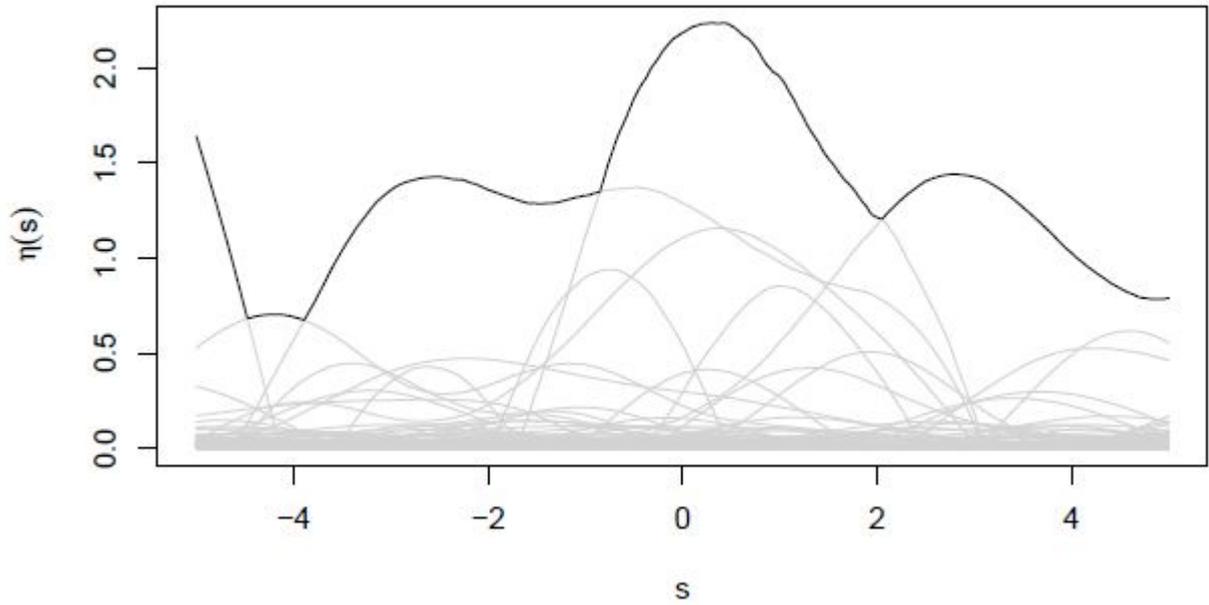


FIGURE 2.1 – Illustration de la caractérisation spectrale du Théorème 2.5. Les courbes grises sont les atomes de la forme  $\zeta_i Y_i$  et la courbe noire la trajectoire du processus max-stable simple.

Cas particulier. Si l'on prend  $Y = \max\{\varepsilon, 0\}$  ou  $Y = \max\{T, 0\}$  où  $\varepsilon$  est un processus Gaussien et  $T$  un processus de Student cela nous conduira au même processus max-stable. En effet, on a

$$T \stackrel{d}{=} \sqrt{\frac{\nu}{\chi^2 \nu}} \varepsilon. \quad (2.32)$$

Remarque. Notons également qu'il existe une autre écriture de la représentation spectrale (2.26) sous la forme

$$\eta = \max_{i \geq 1} \zeta_i f_i. \quad (2.33)$$

où  $\{(\zeta_i, f_i) : i \geq 1\}$  sont les points d'un processus de Poisson sur  $(0, \infty) \times \mathbb{C}_0$  d'intensité  $\zeta^{-2} d\zeta \nu(df)$  et où  $\nu$  est une mesure localement finie définie sur l'espace  $\mathbb{C}_0$  des fonctions positives continues sur  $\chi$  et telles que

$$\int f(s) \nu(df) = 1, \quad s \in \chi. \quad (2.34)$$

Ainsi la représentation (2.26) correspond au cas où  $\nu$  est une mesure de probabilité. On pourra d'ailleurs toujours se ramener à ce cas en appliquant la transformation

$$\begin{aligned} T & : (0, \infty) \times \mathbb{C}_0 \longrightarrow (0, \infty) \times \mathbb{C}_0 \\ (\zeta, f) & \longrightarrow (\zeta/\nu(\chi), f), \end{aligned}$$

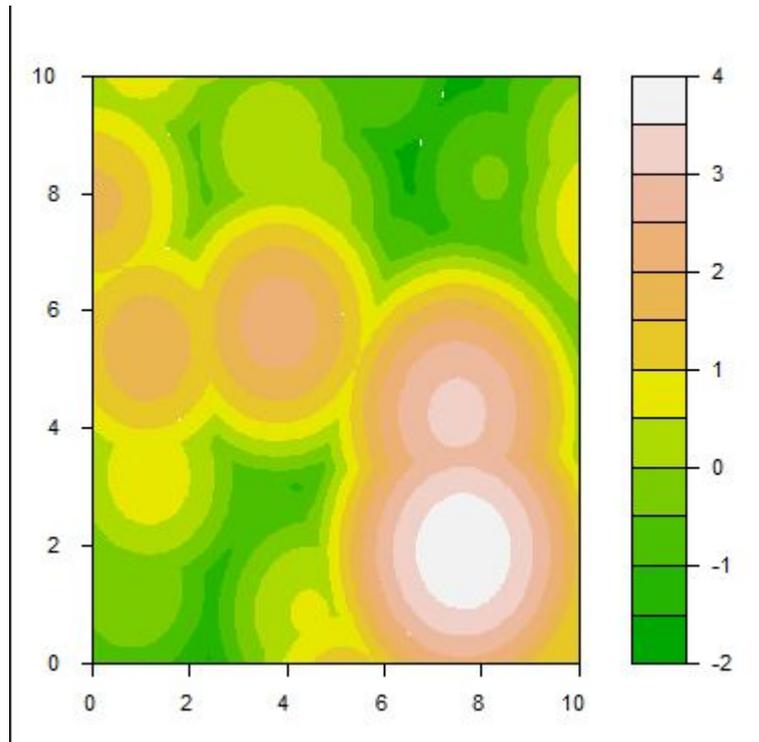


FIGURE 2.2 – Modèle de Smith

où point du processus de Poisson. Ceci est rendu possible (avec nos hypothèses de compacité de  $\chi$ ) puisque  $v(\chi) < \infty$ .

Ainsi on utilisera dans la suite indifféremment l'une ou l'autre.

### 2.3.2 Quelques modèles max-stables

Les représentations spectrales des champs aléatoires max-stables simples stationnaires sont à la base de construction de beaucoup de modèles des champs aléatoires max-stables stationnaires. Dans cette section, on présentera les modèles de ces champs aléatoires les plus utilisés.

#### 2.3.2.1 Modèle de Smith

— Le premier modèle apparu dans la littérature est ce qu'on appelle le processus de Smith (Smith, 1990) [65] aussi parfois fait référence au processus des valeurs extrêmes gaussiennes

$$\eta(s) = \max_{i \geq 1} \zeta_i \varphi(s - U_i; \Sigma), \quad s \in \chi \subset \mathbb{R}^d, \quad (2.35)$$

où  $\{(\zeta_i, U_i) : i \geq 1\}$  sont les points d'un processus de Poisson sur  $(0, \infty) \times \mathbb{R}^d$  avec la mesure d'intensité  $\zeta^{-2} d\zeta du$  et  $\varphi(\Sigma)$  indique la densité d-variable gaussienne

avec moyenne 0 et matrice de covariance  $\Sigma$ . Bien que ce processus soit intéressant pour des raisons historiques, il est rarement utile en raison d'un manque de flexibilité - la forme des densités gaussiennes multivariées étant trop restrictive.

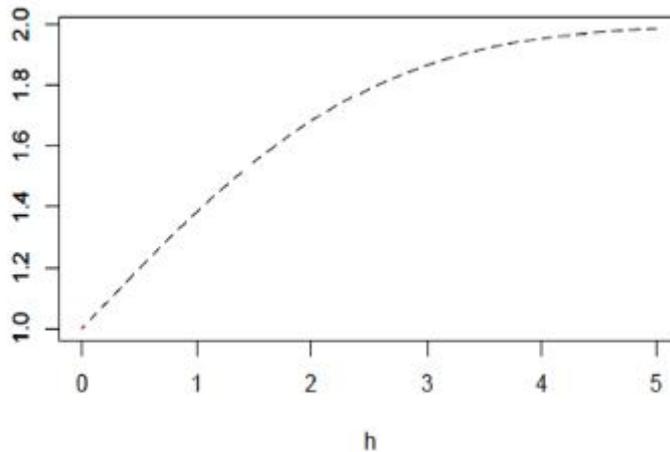


FIGURE 2.3 – Graphe du coefficient extrême du modèle de Smith avec  $\Sigma$  la matrice identité d'ordre 2

**Proposition 2.2** *Richard L. Smith a montré que la loi de probabilité marginale bidimensionnelle du champ aléatoire max-stable simple stationnaire  $\eta = \{\eta(s); s \in \mathbb{R}^d\}$  en deux sites  $s_1 \in \mathbb{R}^d$  et  $s_2 \in \mathbb{R}^d$  est donnée, pour tout  $z_1 \in \mathbb{R}_+^*$  et  $z_2 \in \mathbb{R}_+^*$  par :*

$$\Pr(\eta(s_1) \leq z_1, \eta(s_2) \leq z_2) = \exp - \left\{ \frac{1}{z_1} \Phi \left( \frac{a}{2} + \frac{1}{a} \log \left( \frac{z_2}{z_1} \right) \right) + \frac{1}{z_2} \Phi \left( \frac{a}{2} + \frac{1}{a} \log \left( \frac{z_1}{z_2} \right) \right) \right\}, \quad (2.36)$$

où  $\Phi$  est la fonction de répartition d'une  $N(0, 1)$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad \forall x \in \mathbb{R} \quad (2.37)$$

et  $a^2 = (s_1 - s_2)^t \Sigma^{-1} (s_1 - s_2)$ , la distance de Mahalanobis.

**Corollaire 2.1** *La fonction du coefficient extrême est donc*

$$\theta(h) = 2\Phi \left( \frac{1}{2} \sqrt{h \Sigma^{-1} h} \right) \quad (2.38)$$

Bien que ce modèle soit peu exible il a été largement utilisé au tout début de la géostatistique des extrêmes. En particulier car son interprétation en terme physique de  $R$ . Smith l'a rendu accessible. Voici ces caractéristiques :

- peu réaliste car les fonctions spectrales sont déterministes à une translation près.
- sa simulation est très très simple et rapide - la densité gaussienne possède une borne uniforme.
- bref pas vraiment conseillé pour les applications mais intéressant pour les études de simulations et/ou la théorie.

### 2.3.2.2 Modèle de Schlather

— Le second modèle a été introduit environ 10 ans après dans le document fondateur de Schlather (2002)[66], qui a introduit ce que l'on appelle le processus de Schlather, parfois appelé processus de Gauss extrême. Ce modèle est défini par

$$\eta(s) = \sqrt{2\pi} \max_{i \geq 1} \zeta_i \max\{0, \varepsilon_i(s)\}, \quad s \in \chi \quad (2.39)$$

soit  $\varepsilon_1, \varepsilon_2, \dots$  sont des copies indépendantes d'un processus Gaussien standard de fonction de corrélation  $h \rightarrow \rho(h)$ , tq  $\sqrt{2\pi} \mathbb{E}[\max\{0, \varepsilon_i(s)\}] = 1$ . pour tous  $s \in \chi$

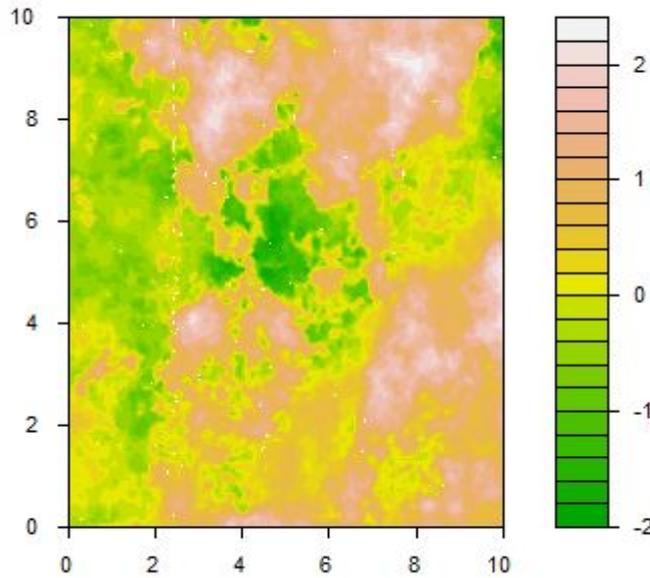


FIGURE 2.4 – Modèle de Schlather.

Avec ces hypothèses supplémentaires, Martin Schlather a montré que la loi de probabilité marginale bivariee du champ aléatoire max-stable simple stationnaire  $\eta = \{\eta(s); s \in \mathbb{R}^d\}$  en deux sites  $s_1 \in \mathbb{R}^d$  et  $s_2 \in \mathbb{R}^d$  est donnée, pour tout  $z_1 \in \mathbb{R}_+^*$  et  $z_2 \in \mathbb{R}_+^*$  par :

$$\Pr(\eta(s_1) \leq z_1, \eta(s_2) \leq z_2) = \exp \left\{ \frac{-1}{2} \left( \frac{1}{z_1} + \frac{1}{z_2} \right) \left( 1 + \sqrt{1 - 2(\rho(h) + 1) \frac{z_1 z_2}{(z_1 + z_2)^2}} \right) \right\}.$$

$$h = \|s_1 - s_2\| \in \mathbb{R}_+. \quad (2.40)$$

- Généralement, le corrélogramme  $\rho$  du champ aléatoire stationnaire  $Y = \{Y(s); s \in \mathbb{R}^d\}$  est choisi parmi l'une des familles paramétriques suivantes :

#### 1. Famille de Whittle -Mathérn :

$$\forall h \in \mathbb{R}_+, \quad \rho(h) = m_1 \frac{2^{1-\ell}}{\Gamma(\ell)} \left( \frac{h}{m_2} \right)^\ell k_\ell \left( \frac{h}{m_2} \right), \quad 0 \leq m_1 \leq 1, \quad m_2 > 0 \text{ et } \ell > 0 \quad (2.41)$$

où  $k_\ell$  est une fonction de Bessel modifiée de deuxième espèce de paramètre  $\ell$  et  $\Gamma$  est la fonction Gamma définie par :

$$\forall s > 0, \Gamma(s) = \int_0^\infty t^{s-1} e^{-t} dt \quad (2.42)$$

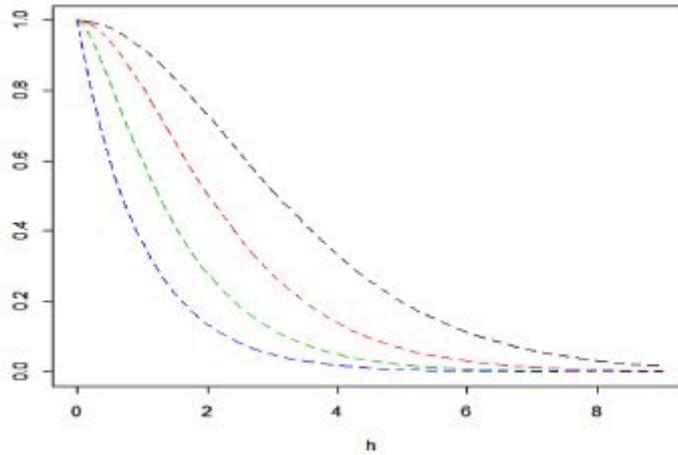


FIGURE 2.5 – Graphes des corrélogrammes de la famille Whittle -Mathérn selon les valeurs de  $h \in \mathbb{R}_+$  pour différentes valeurs de  $\ell$ , avec  $m_1 = m_2 = 1$  : Le noir pour  $\ell = 4$ , le rouge pour  $\ell = 2$ , le vert pour  $\ell = 1$  et le bleu pour  $\ell = 0.5$ .

## 2. Famille de Cauchy :

$$\forall h \in \mathbb{R}_+, \rho(h) = m_1 \left[ 1 + \left( \frac{h}{m_2} \right)^2 \right]^{-\ell}, \quad 0 \leq m_1 \leq 1, m_2 > 0 \text{ et } \ell > 0 \quad (2.43)$$

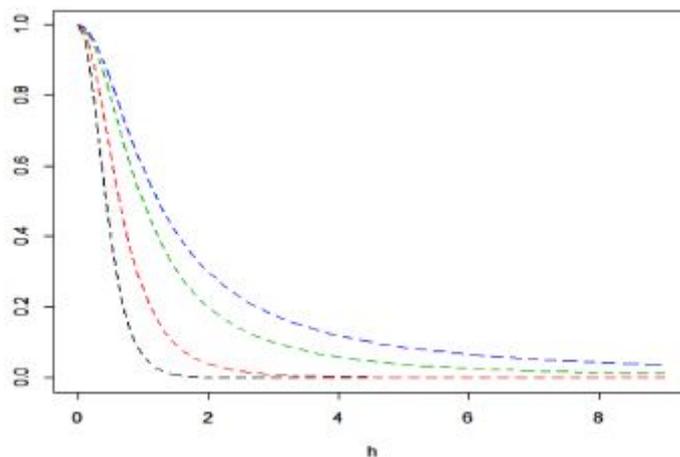


FIGURE 2.6 – Graphes des corrélogrammes de la famille de Cauchy selon les valeurs de  $h \in \mathbb{R}_+$  pour différentes valeurs de  $\ell$ , avec  $m_1 = m_2 = 1$  : Le noir pour  $\ell = 4$ , le rouge pour  $\ell = 2$ , le vert pour  $\ell = 1$  et le bleu pour  $\ell = 0.75$ .

