



Analyse de tendance dans les extrêmes dans un cadre spatio-temporel par dépassement de seuil. Application aux pluies et températures extrêmes au Burkina Faso

Béwentaoré Sawadogo

► To cite this version:

Béwentaoré Sawadogo. Analyse de tendance dans les extrêmes dans un cadre spatio-temporel par dépassement de seuil. Application aux pluies et températures extrêmes au Burkina Faso. Statistiques [math.ST]. Université Paris-Saclay; Université Joseph Ki-Zerbo (Ouagadougou, Burkina Faso), 2022. Français. NNT : 2022UPASM023 . tel-03940857

HAL Id: tel-03940857

<https://theses.hal.science/tel-03940857>

Submitted on 16 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse de Tendances dans les extrêmes dans un
cadre spatio-temporel par dépassements de seuil.
Application aux pluies et températures extrêmes au
Burkina Faso

*Trend detection in extremes : pointwise and spatial
approaches by Peaks-Over-Thresholds. Application
to extreme temperature and precipitation in Burkina
Faso*

**Thèse de doctorat de l'Université Paris-Saclay et
l'Université Joseph Ki-Zerbo**

École doctorale n° 574, mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées/Statistique
Graduate School : Mathématiques, Référent : Faculté des sciences
d'Orsay

Et

École doctorale Sciences et Technologies (EDST)
Spécialité de doctorat : Mathématiques appliquées/Statistique
Unité de Formation et de Recherche en Sciences Exactes et
Appliquées, Département de Mathématiques

Thèse préparée dans les unités de recherche UMR MIA-Paris Saclay et
LANIBIO sous la direction de Liliane BEL, Professeure des Universités,
la co-directeur de Diakarya BARRO, Professeur Titulaire

Thèse soutenue à UJKZ, le 18 Novembre 2022 par

Béwentaoré Sawadogo

Composition du jury :

David Makowski Directeur de Recherche, INRAE, Université Paris Saclay	Président
Jean-Noël Bacro Professeur, Université de Montpellier	Rapporteur
Nicolas Eckert ICPEF INRAE, Université Grenoble Alpes	Rapporteur
Blaise Somé Professeur, Université Joseph Ki-Zerbo (Burkina-Faso)	Rapporteur & Examineur
Sylvie Parey Ingénieure de Recherche, EDF, France	Examinatrice
Liliane Bel Professeure, AgroParisTech, Université Paris Saclay	Directrice de thèse
Diakarya Barro Professeur, Université Thomas Sankara (Burkina-Faso)	Co-directeur de thèse

Partenaires



Dédicace

Je dédie cet travail à mes parents, ma mère Sindinmanegdé NABALOUM et mon père Paligwendé SAWADOGO et aussi à mes enfants, mon fils Natéwendé Abdoul Rachid SAWADOGO et ma fille Sibdou Sakinatou SAWADOGO.

Remerciements

Je tiens à remercier très sincèrement mes deux directeurs de thèse pour la confiance qu'ils m'ont témoignée en acceptant tous les deux assurer la direction scientifique de mes travaux de recherche. J'adresse tout d'abord mes plus sincères remerciements à ma directrice de thèse, Professeure Liliane BEL que j'appelle affectueusement ma "*Maman*". Je lui suis très reconnaissante et redevable pour m'avoir fait bénéficier de son amour pour un travail de qualité, de sa rigueur scientifique, sa grande expérience et de sa constante disponibilité en dépit de ses tâches administratives assez lourdes. Elle a beaucoup oeuvré dans la mise en place de la collaboration entre l'Université Paris-Saclay et l'Université Joseph Ki-Zerbo dont je suis le premier bénéficiaire et la mise en oeuvre intégrale de ce projet de recherche, du point de vue scientifique, pédagogique et administrative. Son amour pour un *travail de qualité* et sa *rigueur scientifique* ont été des indicateurs pour moi tout au long de cette thèse et plus particulièrement dans les moments sombres auxquels un doctorant est souvent confronté. Elle a su me donner les moyens nécessaires pour réussir mon projet doctoral à travers ses conseils, ses orientations et ses choix de formations thématiques et transversales, elle m'a fait participer à des conférences de très haut niveau sur ma thématique recherche et à des séminaires à travers la France. Au delà de la direction de la thèse, elle s'est toujours préoccupée de mon bien être social pendant mes séjours en France, chose dont j'avais beaucoup besoin pour mon adaptation et mon intégration à mon nouvel environnement de vie. Mes remerciements s'adressent également au Professeur Diakarya BARRO qui m'a initié à la recherche en acceptant encadrer mon mémoire de Master de Mathématiques Appliquées Informatiques et Modélisation en Environnement, option statistique et modélisation stochastique. Ce qui a rendu possible l'ouverture vers cette thèse qu'il co-dirige. Il s'est beaucoup impliqué dans la mise en place et l'exécution de ce projet de recherche aussi bien scientifique, pédagogique qu'administrative. Honnêtement c'est un modèle pour moi qui oeuvre inlassablement dans la formation de la relève en matière de ressources humaines dans les universités de l'espace UEMOA et en particulier au Burkina-Faso. A tous les deux merci pour vos différentes qualités conjuguées qui m'ont forgé des compétences aussi bien scientifiques que sociales dont je suis très fier de vous avoir rencontré dans ma vie et de travailler avec vous. Vos précieux conseils et suggestions ont permis à ces recherches de progresser et d'aboutir aux résultats intéressants auxquels nous sommes parvenus. Je souhaiterai que cette relation "Enseignant-Etudiant" du triangle didactique construite au cours de ces années de thèse puisse se perdurer.

Je tiens également à remercier les membres du jury qui m'ont fait l'honneur en consacrant leur temps à l'évaluation de mon manuscrit et en participant au jury de la soutenance. Merci au Professeur Jean-Noël Barro et Nicolas Echert d'avoir accepté rapporter cette thèse que j'ai rencontrés pour la première fois respectivement à Montpellier et à Toulouse. J'ai eu avec eux des échanges assez fructueux et constructifs sur mon projet de recherche et je souhaite que cela se poursuive au delà de la thèse. Je remercie également le Professeur Blaise SOME, directeur du

Laboratoire d'Analyse Numérique d'Informatique et de BIOmathématique (LANIBIO) d'avoir accepté rapporter cette thèse et surtout pour tout ce qu'il a fait pour la promotion de la recherche en mathématiques appliquées en Afrique. Mes remerciements s'adressent également à Madame Sylvie Parey, pour avoir accepté évaluer, examiner mon manuscrit de thèse. Une fois de plus merci pour tes travaux qui nous ont permis de nous frayer des pistes de réflexions assez intéressantes. Merci encore à tous les membres du jury, vos précieuses interrogations et suggestions ont été prises en compte dans la version finale du manuscrit.

Je voudrais adresser mes remerciements les plus sincères à mes deux unités de recherche d'accueil. Le laboratoire MIA-Paris Saclay, AgroParisTech/INRAe de l'Université Paris-Saclay qui a bien voulu m'accueillir dans ces locaux et me donner les conditions nécessaires de travail pour le bon déroulement de mon projet de recherche. Une fois de plus je remercie ces premiers responsables et tout son personnel sans exception pour l'accueil chaleureux qui m'a été réservé et les compléments de financements dont j'ai bénéficié dans le cadre de mes travaux. Je m'en voudrais vraiment de ne pas citer de noms, mais je me réserve le droit de ne pas citer de peur d'en oublier, que vous soyez toutes et tous ici remerciés. Nous comptons poursuivre notre belle collaboration sur les thématiques restées en perspectives au cours de cette thèse. Je remercie également le Laboratoire d'Analyse Numérique d'Informatique et de BIOmathématique (LANIBIO), mon laboratoire d'origine pour m'avoir accueilli en son sein depuis mon Master et me donner les conditions nécessaires de travail. Je remercie très sincèrement ces premiers responsables, mes enseignants et tout le personnel sans exception pour leurs accompagnements dans la mise en oeuvre de ce projet doctoral. Je m'en voudrais vraiment de ne pas citer de noms, mais je me réserve le droit de ne pas citer de peur d'en oublier, que vous soyez toutes et tous ici remerciés.

Je remercie l'Université Paris-Saclay/Institut de Mathématiques d'Orsay (IMO) qui m'a recruté au poste d'Attaché Temporaire d'Enseignements et de Recherche (ATER) pour 2019-2021 parallèlement à mes activités de recherche. Ce qui m'a permis de faire une immersion dans l'enseignement supérieur Français et de côtoyer d'éminents professeurs qui ont bien voulu partager avec moi leurs expériences scientifiques et pédagogiques. Mes remerciements s'adressent également à l'Université Joseph Ki-Zerbo pour avoir garanti ma formation doctorale et oeuvré dans la préparation et la mise en place de la convention de co-tutelle internationale.

Je tiens à témoigner ma profonde gratitude au Service de Coopération et d'Action Culturelle (SCAC) de l'Ambassade de France au Burkina Faso pour m'avoir octroyé une bourse SCAC qui m'a permis de mener mes études doctorales et de vivre pendant quelques années en France. Cette bourse m'a apporté de solides compétences théoriques et pratiques en mathématiques appliquées option Statistique avec des applications concrètes aux phénomènes environnementaux et climatiques extrêmes dans le contexte actuel du changement climatique ; des compétences pédagogiques dans l'enseignement supérieur qui vont sans doute me permettre de réajuster mes pratiques pédagogiques en tant qu'enseignant du second degré et de mieux les adapter à l'évolution technologique ; des compétences diverses. Il faut noter que cet accompagnement a apporté une contribution significative en matière de

Remerciements

statistique, de diversification des axes de recherche et la recherche de haut cadre pour l'enseignement supérieur pour le Burkina Faso.

Je remercie l'état burkinabé à travers le Ministère de l'Education Nationale de l'Alphabétisation et de la Promotion des Langues Nationale (MENAPLN), pour la mise en position de stage de spécialisation et ses prolongations qui m'a été accordé pour la préparation de cette thèse. Je remercie l'Agence Nationale de la Météorologie du Burkina-Faso (ANAM) pour la mise à disposition des données climatologiques. Je profite de cette occasion pour adresser mes vifs remerciements particulièrement au Directeur de la Climatologie et du Réseau Météorologique, Monsieur Aimé Evariste OUEDRAOGO pour son soutien sans faille.

Je ne saurai terminer sans remercier très sincèrement tous les membres de ma famille, mon père, mes mamans, mes frères et soeurs pour m'avoir trouvé tous les moyens de mener à bien mes études et pour m'avoir soutenu dans mes moments difficiles et de joies. Merci d'avoir été présents quand rien ne semblait aller, vous avez beaucoup travaillé à me trouver des solutions souvent pas évidentes. Je tiens à exprimer à toutes et à tous ma vive reconnaissance. Je voudrais également saluer le dynamisme de ma compagne Clémentine OUEDRAOGO qui a su s'adapter pendant mes longues périodes d'absence au Burkina Faso. Une fois de plus je te dis sincèrement merci pour ta patience, ta présence à mes côtés et la lourde charge d'éducation des enfants que tu as pleine assumée pendant mes absences. Je remercie également ma famille adoptive, en particulier Awa NAMANEGUE et mon tuteur feu Boukary BIKIENGA, qui ont su m'encourager et me soutenir depuis un certain temps. Par contre j'ai malheureusement perdu mon tuteur au cours de ces dernières années de thèse. Que âme de Monsieur Boukary BIKIENGA repose en paix.

Pour terminer je ne peux pas m'empêcher de citer ma famille d'accueil en France : Je remercie très sincèrement une fois de plus Luc et Liliane pour leurs soutiens multiformes lors de mes séjours en France. Je remercie également le Professeur Xavier Guyon à Sorbonne Université qui a bien voulu nous accompagner dans la mise en œuvre de cet projet.

Productions scientifiques

Publications scientifiques

- **Béwentaoré, Sawadogo, Barro, Diakarya.** *Space-Time Trend Detection and Dependence Modeling in Extreme Event Approaches by Functional Peaks-Over-Thresholds : Application to Precipitation in Burkina Faso.* Cet article est publié dans "International Journal of Mathematics and Mathematical Sciences" : <https://doi.org/10.1155/2022/2608270> ;
- **Béwentaoré Sawadogo, Liliane Bel, Diakarya Barro.** *A spatial approach of the POT method in trend detection of extreme temperatures and precipitations in Burkina Faso.* Cet papier est en attente d'être soumis ;
- **Bel, L and Gomez-Garcia, J and Sawadogo, B.** *Spatial return level surfaces for non-stationary spatio-temporal processes.* Article en finition.

Communications orales

- **Bel, L and Gomez-Garcia, J and Sawadogo, B.** *Spatial return level surfaces for non-stationary spatio-temporal processes.* METMA X, pages 71, 2022
- **Béwentaoré Sawadogo, Liliane Bel, Diakarya Barro.** *Space-time trends detection and dependence modelling approaches by functional Peaks-Over-Thresholds. Application to precipitation in Burkina Faso.* 52es Journées de Statistique de la SFDS, Communication acceptée, du 07 au 11 Juin 2021, Nice, France ;
- **Béwentaoré Sawadogo, Liliane Bel, Diakarya Barro.** *Trend detection in extremes : pointwise and spatial approaches by Peaks-Over-Thresholds. Application to extreme temperature and precipitation in Burkina Faso.* 52es Journées de Statistique de la SFDS, Communication acceptée, du 07 au 11 Juin 2021, Nice, France ;
- **Béwentaoré Sawadogo, Liliane Bel, Diakarya Barro.** *Modèle pot non-stationnaire pour l'analyse des températures et des précipitations extrêmes au Burkina Faso.* 51es Journées de Statistique de la SFDS, 2019
- **Béwentaoré Sawadogo, Liliane Bel, Diakarya Barro.** *Non stationary POT model for extreme temperature and precipitation analysis in Burkina Faso.* The 14th International Meeting on Statistical Climatology (IMSC), poster accepté, du 24-28 Juin 2019 à Toulouse, France ;
- MathDACC 2017. Ecole Chercheur du Laboratoire des Sciences du Climat et de l'Environnement (LSCE) sur le thème : Mathematics for Climate Change

Detection and Attribution. Aussois (France), du 24 au 29 Septembre 2017

- Premières journées scientifiques de l'Ecole doctorale Sciences et technologies.
Analyse de tendances dans les extrêmes dans un cadre spatio-temporel par dépassement de seuil. Application aux pluies extrêmes aux Burkina Faso, poster accepté, Ouagadougou du 25 au 28 Octobre 2017