3. Famille stable :

$$\forall h \in \mathbb{R}_+, \rho(h) = m_1 \cdot \exp\left\{-\left(\frac{h}{m_2}\right)^\ell\right\}, \quad 0 \leq m_1 \leq 1, \quad m_2 > 0 \text{ et } 0 < \ell \leq 2 \quad (2.44)$$

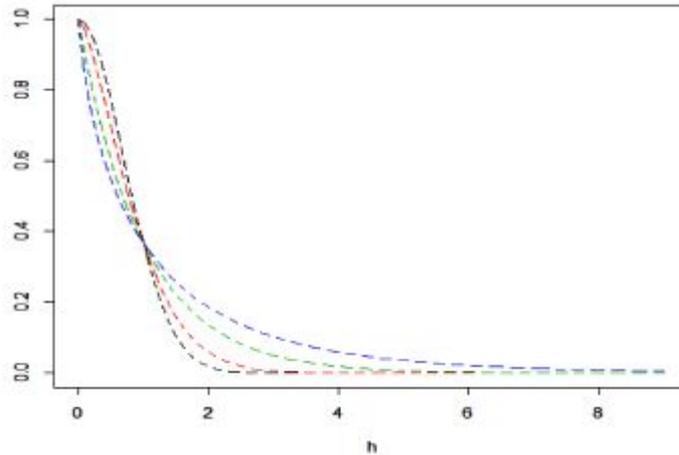


FIGURE 2.7 – Graphes des corrélogrammes de la famille stable selon les valeurs de  $h \in \mathbb{R}_+$  pour différentes valeurs de  $\ell$ , avec  $m_1 = m_2 = 1$  : Le noir pour  $\ell = 2$ , le rouge pour  $\ell = 1.5$ , le vert pour  $\ell = 1$  et le bleu pour  $\ell = 0.75$ .

**Corollaire 2.2** *La fonction du coefficient extrême est donc*

$$\theta(h) = 1 + \sqrt{\frac{1 - \rho(h)}{2}}. \quad (2.45)$$

Voici ses caractéristiques :

- Plus flexible que le modèle précédent du fait des fonctions spectrales aléatoires
- mais n’atteint jamais l’indépendance car nécessairement comme fonction *def. pos.*  $\rho(h) \rightarrow 0$  lorsque  $h \rightarrow \infty$  impliquant  $\theta(h) \rightarrow 1 + \sqrt{1/2} \approx 1.71$ .
- Ce processus se simule assez bien (sauf sur les grilles larges mais c’est toujours le cas pour les grilles larges) mais les simulations ne sont pas exactes. En effet, le processus Gaussien n’est évidemment pas uniformément bornée (on prendra comme borne pseudo uniforme  $M = 3$  ou  $4$ ).

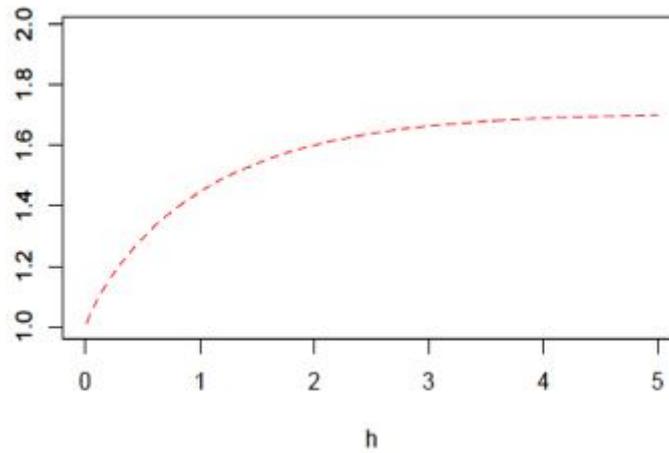


FIGURE 2.8 – Graphe du coefficient extrême du modèle de Schlather où  $\rho$  le corrélogramme de Whittle-Mathérn avec  $m_1 = m_2 = \ell = 1$ .

### 2.3.2.3 Modèle de Brown-Resnick

— Ensuite est venu ce qu'on appelle le processus de Brown-Resnick [Brown and Resnick, 1977; Kabluchko et al., 2009][4][37]. Ce modèle est défini par :

$$\eta(s) = \max_{i \geq 1} \zeta_i \exp \left\{ \varepsilon(s) - \frac{\sigma^2(s)}{2} \right\}, \quad (2.46)$$

où  $\varepsilon$  est un processus Gaussien centré à accroissements stationnaires.

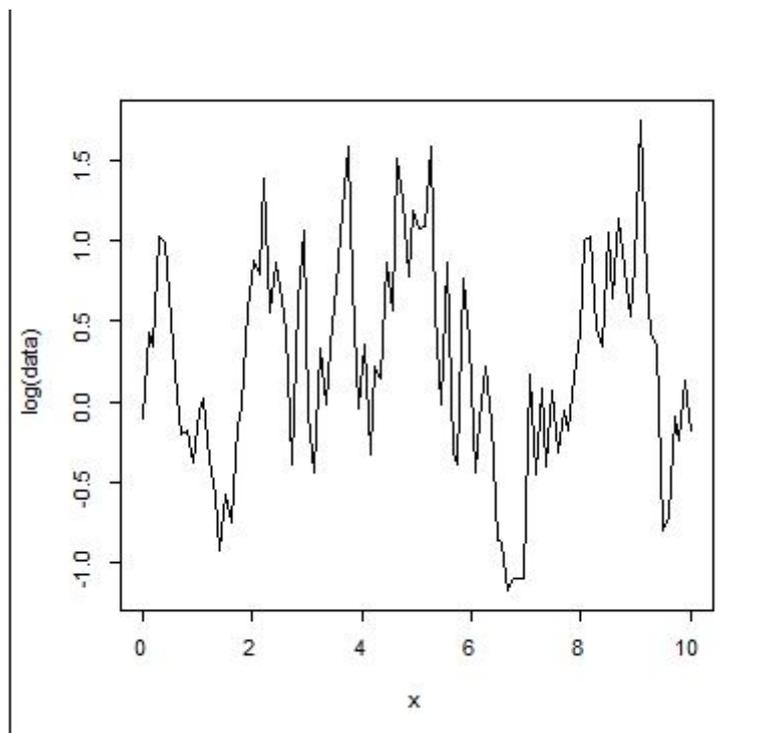


FIGURE 2.9 – Modèle de Brown-Resnick.

**Remarque 2.2** On rappelle que la transformée de Laplace d'une  $X \sim N(\mu, \sigma^2)$  vaut

$$E \{ \exp(tX) \} = \exp \left( \mu t + \frac{\sigma^2 t^2}{2} \right). \quad (2.47)$$

d'où l'expression de la fonction spectrale  $Y$  pour avoir  $\mathbb{E}\{Y(s)\} = 1$  pour tout  $s \in \chi$ .

**Proposition 2.3** ([Kabluchko et al., 2009]) Bien que  $\exp \left\{ \varepsilon(s) - \frac{\sigma^2(s)}{2} \right\}$  ne soit pas stationnaire, De plus ce processus max-stable est complètement caractérisé par le variogramme de  $\varepsilon$ .

**Proposition 2.4**

$$\Pr \{ \eta(s_1) \leq z_1, \eta(s_2) \leq z_2 \} = \exp \left\{ \begin{array}{l} -\frac{1}{z_1} \Phi \left( \frac{a}{2} + \frac{1}{a} \log \left( \frac{z_2}{z_1} \right) \right) \\ -\frac{1}{z_1} \Phi \left( \frac{a}{2} + \frac{1}{a} \log \left( \frac{z_2}{z_1} \right) \right) \end{array} \right\} \quad (2.48)$$

où  $a^2 = 2\gamma(s_1 - s_2)$  avec  $\gamma$  le semi-variogramme de  $\varepsilon$ , *i.e.*

$$\gamma(h) = \frac{1}{2} \text{var} \{ \varepsilon(o) - \varepsilon(h) \}. \quad (2.49)$$

**Corollaire 2.3** La fonction du coefficient extrême est donc

$$\theta(h) = 2\Phi \left\{ \sqrt{\frac{\gamma(h)}{2}} \right\} \quad (2.50)$$

Un choix très utilisé (mais ce n'est pourtant pas le seul) est de prendre  $\varepsilon$  comme un mouvement Brownien fractionnaire, *i.e.*, de variogramme proportionnel à  $h^\alpha$ ,  $0 < \alpha \leq 2$ .

Voici les caractéristiques de ce modèle :

- Il est très flexible.
- Si le variogramme n'est pas borné alors l'indépendance est atteinte lorsque  $h \rightarrow \infty$ .
- Pour le choix d'un mouvement Brownien fractionnaire, le cas  $\alpha = 2$  correspond à un modèle de Smith isotrope.
- Il est difficile à simuler, cf. remarque suivante-quoique les dernières avancées le rende plus sympathique.

**Remarque 2.3** Si dans la caractérisation spectrale d'un Brown-Resnick  $\varepsilon$  est un mouvement Brownien fractionnaire, *i.e.*,

$\gamma(h) = h^\alpha$ ,  $0 < \alpha \leq 2$ , nous savons que ses trajectoires sont presque sûrement continue au sens de Hölder, *i.e.*, pour tout  $\varepsilon > 0$ , il existe une constante  $c > 0$  telle que

$$\| \varepsilon(s_1) - \varepsilon(s_2) \| \leq c \| s_1 - s_2 \|^{2-\varepsilon}, p.s., \quad (2.51)$$

Ceci suggère que le processus  $\exp \{ \varepsilon(s) - \gamma(s) \} \rightarrow 0$  lorsque  $\|s\| \rightarrow \infty$ .

2.3.2.4 Modèle de Extremal-t

— Enfin un dernier modèle, connue sous le nom de processus extremal-t [Nikoloulou-poulos et al., 2009; Davison et al., 2012; Opitz, 2013; Ribatet, 2013][44][13][45][61], qui s'écrit :

$$\eta(s) = c_v \max_{i \geq 1} \zeta_i \max\{0, \varepsilon(s)\}^v, \text{ où } c_v = \sqrt{\pi} 2^{-(v-2)} \Gamma\left(\frac{v+1}{2}\right)^{-1}, \quad v \geq 1, \quad (2.52)$$

et généralise donc le processus de Schlather puisque  $\varepsilon$  est un processus Gaussien standard de fonction de corrélation  $\rho$ .

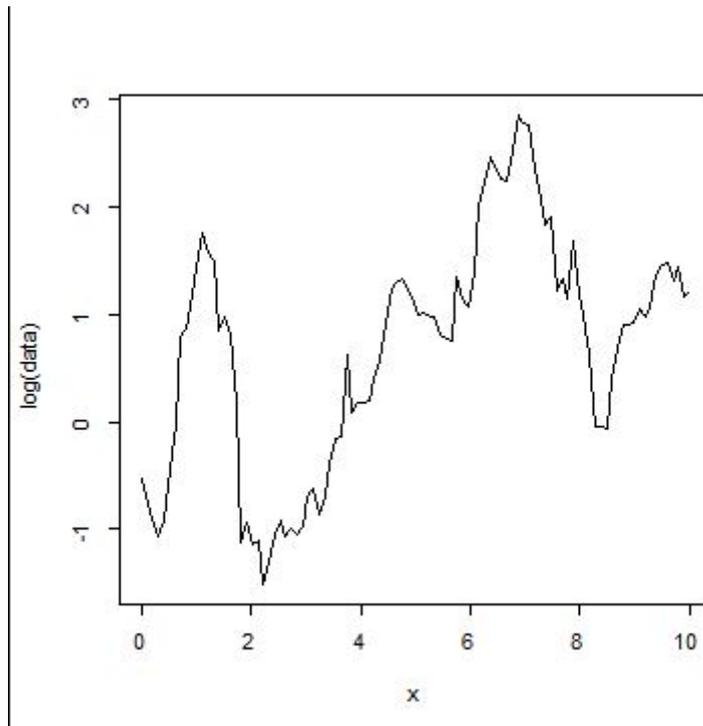


FIGURE 2.10 – Modèle de Extremal-t.

**Proposition 2.5**

$$\Pr\{\eta(s_1) \leq z_1, \eta(s_2) \leq z_2\} = \exp \left[ \begin{array}{l} -\frac{1}{z_1} T_{v+1} \left\{ -\frac{\rho(s_1-s_2)}{b} + \frac{1}{b} \left(\frac{z_2}{z_1}\right)^{1/v} \right\} \\ -\frac{1}{z_2} T_{v+1} \left\{ -\frac{\rho(s_1-s_2)}{b} + \frac{1}{b} \left(\frac{z_1}{z_2}\right)^{1/v} \right\} \end{array} \right], \quad (2.53)$$

où  $T_v$  est la f.d.r. d'une Student a  $v$  d.d.l. et  $b^2 = \{1 - \rho(s_1 - s_2)\}^2 / (v + 1)$ .

**Corollaire 2.4** La fonction du coefficient extrême est donc

$$\theta(h) = 2\Phi\left(\frac{\sqrt{\delta^2(1 - \rho(h))}}{2}\right) \quad (2.54)$$

Voici ces caractéristiques :

- il est aussi exible que Brown-Resnick (mais à un paramètre supplémentaire souvent difficile à estimer  $\nu$ );
- sa simulation se fait plus facilement que Brown-Resnick sauf lorsque est grand (la pseudo borne uniforme est alors très grande);
- l'indépendance est virtuellement atteinte lorsque est grand, disons  $\nu \geq 10$ .

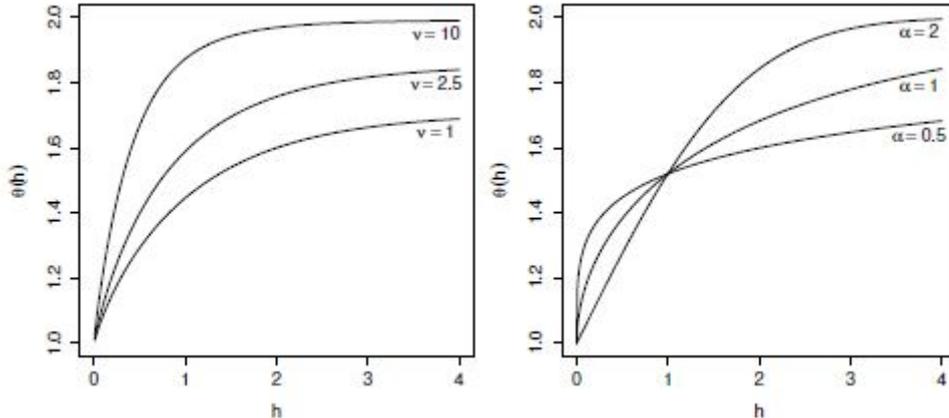


FIGURE 2.11 – Fonctions des coefficients extrêmes. Gauche : processus Extremal-t avec  $\rho(h) = \exp(-h)$  et degrés de liberté  $n = 1, 2 : 5, 10$ , et À droite : procédé Brown Resnick avec  $\gamma(h) = h^a$  et  $a = 0.5, 1, 2$  : Pour le modèle extremal-t le cas  $n = 1$  correspond à un processus de Schlather. Pour le modèle de Brown Resnick, le cas  $a = 2$  correspond à un processus de Smith avec une matrice de covariance égale à la matrice d'identité.

## 2.4 Application aux données spatiales

Nous examinons maintenant l'application de ces idées à des données spatiales du type décrit dans l'introduction. L'attention est limitée à l'approche «traditionnelle» de la valeur extrême basée sur les maxima annuels c'est-à-dire que nous supposons que les données consistent en des maxima annuels sur un ensemble de sites. Clairement il y a beaucoup de possibilité de développer des méthodes de seuil dans ce contexte, mais cela ne sera pas pris en compte ici.

Le but de la présente section est de proposer une procédure statistique pour ajuster les modèles décrits à bien des égards, il s'agit d'une procédure hautement ad hoc, qui a été saluée par toutes les formes d'optimalité. c'est aussi une question pour des recherches ultérieures Notre objectif principal est de démontrer que les modèles développés sont au moins potentiellement applicables aux données réelles. Dans la section suivante leur performance sur certaines données réelles sera prise en compte.

La procédure proposée est la suivante :

### 2.4.1 Etape 1 : La distribution des valeurs extrêmes généralisées

$$F(x; \mu, \sigma, \xi) = \exp \left[ - \left\{ 1 - \frac{\xi(x - \mu)}{\sigma} \right\}_+^{\frac{1}{\xi}} \right] \quad (2.55)$$

est adapté aux données de chaque site, puis la transformation intégrale de probabilité

$$y = \left\{ 1 - \frac{\xi(x - \mu)}{\sigma} \right\}_+^{-\frac{1}{\xi}} \quad (2.56)$$

utilisé pour transformer les données sur chaque site afin qu'ils aient une fonction de distribution Fréchet standard  $(F(y) = e^{-\frac{1}{y}})$ . Pour la présente étude, cela se fait séparément sur chaque site bien que dans le contexte des méthodes régionales en hydrologie, il serait judicieux d'envisager également des procédures dans lesquelles tout ou partie des paramètres  $\mu, \sigma, \xi$  sont communs d'un site à l'autre. L'ajustement s'effectue par maximum de vraisemblance, bien qu'il existe à nouveau d'autres procédures telles que la méthode des moments de probabilité (Hosking, Wallis et Wood, 1985)[35], qui pourraient être envisagées. Nous supposons désormais que les données sur le site  $t_i$  ( $i = 1, \dots, p$ ) consistent en observations normalisées  $\{y_{ni}, 1 \leq n \leq N\}$  avec une distribution marginale de Fréchet standard.

### 2.4.2 Étape 2 : Estimer le coefficient extrême entre chaque paire de sites

La seule nouveauté dans le concept de coefficient extrême est son nom. L'idée a été proposée par Tiago de Oliveira (voir de Haan, 1985)[16] comme un indice de la dépendance de la valeur extrême entre deux variables, et est également apparue sous différentes formes dans la littérature hydrologique. Le nom choisi ici est conçu comme un lien entre le coefficient de corrélation (dont il est en quelque sorte un analogue des valeurs extrêmes) et l'indice extrême (Leadbetter, 1983)[38], qui est un concept similaire pour mesurer l'effet de la dépendance dans un processus stochastique stationnaire.

L'utilisation directe du coefficient de corrélation dans la théorie des valeurs extrêmes est moins souhaitable, car le coefficient de corrélation n'a pas la propriété d'invariance dans les transformations marginales, et certains des cas que nous considérons (y compris le Fréchet) n'ont même pas de variance infinie.

Supposons que  $(X_1, X_2)$  ait une distribution à deux valeurs extrêmes avec une fonction de distribution marginale commune  $F$ ; le coefficient extrême  $\theta$  entre  $X_1$  et  $X_2$  est défini par la relation

$$Pr\{\max(X_1, X_2) \leq x\} = F^\theta(x)$$

. En termes de fonction de dépendance, nous avons  $\theta = 2A(1/2)$ . En ce qui concerne la notion de "nombre équivalent de sites indépendants" décrite à la section 1, on peut constater que le coefficient extrême est simplement celui qui n'existe que sur deux sites.

Enfin, il convient de noter que Buishand (1984)[3] a défini une fonction  $\theta(x)$  par

$$Pr\{\max(X_1, X_2) \leq x\} = F^{\theta(x)}(x)$$

il s'agit d'une constante lorsque la distribution conjointe est effectivement une distribution de valeurs extrêmes bivariée mais que, dans la pratique, on constate souvent qu'elle n'est pas constante, phénomène également noté par Tawn (1990b, 1990c) [68][69] dans plusieurs séries de données réelles.

Si la distribution marginale est l'unité Fréchet, alors  $\frac{1}{X_1}$  et  $\frac{1}{X_2}$  ont des distributions exponentielles unitaires et  $\frac{1}{\max(X_1, X_2)}$  a une distribution exponentielle de moyenne  $\frac{1}{\theta}$ . Ceci suggère

un estimateur évident. Revenons à notre contexte actuel dans lequel nous avons transformé les données  $\{Y_{ni}, 1 \leq n \leq N, 1 \leq i \leq p\}$ ,

et notant que l'ajustement du maximum de vraisemblance des distributions marginales aura permis d'assurer que  $\sum_n Y_{ni}^{-1} = \sum_n Y_{nj}^{-1} = N$ , l'estimateur naturel du coefficient extrémal  $\theta_{ij}$  entre les sites  $i$  et  $j$  est

$$\theta_{ij} = \frac{N}{\left\{ \sum_{n=1}^N \min(Y_{ni}^{-1}, Y_{nj}^{-1}) \right\}} \quad (2.57)$$

Ces quantités sont appelées estimations brutes des coefficients extrêmes, car elles ne sont basées sur aucun modèle.

Les propriétés théoriques des extrêmes bivariés montrent que  $1 \leq \theta_{ij} \leq 2$ , les cas extrêmes 1 et 2 correspondant respectivement à la dépendance et à l'indépendance complètes. L'estimateur  $\theta_{i,j}$  pourrait prendre n'importe quelle valeur entre 1 et  $+\infty$ , le premier seulement si  $Y_{ni} = Y_{nj}$  pour tout  $n$ , mais les cas pour lesquels  $\theta_{ij} > 2$  correspondent à une dépendance négative entre les sites et sont relativement rares dans un contexte de valeurs extrêmes.

Il convient de noter que le concept de coefficient extrême à une extension évidente à plus de 2 variables : le coefficient extrême de  $k$  variables  $X_1, \dots, X_k$ , supposées extrêmes multivariées avec une fonction de distribution marginale commune  $F$ , est défini par la relation

$$Pr\{\max(X_1, \dots, X_k) \leq x\} = F^{\theta}(x).$$

Dans ce cas, la plage théorique est  $1 \leq \theta \leq k$

### 2.4.3 Étape 3 : Estimer l'erreur de chaque type

Cela doit prendre en compte les étapes 1 et 2, car la marge marginale la transformation affecte sensiblement l'erreur standard. Pour voir cela, considérez le cas extrême dans lequel

$Y_{i,n} = Y_{j,n}$  pour tout  $n$  La transformation marginale assure  $\sum_n Y_{i,n}^{-1} = \sum_n Y_{j,n}^{-1} = N$  et donc  $\tilde{\theta}_{ij} = 1$  en d'autres termes, la norme erreur est 0. Sans la transformation marginale,  $\frac{1}{\tilde{\theta}_{i,j}}$  serait la moyenne de  $N$  variables indépendantes exponentielles et donc avoir la variance  $\frac{1}{N}$ .

Face à cette difficulté, il n'y a pas de formule approximative simple pour l'erreur standard. La procédure la plus raisonnable semblerait être une basé sur le ré-échantillonnage, c'est-à-dire jackknife ou bootstrap. Pour la présente étude, le l'estimateur jackknife de l'erreur type (Efron 1982) [25] a été utilisé; cela mène à la formule de l'erreur type

$$\left\{ \frac{N-1}{N} \sum_{n=1}^N \left( \tilde{\theta}_{ij}^{(n)} - \tilde{\theta}_{ij} \right)^2 \right\}^{-1},$$

dans lequel  $\tilde{\theta}_{ij}^{(n)}$  est l'estimation obtenue à partir du seul échantillon de jackknife de dont les observations de l'année  $n$  sont omises Bien sûr, le marginal les paramètres doivent être ré-estimés pour chaque échantillon Jackknife.

#### 2.4.4 Étape 4 : Ajuster un modèle max-stable

L'algorithme proposé ici est une somme de carrés de résidus pondérés : définissant

$$r_{ij} = \left( \tilde{\theta}_{ij} - \hat{\theta}_{ij} \right) / s_{ij}$$

où  $\hat{\theta}_{ij}$  est la valeur ajustée d'un modèle,  $\tilde{\theta}_{ij}$  est l'estimation brute calculée en Étape 2, et  $s_{ij}$  l'erreur type de l'étape 3, les paramètres du modèle sont choisis minimiser  $\sum_{i,j} r_{i,j}^2$ .

Le critère de la somme pondérée des carrés a été proposé par S.J. Neil, dans le cadre d'un M.Phil. thèse à l'Université de Surrey, après une tentative antérieure basée sur les moindres carrés non pondérés a été jugée non satisfaisante dans les régions de forte dépendance, quand  $\tilde{\theta}_{ij}$  est proche de 1. Néanmoins, il est encore vraiment un critère ad hoc; c'est une question ouverte s'il est possible trouver un moyen d'approximer la fonction de vraisemblance.

#### 2.4.5 Étape 5. Tester l'ajustement du modèle

Une méthode de vérification de l'adéquation est la plus évidente qui consiste à examiner la pondération des résidus de l'étape 4 pour les valeurs observées, ou pour toute preuve systématique de non-aléatoire comportement (tracé des résidus contre des valeurs ajustées ou contre d'autres variables telles que les distances entre les points). Une deuxième idée, notant que l'algorithme d'estimation repose entièrement sur des dépendances par paires, est pour tester l'ajustement sur les dépendances d'ordre supérieur. Cela peut être fait en calculant les coefficients extrémaux d'ordre supérieur mentionnés à la

fin de l'étape 2, avec leurs erreurs types, et en les comparant aux valeurs ajustées basées sur le modèle estimé.

## 2.5 Mesure de dépendance pour les extrêmes spatiaux

Lorsque l'on s'intéresse au comportement des extrêmes dans un cadre multivarié tel les champs aléatoires, la caractérisation de la dépendance des réalisations extrêmes est une notion fondamentale pour toute inférence statistique sur la distribution jointe des extrêmes. Quand les données ne sont pas extrêmes, les outils permettant de mesurer cette dépendance sont la fonction de covariance, la fonction de corrélation et le variogramme. Mais dès que notre intérêt se porte sur les valeurs extrêmes, ces outils ne seront plus d'une grande utilité et ils peuvent, même, ne pas exister. Ainsi, des différentes caractérisations ont été proposées dans la littérature qui ont été à la base de différents travaux pour la caractérisation de la dépendance spatiale des extrêmes.

### 2.5.1 Madogramme

Comme on l'a déjà dit auparavant, en géostatistique le variogramme est un bon outil pour mesurer la dépendance spatiale entre deux réalisations d'un champ aléatoire. Cependant, quand il s'agit des observations extrêmes, son usage est très limité car il se peut qu'il n'existera pas pour de telles observations. Pour voir cela, considérons un champ aléatoire max-stable avec des lois de probabilité marginales univariées de Fréchet unitaires. Pour un tel champ aléatoire, l'espérance et la variance ne sont pas finies et par conséquent le variogramme n'existe pas théoriquement. D'où la nécessité d'introduire de nouveaux outils qui remplacent le variogramme. En 1987, dans le cadre de la géostatistique, Georges Matheron a introduit un nouveau outil similaire au variogramme qu'on appelle « madogramme », ou quelques fois rencontré sous le nom de « variogramme au premier ordre ». Dans l'article Cooley, Naveau & Poncet (2006)[10], Dan Cooley, Philippe Naveau et Paul Poncet ont montré qu'il y a une forte relation entre le coefficient extrémal bivarié d'un champ aléatoire max-stable et son madogramme, lorsque ce dernier existe. Et par conséquent, ils ont utilisé le madogramme comme un moyen pour estimer le coefficient extrémal bivarié d'un champ aléatoire max-stable.

**Définition 2.6 (Cooley, Naveau & Poncet (2006)) [10]**

Soit  $Y = \{Y(s); s \in \mathbb{R}^d\}$  un champ aléatoire stationnaire défini sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathcal{P})$  à espace d'état  $\mathbb{R}$  tel que son espérance est finie. On appelle « madogramme » du champ aléatoire  $Y$  la fonction, qu'on note  $M$ , définie sur  $\mathbb{R}^d$  dans  $\mathbb{R}^+$  par :

$$\forall h \in \mathbb{R}^d, M(h) = \frac{E(|Y(s+h) - Y(s)|)}{2}, s \in \mathbb{R}^d \quad (2.58)$$

Dans la proposition suivante, on va essayer de récapituler les propriétés de base, qui sont faciles à vérifier, du madogramme d'un champ aléatoire stationnaire :

**Proposition 2.6 (Cooley, Naveau & Poncet (2006)) [10]**

Soit  $M$  le madogramme d'un champ aléatoire stationnaire  $Y = \{Y(s); s \in \mathbb{R}^d\}$  défini sur un espace probabilisé  $(\Omega; \mathcal{A}; \mathcal{P})$  à espace d'état  $\mathbb{R}$ . On a :

1.

$$M(0_{\mathbb{R}^d}) = 0 \quad (2.59)$$

2.  $M$  est une fonction positive :

$$\forall h \in \mathbb{R}^d, M(h) \geq M(0_{\mathbb{R}^d}) = 0 \quad (2.60)$$

3.  $M$  est une fonction paire :

$$\forall h \in \mathbb{R}^d, M(h) = M(-h) \quad (2.61)$$

$$\forall (h, k) \in \mathbb{R}^d \times \mathbb{R}^d, M(h+k) \leq M(h) + M(k) \quad (2.62)$$

### 2.5.2 Madogramme et coefficient extrémal

Dans cette section, on va voir la relation qui existe entre le madogramme d'un champ aléatoire max-stable stationnaire et son coefficient extrémal bivarié. En fait, il y a le même problème avec le madogramme que celui avec le variogramme lorsque les lois de probabilité marginales du champ aléatoire max-stable sont de Fréchet unitaires. Dans ce cas, l'espérance d'un tel champ aléatoire est infinie, et par conséquent son madogramme n'existe pas. Cela a mené, en 2006, Dan Cooley, Philippe Naveau et Paul Poncet, (2006) [10] à standardiser, sans perte de généralisation et de la max-stabilité, les lois de probabilité marginales univariées d'un champ aléatoire max-stable en les transformant en distributions des valeurs extrêmes généralisées identiques avec un paramètre de forme  $\xi < 1$ . Dans ce cas, l'espérance d'un tel champ aléatoire max-stable standardisé est finie, ce qui entraîne l'existence de son madogramme. Bien sûr, il est possible de suivre la même stratégie pour assurer l'existence du variogramme d'un champ aléatoire max-stable, mais on va voir qu'en travaillant avec le madogramme est plus convenable car il a des liens forts avec le coefficient extrémal.

Soit  $Z = \{Z(s); s \in \mathbb{R}^d\}$  un champ aléatoire max-stable stationnaire ayant des distributions des valeurs extrêmes généralisées identiques comme lois de probabilité marginales univariées de paramètre de forme  $\xi \in \mathbb{R}$ , de paramètre de localisation  $\mu \in \mathbb{R}$  et de paramètre d'échelle  $\sigma > 0$ , c'est-à-dire :

$$\forall z \in \mathbb{R}, \Pr[Z(s) \leq z] = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]_+^{-1/\xi}, s \in \mathbb{R}^d \right\} \quad (2.63)$$

avec :

$$\left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]_+ = \begin{cases} 1 + \xi \left( \frac{z - \mu}{\sigma} \right) & \text{si } 1 + \xi \left( \frac{z - \mu}{\sigma} \right) > 0 \\ 0 & \text{si } 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \leq 0 \end{cases} \quad (2.64)$$

Dans la proposition suivante, on va présenter un résultat intéressant qui nous donne

une méthode directe pour estimer le coefficient extrémal bivarié d'un champ aléatoire max-stable stationnaire.

**Proposition 2.7 (Cooley, Naveau & Poncet (2006)) [10]**

Soient  $Z = \{Z(s); s \in \mathbb{R}^d\}$  le champ aléatoire max-stable stationnaire défini ci-dessus et  $\theta$  son coefficient extrémal bivarié. Si le paramètre de forme  $\xi$  est strictement inférieur à 1 ( $\xi < 1$ ), alors le madogramme  $M$  du champ aléatoire  $Z$  et son coefficient extrémal  $\theta$  en deux sites d'observations  $s_1 \in \mathbb{R}^d$  et  $s_2 \in \mathbb{R}^d$  vérifient la relation suivante :

$$\theta(s_2 - s_1) = \begin{cases} u_\beta(z) \left[ \mu + \frac{M(s_2 - s_1)}{\Gamma(1 - \xi)} \right], & \text{si } \xi \neq 0 \\ \exp\left\{ \frac{M(s_2 - s_1)}{\sigma} \right\}, & \text{si } \xi = 0 \end{cases} \quad (2.65)$$

avec

$$\forall z \in \mathbb{R}, u_\beta(z) = \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]_+^{-1/\xi} \quad \text{et} \quad \forall y > 0, \Gamma(y) = \int_0^{+\infty} t^{y-1} e^{-t} dt \quad (2.66)$$

**Remarque :** Dans le cadre de la proposition ci-dessus et afin de voir l'application directe de son résultat, on a :

1. Lorsque les lois de probabilité marginales univariées du champ aléatoire max-stable stationnaire  $Z = \{Z(s); s \in \mathbb{R}^d\}$  sont de Weibull standards  $\xi = -1, \mu = 0, \sigma = 1$ , on a : d'après la formule 2.65, on a

$$M(s_2 - s_1) = 1 - \frac{1}{\theta(s_2 - s_1)}$$

2. Lorsque les lois de probabilité marginales univariées du champ aléatoire max-stable stationnaire  $Z = \{Z(s); s \in \mathbb{R}^d\}$  sont de Gumbel standards  $\xi = 0, \mu = 0, \sigma = 1$ , comme c'est le cas du modèle de Brown-Resnick, on a d'après la formule 2.65 :

$$M(s_2 - s_1) = \ln \theta(s_2 - s_1)$$

### 2.5.3 F-madogramme

De ce qu'on a vu jusque là, on constate qu'il y a un lien fort entre le coefficient extrémal bivarié d'un champ aléatoire max-stable stationnaire standardisé et son madogramme. Ce qui nous a donné une méthode directe pour estimer ce coefficient extrémal. Mais lorsque les lois de probabilité marginales univariées d'un champ aléatoire max-stable stationnaire sont de Fréchet unitaires, le madogramme tel qu'il est défini ci-dessus ne sera pas d'une grande utilité, car ce madogramme n'existe pas

( $\xi = 1$ ). Dans le même article Cooley, Naveau & Poncet (2006) [10], Dan Cooley, Philippe Naveau et Paul Poncet ont proposé, au lieu du madogramme, le madogramme modifié qu'ils ont appelé « le **F-madogramme** ».

Soit  $Z = \{Z(s); s \in \mathbb{R}^d\}$  un champ aléatoire max-stable stationnaire ayant des distributions des valeurs extrêmes généralisées identiques comme lois de probabilité marginales univariées de paramètre de forme  $\xi \in \mathbb{R}$ , de paramètre de localisation  $\mu \in \mathbb{R}$  et de paramètre d'échelle  $\sigma > 0$ .

**Définition 2.7 (Cooley, Naveau & Poncet (2006)) [10]**

Soit  $Z = \{Z(s); s \in \mathbb{R}^d\}$  le champ aléatoire max-stable stationnaire défini ci-dessus. On appelle **F-madogramme** du champ aléatoire max-stable  $Z$  la fonction, qu'on note  $V_F$ , définie sur  $\mathbb{R}^d$  dans  $\mathbb{R}^+$  par :

$$V_F(h) = \frac{1}{2} E[|F\{Z(s+h) - FZ(s)\}|], \quad \forall s, h \in \mathbb{R}^d \quad (2.67)$$

avec

$$\forall z \in \mathbb{R}, F(z) = \Pr[Z(s) \leq z] = \exp\left\{-\left[1 + \xi \left(\frac{z - \mu}{\sigma}\right)\right]_+^{-1/\xi}, s \in \mathbb{R}^d\right\}. \quad (2.68)$$

Dans la proposition suivante, on va présenter un résultat qui nous montre le lien entre le coefficient extrémal bivarié d'un champ aléatoire max-stable stationnaire et son F-madogramme :

**Proposition 2.8 (Cooley, Naveau & Poncet (2006) et Ribatet (2009)) [10][62]**

Soient  $Z = \{Z(s); s \in \mathbb{R}^d\}$  le champ aléatoire max-stable stationnaire défini ci-dessus, son coefficient extrémal bivarié et  $V_F$  son F-madogramme. (Cooley et al., 2006)[10] On a :

$$2V_F(s_1 - s_2) = \frac{\theta(s_2 - s_1) - 1}{\theta(s_2 - s_1) + 1}, \quad s_1 \in \mathbb{R}^d \text{ et } s_2 \in \mathbb{R}^d \quad (2.69)$$

Où, ce qui est équivalent :

$$\theta(s_2 - s_1) = \frac{1 + 2V_F(s_1 - s_2)}{1 - 2V_F(s_1 - s_2)}, \quad s_1 \in \mathbb{R}^d \text{ et } s_2 \in \mathbb{R}^d \quad (2.70)$$

La figure 2.2 illustre graphiquement les estimations empiriques pour le F-madogramme et les coefficients extrêmes extrêmes par paires correspondants pour les données de précipitations suisses. disponible gratuitement dans le package SpatialExtremes (Ribatet et al., 2013)[59]. Comme On peut s'attendre à ce que la dépendance spatiale diminue avec l'augmentation du décalage temporel, mais semble persister au-delà de 100 km.