Table des figures

1.1	Effet des paramètres sur la fonction de répartition (ligne2) et la densité de probabilité(ligne1) d'une distribution GEV. Dans chaque figure, deux paramètres sont maintenus constants, à la troisième, trois valeurs différentes sont attribuées. Dans (a) le paramètre de localisation est modifié, dans (b) l'échelle et dans (c) la forme.	10
1.2	Maxima annuel de précipitations (mm) enregistrés à la station synoptique de Ouagadougou.	14
1.3	Dépassements supérieurs à $u = 20\text{mm}$ de pluies enregistrés à la station synoptique de Ouagadougou.	15
1.4	Effet des paramètres sur la fonction de répartition (ligne2) et la densité de probabilité(ligne1) d'une distribution GPD. Dans chaque figure, deux paramètres sont maintenus constants, à la troisième, trois valeurs différentes sont attribuées. Dans (a) le paramètre de localisation est modifié, dans (b) l'échelle et dans (c) la forme.	17
1.5	Panneau de sélection des seuils de pluies enregistrés pour la station synoptique de Ouagadougou. Mean excess plot (gauche) et stabilité des paramètres (droite).	22
1.6	Courbe de quelques fonctions de corrélation isotrope.	31
2.1	Emplacement des différentes stations synoptiques	54
2.2	Choix du seuil u pour Tmax par la méthode "Mean Excess Plot".	61
2.3	Tendance des occurrences des températures minimales et maximales.	63
2.4	Tendance des occurrences des précipitations extrêmes pour les stations de Ouaga au Centre et Ouahigouya au Nord du pays.	67
2.5	Krigeage des niveaux de retour non-stationnaires de Tmax et Tmin pour des périodes de retour $m = 20$ et $m = 100$ ans	68
2.6	Ecart-type de krigeage des niveaux de retour non-stationnaires de Tmax et Tmin pour des périodes de retour $m = 20$ et $m = 100$ ans	69
2.7	Krigeage des niveaux de retour non-stationnaires des précipitations extrêmes pour des périodes de retour $m = 50$ et $m = 100$ ans	70
3.1	Spatial distribution of synoptic sampling stations associated with climatic regimes, source: National Agence Meteorology of Burkina Faso.	82
3.2	Optimal trends of the occurrences of Tmax (line 1), Tmin (line 2) and Precip (line 3) of some reference stations.	84
3.3	Evaluation of significant trend coefficients for the variability and intensity of extreme.	84
3.4	Adjustment of the semi-variogram calculated for each pair of sites, for minimum and maximum temperatures.	86
3.5	Adjustment of the extremogram calculated for each pair of sites, for minimum and maximum temperatures.	87
3.6	Repartition of the grid points in ten clusters.	87

3.7	Simulate Spatiotemporal processes for Tmax(Top) and Tmin (bottom) by standard ℓ -Pareto processes approach for each step t	88
3.8	non-stationary return levels for Tmax (Top) and Tmin (bottom) by standard ℓ -Pareto processes approach.	89
4.1	Allure de la fonction de tendance Log-linéaire à droite et celle de la fonction de tendance Linéaire simple à gauche pour différentes valeur de θ	97
4.2	Paramètres marginaux a_Z et θ avec une extrapolation des paramètres $a_{t,s}$ et $b_{t,s}$ à une date futur $t = 10$ ans.	101
4.3	Ajustement de la fonction d'extrémogramme à gauche et la fonction semi-variogramme à droite appliquée aux pseudo observations. . . .	101
4.4	Surface de retour des températures maximales associées aux fonctions linéaire et quadratique dans le cas d'un ajustement Brown-Resnick. . . .	102
5.1	Spatial distribution of synoptic sampling stations associated with climatic regimes, source: National Meteorology Agency of Burkina Faso.	113
5.2	Estimated functions a_n (5.2a) and b_n (5.2b) of the generalized ℓ -Pareto process for modelling extreme precipitation, and θ for Log-linear (5.2c) and simple linear (5.2d) trends. Estimates are obtained by shifting the local empirical quantiles $u_{0.80}\{\ell(Z(s))\}$ by $\hat{b}_n = 33.4mm$	114
5.3	q-q plot of the distributions of the local tails of six synoptic positions including two stations for each climate region. left column: Sudanian climate; middle column: Sudano-Sahelian climate; right column: Sahelian climate. The threshold is taken equal to $b_n(s)$ for each station.	114
5.4	Local fitting of the log-linear trend model c_θ^1 for some of the selected sites with two stations for each climate region	115
5.5	Isotropic extremogram $\pi(h) = \lim_{q \rightarrow 1} P(Z(s+h) > u_q(s+h) \mid \{Z(s) > u_q(s), \ell(Z) > u\})$ (Figure 5.5a) and variogram (Figure 5.5b) estimated for risk functional $\ell(Z) = \max_{s \in S} Z(s)$ of distance between locations s and $s+h$. Blue line: empirical extremogram and red line: fitted dependence model.	116
5.6	Trends of stationary and non-stationary return levels at reference stations computed with a log-linear tail trend function.	117
5.7	Maps of the non-stationary 50-years return levels(left) and 100-years return level (right) obtained by extrapolating the information from the log-linear tail trend function and the dependence structure. . . .	118

Liste des tableaux

1.1	Familles paramétriques de fonctions de corrélation et de semi-variogramme isotrope, où K_v est la fonction de Bessel d'ordre v , $\Gamma(v)$ est la fonction Gamma.	31
2.1	Caractéristiques des stations météorologiques.	54
2.2	Seuils optimaux u (en ° C) estimés pour Tmin et Tmax.	61
2.3	Paramètres estimés par la méthode du maximum de vraisemblance pour Tmin et Tmax.	62
2.4	Degré de la tendance optimale du paramètre d'échelle ($\sigma_{\beta,u}(t)$) et l'intensité du processus de Poisson non stationnaire ($\lambda_\alpha(t)$).	62
2.5	p-value du test du rapport de vraisemblance des tendances de la fonction d'intensité λ_α et d'échelle σ_β pour Tmax.	63
2.6	p-value du test du rapport de vraisemblance des tendances de la fonction d'intensité λ_α et d'échelle σ_β pour Tmin.	64
2.7	Niveau de retour non- stationnaire(en [°]C) et intervalle de confiance à 95% pour Tmax calculés par la méthode de bootstrap.	64
2.8	Niveau de retour non- stationnaire (en [°]C) et intervalle de confiance à 95% pour Tmin calculés par la méthode de bootstrap.	65
2.9	Paramètres estimés des fonctions d'intensité λ_α et d'échelle σ_β pour les précipitations	65
2.10	Degrés de la tendance optimale du paramètre d'échelle ($\sigma_\beta(t)$) et l'intensité du processus de Poisson ($\lambda_\alpha(t)$) pour les précipitations.	66
2.11	p-value des tests du rapport de vraisemblance pour les précipitations au niveau de confiance de 95%.	66
2.12	Niveau de retour non- stationnaire (en mm de pluies) et intervalle de confiance à 95% pour les précipitations calculés par la méthode de bootstrap.	67
3.1	Degree of the optimal trend of the scale parameter ($\sigma_u(s, t)$) and the intensity of the non-stationary Poisson process ($\lambda(s, t)$).	83
3.2	Pointwise return level associated with the return period $m = 50$ years for the different synoptic stations.	85
3.3	Dependence structure parameters estimated by gradient score rule method and their confidence interval obtained by bootstrap	86