**Remarque :** De la proposition ci-dessus, on constate que pour les champs aléatoires max-stables stationnaires ayant des lois de probabilité marginales univariées de Fréchet unitaires, comme celles des modèles de Smith et de Schlather, il serait plus convenable d'utiliser le F-madogramme pour l'estimation du coefficient extrémal bivarié au lieu du madogramme.

## 2.6 Inférence

Dans cette section, on va essayer de donner quelques méthodes d'estimation du coefficient extrémal bivarié d'un champ aléatoire max-stable ainsi que les paramètres de sa

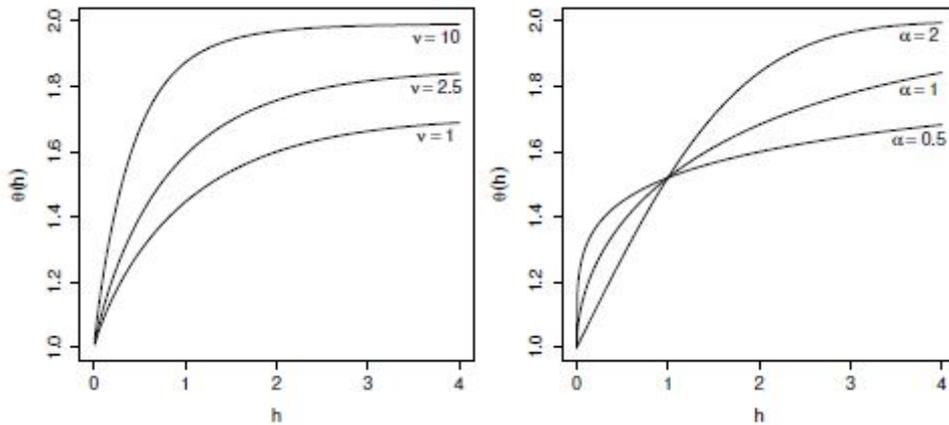


FIGURE 2.12 – F-Madogramme empirique (à gauche) et coefficients extrêmes correspondants (à droite) pour les données de précipitations suisses du paquet SpatialExtremes (Ribatetet al., 2013). Les points gris sont des estimations par paires, tandis que les noirs sont des estimations combinées obtenues à partir de 200 cellules.

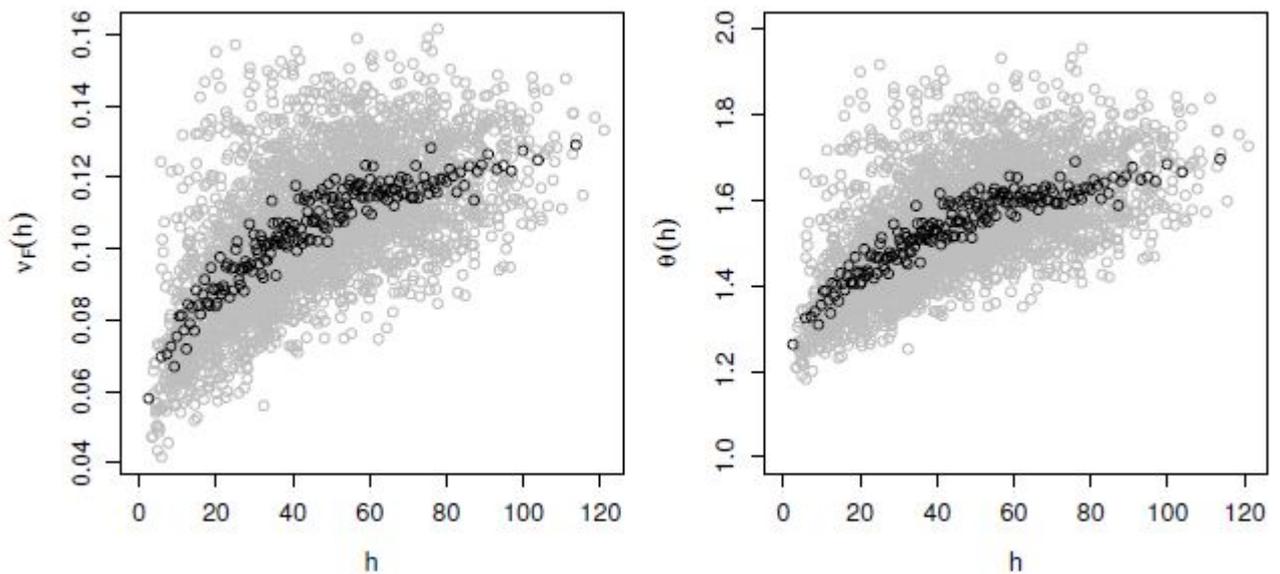


FIGURE 2.13 – Estimation empirique d'un madogramme

loi de probabilité marginale bivariée :

### 2.6.1 Estimation du coefficient extrémal

#### 1. Approche Smith

L'estimation du coefficient extrémal par l'approche Smith est une méthode proposée par Richard L. Smith dans son article

[Smith (1990)][65]. Elle sert à estimer le coefficient extrémal bivarié d'un champ aléatoire max-stable simple. Cette méthode peut être résumée comme suit :

Soit  $Y = \{Y(s); s \in \mathbb{R}^d\}$  un champ aléatoire max-stable simple, c'est-à-dire que ses lois de probabilité marginales univariées

sont des lois de probabilité de Fréchet standards. Et soient  $Y_1, Y_2, \dots, Y_n$   $n$  copies indépendantes du champ aléatoire

max-stable simple  $Y$ . Alors, pour tout  $s \in \mathbb{R}^d$ , la variable aléatoire  $1/Y(s)$  suit une loi de probabilité exponentielle standard  $\varepsilon(1)$ . Et, pour tout  $s_1 \in \mathbb{R}^d$  et  $s_2 \in \mathbb{R}^d$ , la variable aléatoire  $\min[1/Y(s_1), 1/Y(s_2)]$  suit une loi de probabilité exponentielle de paramètre  $\theta(s_2 - s_1)$ , où  $\theta(s_2 - s_1)$  est le coefficient extrémal bivarié du champ aléatoire max-stable  $Y = \{Y(s); s \in \mathbb{R}^d\}$  entre les deux sites d'observations  $s_1$  et  $s_2$ .

Richard L. Smith a proposé un estimateur  $\widehat{\theta}(s_2 - s_1)$  du coefficient extrémal bivarié  $\theta(s_2 - s_1)$  donné par :

$$\widehat{\theta}(s_2 - s_1) = \frac{n}{\sum_{i=1}^n \min[1/Y_i(s_1), 1/Y_i(s_2)]} \quad (2.71)$$

#### 2. Approche F-madogramme

L'estimation du coefficient extrémal par l'approche F-madogramme est une méthode proposée par Dan Cooley, Philippe Naveau et Paul Poncet dans leur article Cooley, Naveau & Poncet (2006) [10]. Cette méthode est basée sur les deux résultats 2.67 et 2.70 et elle sert à estimer le coefficient extrémal bivarié d'un champ aléatoire max-stable ayant des lois de probabilité marginales univariées identiques. Donc elle peut être utilisée pour l'estimation du coefficient extrémal bivarié du modèle de Smith, du modèle de Schlather, ainsi que celui du modèle de Brown-Resnick.

Brown-Resnick.

Pour résumer, soit  $Y = \{Y(s); s \in \mathbb{R}^d\}$  un champ aléatoire max-stable ayant des lois de probabilité marginales univariées identiques. Et soient  $Y_1, Y_2, \dots, Y_n$ ,  $n$  copies indépendantes du champ aléatoire max-stable  $Y$ . Notons, pour  $s_1$  et  $s_2$  dans  $\mathbb{R}^d$ ,  $\theta(s_2 - s_1)$  le coefficient extrémal bivarié du champ aléatoire max-stable  $Y$  entre les deux sites d'observations  $s_1$  et  $s_2$ . Alors, d'après 2.70 on a :

$$\widehat{\theta}(s_2 - s_1) = \frac{1 + 2\widehat{V}_F(s_2 - s_1)}{1 - 2\widehat{V}_F(s_2 - s_1)} \quad (2.72)$$

Avec, d'après 2.67 :

$$\widehat{V}_F(s_2 - s_1) = \frac{1}{2n} \sum_{i=1}^n |\widehat{F}[Y_i(s_2)] - \widehat{F}[Y_i(s_1)]| \quad (2.73)$$

Où  $\widehat{F}[Y_i(s_j)]$  est la fonction de répartition empirique associée à la variable aléatoire  $Y(s_j)$  basée sur l'échantillon  $Y_1(s_1), Y_2(s_2), \dots, Y_n(s_j)$

### 3. Approche madogramme

L'estimation du coefficient extrémal par l'approche madogramme est une méthode proposée, elle aussi, par Dan Cooley, Philippe Naveau et Paul Poncet dans leur article Cooley, Naveau & Poncet (2006)[10]. Cette méthode est basée sur les deux résultats 2.58 et 2.65 et elle sert à estimer le coefficient extrémal bivarié d'un champ aléatoire max-stable ayant des GEVD identiques comme lois de probabilité marginales univariées dont le paramètre de forme  $\xi < 1$ . Donc elle peut être utilisée pour l'estimation du coefficient extrémal bivarié du modèle de Brown-Resnick.

Pour cela, soit  $Y = \{Y(s); s \in \mathbb{R}^d\}$  un champ aléatoire max-stable ayant des GEVD identiques comme lois de probabilité marginales univariées dont le paramètre de forme  $\xi < 1$ . Et soient  $Y_1, Y_2, \dots, Y_n$   $n$  copies indépendantes du champ aléatoire max-stable  $Y$ . Notons, pour  $s_1$  et  $s_2$  dans  $\mathbb{R}^d$ ,  $\theta(s_1 - s_2)$  le coefficient extrémal bivarié du champ aléatoire max-stable  $Y$  entre les deux sites d'observations  $s_1$  et  $s_2$ .

Alors, d'après 2.65 on a :

$$\widehat{\theta}(s_2 - s_1) = \begin{cases} u_\beta(s_2 - s_1) \left[ \mu + \frac{\widehat{M}(s_2 - s_1)}{\Gamma(1 - \xi)} \right] & \text{si } \xi \neq 0 \\ \exp \left\{ \frac{\widehat{M}(s_2 - s_1)}{\sigma} \right\} & \text{si } \xi = 0 \end{cases} \quad (2.74)$$

Et lorsque le champ aléatoire max-stable  $Y = \{Y(s); s \in \mathbb{R}^d\}$  est un modèle de Brown-Resnick, on aura :

$$\widehat{\theta}(s_2 - s_1) = \exp \left[ \widehat{M}(s_2 - s_1) \right] \quad (2.75)$$

Avec, d'après (2.58) :

$$\widehat{M}(s_2 - s_1) = \frac{1}{2n} \sum_{i=1}^n |Y_i(s_2) - Y_i(s_1)| \quad (2.76)$$

**Remarque :** Pour mettre en application les résultats précédents, il est souvent préférable de suivre la procédure suivante : Supposons qu'on a  $m$  sites d'observation où, sur une période de temps donnée, on a récolté sur chaque site un ensemble d'observations. Ensuite, on répartit l'ensemble de données de chaque site en  $n$  sous ensemble, où  $n$  est souvent pris comme étant le nombre d'années de la période de temps sur laquelle on a relevé les observations. Après, sur chaque site d'observation, on construit un autre ensemble de données composé de maximum de données relevées durant chacune des  $n$  an-

nées d'observation. Ainsi, on pourra dire qu'on a construit, approximativement,  $n$  champs aléatoires max-stables  $Y_1, Y_2, \dots, Y_n$  indépendants et identiquement distribués, avec :

$$\forall i \in \{1, 2, \dots, n\} \quad Y_i = \{Y_1(s_1), Y_2(s_2), \dots, Y_i(s_j)\}$$

Où, pour  $j \in \{1, 2, \dots, m\}$ ,  $Y_i(s_j)$  représente le maximum des données récoltées sur le  $j^{\text{eme}}$  site durant la  $i^{\text{eme}}$  année.

Remarque : Parler vite fait des moindres carrés qui sont possibles mais pas trop utilisés.

## 2.6.2 Le cas simple

On sait que

$$p\{\eta(s_1) \leq z_1, \dots, \eta(s_k) \leq z_k\} = \exp\{-V_s(z_1, \dots, z_k)\} \sum_{\tau \in P_k} \omega(\tau). \quad (2.77)$$

Du coup la densité associée est de la forme

$$f(z_1, \dots, z_k) = \exp\{-V_s(z_1, \dots, z_k)\} \sum_{\tau \in P_k} \omega(\tau), \quad (2.78)$$

où  $P_k$  est l'ensemble des partitions de  $\{s_1, \dots, s_k\}$  et

$$\omega(\tau) = (-1)^{|\tau|} \prod_{j=1}^{|\tau|} \frac{\partial^{|\tau_j|}}{\partial z_{\tau_j}} V(z_1, \dots, z_k), \quad (2.79)$$

où  $\tau = (\tau_1, \dots, \tau_\ell)$ ,  $\ell = |\tau|$  et  $\partial^{|\tau_j|} / \partial z_{\tau_j}$  représente les dérivées partielles par rapport aux élément de  $\tau_j$ .

### Définition 2.8 ([Lindsay, 1988]) [52]

Soit  $\{f(z; \psi) : z \in Z, \psi \in \Psi\}$  un modèle statistique paramétrique avec  $Z \subseteq \mathbb{R}^K$  et  $\Psi \subseteq \mathbb{R}^p$ ,  $K \geq 1, p \geq 1$ . Considérons un ensemble d'événements  $\{A_i : A_i \subseteq Z, i \in I\}$  où  $I \subset \mathbb{N}$  et une tribu sur  $Z$ . Une (log) vraisemblance composite est définie par

$$\ell_c(\psi; z) = \sum_{i \in I} \omega_i \ln f(y \in A_i; \psi), \quad (2.80)$$

où  $\{\omega_i : i \in I\}$  est un ensemble de poids (positifs).

- Parler de la vraisemblance indépendante connue aussi sous Bases pseudo-likelihood dans la littérature des champs de Markov.
- Parler de la vraisemblance conditionnelle, i.e,

$$A_{i,j} = \{z_i \mid z_j, 1 \leq i, j \leq k, i \geq j\}.$$

On pourra prendre le cas d'une Gaussienne  $k$ -variée par exemple.

Il est important de garder en mémoire le fait que **les vraisemblances composites ne sont pas de véritables vraisemblances** (sauf cas triviaux) et **ne relèvent pas d'un modèle statistique**. Cela provient de la théorie des équations d'estimation non biaisées.

Rappel : Dans un modèle régulier  $\{f(z; \psi)\}$ , le maximum de vraisemblance est la solution de **l'équation du score**

$$\nabla \ell(\psi; z) = 0 \iff \sum_{m=1}^n \nabla \ln f(z_m; \psi) = 0,$$

qui forme une équation d'estimation non biaisée puisque, d'après la 1<sup>ère</sup> identité de Bartlett,

$$E[\nabla \ln f(z; \psi_0)] = 0.$$

En utilisant des vraisemblances composites, on a

$$\nabla \ell_c(\psi; z) = 0 \iff \sum_{m=1}^n \sum_{i \in I} \omega_i \nabla \ln f(z_m \in A_i; \psi) = 0,$$

et cette dernière équation forme une combinaison linéaire d'équation d'estimation non biaisée puisque pour chaque  $i \in I$

$$E[\nabla \ln f(Z \in A_i; \psi_0)] = 0.$$

avec la contrainte additionnelle que le paramètre doit être identifiable à partir des événements  $A_i$ .

**Théorème 2.6** *Sous les mêmes hypothèses de régularité que pour le maximum de vraisemblance et sous la contrainte d'identifiabilité sur  $\psi$*

$$\sqrt{n}(\widehat{\psi} - \psi_0) \xrightarrow{d} N\{0, H(\psi_0)^{-1} J(\psi_0) H^{-1}(\psi_0)\}, \quad n \rightarrow \infty,$$

avec

$$H(\psi_0) = E[\nabla^2 \ell_c\{\psi_0; \eta(s)\}], \quad J(\psi_0) = \text{var}[\nabla \ell_c\{\psi_0; \eta(s)\}].$$

Remarque : Dans le théorème précédent si l'on substitue la vraisemblance composite par une «vraie» vraisemblance on trouve, par la 2<sup>ème</sup> identité de Bartlett,

$$H(\psi_0) + J(\psi_0) = 0,$$

et donc que

$$\sqrt{n}(\widehat{\psi} - \psi_0) \xrightarrow{d} N\{0, -H(\psi_0)^{-1}\}, \quad n \rightarrow \infty$$

Si n'importe quelle forme de vraisemblance composite peut-être prise, pour les processus max-stables on prendra bien souvent l'estimateur du maximum **de vraisemblance**

par paire [Padoan et al., 2010][51]

$$\widehat{\psi}_p = \arg \max_{\psi \in \Psi} \sum_{m=1}^n \sum_{i=1}^{k-1} \sum_{j=i+1}^k \omega_{i,j} \ln f\{\eta_k(s_i), \eta_k(s_j); \psi\}.$$

### 2.6.3 Le cas non simple

En pratique notre processus max-stable ne sera pas simple mais aura des marges *GEV*. Plus précisément,

$$Z(s) \sim \text{GEV}\{\mu(s), \sigma(s), \xi(s)\}, \quad s \in \mathcal{X}.$$

En utilisant une anamorphose on peut toujours se ramener au cas simple, i.e., considérer le processus

$$\eta(s) = t_s\{Z(s)\}. \quad s \in \mathcal{X}.$$

avec

$$\begin{aligned} t_s &: \mathbb{R} \longrightarrow (0, \infty) \\ Z(s) &\longrightarrow \left\{ 1 + \xi(s) \frac{Z(s) - \mu(s)}{\sigma(s)} \right\}_+^{1/\xi(s)} \end{aligned}$$

Cela ne coûte quasiment rien dans notre approche par vraisemblance par paire puisque cette dernière s'écrit alors

$$\ell_p\{\psi; Z(s)\} = \sum_{m=1}^n \sum_{1 \leq i \leq j \leq k} \omega_{i,j} \left[ \begin{array}{c} \ln f\{\eta_k(s_i), \eta_k(s_j); \psi\} \\ + \ln J\{(Z_k(s_i), Z_k(s_j)) \longrightarrow (\eta_k(s_i), \eta_k(s_j))\} \end{array} \right],$$

où  $J\{(Z_k(s_i), Z_k(s_j)) \longrightarrow (\eta_k(s_i), \eta_k(s_j))\}$  est le Jacobien de la transformation (ponctuelle) du processus  $Z$  vers  $\eta$ , i.e.,

$$\begin{aligned} J\{(Z_k(s_i), Z_k(s_j)) \longrightarrow (\eta_k(s_i), \eta_k(s_j))\} &= \\ \prod_{m \in \{i,j\}} \frac{1}{\sigma(s_m)} \left\{ 1 + \xi(s_m) \frac{Z_k(s_m) - \mu(s_m)}{\sigma(s_m)} \right\}_+^{-1+1/\xi(s_m)}. \end{aligned}$$

Remarque : Parler des surfaces de réponses pour  $s \longrightarrow \{\mu(s), \sigma(s), \xi(s)\}$

### 3.1 Introduction

Les événements naturels tels les canicules, les précipitations, les marées et les tempêtes constituent des vrais dangers lorsqu'ils sont extrêmes. Bien que ce soit difficile d'attribuer un événement particulier, tel l'Ouragan Katrina ou l'inondation de 2010 au Pakistan, aux effets du changement de climat, les données d'observation et les modèles climatiques existants suggèrent que les dimensions de telles catastrophes augmenteront dans le futur. Les conséquences potentielles incluent des augmentations des tempêtes sévères, des inondations, des feux dans la nature, des échecs de la récolte, des déplacements de la population et augmentation de la mortalité. À part leurs impacts directs, de tels événements auront aussi des effets indirects tels que l'augmentation des coûts pour fortifier les infrastructures et l'augmentation des primes d'assurance. Il y a donc un besoin pressant pour une meilleure compréhension des extrêmes spatiaux et une estimation plus détaillée de leurs conséquences. Et sur les dernières années, le sujet est devenu une interface active entre les scientifiques du climat, du social et des statistiques, en interaction avec les actionnaires tels les compagnies d'assurances et les fonctionnaires de la santé.

La théorie des valeurs extrêmes spatiale est un outil pertinent pour décrire de tels événements et prévenir leurs conséquences. Comme exemple d'application, on va utiliser les données « rainfall » incluses dans le package « SpatialExtremes » du logiciel R.

R est un logiciel de traitement statistique des données et un langage de programmation. Il fonctionne sous la forme d'un interpréteur de commandes. Il dispose d'une bibliothèque très large de fonctions statistiques, d'autant plus large qu'il est possible d'en intégrer de nouvelles par le système des "packages", des modules externes compilés (sous forme de DLL sous Windows) que l'on peut télécharger gratuitement sur internet. R propose également une palette étendue de fonctionnalités graphiques. Il est possible d'utili-

ser R en mode interactif sans jamais avoir à programmer.

## 3.2 Qu'est-ce que le package `SpatialExtremes` ?

Le package `SpatialExtremes` est un package complémentaire du système de calcul statistique R [R Development Core Team, 2007][70]. Il fournit des fonctions pour l'analyse des extrêmes spatiaux à l'aide de processus (actuellement) max-stables. Tous les commentaires, critiques et requêtes sur le package ou la documentation associée sont reçus avec gratitude.

Je suppose que vous êtes déjà familier avec le logiciel R et que ce dernier est déjà installé sur votre ordinateur. Pour analyser les extrêmes spatiaux, nous allons utiliser une librairie supplémentaire : le package `SpatialExtremes`. Avant toute chose, il faudra donc l'installer ;`installSpatialExtremes,eval=FALSE`;`install.packages("SpatialExtremes")`

@ mais aussi la charger

```
library(SpatialExtremes)
```

Bien entendu cette librairie est entièrement documentée et il peut être profitable d'aller jeter un oeil à la documentation. Un bon point de départ pourrait être

```
help(SpatialExtremes)
```

## 3.3 Description des données

Les données qui seront traitées dans cette section sont des données qui correspondent aux quantités maximales des précipitations journalières en millimètres récoltées sur 79 sites en Suisse durant les étés des années 1962 à 2008. En fait, les données `rainfall` du package `SpatialExtremes` comportent deux tableaux de données appelés « coord » et « rain ». Le tableau `coord` correspond aux positions géographiques des 79 sites sur lesquels on a relevé les données et il comporte 79 lignes et 3 colonnes, où la première colonne correspond aux longitudes, la deuxième aux latitudes et la troisième aux altitudes en mètres des 79 sites et chaque ligne correspond à un site. Et le tableau `rain` correspond aux données qu'on va traiter dans cette section et il comporte 47 lignes et 79 colonnes, où chaque colonne correspond à un site et chaque ligne à une année. Figure montre les positions de ces sites sur la carte géographique.

### 3.3.1 code R

```
rm(list=ls())
```

```
data(rainfall)##rainfalldataset
```

```
data(swissalt)##elevationdataforSwitzerland
```

```
par(mar=rep(0,4),ps=16)
```

```
image(lon.vec,lat.vec,alt.mat,col=terrain.colors(64),asp=1,bty="n",xlab="n",ylab="n",axes=FALSE)
```

```
swiss(add=TRUE,city=TRUE)
```

```
points(coord,pch=15)
```

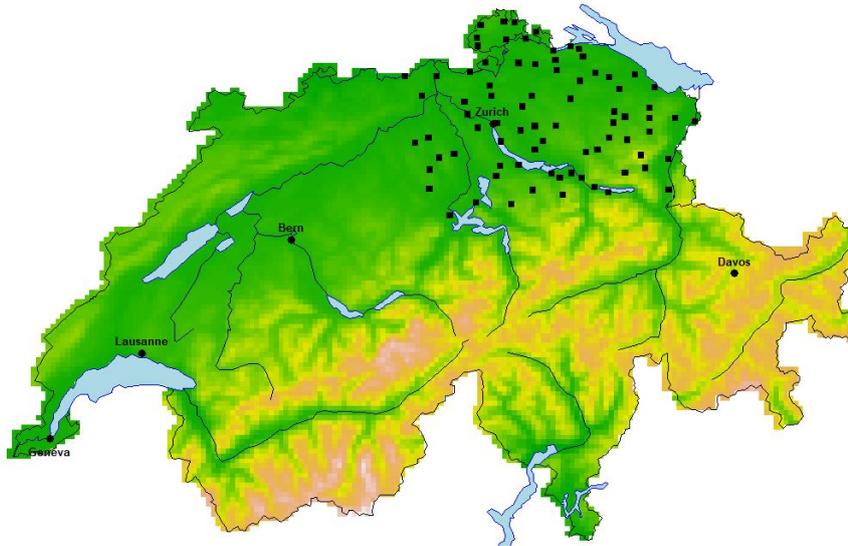


FIGURE 3.1 – Plan de la Suisse avec 79 stations pluviométriques

pour faire une simple carte, en va utiliser **Le Package "maps"**. une fois installés , nous utiliserons la fonction `map()` pour crier notre carte

par exemple la carte d'Algerie sur R

```
library(maps) map('worldHires')# ce code donne une carte du monde map('worldHires',  
"algeria", col='gray', fill=T)
```



FIGURE 3.2 – Plan de l'Algérie

pour obtenir directement les régions de la carte d'Algérie en va utiliser **Le Package "raster"**

```
install.packages("raster") library(raster) adm=getData('GADM',country='DZA', level=2)
plot(adm_al)
```

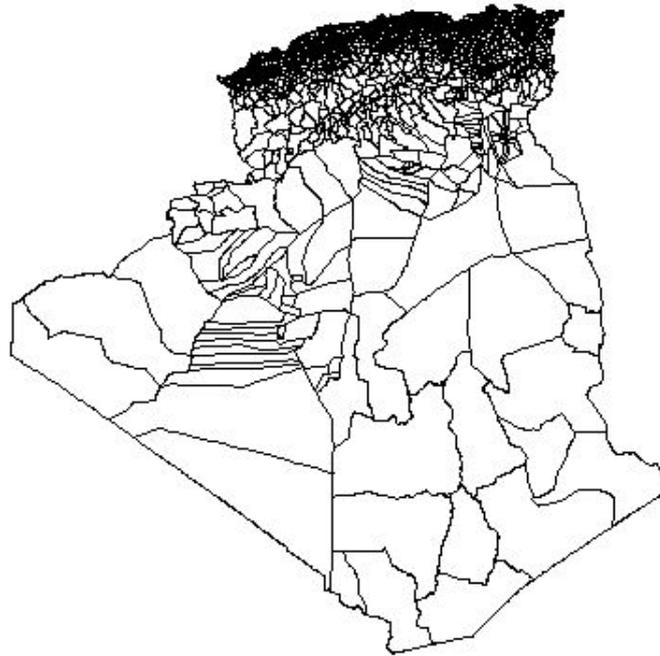


FIGURE 3.3 – Plan de Algerie avec régions

Installez le paquet raster et chargez-le d’abord. Pour pouvoir utiliser la fonction `getData()` afin d’acquérir des données sur l’amd global. limites, nous devons spécifier trois arguments :

1. Sélectionner le jeu de données : le premier argument spécifie le jeu de données. «GADM» renvoie les limites administratives globales.
2. Select country : Le deuxième argument fournit le nom du pays des limites en utilisant son code pays
3. Spécifiez le niveau : le troisième argument spécifie le niveau de la subdivision administrative (0 = pays, 1 = subdivision du premier niveau).

Comme vous pouvez le constater, l’Autriche n’est pas entièrement couverte par cette tuile. Permet de télécharger deux autres tuiles orientales et de les mosaïquer pour obtenir toute l’étendue de l’Autriche. Vous pouvez obtenir plus d’informations sur l’étendue des carreaux [ici](#).

### 3.4 Simulation de max-stables

La but de cette section est de décrire des approches pour simuler des processus max-stable simples. D'après la représentation spectrale, on a

$$\eta(s) = \max_{i \geq 1} \zeta_i Y_i,$$

et c'est donc plutôt mal barré pour la simulation car il faut donc obtenir le maximum ponctuel sur une infinité de fonctions de la forme  $s \rightarrow \zeta_i Y_i(s)$ ...Oublions ce "détail un instant". Hormis ce détail fâcheux, il faut donc savoir :

1. simuler  $Y$  sur  $\chi$ ;
2. générer les points d'un processus de Poisson sur  $(0, \infty)$  d'intensité  $\zeta^{-2} d\zeta$ .

Généralement le premier point ne posera pas de problème et reviendra souvent à simuler un processus Gaussien . Comment faire pour le second point ?

Première idée est une méthode du rejet pour les processus de Poisson, i.e., on génère le nombre de points puis rejet classique, mais ici ce n'est pas possible car  $\Lambda \{(0, \infty)\} = \infty$ .

Une deuxième possibilité consiste à appliquer une transformation sur les points.

**Proposition 3.1** Soit  $\{X_i : i \geq 1\}$  les points d'un processus de Poisson standard, i.e., d'intensité unitaire, sur  $(0, \infty)$ . Alors

$$\{\zeta_i : i \geq 1\} := \left\{ \frac{1}{X_i} : i \geq 1 \right\}$$

sont les points d'un processus de Poisson sur  $(0, \infty)$  d'intensité  $\zeta^{-2} d\zeta$ .

Démonstration.

On se souviendra que si l'on applique une bijection aux point d'un processus de Poisson on reste poissonnien. Soit  $\tilde{\Lambda}$  la mesure d'intensité de ce nouveau processus. On a pour  $0 < a < b < \infty$

$$\tilde{\Lambda} \{(a, b)\} = \Lambda \left\{ \left( \frac{1}{b}, \frac{1}{a} \right) \right\} = \frac{1}{a} - \frac{1}{b} = \int_a^b \zeta^{-2} d\zeta.$$

Rappelons de plus qu'un processus de Poisson standard peut s'écrire ( à un ordonnancement près des points)

$$\left\{ \sum_{j=1}^i E_j : i \geq 1 \right\}, \quad E_1, E_2, \dots \sim^{iid} \exp(1).$$

donc on peut faire les etapes 1 et 2 requises pour la simulation. Reste à traiter le problème du maximum pris sur une infinité de fonctions.

**Remarque 3.1** Remarquons que l'on peut sans aucun problème considérer que les points sont ordonnés par ordre décroissant puisque cela ne changera rien dans

$$\max_{i \geq 1} \zeta_i Y_i.$$

De plus  $\zeta_i \downarrow 0$  p.s. lorsque  $i \rightarrow \infty$ .

alors on a deux situation qui sont possibles :

1. Si le processus  $Y$  est uniformément borné par une constante  $0 < M < 1$ , ie ,  $p\{Y(s) < M\} = 1$  pour tout  $s \in \mathcal{X}$ , alors

$$\zeta_i Y_i \leq \zeta_i M \downarrow 0, \quad i \rightarrow \infty.$$

Ceci implique qu'il existe  $i_0 \in \mathbb{N}$  (aléatoire) tel que

$$\max_{i \geq 1} \zeta_i Y_i = \max_{i=1, \dots, i_0} \zeta_i Y_i.$$

2. Si le processus  $Y$  n'est pas uniformément borné alors on est foutu pour la simulation exacte et on obtiendra uniquement des simulations approchées avec une précision arbitraire en prenant une **pseudo borne uniforme**, i.e., telle que  $p\{\eta(s) < M\} = 1 - \varepsilon$  [Schlather, 2002][66].

---

Simulation (non exacte) d'un processus max-stable simple  $\eta$  à partir de la borne (pseudo) uniforme  $M > 0$ .

1. Poser  $\eta \leftarrow 0, E \leftarrow 0, flag \leftarrow TRUE$ ;
  2. **while** flag = TRUE **do**
  3.  $\left[ \begin{array}{l} \text{Générer } E \leftarrow E + Exp(1) \text{ et } Y \sim \sigma; \\ \text{Poser } \max(\eta, Y/E); \\ \text{if } \min \eta > M/E \text{ then} \\ \quad \lfloor \text{poser } flag \leftarrow FALSE; \end{array} \right.$
  4. Renvoyer  $\eta$ ;
- 

A l'aide du package **SpatialExtremes** (qu'il faudra avant toute chose installer et charger), nous pouvons simuler les processus max-stables. C'est la fonction **rmaxstab** qui s'occupe de cela. Voici comment cela fonctionne

Préparation de la fenêtre graphique *par(mfrow = c(1, 3))* On definit  $X$  comme une grille sur  $[-5, 5]$ . `s <- seq(-5, 5, length=500)`

**Simulation d'un processus de Schlather** avec une fonction de corrélation de type exponentiel puissance, i.e.,

$$\rho(h) = \exp\left(-\left(\frac{h}{2}\right)^{1.5}\right)$$

`eta <- rmaxstab(1, s, "powexp", nugget = 0, range = 2, smooth = 1.5)`

`plot(s, eta, type="l")`

---

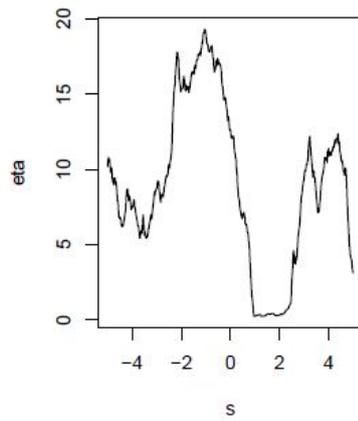


FIGURE 3.4 – Simulation de processus Schlather

**Processus de Smith :** sur une grille régulière de  $[0, 5] \times [2, 12]$

```
s.x<-seq(0,5,length=100) s.y<-seq(2,12,length=100) s<-cbind(s.x,s.y) η<-rmaxstab(1,s,"gauss",cov1
0.5,cov22=2,grid=TRUE)
image(s.x,s.y,η)
```

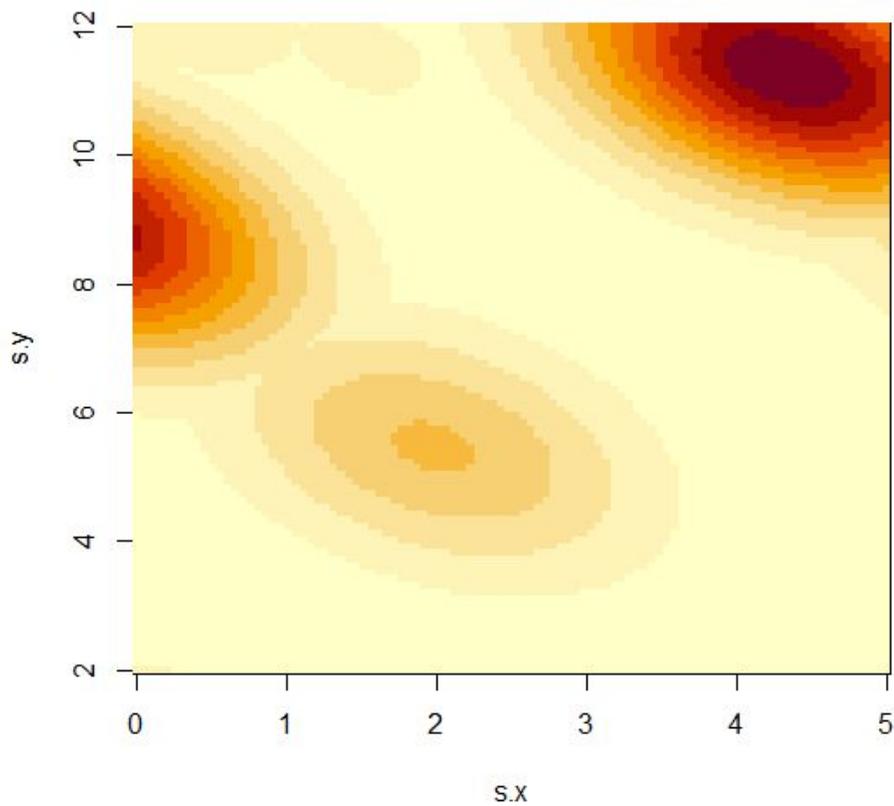


FIGURE 3.5 – Processus de Smith

**Processus de Schlather :**

```
η<- rmaxstab(1, s, "powexp", nugget = 0, range = 2, smooth = 1.5, grid=TRUE)
```

image(s.x, s.y,  $\eta$ )