Table des matières

Dédicace	iii
Remerciements	v
Productions scientifiques	xii
Introduction	4
1 Cadre de modélisations ponctuelles et spatiales par dépassement de seuils	7
1.1 Extrêmes univariés	7
1.1.1 Approche des maxima par blocs	7
1.1.2 Approche des dépassements de seuils	14
1.1.3 Niveau de retour	22
1.2 Extrêmes multivariés	23
1.2.1 Variation régulière multivariée	24
1.2.2 Extension GEV multivariée	25
1.2.3 Extension POT multivariée	27
1.3 Extrêmes fonctionnels	27
1.3.1 Processus spatiaux et spatio-temporels	27
1.3.2 Processus max-stables	32
1.3.3 Processus ℓ -Pareto et extensions	38
1.4 Conclusion	48
2 Approches POT ponctuelles non-stationnaires et applications	51
2.1 Introduction	51
2.2 Présentation des données	53
2.3 Modèle POT non stationnaire	55
2.3.1 Distribution de Paréto Généralisée	55
2.3.2 Processus de Poisson non-homogène	55
2.3.3 Niveaux de retour non stationnaires	56
2.3.4 Inférence statistique	59
2.4 Résultats	60
2.4.1 Application aux données de températures	60
2.4.2 Application aux données pluviométriques	65
2.4.3 Krigeage des niveaux de retour	67
2.5 Discussion et perspectives	70
3 A spatial approach of the POT method in trend detection of extreme temperatures and precipitations in Burkina Faso	75
3.1 Introduction	75
3.2 Pointwise Peaks-Over-Thresholds	77
3.2.1 Stationary threshold exceedances modeling	77

3.2.2	Non-stationary pointwise POT-PP modeling	78
3.3	Non-stationary Spatial ℓ -POT modeling	79
3.4	Application to climate data in Burkina Faso	81
3.4.1	Data set	81
3.4.2	Non-stationary marginal modeling	82
3.4.3	Spatial and temporal dependence	85
3.5	Discussion and perspectives	90
4	Modélisation hétéroscédastique et application à la recherche de surfaces de retour	93
4.1	Introduction	93
4.2	Modèles hétéroscédastiques	94
4.2.1	Hétéroscédasticité de queue et pseudo observations	94
4.2.2	Estimation de la dépendance spatiale et marginale	96
4.2.3	Surface de retour	98
4.3	Application aux données de température	100
4.4	Discussion et perspectives	102
5	Space-time trend detection and dependence modelling in extreme events approaches by functional Peaks-Over-Thresholds. Application to precipitation in Burkina Faso	105
5.1	Introduction	105
5.2	Methodology	107
5.2.1	Space-time trends modelling	107
5.2.2	Non-stationary return period and return level	108
5.2.3	Spatial POT modelling	110
5.2.4	Statistical inference	111
5.3	Application to extreme precipitation in Burkina Faso	112
5.3.1	Data set analysis	112
5.3.2	Marginal characteristics	113
5.3.3	Estimated spatial dependence model	115
5.3.4	Non-stationary return level results	116
5.4	Discussion and conclusion	118
1.A	Proof of proposition 5.2.1	120
1.B	Overview on generalized ℓ -Pareto process	120
1.C	Asymptotic distribution of ℓ -Pareto Process	122
	Conclusions et perspectives	131

Introduction

La Théorie des Valeurs Extrêmes (TVE) offre un cadre mathématique rigoureux pour décrire les comportements des queues des distributions, ce qui permet en particulier d'estimer, le risque dans de nombreuses disciplines appliquées comme l'environnement et le climat. Introduite à la fin des années 1920 avec les travaux fondateurs de [Fisher et Tippet \(1928\)](#), elle a depuis connu de nombreux développements, et continue actuellement à faire l'objet de recherches actives. Une des branches de cette théorie est basée sur une approximation asymptotique des lois de maxima (ou minima) convenablement renormalisés de vecteurs aléatoires dont les composantes sont des variables supposées *iid* et donne les distributions des valeurs extrêmes (GEV : Generalized Extreme Value [Fisher et Tippet \(1928\)](#)). Une approche alternative aux maxima par bloc est celle basée sur les dépassements de seuils suffisamment élevés (GPD : Generalized Pareto Distribution [Pickands \(1975\)](#), [Balkema et De Haan \(1974\)](#)). Ces deux approches se sont généralisées au contexte multivarié et fonctionnel pour donner respectivement la classe des processus max-stables ([De Haan et al. \(1984\)](#), [de Haan et Ferreira \(2006\)](#)) et ℓ -Pareto ([Ferreira et al. \(2014\)](#), [Dombry et Ribatet \(2015\)](#), [de Fondeville et Davison \(2020\)](#)).

Le but de ce travail porte sur la modélisation statistique des tendances des extrêmes climatiques et son utilisation pour la caractérisation des aléas climatiques en Afrique sub-saharienne. En effet, les événements climatiques extrêmes causent chaque année à travers le monde des pertes en vies humaines ainsi que de nombreux dégâts matériels. Ce ne sont certes pas des phénomènes nouveaux, mais ils ont pris une ampleur particulièrement importante au cours de ces trente dernières années. En effet, selon des rapports d'évaluation du changement climatique du GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat), la température moyenne de l'air augmente d'environ $0.2 \pm 0.1^\circ\text{C}$ tous les dix ans ([IPCC \(2014\)](#)). Toujours selon le dernier rapport ([IPCC \(2022\)](#)) au delà de l'aspect entropique, l'origine humaine des changements climatiques est sans équivoque. Cette influence humaine se traduit notamment par une acidification des océans et la hausse des niveau marin. En 2021 le réchauffement global a atteint $1.09 \pm 0.1^\circ\text{C}$ et pourrait dépasser $1.5 \pm 0.1^\circ\text{C}$ entre 2030 et 2052 ; $2.7 \pm 0.1^\circ\text{C}$ d'ici la fin siècle. Ainsi le réchauffement se manifestera dans la plupart des régions du monde, par une recrudescence des chaleurs extrêmes, des précipitations intenses et des risques de sécheresse. En Afrique sub-sahélienne, la fin de la décennie 2000 et la première moitié de celle de 2010, ont été marquées, par des événements extrêmes aux forts impacts socio-économiques et environnementaux, jamais enregistrés depuis plus de 30 ans (Centre Régional AGRHYMET, 2010). En 2012 par exemple, 363 personnes sont mortes au Nigéria dans la pire des inondations qu'a connues le pays. Il y a eu environ 2 millions de déplacés et 618 000 maisons détruites. Au Niger, les inondations ont provoqué également des dégâts considérables : 65 décès et 125 000 personnes sans abri. Sur la même période, des dégâts ont été aussi enregistrés dans d'autres pays d'Afrique de l'Ouest, notamment au Sénégal, au Mali, au Burkina Faso et au Sierra Leone.