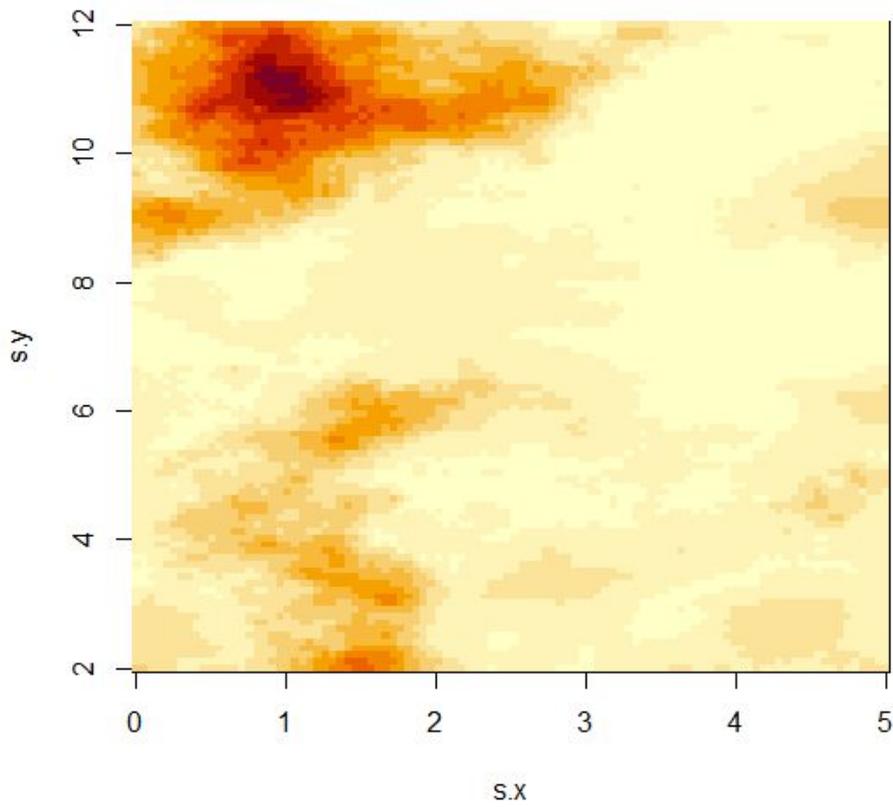


FIGURE 3.6 – Processus de Schlather

### Processus de Extremal-t

```
 $\eta$  <- rmaxstab(1, s, "whitmat", nugget = 0, range = 2, smooth = 1.5, grid=TRUE)
```

image(s.x, s.y,  $\eta$ )

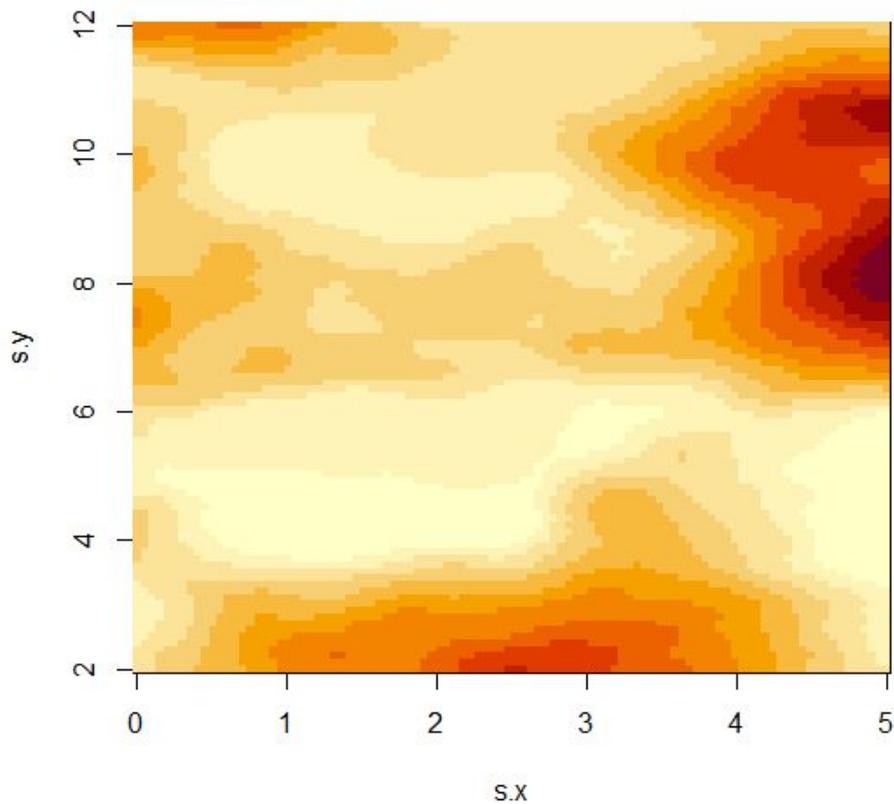


FIGURE 3.7 – Processus de Extremal-t

**Une représentation** 3d de ce même processus (pour le fun mais inutile) et à l'échelle log

```
col<-rev(heat.colors(64)) zfacet<- $\eta[-1,-1]$ + $\eta[-1,-100]$ + $\eta[-100,-1]$ + $\eta[-100,-100]$ 
facetcol<-cut(log(zfacet),64)
```

```
persp(s.x,s.y,log(eta),col=col[facetcol],theta=-45,phi=30,
```

*border = NA*)

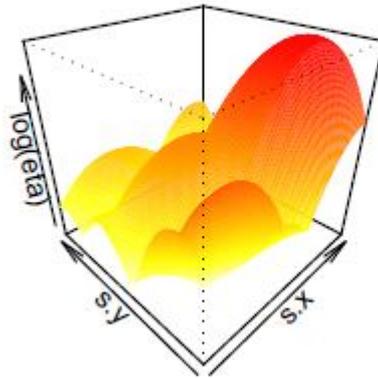


FIGURE 3.8 – Représentation 3D de Smith

## 3.5 Modélisation *GEV* des précipitations des maxima annuels et sélection de modèle pour la prédiction de quantiles élevés

### 3.5.1 Introduction

Les extrêmes climatiques constituent depuis longtemps un sujet d'intérêt, en particulier dans les dernières années.

décennies. La théorie des valeurs extrêmes (*EVT*) se concentre sur la modélisation et l'évaluation statistique des événements extrêmes. L'un des principaux problèmes rencontrés dans l'*EVT* est l'estimation d'un quantile extrême, ou de manière équivalente, d'un niveau de retour associé à une période de retour importante. Le niveau de retour de  $T$  années correspond à la valeur qui sera dépassée en moyenne une fois tous les  $T$  ans.

Ce document consiste en un rapport sur la participation au challenge de donné . La méthodologie générale que nous avons utilisée consiste à prendre des maxima annuels pour chaque mois et année (une observation par mois, mois et année) et pour s'adapter à plusieurs modèles de valeurs extrêmes généralisées (*GEV*) avec des covariables. Dans ces types de modèles, les paramètres *GEV* dépendent de la latitude, longitude et le mois alors que les observations sont considérées comme indépendantes, c'est-à-dire ignorer complètement la dépendance spatiale entre les stations. Pour le challenge data participation, la sélection du modèle a été effectuée via le simple système d'information Akaike critère (*AIC*). Des discussions avec les autres équipes participantes ont suggéré que  $k$  fold la validation croisée devrait mieux fonctionner car elle intègre le critère sur lequel les différentes méthodes sont classées. En conséquence, nous avons depuis réalisé le mo-

dèle sélection basée sur la validation croisée k-fold et il s'avère que le modèle sélectionné fournit meilleures prévisions sur le test de l'échantillon.

L'estimation est réalisée avec le paquet `extRemes` dans *R* (Gilleland et Katz (2016))[29]. Ce paquet a été conçu pour analyser les événements extrêmes et leur change avec le temps. Il se concentre sur la géophysique avec des applications météorologiques et climatiques (Gilleland et Katz (2011))[30]. Il incorpore GEV, Pareto généralisé (GP) et des modèles de processus ponctuels (PP). Les fonctions utilisées sont `fevd` pour estimer le modèle (avec des covariables) et `calculez AIC et return.level` pour estimer les quantiles.

### 3.5.2 Les modèles

Nous comparons plusieurs modèles pour la dépendance des paramètres *GEV* en tant que fonctions des covariables. Pour cela, nous utilisons des splines naturelles (James et al. (2013))[36],

qui offrent un moyen simple et flexible pour la modélisation de surface de réponse. Les splines sont fonctions polynomiales par morceaux, telles que présentées dans la définition suivante.

**Définition 3.1** *Laisser  $\xi_1 < \xi_2 < \dots < \xi_n$  être  $N$  nœuds. Une spline de degré  $m$  est une fonction de la classe  $C^{(m-1)}$  dont la restriction sur chacun des intervalles  $(-\infty, \xi_1), (\xi_1, \xi_2), \dots, (\xi_N, +\infty)$  est une fonction polynomiale de degré au plus  $m$ .*

Une spline naturelle est une spline linéaire sur les intervalles de limites  $(-\infty, \xi_1)$  et  $(\xi_N, +\infty)$ .

Par souci de simplicité, nous supposons le paramètre  $\gamma$  de forme de la constante de distribution de *GEV* (ne dépendant pas des covariables); son estimation est en effet assez complexe. Le paramètre de localisation  $\mu$  et le paramètre d'échelle  $\sigma$  sont exprimés en fonction des covariables : mois, latitude et longitude. Nous définissons trois types de modèles pour les covariables des paramètres de *GEV*.

### 3.5.3 Critère d'information Akaike

Dans cette section, nous présentons la sélection du modèle *AIC* qui a été utilisée pour notre participation au challenge de données dans *EVA2017*.

#### 3.5.3.1 Sélection du modèle

*L**AIC* a été introduite pour la première fois par Akaike (1973). Il mesure la qualité d'un modèle en faisant un compromis entre sa complexité et sa qualité d'ajustement. Il est défini comme

$$AIC = 2K - 2\log(L),$$

où  $K$  est le nombre de paramètres estimés et  $\log(L)$  est la fonction de maximum vraisemblance du modèle.

Nous avons également essayé d'introduire l'effet de l'année dans les modèles, mais nous n'avons obtenu aucune amélioration de leurs performances en ajoutant ce facteur.

pour les sites observés (*groupe C1*) et les sites non observés (*groupe C2\C1*) dont les niveaux de retour estimés

Les performances des prédictions quantiles  $\widehat{q}$  sont évaluées à l'aide de la fonction de score suivante :

$$S_i(\widehat{q}) = \sum_{s \in C_i} \sum_{j=1}^{12} S_{s,j}(\widehat{q}_{s,j}), \quad i = 1, 2, \quad (\text{"Equation de score"})$$

avec  $\widehat{q}_{s,j}$  est le niveau de retour estimé pour la station  $s$  et le mois  $j$

$C_1$  et  $C_2$  sont des ensemble des stations, La fonction  $S_{s,j}(\widehat{q}_{s,j})$  est définie comme

$$S_{s,j}(\widehat{q}_{s,j}) = \sum_{\text{jour } t \text{ de la période de test du mois } j} \ell(P_{s,t}, \widehat{q}_{s,j}),$$

où  $\ell$  est la fonction de perte quantile

$$\ell(x, y) = \alpha(x - y)1_{x > y} + (1 - \alpha)(y - x)1_{y \geq x},$$

tq avec  $x, y \in \mathbb{R}$  et  $P_{s,t}$  est la précipitation dans la station  $s$  et  $t$  le jour de la période  $j$ . Le score final est comparé au score d'un repère prédictif des quantiles par les maxima de l'échantillon d'apprentissage pour les stations de  $C_1$  et par la moyenne des maxima mensuels des stations de l'échantillon d'instruction.

### 3.5.4 Méthode de validation croisée à en $k$ fold

Sur la base de discussions avec les autres équipes participant au défi des données, nous avons envisagé la sélection de modèle sur la base de la méthode de validation croisée des  $k$ -fold. Il est en effet judicieux d'incorporer le score utilisé pour comparer les différentes méthodes dans la procédure de sélection du modèle.

#### 3.5.4.1 Sélection du modèle

La validation croisée est une méthode utilisée pour évaluer la performance prédictive d'un modèle statistique donné. Il consiste à supprimer une partie des données appelée ensemble de test et d'utiliser la partie restante comme un ensemble d'apprentissage. Le modèle est ajusté sur l'ensemble d'apprentissage et l'erreur de prédiction pour le modèle appris est ensuite évaluée sur l'ensemble d'essai (James et al. (2013))[36]. Il existe plusieurs approches pour la validation croisée : l'approche jeu de validation, la validation croisée "laissez-un-un" et la validation croisée " $k$  - fold". La principale différence entre les trois méthodes est la manière de scinder les observations dans les ensembles de test et de formation. Nous considérons ici la validation croisée des  $k$  fold : après divisant l'en-

semble de données en  $k$  fold, nous calculons l'erreur sur la base des observations du pli maintenu (James et al. (2013))[36]. Nous répétons la même procédure  $k$  fois et à chaque fois, un pli joue le rôle d'un ensemble de test et  $k - 1$  plis jouent le rôle d'un ensemble de formation. Enfin, l'erreur globale est la somme des  $k$  erreurs calculés.

### 3.6 Application

Les données qui seront traitées dans cette section sont des données qui correspondent aux quantités maximales des précipitations Mensiel millimètres récoltées sur 4 sites en Algerie durant les étés des années 1975 à 1982. En fait, les données du package Spatia-Extremes comportent deux tableaux de données appelés « coord » et « rain ». Le tableau coord correspond aux positions géographiques des 4 sites sur lesquels on a relevé les données et il comporte 4 lignes et 3 colonnes, où la première colonne correspond aux longitudes, la deuxième aux latitudes et la troisième aux altitudes en mètres des 4 sites et chaque ligne correspond à un site. Et le tableau rain correspond aux données qu'on va traiter dans cette section et il comporte 12 lignes et 4 colonnes, où chaque colonne correspond à un site et chaque ligne à un mois.

TABLE 3.1 – Tablau des coordonnées

coord	x	y	z
FOUKA	504,1	374,4	25
AMEUR	488,1	352,75	120
FER.A	509.95	373.2	10
BLIDA	509,8	352,75	210

TABLE 3.2 – Tablau des données

rain	FOUKA	AMEUR	FER.A	BLIDA
sep	7,38	9,66	6,87	12,1
oct	29,2	12,6	23,22	25,51
ven	31,05	28,12	26,62	32,08
dec	30,52	23,47	24,88	32,08
janv	28,02	24,52	19,18	27,85
fer	23,06	20,2	19,1	23,2
mars	17,38	28,86	24,16	36,35
avril	17,3	19,42	17,5	31,86
mai	18,67	14	13,22	20,15
juin	2,62	4,25	1,41	5,43
juillet	1,55	0,12	2,43	1,9
aout	6,62	1,12	8,58	4,26

### 3.7 Transformation des données

Notre but dans cette analyse est de poser un modèle des processus max-stables approprié aux données qu'on a. Afin d'atteindre cet objectif, on suppose que les quantités maximales des précipitations mensiel de chaque site durant les 8 années sont des réalisations d'une certaine variable aléatoire suivant une *GEV*. Après avoir posé la *GEV* appropriée à chaque site, et cela en estimant ses paramètres, on doit transformer, sans perte de généralisation, les données de chaque site en des données issues d'une variable aléatoire suivant une distribution des valeurs extrêmes de Fréchet standard. Bien sûr, on peut envisager d'autres distributions des valeurs extrêmes comme celle de Gumbel ou celle de Weibull.

L'expression de la fonction de répartition d'une *GEV* et l'expression de sa fonction de densité sont connues voir CHAPITRE 1, Sous l'hypothèse d'indépendance des données de chaque site d'observation, on peut construire la fonction de vraisemblance des quantités maximales des précipitations mensiel de chacun des 4 sites. Ensuite on estimera les paramètres de chaque *GEV* qui correspond à chaque site en maximisant le fonction de vraisemblance qui correspond à ce site. Ainsi, on aura  $3 \times 4 = 12$  paramètres à estimer. Pour cela, on utilisera l'instruction "gev.fit" du package « ismev » de logiciel R.

Après avoir posé la *GEV* appropriée à chaque site d'observation, on transformera les données de chacun de ces sites en des données issues d'une variable aléatoire suivant une distribution des valeurs extrêmes de Fréchet standard. Et cela en utilisant la transformation :

$$y \rightarrow z = \left[ 1 + \xi \left( \frac{y - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}}$$

Où les paramètres  $\xi \in \mathbb{R}$ ,  $\mu \in \mathbb{R}$  et  $\sigma \in \mathbb{R}_+^*$  sont respectivement le paramètre de forme, de localisation et d'échelle de la *GEV*. Pour cela, on utilisera l'instruction « gev2frech » du package SpatialExtremes.

Ainsi, on construira un nouveau tableau de données composé de 12 lignes et 4 colonnes, où chacune de ses composantes correspond à une réalisation d'une variable aléatoire suivant une distribution des valeurs extrêmes de Fréchet standard.