Au Burkina Faso, le Ministère de l'Environnement de l'Économie Verte et du

TABLE DES MATIÈRES

Changement Climatique dans son Programme d'Action National d'Adaptation aux variabilités et changements climatiques (PANA, 2006), a montré que les températures et les précipitations sont soumises à une forte variabilité spatiale et temporelle avec une tendance à la baisse des cumuls pluviométriques et des pluies journalières et une légère tendance à la hausse des températures. Des travaux réalisés à cet effet sur le Burkina Faso par Diallo (2010) ont permis également de montrer que la pluviométrie a baissé globalement et que la température est en hausse avec des écarts de 0.74°C, 1.05°C et 1.24°C respectivement sur la zone soudanienne au Sud, la zone soudano-sahélienne au centre et la zone sahélienne au Nord. Selon les dernières publications de l'Agence Nationale de la Météorologie (ANAM), la pluviométrie a amorcé un retour vers des périodes humides depuis la fin des années 1980 et les décennies 1990 et 2000. Le retour des pluies est plus lié à une fréquence élevée des événements de forte intensité pluviométrique qu'à une augmentation du nombre de jours pluvieux. Les souvenirs les plus marquants remontent au 1er Septembre 2009 où Ouagadougou et ses environs ont enregistré une pluie de 261.3 mm. Cette pluie a causé plus de 150000 sinistrés, la destruction de plusieurs ponts et plus de 9300 hectares de cultures ont été inondés sur l'ensemble du pays.

Une des questions souvent évoquées dans le cadre du changement climatique est celle de son influence sur les extrêmes météorologiques en particulier pour les précipitations et les températures. Cette question est très largement débattue dans la communauté des climatologues (Zhang *et al.* (2007), Donat *et al.* (2016), Grotjahn *et al.* (2016)) qui utilisent des données historiques en différentes parties du monde et divers indices statistiques pour argumenter pour ou contre une tendance significative dans les extrêmes dans un sens ou dans l'autre. Les statisticiens se sont penchés depuis longtemps sur la question de détecter une tendance voire des ruptures dans les extrêmes de séries temporelles, avec des applications qui dépassent le cadre du climat (Madsen *et al.* (2014), Cheng *et al.* (2014), de Haan *et al.* (2015), Eastoe et Tawn (2009)). Les approches sont diverses mais en général elles sont basées sur un modèle paramétrique de tendance et la mise en oeuvre d'un test sur le paramètre. Une des approches dite paramétrique ponctuelle des extrêmes (Davison et Smith (1990), Parey *et al.* (2007, 2010)), consiste à inclure des covariables dans les paramètres des distributions des valeurs extrêmes ($GPD(\sigma_u(t), \gamma_u(t))$ et $GEV(\mu(t), \sigma(t), \gamma(t))$), avec une paramétrisation qui peut être polynomiale, linéaire, linéaire par morceaux, semi-paramétrique ou non-linéaire. Une autre alternative consiste à transformer une série climatique non-stationnaire en une série stationnaire (Eastoe et Tawn (2009), Dacunha-Castelle *et al.* (2015), Mentaschi *et al.* (2016)) en utilisant des techniques de stationnarisation pour laquelle la théorie des valeurs extrêmes classiques peut être appliquée. On peut alors retourner aux données non-stationnaires en utilisant la transformation inverse de stationnarisation et les tendances identifiées dans les paramètres. Notons que même si des covariables d'espace sont généralement prises en compte dans les paramètres de ces modèles, il faut rappeler que ces approches ne prennent pas de façon rigoureuse la structure de dépendance spatiale. Pour prendre en compte de façon explicite la structure de dépendance spatiale, des modèles spatiaux max-stables ont été introduit principalement par De Haan *et al.* (1984), de Haan et Ferreira (2006). Ainsi depuis l'essor de ces modèles max-stables qui permettent en particulier d'appréhender l'aspect spatial des processus, de nom-

TABLE DES MATIÈRES

breux auteurs ont proposé des modèles pour les extrêmes dans le domaine du climat (Davison *et al.* (2012), Gaume *et al.* (2013), Bechler *et al.* (2015), Blanchet *et al.* (2018), Mefleh *et al.* (2020)). Ces études se font en général dans un cadre stationnaire et spatial, la dimension temporelle tenant lieu de répétitions indépendantes, de plus ces approches sont très lourdes en temps de calculs en grand dimension et ne sont pas adaptées pour modéliser les dépassements de seuil. Cela a motivé le développement des modèles spatiaux des dépassements de seuil par Ferreira *et al.* (2014, 2017) avant d'être généralisé par Dombry et Ribatet (2015), de Fondeville et Davison (2020). Là encore ces développements sont traités dans un cadre stationnaire et spatial, nonobstant quelques travaux commencent à apparaître dans un cadre réellement spatio-temporel (Khan *et al.* (2007), Xia *et al.* (2012)).

La modélisation et la prévision précise de l'intensité, l'étendue et l'occurrence dans le temps et dans l'espace des événements météorologiques extrêmes sont des tâches particulièrement difficiles dans un contexte de changement climatique. Elles s'avèrent essentielles pour l'évaluation des risques dans de nombreuses disciplines appliquées, y compris les sciences de l'environnement et du climat. Cette évaluation requiert aujourd'hui des approches statistiques flexibles capables de saisir l'influence du changement climatique sur la fréquence et l'intensité des événements climatiques extrêmes. Quelques réponses à ce sujet ont été proposées ces dernières années pour la détection, l'attribution et l'analyse des tendances des événements climatiques extrêmes avec des approches ponctuelles et des approches spatiales ou spatio-temporelles (Bechler *et al.* (2015), Diakarya *et al.* (2017), de Fondeville et Davison (2020); Cabral *et al.* (2020)). Il reste néanmoins encore beaucoup de choses à développer aussi bien du point de vue de la méthodologie que des applications. La théorie des valeurs extrêmes spatiales permet de modéliser et de prédire la fréquence des événements extrêmes dans un contexte spatial comme, par exemple, les précipitations extrêmes, les températures extrêmes ou les fortes concentrations de pollution dans l'air. Elle est bien adaptée aux séries chronologiques, lorsque l'objet spatial étudié est observé dans le temps. Cependant, dans certains cas, il est impossible d'accéder à ce type de données qui nécessitent généralement des moyens techniques et financiers : seuls un ou quelques enregistrements sont disponibles.

Dans le cadre de cette thèse, on envisage d'étudier la présence de tendance dans les extrêmes en intégrant l'aspect spatial, donc en se plaçant dans un cadre spatio-temporel. Plutôt que de travailler en modélisant les maxima on regardera les dépassements de seuil et on utilisera les travaux récents sur les processus de Pareto (Ferreira *et al.* (2014)) et leur généralisation aux ℓ -excès (Dombry et Ribatet (2015), de Fondeville et Davison (2020)) afin de proposer des méthodes statistiques qui prennent en compte non seulement les non-stationnarités marginales mais aussi la structure de dépendance spatiale. On pourra ainsi tenter de répondre à des questions du type "l'évènement $E(A) = \{\ell(X(s)) > u, s \in A\}$, A étant une région, a-t-il évolué significativement dans le temps et comment ?"; "Comment la variabilité à différentes échelles temporelles et spatiales mentionnée précédemment affecte les queues de distributions des températures et des précipitations ?" Nous envisageons également d'étendre les mesures risques développées dans un cadre ponctuel par Parey *et al.* (2010) et Cooley (2013) aux processus stochastiques spatio-temporels.

L'application des modèles développés portera sur les données réelles de températures et de précipitations quotidiennes du Burkina Faso. Le Burkina Faso dispose d'une bonne base de données climatologiques avec des observations remontant à 1902 pour certains paramètres météorologiques. La couverture spatiale s'est accrue dans le temps pour arriver à une quarantaine de stations actuellement ce qui devrait permettre d'envisager une modélisation raisonnable.

Le premier chapitre présente de façon succincte les notions essentielles et les principaux résultats des modèles des valeurs extrêmes dans un cadre ponctuel, multivarié et fonctionnel avant la mise en relief des approches les plus récentes dans le cadre fonctionnel des dépassements de seuil.

Le chapitre 2 vise à mettre au point une approche ponctuelle paramétrique non-stationnaire de détection de tendances dans les extrêmes par la méthode des dépassements de seuil (Parey *et al.* (2007), Parey *et al.* (2010)). Il sera question de prendre en compte la non-stationnarité dans les paramètres d'échelle de la distribution des dépassements de seuil et l'intensité de la loi des occurrences des événements extrêmes pour évaluer l'évolution des tendances temporelles des niveaux de retour. La structure spatiale des niveaux de retour non-stationnaires sera évaluée en utilisant la méthode du krigeage spatial.

Dans le chapitre 3 nous présentons une approche spatio-temporelle de détection de tendances dans les extrêmes. Dans cette approche, au lieu d'évaluer la structure de dépendance en utilisant le krigeage spatial, nous ajustons cette structure de dépendance spatiale par un processus spatial attaché aux dépassements de seuil. C'est à dire qu'on reprend l'approche ponctuelle paramétrique non-stationnaire décrite dans le chapitre 2 tout en ajustant un processus ℓ -Pareto (Ferreira *et al.* (2014), Dombry et Ribatet (2015)) aux marginales standardisées.

Le chapitre 4 présente un modèle spatio-temporel hétéroscédastique de détection de tendances, au lieu d'ajuster un processus ℓ -Pareto au marginal standard, nous utilisons ici, une relation entre le processus non-stationnaire et un processus spatial latent, via une fonction de scédasticité introduite dans le cadre des extrêmes hétéroscédastiques (Ferreira *et al.* (2017), Mefleh *et al.* (2020)). Une extension du niveau de retour non-stationnaire ENE aux processus stochastiques spatio-temporels est proposée afin de rechercher les surfaces de retour non-stationnaires.

Dans le chapitre 5, nous présentons une approche spatio-temporelle de détection de tendance en exploitant la théorie des extrêmes hétéroscédastiques. Dans cette approche au lieu de marginaliser et d'ajuster un processus ℓ -Pareto standard aux données marginales, nous modélisons directement la structure de dépendance par un processus ℓ -Pareto généralisé (de Fondeville et Davison (2020)) construite à partir du processus ℓ -Pareto standard. Il s'agira de prendre en compte la non-stationnarité dans les distributions marginales et de réadapter les définitions des niveaux de retour non-stationnaires ENE et EWT , en introduisant une fonction de scédasticité et la loi des ℓ -dépassements. La relation entre le processus spatio-temporel non-stationnaire et un processus spatial latent sera mise également à profit pour évaluer les futurs quantiles extrêmes.

Cadre de modélisations ponctuelles et spatiales par dépassement de seuils

L'approche des dépassements de seuil ou Peak-Over-Threshold (POT) est l'une des alternatives permettant de décrire les comportements de queue des distributions d'une variable aléatoire. Introduite par [Balkema et De Haan \(1974\)](#) et [Pickands \(1975\)](#) dans le cadre univarié pour des suites de variables aléatoires supposées indépendantes et identiquement distribuées (i.i.d), le modèle POT a d'abord été étendu aux environnements multivariés ([Rootzén et al. \(2006\)](#), [Rootzén et al. \(2018a\)](#), [Rootzén et al. \(2018b\)](#)) avant d'être généralisé récemment aux données fonctionnelles ([Ferreira et al. \(2014\)](#), [Dombry et Ribatet \(2015\)](#), [de Fondeville et Davison \(2020\)](#)) pour donner naissance à la famille des processus ℓ -Pareto généralisés. Cette nouvelle classe de processus spatiaux est beaucoup plus adaptée pour modéliser la structure de dépendance spatiale des excès de seuil. Ce chapitre introduit ces principaux résultats de façon succincte dans le cadre univarié, multivarié et spatial. Un accent particulier sera accordé à la modélisation multivariée et spatiale des dépassements de seuils. Mais avant de parler proprement de ce sujet nous referons un petit tour sur l'approche des valeurs extrêmes généralisées (GEV : Generalized Extreme Values) et ses extensions. Les processus max-stables relient deux domaines de la statistique, la théorie des valeurs extrêmes et la géostatistique, de sorte que les deux méritent d'être succinctement présentés.

1.1. Extrêmes univariés

1.1.1. Approche des maxima par blocs

a. Les trois types extrêmes

Soient $Z_1, \dots, Z_n$ des copies indépendantes distribuées de manière identique d'une variable aléatoire Z de fonction de répartition commune F et de point terminal droit z_F . On désigne par $M_n = \max_{1 \leq i \leq n} (Z_i)$ la variable aléatoire représentant la série des maxima de blocs et de fonction de distribution F_{M_n} . On établit que $F_{M_n} = F^n$. Cependant la distribution F en général n'est pas connue, et il faut approximer F_{M_n} avec une distribution pertinente. On considère plutôt des familles de distributions qui assurent que F_{M_n} converge vers les queues de la distribution F . Le point terminal droit z_F du support de Z n'est pas nécessairement fini, de sorte que la distribution de M_n dégénère lorsque $n \rightarrow \infty$, en une distribution avec une masse ponctuelle à

l'extrémité supérieure z_F de la distribution sous-jacente F telle que :

$$\lim_{n \rightarrow \infty} P(M_n \leq z) = \lim_{n \rightarrow \infty} F^n(z) = \begin{cases} 0, & \text{si } F(z) < 1 \\ 1, & \text{si } F(z) = 1 \end{cases}, \quad (1.1)$$

donc $M_n \xrightarrow{d} z_F$, où $z_F = \sup \{z : F(z) < 1\}$. Il faut noter que ce résultat est très peu informatif sur le comportement du maximum et il est préférable d'obtenir une loi limite non-dégénérée. Pour ce faire, l'idée est de considérer, non pas le maximum en tant que tel dans l'équation (1.1), mais une renormalisation de ce dernier. La renormalisation la plus simple consiste à faire une transformation linéaire, à l'image de celle utilisée dans le Théorème Central Limite (TCL). Nous rappelons brièvement ici que dans l'étude de la moyenne d'un échantillon, le TCL montre que la distribution de $S_n = \sum_{i=1}^n Z_i$, pour tout $n \in \mathbb{N}^*$ converge vers une normale standard lorsque $n \rightarrow \infty$ dans les conditions faibles et de normalisation affine appropriées, i.e.,

$$\text{Pour tout } z \in \mathbb{R}, \lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2} dt. \quad (1.2)$$

Parallèlement au TCL, la distribution de M_n est étudiée pour des suites de normalisation convenablement choisies de sorte à éviter le problème de dégénérescence mentionnée dans (1.1). En effet, les premiers résultats concernant la distribution asymptotique du maximum d'un échantillon de variables aléatoires *i.i.d* convenablement normalisés sont dus à [Fisher et Tippett \(1928\)](#). Ils ont ensuite été formalisés par [Gnedenko \(1943\)](#) au moyen d'un théorème connu sous le nom du théorème des trois types extrêmes unifiés.