#### 3.7.1 Les données

Nous sauvegardons les données dans un fichier Excel sous type .CSV (séparateur point-virgule). Ensuite, dans logiciel Rstudio et en utilisant le package "Rcmdr", nous utilisons le code suivant : `data=read.csv2(file="EEE.CSV") data`

	xxx	FOUKA	AMEUR	FER.A	BLIDA
1	sept	7.38	9.66	6.87	12.10
2	oct	29.20	12.60	23.22	25.51
3	nev	31.05	28.12	26.62	32.08
4	dec	30.52	23.47	24.88	32.08
5	janv	28.02	24.52	19.18	27.85
6	fev	23.06	20.20	19.10	23.20
7	mars	17.38	28.86	24.16	36.35
8	avril	17.30	19.42	17.50	31.86
9	mai	18.67	14.00	13.22	20.15
10	juin	2.62	4.25	1.41	5.43
11	juillet	1.55	0.12	2.43	1.90
12	aout	6.62	1.12	8.58	4.26

FIGURE 3.9 – Représentation des donn brutesdes pricipitations

data\$FOUKA data\$AMEUR data\$FER.A data\$BLIDA

```
> data$FOUKA
[1] 7.38 29.20 31.05 30.52 28.02 23.06 17.38 17.30 18.67 2.62 1.55 6.62
> data$AMEUR
[1] 9.66 12.60 28.12 23.47 24.52 20.20 28.86 19.42 14.00 4.25 0.12 1.12
> data$FER.A
[1] 6.87 23.22 26.62 24.88 19.18 19.10 24.16 17.50 13.22 1.41 2.43 8.58
> data$BLIDA
[1] 12.10 25.51 32.08 32.08 27.85 23.20 36.35 31.86 20.15 5.43 1.90 4.26
```

FIGURE 3.10 – Données par stations

## 3.8 Analyses des données

### 3.8.1 boxplot

```
par(mfrow = c(1,1)) boxplot(data$FOUKA,data$AMEUR,data$FER.A,data$BLIDA,names=c("FOUKA", "AMEUR", "FER.A", "BLIDA"))
```

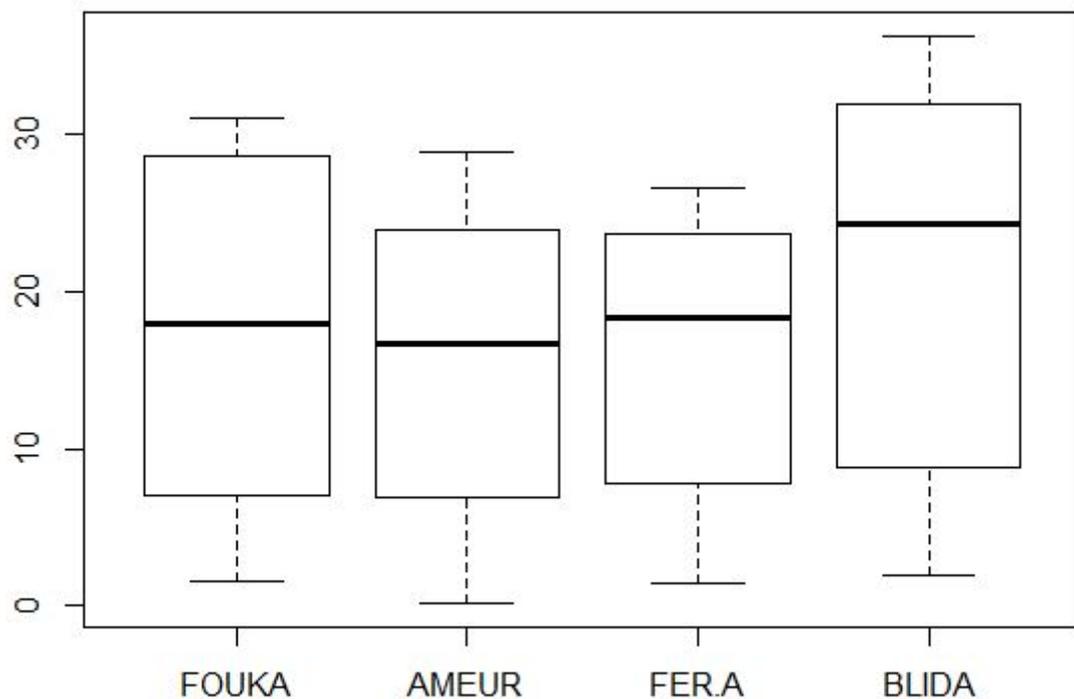


FIGURE 3.11 – Résumé des données par les boites à moustaches

### 3.8.2 Résumé descriptif de la série pluviométrique des quatre stations

TABLE 3.3 – Tableau de summary

	<i>Min</i>	<i>Ist ou</i>	<i>Median</i>	<i>mean</i>	<i>3rd ou</i>	<i>Max</i>
<i>FOUKA</i>	1.55	7.19	18.02	17.78	28.32	31.05
<i>AMOURE</i>	0.120	8.307	16.710	15.528	23.733	28.860
<i>FER.A</i>	1.410	8.152	18.300	15.598	23.455	26.620
<i>BLIDA</i>	1.90	10.43	24.36	21.06	31.91	36.35

### 3.8.3 Test d'homogénéité

TABLE 3.4 – Test d'homogénéité

P-value	chisq.test	wilxon.test
FOUKA	1.53e-11	0.0004883
AMOURE	3.167e-11	0.0004883
FER.A	5.071e-08	0.0004883
BLIDA	3.944e-12	0.002516

tous les valeur de P-value inférieurs à 0.01, indiquent que les données des 4 station sont homogènes

### 3.9 Ajustement de loi GEV

Pour transformer les données du tableau data, qu'on a appelé volontairement «y» pour faciliter les manipulations, on suit les étapes suivantes :

1. On estime le paramètre de localisation, le paramètre d'échelle et le paramètre de forme de la GEV de chaque site en utilisant le code R suivant :

```
insttal.packages("ismev") library(ismev) gev.fit(data$FOUKA) gev.fit(data$BLIDA)
```

TABLE 3.5 – Estimateurs MLE des paramètre de GEV estimées

	$\mu$	$\sigma$	$\xi$
FOUKA	15.52	19.06	-1.22
AMOUR	16.88	12.61	-1.05
FER.A	14.87	12.31	-1.04
BLIDA	20.40	14.15	-0.87

```
library(evd)
```

```
data<- rgev(100, loc = 0.13, scale = 1.1, shape = 0.2)
```

```
M1<- fgev(data) par(mfrow = c(2,2)) plot(M1)
```

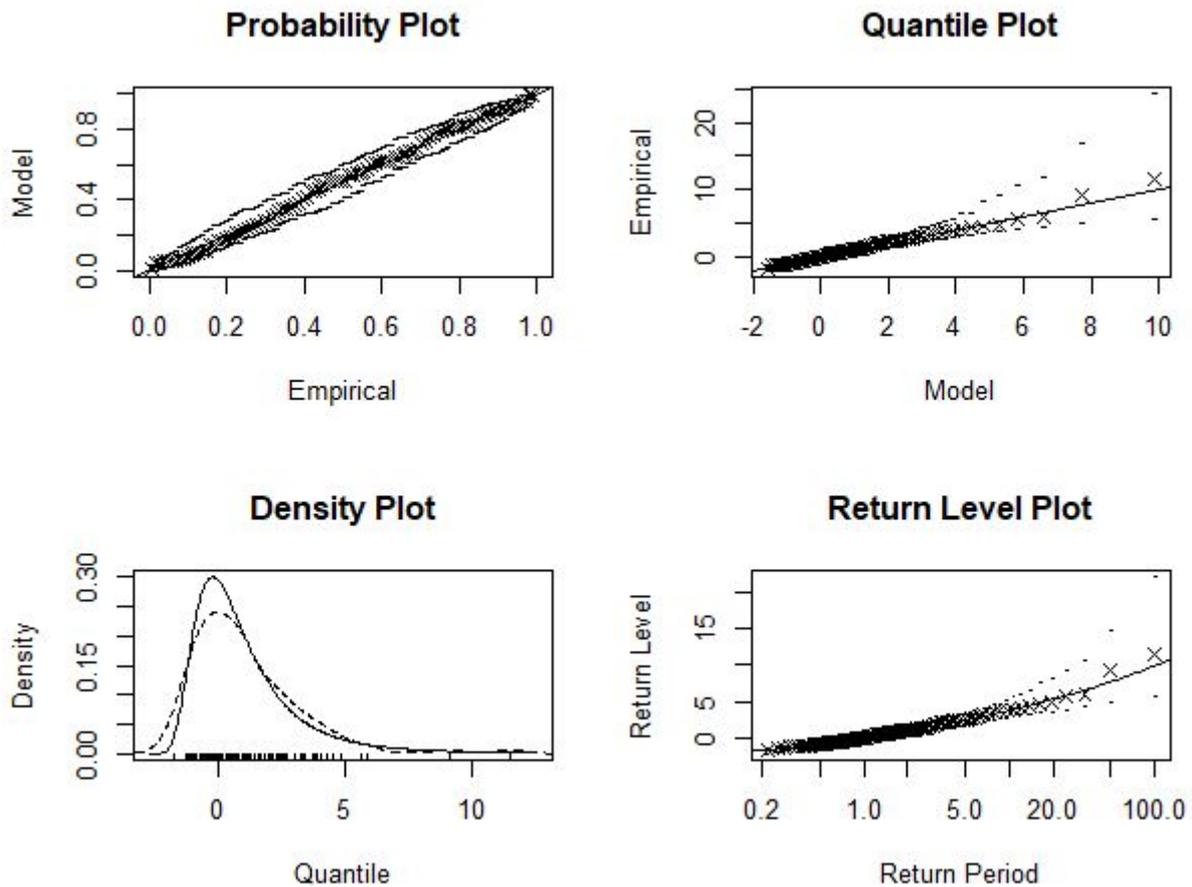


FIGURE 3.12 – Représentation des distributions GEV

On construira une matrice « Z » qui contient tous les paramètres estimés. C'est-à-dire que la matrice Z comportera 3 lignes et 4 colonnes, où chaque colonne contiendra les paramètres estimés de la GEV d'un site. La première ligne comportera les paramètres de localisation, la deuxième comportera les paramètres d'échelle et la troisième contiendra les paramètres de forme.

TABLE 3.6 – Valeurs des paramètres estimés de la distribution GEV

	$\mu$	$\sigma$	$\xi$
FOUKA	15.52	19.06	-1.22
AMOUR	16.88	12.61	-1.05
FER.A	14.87	12.31	-1.04
BLIDA	20.40	14.15	-0.87

Une fois que la matrice z est construite, on construira une autre matrice « p » composée de 12 lignes et 4 colonnes et qui contiendra les données transformées. Le Code R utilisé est le suivant :

```
p<-matrix(rep(0,48), nrow = 12, ncol = 4) for (i in 1 :4) p[,i] gev2frech(y[,i], Z[1,i], Z[2,i], Z[3,i], emp = FALSE)
```

Ainsi, les composantes de la matrice  $p$  représenteront des réalisations d'une variable aléatoire suivant une loi de probabilité de Fréchet standard

```
> for (i in 1:4) p[,i] <- gev2frech(y[,i], Z[1,i], Z[2,i], Z[3,i], emp = FALSE)
> p
      [,1]      [,2]      [,3]      [,4]
[1,] 0.7090994 0.6386937 0.6086722 0.6225500
[2,] 5.5214127 0.7480340 3.2390206 1.5426483
[3,] 66.7096090 13.6924051 113.2038245 4.2868208
[4,] 14.0273244 2.1335813 6.0308797 4.2868208
[5,] 3.7420798 2.6192838 1.5454920 2.0220879
[6,] 1.7162657 1.3609367 1.5298596 1.2425439
[7,] 1.1094920 305.5613846 4.3813714 93.3005134
[8,] 1.1042336 1.2539594 1.2733011 4.0619700
[9,] 1.2027044 0.8148740 0.8820720 0.9826194
[10,] 0.6105370 0.5043775 0.4817810 0.4723436
[11,] 0.5923826 0.4351712 0.5012293 0.4176456
[12,] 0.6910354 0.4500831 0.6637869 0.4527941
```

FIGURE 3.13 –

### 3.10 Modèle des données

Parmi les modèles des max-stables on a choisi le modèle de Schlather pour modéliser les quantités maximales des précipitations mensiel. On a préféré le modèle de Schlather car ce modèle est flexible et a une interprétation simple voir Ribatet (2009)[62]. Bien sûr, si cela paraît plus raisonnable, le modèle de Schlather pourrait être remplacé par un autre modèle tel le modèle de Smith ou celui de Brown-Resnick, et les mêmes techniques statistiques seront valables pour ce choix alternatif.

La loi de probabilité marginale bivariée du modèle de Schlather est donnée par sa fonction de répartition. Et la fonction de densité marginale bivariée de ce modèle de la thèse Gholam Rezaee (2010)[58].

### 3.11 Coefficient extrême

Lorsque il s'agit des processus max-stables, c'est-à-dire que les réalisations sont extrêmes, Alors les statisticiens ont proposé un outil pour mesurer la dépendance des extrêmes qui est le coefficient extrême. Différentes approches pour estimer le coefficient extrême. Sous l'hypothèse que les données transformées sont des réalisations d'un modèle de Schlather, on estimera le coefficient extrême par l'approche F-madogramme. Pour cela, on utilisera l'instruction « fmadogram » du package SpatialExtremes.

1. On construit une matrice «  $M$  » qui contiendra seulement les longitudes et les latitudes des 4 sites. C'est-à-dire que  $M$  comportera 4 lignes et 2 colonnes, où la première colonne contiendra les longitudes et la deuxième colonne contiendra les latitudes chaque ligne correspondra à un site. Le code R qu'on utilisera est le suivant :

**Code R**

Pour estimer les valeurs du coefficient extrémal associé aux données de la matrice  $p$ , on suit les étapes suivantes :

```
h=c(374.40,352.75,373.20,352.75) i=c(504.10,488.10,509.95,509.80) M<-matrix(c(i,h),nrow=4,ncol=4)
```

TABLE 3.7 – Tableau des coordonnées

coord	x	y	z
FOUKA	504,1	374,4	25
AMEUR	488,1	352,75	120
FER.A	509.95	373.2	10
BLIDA	509,8	352,75	210

2. Une fois la matrice  $M$  est construite, on calculera les valeurs du F-madogramme et celles du coefficient extrémal associé aux données en utilisant le code R suivant :

```
fmadogram(p, M)
```

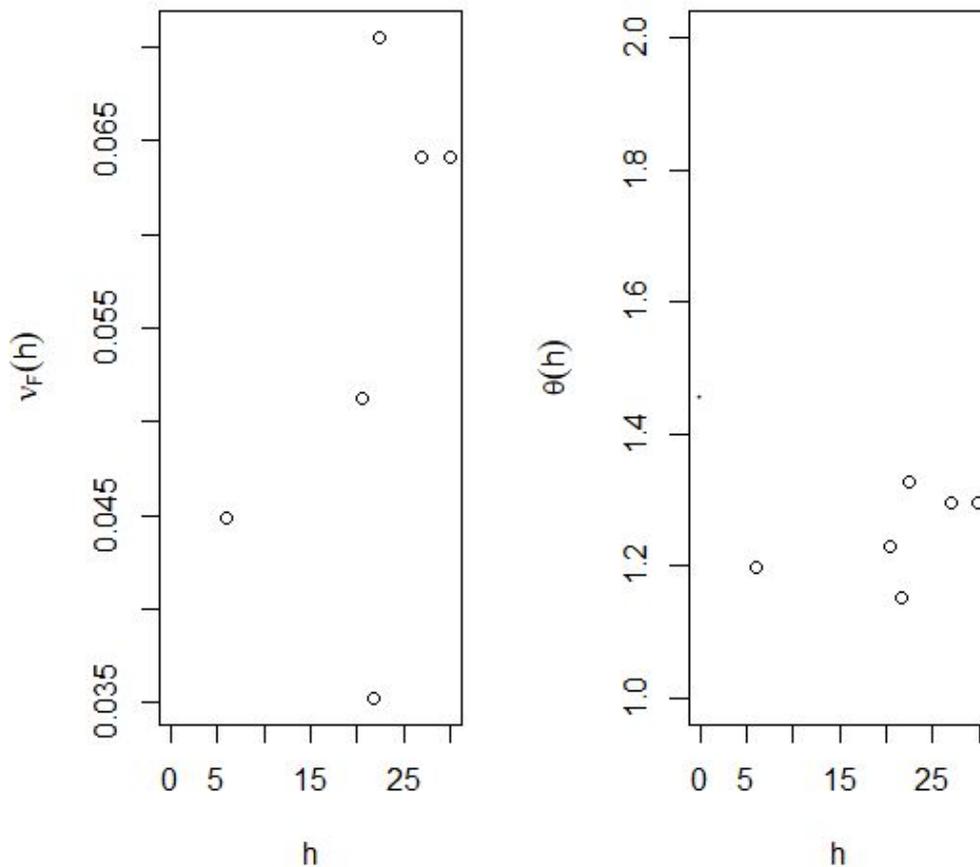


FIGURE 3.14 – Valeurs estimées du coefficient Extrémal par l’approche F-madogramme en fonction de la distance euclidienne ( $h$  représente la distance euclidienne entre deux sites d’observation).

### 3.12 Estimation des paramètres des modèles

Après la transformation des données en des données issues d'une variable aléatoire ayant une distribution des valeurs extrêmes de Fréchet Standard et avoir choisi le modèle de Schlather pour modéliser les quantités maximales des précipitations mensiel, on doit estimer les paramètres de ce modèle pour atteindre l'objectif de cette analyse qui est de trouver le modèle approprié aux données qu'on a. Pour cela, on va utiliser «la méthode de vraisemblance par paires»

On a choisi de travailler avec la famille des corrélogrammes de Whittle-Mathérn car elle engendre beaucoup d'autres familles de corrélogrammes. Bien sûr, on peut travailler avec d'autres familles telle la famille stable ou la famille de Cauchy et les techniques statistiques resteront les mêmes.