Théorème 1.1.1. ([Fisher et Tippett \(1928\)](#), [Gnedenko \(1943\)](#))

Soit $(Z_i; i \geq 1)$ une suite de variables aléatoires *i.i.d* de fonction de répartition F . Supposons qu'il existe des suites normalisantes réelles $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ telles que, pour tout $z \in \mathbb{R}$:

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \xrightarrow{\text{loi}} G(z), \quad n \rightarrow \infty, \quad (1.3)$$

où G est une distribution non dégénérée. Alors G appartient à l'une des trois classes suivantes :

- $\Lambda(z) = \exp\left\{-\exp\left(-\frac{z-b}{a}\right)\right\}$, avec $z \in \mathbb{R}$ (Gumbel);
- $\Phi_\alpha(z) = \begin{cases} \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\} & \text{si } z > b \\ 0 & \text{si } z \leq b \end{cases}$ (Fréchet);
- $\Psi_\alpha(z) = \begin{cases} \exp\left\{-\left(-\frac{z-b}{a}\right)^\alpha\right\} & \text{si } z < b \\ 1 & \text{si } z \geq b \end{cases}$ (Weibull);

où $\alpha, a > 0$ et $b \in \mathbb{R}$.

1.1. EXTRÊMES UNIVARIÉS

Les trois lois Λ , Φ_α et Ψ_α sont dites distributions de valeurs extrêmes. Leurs formes réduites sont obtenues en les standardisant i.e. en posant $b = 0$ et $a = 1$. Une distribution de valeurs extrêmes est donc, à une translation et un changement d'échelle près, une distribution de Gumbel, Fréchet ou de Weibull. On montre que chacune des trois distributions de valeurs extrêmes peut s'obtenir par transformation fonctionnelle de l'autre. Plus précisément, elles sont liées entre elles par la relation fonctionnelle

$$Si \ Z \rightsquigarrow \Phi_\alpha \iff \ln Z^\alpha \rightsquigarrow \Lambda \iff -\frac{1}{Z} \rightsquigarrow \Psi_\alpha. \quad (1.4)$$

Von Mises (1954) et Jenkinson (1955) proposent une famille paramétrique, la distribution généralisée des valeurs extrêmes (Generalized Extreme Values : GEV, en anglais), qui permet d'unifier les trois types extrêmes d'après le résultat suivant (Galambos (1978))

$$G_\gamma(z) = \begin{cases} \exp \left\{ - \left[1 + \gamma \left(\frac{z - \mu}{\sigma} \right) \right]_+^{-1/\gamma} \right\} & si \ \gamma \neq 0 \\ \exp \left\{ - \exp \left[- \left(\frac{z - \mu}{\sigma} \right) \right] \right\} & si \ \gamma = 0 \end{cases}; \quad (1.5)$$

où $z \in \{z \in \mathbb{R} : 1 + \gamma(z - \mu)/\sigma \geq 0\}$, $\mu \in \mathbb{R}$, $\sigma > 0$ et $\gamma \in \mathbb{R}$.

Dans cette famille paramétrique de fonctions de distribution, les paramètres σ et μ sont respectivement les paramètres d'échelle et de localisation. Le régime de décroissance de queue est déterminé par le paramètre de forme ou indice de valeurs extrêmes γ . Selon le signe de γ , on aboutit à trois types de distributions des valeurs extrêmes. Le type Fréchet correspond à $\gamma > 0$ et est caractérisé par une décroissance polynomiale de la queue dont la limite supérieure de son support est infinie. Pour $\gamma < 0$, la distribution des valeurs extrêmes généralisée est bornée par $\mu - \sigma/\gamma$, ce qui donne le type de Weibull. La classe Gumbel est interprétée comme la limite des distributions de Weibull et Fréchet lorsque $\gamma \rightarrow 0$. La figure 1.1 illustre l'effet des paramètres sur la fonction de répartition et la densité de probabilité d'une distribution *GEV*.

b. Domaine d'attraction max-stable

Une des notions clés dans l'analyse des extrêmes que nous ne pouvons pas passer sous silence est la caractérisation des domaines d'attraction des types extrêmes. En effet, étant donné une loi extrême G_γ , on cherche les conditions suffisantes et nécessaires pour que la loi du maximum d'une suite de variables aléatoires *i.i.d* de loi F converge vers G_γ . La recherche de ces conditions où l'étude de ces domaines d'attraction fait appel aux classes de fonctions à variation régulière. Nous commençons par introduire de façon succincte quelques résultats importants sur la notion de distributions de même type et celle de la variation régulière indispensable pour la caractérisation des domaines d'attraction.

Définition 1.1.1. On dit que deux variables aléatoires Z et Y sont de même type, s'ils ont la même loi à une transformation affine près, i.e s'il existe des constantes $a > 0$ et $b \in \mathbb{R}$ telles que : $Z \stackrel{loi}{=} aY + b$.

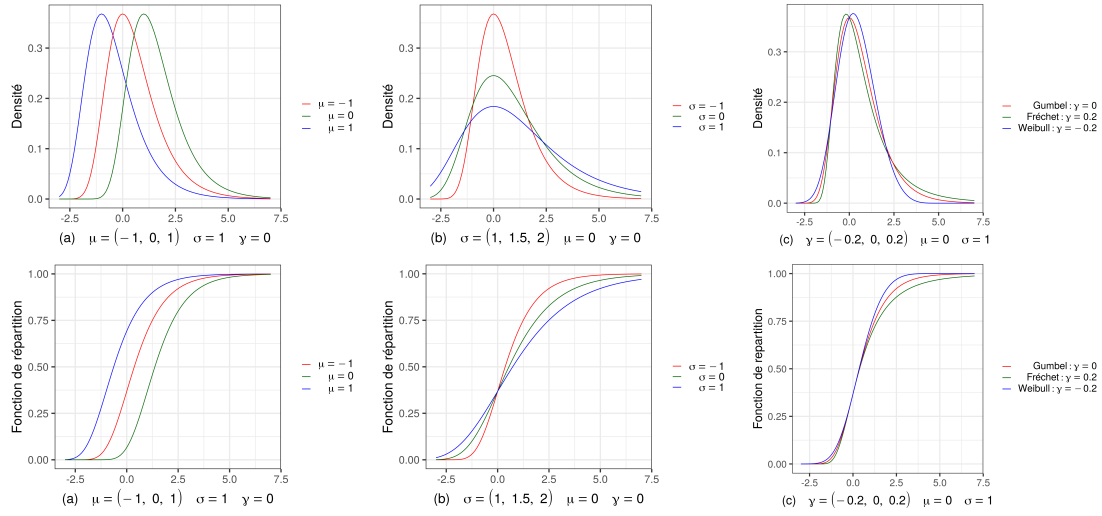


FIGURE 1.1 – Effet des paramètres sur la fonction de répartition (ligne2) et la densité de probabilité (ligne1) d’une distribution GEV. Dans chaque figure, deux paramètres sont maintenus constants, à la troisième, trois valeurs différentes sont attribuées. Dans (a) le paramètre de localisation est modifié, dans (b) l’échelle et dans (c) la forme.