Donc, notre objectif est d'estimer les paramètres  $0 \leq m_1 \leq 1$ ,  $m_2 > 0$  et  $\rho > 0$  en employant la méthode de vraisemblance par paires. Pour cela, on utilise l'instruction « fitmaxstab » du package SpatialExtremes

**Code R**

Pour estimer les paramètres du modèle de Schlather approprié à nos données et comparer son coefficient extrémal avec le coefficient extrémal estimé, on suit les étapes suivantes :

On estime les paramètres du corrélogramme du modèle de Schlather par la méthode de vraisemblance par paires. Le code R utilisé est le suivant :

**Code R**

```
Schlather <-fitmaxstab(p, M, "whitmat")
```

TABLE 3.8 – Les paramètres estimés de modèle de schlather

	nugget	range	moth
dependence parametres	0.1334	2.5837	54.1107
standard Errors	0.127	460.261	19010.930

TABLE 3.9 – Tableau de varcov de modèle de Schlather

vercov	nugget	range	moth
nugget	1.614e-02	-4.532e+01	1.870e+03
range	-4.532e+01	2.118e+05	-8.750e+06
moth	1.870e+03	-8.750e+06	3.614e+08

```
smith <-fitmaxstab(p, m, "cauchy")
```

TABLE 3.10 – Tableau du paramètres du modèle de Smith

	nugget	range	moth
dependence parametres	1.533e-06	3.891e+00	1.451e-01
standard Errors	3.6906	84.7894	0.6851

TABLE 3.11 – Tableau de varcovar de modèle de Smith

varcov	nugget	range	moth
nugget	13.6203	312.7227	2.4750
range	312.7227	7189.2353	57.1397
moth	2.4750	57.1397	0.4693

brown<-fitmaxstab(p, m,"brown")

TABLE 3.12 – Tableau estimés de modèle de Brown

	range	moth
dependence parametres	16.396	1.536
standard Errors	3.7947	0.1517

TABLE 3.13 – Tableau de varcovar de modèle de Brown

varcov	range	moth
range	14.39942	0.25334
moth	0.25334	0.02301

extremalt <- fitmaxstab(p, m, "powexp")

TABLE 3.14 – Tableau estimés de modèle de Extrémal-t

	nugget	range	moth
dependence parametres	2.192e-06	1.791e+02	4.159e-01
standard Errors	5.724	611.305	7.090

TABLE 3.15 – Tableau de varcovar de modèle de Extrémal-t

varcov	nugget	range	moth
nugget	32.77	-2338.58	40.53
range	-2338.58	373693.43	-3016.44
moth	40.53	-3016.44	50.26

### 3.12.1 AIC

TABLE 3.16 – Le AIC des trois modèle

schlater	smith	brown	extremalt
617.2358	617.3291	618.9964	617.8073

On compare le coefficient extrémal du modèle estimé avec le coefficient extrémal estimé précédemment par l'approche F-madogramm Le code R utilisé est le suivant :

fmadogram(fitted = Schlather, which = "ext")

fmadogram(fitted = Smith, which = "ext")

fmadogram(fitted = brown, which = "ext")

fmadogram(fitted = extremalt, which = "ext")

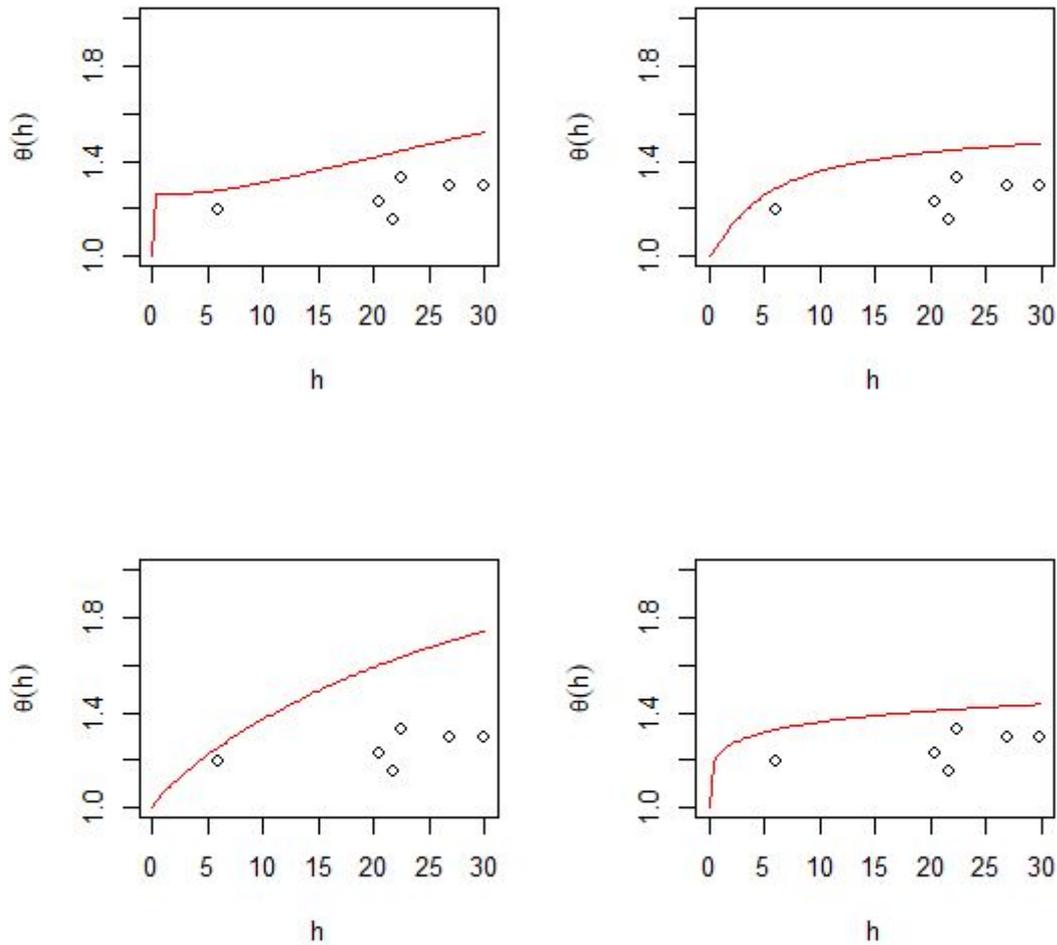


FIGURE 3.15 – Les points représentent les valeurs estimées du coefficient extrême par l’approche F-madogramme et la ligne rouge représente le graphe du coefficient extrême du modèle (schlather,smith,brown,extrême-t) approprié aux données ( $h$  représente la distance euclidienne entre deux sites d’observation)

### 3.13 Conclusion

Les processus max-stables sont une généralisation des distributions des valeurs extrêmes et, par conséquent, sont des candidats idéaux pour la modélisation des extrêmes spatiaux. Dans ce chapitre, qui est le cœur du sujet de ce mémoire, on a essayé de donner une introduction à la théorie des valeurs extrêmes spatiales, où les processus max-stables occupent le trône de cette théorie. Le but principal de cette étude est de présenter quelques-uns des modèles, les plus utilisés en applications, des processus max-stables et leurs propriétés, ainsi que les méthodes statistiques qui permettent d’estimer leurs paramètres pour terminer avec un exemple d’application où on a illustré la manière dont on peut mettre en application tous les résultats vus dans cette étude. Comme on peut le re-

marquer, l'approche statistique qui nous a permis de mettre en application la théorie des valeurs extrêmes spatiales dans ce mémoire est «l'approche maximums par blocs », alors, avant de terminer ce travail, on signale qu'on peut envisager une autre approche qui est «l'approche excès au-delà d'un seuil». Cependant, on veut aussi signaler qu'il y a beaucoup de questions non traitées dans ce travail, surtout ce qui concerne l'aspect statistique de la théorie des valeurs extrêmes spatiale, comme l'approche bayésiennes, l'approche par copules, les outils de diagnostics, la prédiction, etc. qui sont des sujets d'actualité en plein essor.

## BIBLIOGRAPHIE

- [1] Bacro, Jean-Nol. Modélisations stochastiques et valeurs extrêmes. *La Houille Blanche* 5 (2006) : 69-73.
- [2] Beirlant, Jan, Petra Vynckier, and Jozef L. Teugels. Tail index estimation, pareto quantile plots regression diagnostics. *Journal of the American statistical Association* 91.436 (1996) : 1659-1667.
- [3] Buishand, T. A. "Bivariate extreme-value data and the station-year method." *Journal of Hydrology* 69.1-4 (1984) : 77-95.
- [4] Brown, Bruce M., and Sidney I. Resnick. "Extreme values of independent stochastic processes." *Journal of Applied Probability* 14.4 (1977) : 732-739.
- [5] Basrak, Bojan. "Fisher-tippett theorem." *International Encyclopedia of Statistical Science* (2011) : 525-526.
- [6] Chandler, Richard E., and Steven Bate. "Inference for clustered data using the independence loglikelihood." *Biometrika* 94.1 (2007) : 167-183.
- [7] Coles, Stuart, et al. *An introduction to statistical modeling of extreme values*. Vol. 208. London : Springer, 2001.
- [8] Coles, Stuart G., and Jonathan A. Tawn. "Modelling extremes of the areal rainfall process." *Journal of the Royal Statistical Society : Series B (Methodological)* 58.2 (1996) : 329-347.
- [9] Coles, Stuart, and Luis Pericchi. "Anticipating catastrophes through extreme value modelling." *Journal of the Royal Statistical Society : Series C (Applied Statistics)* 52.4 (2003) : 405-416.
- [10] Cooley, Dan, Philippe Naveau, and Paul Poncet. "Variograms for spatial max-stable random fields." *Dependence in probability and statistics*. Springer, New York, NY, 2006. 373-390.
- [11] Csorgo, Sandor, Paul Deheuvels, and David Mason. "Kernel estimates of the tail index of a distribution." *The Annals of Statistics* (1985) : 1050-1077.

- [12] Danielsson, Jon, et al. "Using a bootstrap method to choose the sample fraction in tail index estimation." *Journal of Multivariate analysis* 76.2 (2001) : 226-248.
- [13] Davison, Anthony C., Simone A. Padoan, and Mathieu Ribatet. "Statistical modeling of spatial extremes." *Statistical science* 27.2 (2012) : 161-186.
- [14] Davis, Richard, and Sidney Resnick. "Tail estimates motivated by extreme value theory." *The Annals of Statistics* 12.4 (1984) : 1467-1487.
- [15] De Haan, Laurens. "A spectral representation for max-stable processes." *The annals of probability* 12.4 (1984) : 1194-1204.
- [16] Haan, Laurentius Franciscus Maria. *Extremes in higher dimensions : the model and some statistics*. Erasmus University, 1985.
- [17] De Haan, L., and A. Ferreira. "Extreme Value Theory : An Introduction Springer Science+ Business Media." LLC, New York (2006).
- [18] Deheuvels, Paul, Erich Haeusler, and David M. Mason. "Almost sure convergence of the Hill estimator." *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 104. No. 2. Cambridge University Press, 1988.
- [19] Dekkers, Arnold LM, and Laurens De Haan. "On the estimation of the extreme-value index and large quantile estimation." *The Annals of Statistics* 17.4 (1989) : 1795-1832.
- [20] Delmas, Jean-François, and Benjamin Jourdain. "Lois de valeurs extrêmes." *Modèles aléatoires : Applications aux sciences de l'ingénieur et du vivant* (2006) : 303-341.
- [21] Dekkers, Arnold LM, and Laurens De Haan. "On the estimation of the extreme-value index and large quantile estimation." *The Annals of Statistics* 17.4 (1989) : 1795-1832.
- [22] Davison, Anthony C., and Richard L. Smith. "Models for exceedances over high thresholds." *Journal of the Royal Statistical Society : Series B (Methodological)* 52.3 (1990) : 393-425.
- [23] Dombry, Clément, Sebastian Engelke, and Marco Oesting. "Exact simulation of max-stable processes." *Biometrika* 103.2 (2016) : 303-317.
- [24] Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. "Modelling extremal events, volume 33 of *Applications of Mathematics*." (1997).
- [25] Efron, Bradley. *The jackknife, the bootstrap, and other resampling plans*. Vol. 38. Siam, 1982.
- [26] Fisher, Ronald Aylmer, and Leonard Henry Caleb Tippett. "Limiting forms of the frequency distribution of the largest or smallest member of a sample." *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 24. No. 2. Cambridge University Press, 1928.
- [27] Feller, William. "The hypergeometric series." *An introduction to probability theory and its applications* 1 (1968) : 41-45.

- [28] Gnedenko, B. "Sur La Distribution Limite Du Terme Maximum D'Une Série Aléatoire The Annals of Mathematics 44 (3), 423–453. Translated and reprinted in : Breakthroughs in Statistics, Vol. I, 1992, eds. S. Kotz and NL Johnson." (1943) : 195-225.
- [29] Gilleland, Eric, and Richard W. Katz. "extRemes 2.0 : an extreme value analysis package in R." *Journal of Statistical Software* 72.8 (2016) : 1-39.
- [30] Gilleland, Eric, and Richard W. Katz. "New software to analyze how extremes change over time." *Eos, Transactions American Geophysical Union* 92.2 (2011) : 13-14.
- [31] Kortemme, Tanja, Marina Ramírez-Alvarado, and Luis Serrano. "Design of a 20-amino acid, three-stranded  $\beta$ -sheet protein." *Science* 281.5374 (1998) : 253-256.
- [32] Haeusler, E., and Jozef L. Teugels. "On asymptotic normality of Hill's estimator for the exponent of regular variation." *The Annals of Statistics* 13.2 (1985) : 743-756.
- [33] Hall, Peter. "Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems." *Journal of multivariate analysis* 32.2 (1990) : 177-203.
- [34] Hill, Bruce M. "A simple general approach to inference about the tail of a distribution." *The annals of statistics* (1975) : 1163-1174.
- [35] Hosking, Jonathan Richard Morley, James R. Wallis, and Eric F. Wood. "Estimation of the generalized extreme-value distribution by the method of probability-weighted moments." *Technometrics* 27.3 (1985) : 251-261.
- [36] James, Gareth, et al. *An introduction to statistical learning*. Vol. 112. New York : springer, 2013.
- [37] Kabluchko, Zakhar, Martin Schlather, and Laurens De Haan. "Stationary max-stable fields associated to negative definite functions." *The Annals of Probability* 37.5 (2009) : 2042-2065.
- [38] Leadbetter, M. Ross. "Extremes and local dependence in stationary sequences." *Probability Theory and Related Fields* 65.2 (1983) : 291-306.
- [39] Marimoutou, Velayoudoum, Bechir Raggad, and Abdelwahed Trabelsi. "Extreme value theory and value at risk : application to oil market." *Energy Economics* 31.4 (2009) : 519-530.
- [40] Matthys, Gunther, and Jan Beirlant. "Adaptive threshold selection in tail index estimation." *Extremes and Integrated Risk Management* (2000) : 37-49.
- [41] McNeil, Alexander J., and Rüdiger Frey. "Estimation of tail-related risk measures for heteroscedastic financial time series : an extreme value approach." *Journal of empirical finance* 7.3-4 (2000) : 271-300.
- [42] Mason, David M. "Laws of large numbers for sums of extreme values." *The Annals of Probability* 10.3 (1982) : 754-764.

- [43] Mises, R. von. "La distribution de la plus grande de  $n$  valeurs." *Rev. Math. Union Interbalcanique* 1 (1936) : 141-160.
- [44] Nikoloulopoulos, Aristidis K., Harry Joe, and Haijun Li. "Extreme value properties of multivariate  $t$  copulas." *Extremes* 12.2 (2009) : 129-148.
- [45] Opitz, Thomas. "Extremal  $t$  processes : Elliptical domain of attraction and a spectral representation." *Journal of Multivariate Analysis* 122 (2013) : 409-413.
- [46] Pickands III, James. "Statistical inference using extreme order statistics." *the Annals of Statistics* 3.1 (1975) : 119-131.
- [47] Pickands, James. "Multivariate extreme value distribution." *Proceedings 43th, Session of International Statistical Institution, 1981* (1981).
- [48] Padoan, Simone A., Mathieu Ribatet, and Scott A. Sisson. "Likelihood-based inference for max-stable processes." *Journal of the American Statistical Association* 105.489 (2010) : 263-277.
- [49] Penrose, Mathew D. "Semi-min-stable processes." *The Annals of Probability* (1992) : 1450-1463.
- [50] Pickands III, James. "Statistical inference using extreme order statistics." *the Annals of Statistics* 3.1 (1975) : 119-131.
- [51] Padoan, Simone A., Mathieu Ribatet, and Scott A. Sisson. "Likelihood-based inference for max-stable processes." *Journal of the American Statistical Association* 105.489 (2010) : 263-277.
- [52] Lindsay, B "Statistical Inference from Stochastic Processes. American Mathematical Society, Providence" *Composite likelihood methods.* (1988).
- [53] Reiss, Rolf-Dieter, Michael Thomas, and R. D. Reiss. *Statistical analysis of extreme values. Vol. 2.* Basel : Birkhäuser, 2007.
- [54] Resnick, Sidney I. *Extreme values, regular variation and point processes.* No. 04; QA274. 42, R4.. 1987.
- [55] Raggad, Bechir. *Gestion des risques : théorie et application au marché pétrolier.* Diss. Aix-Marseille 2, 2007.
- [56] Reiss, Rolf-Dieter, and Michael Thomas. "Statistical analysis of extreme values with applications to insurance." *Finance, Hydrology and Other Fields* (2007) : 49-50.
- [57] Resnick, Sidney, and Cătălin Stărică. "Smoothing the Hill estimator." *Advances in Applied Probability* 29.1 (1997) : 271-293.
- [58] Rezaee, Mohammad Mehdi Gholam. *Geostatistics of Extremes : A Composite Likelihood Approach.* Diss. Verlag nicht ermittelbar, 2010.
- [59] Ribatet, Mathieu, and Mohammed Sedki. "Extreme value copulas and max-stable processes." *Journal de la Société Française de Statistique* 154.1 (2013) : 138-150.

- [60] Ribatet, Mathieu, Clément Dombry, and Marco Oesting. "Spatial extremes and max-stable processes." *Extreme value modeling and risk analysis, methods and applications*. CRC Press, Boca Raton (2016) : 179-194.
- [61] Ribatet, Mathieu. "Spatial extremes : Max-stable processes at work." *Journal de la Société Française de Statistique* 154.2 (2013) : 156-177.
- [62] Ribatet, Mathieu. "A user's guide to the SpatialExtremes package." EPFL, Lausanne, Switzerland (2009).
- [63] Smith, Richard L. "Estimating tails of probability distributions." *The annals of Statistics* 15.3 (1987) : 1174-1207.
- [64] Smith, Richard L. "Statistics of extremes, with applications in environment, insurance, and finance." *Extreme values in finance, telecommunications, and the environment*. Chapman and Hall/CRC, 2003. 20-97.
- [65] Smith, Richard L. "Max-stable processes and spatial extremes." Unpublished manuscript 205 (1990).
- [66] Schlather, Martin. "Models for stationary max-stable random fields." *Extremes* 5.1 (2002) : 33-44.
- [67] Schlather, Martin, and Jonathan A. Tawn. "A dependence measure for multivariate and spatial extreme values : Properties and inference." *Biometrika* 90.1 (2003) : 139-156.
- [68] Tawn, J. A. "Discussion of paper by AC Davison and RL Smith." *Journal of the Royal Statistical Society. Series B* 52.3 (1990) : 393-442.
- [69] Tawn, J.A. (1990c), Estimating probabilities of extreme sea levels. Submitted.
- [70] Team, Core Writing, Rajendra K. Pachauri, and Andy Reisinger. "Climate change 2007 : synthesis report." Geneva, Switzerland : IPCC 104 (2007).
- [71] Varin, Cristiano, and Paolo Vidoni. "A note on composite likelihood inference and model selection." *Biometrika* 92.3 (2005) : 519-528.
- [72] Wadsworth, Jennifer L., and Jonathan A. Tawn. "Dependence modelling for spatial extremes." *Biometrika* 99.2 (2012) : 253-272.