Dans ce cas, si F et H sont les distributions respectives des variables Z et Y , alors on a : $F(ay + b) = H(y)$ pour tout $y \in \mathbb{R}$. Le théorème suivant établit l’égalité en type entre les limites de fonction de répartition obtenues par différentes normalisations affines.

Théorème 1.1.2. Soit $(Z_n, n \geq 1)$ une suite de variables i.i.d de fonction de répartition F^n et X, Y deux variables aléatoires de fonctions de répartition non dégénérées respectives G et H . Soient $(a_n)_{n \geq 1} > 0$, $(d_n)_{n \geq 1} > 0$, $(b_n)_{n \geq 1} \in \mathbb{R}$ et $(c_n)_{n \geq 1} \in \mathbb{R}$ des suites de constantes telles que :

$$\text{Pour tout } z \in \mathcal{C}(G) = \{z \in \mathbb{R}, G \text{ continue en } z\}, \lim_{n \rightarrow \infty} F^n(a_n z + b_n) = G(z) \quad (1.6)$$

Alors

1. On a :

$$\text{Pour tout } z \in \mathcal{C}(H) = \{z \in \mathbb{R}, H \text{ continue en } z\}, \lim_{n \rightarrow \infty} F^n(c_n z + d_n) = H(z) \quad (1.7)$$

si et seulement s’il existe des constantes réelles $A > 0$ et B telles que :

$$\lim_{n \rightarrow \infty} \frac{a_n}{d_n} = A \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{b_n - c_n}{d_n} = B. \quad (1.8)$$

2. si (1.7) est vérifiée, alors pour tout $z \in \mathbb{R}$, $H(z) = G(Az + B)$.

Pour une preuve du théorème 1.1.2, voir par exemple [Resnick \(1987\)](#), chapitre 0 ou [Leadbetter \(1983\)](#), chapitre 1. A présent nous introduisons quelques résultats

1.1. EXTRÊMES UNIVARIÉS

principaux (définitions et propriétés générales) d'une notion centrale permettant de caractériser les domaines d'attraction, qu'est la variation régulière. Introduite par Jovan Karamata en 1930, la théorie de la variation régulière est un outil indispensable dans l'analyse des événements rares, pour plus de détails on pourra se référer à [Bingham \(1987\)](#). Dans toute la suite nous notons par $F^\leftarrow$ l'inverse généralisée où encore fonction quantile associée à la distribution F , définie par $F^\leftarrow(q) = \inf\{z \in \mathbb{R}, F(z) \geq q\}$, $0 < q \leq 1$. On définit également la fonction quantile de queue par $U(z) = F^\leftarrow(1 - 1/z)$, $z > 0$.

Définition 1.1.2. Une fonction $F : [a; +\infty[\rightarrow \mathbb{R}_+(a \geq 0)$ mesurable est à variation régulière à l'infini d'indice $\alpha \in \mathbb{R}$ noté $F \in RV_\alpha$, si pour tout $z > 0$,

$$\lim_{t \rightarrow \infty} \frac{F(tz)}{F(t)} = z^\alpha. \quad (1.9)$$

En particulier si $F \in RV_0(\alpha = 0)$ alors F est une fonction à variation lente à l'infini. La notion de variation régulière peut également être définie à l'origine ([Maric \(2000\)](#)). La définition suivante donne cette version alternative :

Définition 1.1.3. Une fonction $F : [0; a[\rightarrow \mathbb{R}_+(a > 0)$ mesurable est à variation régulière d'indice $\alpha \in \mathbb{R}$ au voisinage de 0 noté $F \in RV_\alpha^0$, si pour tout $z > 0$,

$$\lim_{t \rightarrow 0} \frac{F(tz)}{F(t)} = z^\alpha. \quad (1.10)$$

c'est à dire $F(1/z)$ est à variation régulière d'indice $-\alpha$ à l'infini.

Les résultats essentielles d'une fonction dite à variation régulière sont rappelés par les théorèmes suivants :

Théorème 1.1.3. ([Bingham \(1987\)](#))

Soit $F : [a; +\infty[\rightarrow \mathbb{R}_+(a \geq 0)$ une fonction mesurable et $\alpha \in \mathbb{R}$. Alors les assertions suivantes sont équivalentes

1. $F \in RV_\alpha$.
2. Il existe $L \in RV_0$ telle que $\forall z > 0, F(z) = z^\alpha L(z)$.
3. Il existe une fonction positive h telle que

$$\text{Pour tout } z > 0, \lim_{t \rightarrow \infty} \frac{F(tz)}{F(t)} = h(z) = z^\alpha$$

Théorème 1.1.4. ([de Haan et Ferreira \(2006\)](#), représentation de Karamata)

Une fonction F est à variation régulière à l'infini d'indice $\alpha \in \mathbb{R}$ si et seulement si elle peut être représentée de la façon suivante :

$$\text{Pour tout } z \geq b > a, F(z) = c(z) \exp \left\{ \int_b^z \frac{\delta(u)}{u} du \right\} \quad (1.11)$$

où c et δ sont des fonctions mesurables sur $]a, \infty[$ ($a \geq 0$) telles que $\lim_{z \rightarrow \infty} c(z) = c_0 \in \mathbb{R}_+^*$ et $\lim_{z \rightarrow \infty} \delta(z) = \alpha$

Soit maintenant $(\mathbb{E}, \mathcal{E})$ un espace localement compact avec une base dénombrable. On désigne par $\mathcal{M}_+(\mathbb{E}) = \{\text{toutes les mesures de Radon sur } \mathbb{E}\}$, i.e pour tout $\lambda \in \mathcal{M}_+(\mathbb{E})$ est une mesure sur $\mathcal{E}$ et $\lambda(K) < \infty$ pour tout compact $K \in \mathcal{E}$. Désignons par $\mathcal{C}_K^+(\mathbb{E}) = \{f : \mathbb{E} \rightarrow \mathbb{R}_+, f \text{ est continue à support compact}\}$

Définition 1.1.4. Soit $(\lambda_n)_{n \geq 0}$ une suite de mesure définies dans $\mathcal{M}_+(\mathbb{E})$ et soit $\lambda \in \mathcal{M}_+(\mathbb{E})$. On dit que λ_n converge vaguement vers λ et on note $\lambda_n \xrightarrow{v} \lambda$ pour tout $f \in \mathcal{C}_K^+(\mathbb{E})$,

$$\lambda_n(f) = \int_{\mathbb{E}} f d\lambda_n \rightarrow \int_{\mathbb{E}} f d\lambda. \quad (1.12)$$

Proposition 1.1.1. (RV de queues de distribution univariée, [Resnick \(2007\)](#))
Soit Z une variable aléatoire de fonction distribution F et sa fonction de survie $\bar{F}(z) = 1 - F(z)$. Les assertions suivantes sont équivalentes :

1. $\bar{F} \in RV_{-\alpha}$, $\alpha > 0$,
2. Il existe une suite $(b_n)_{n \geq 0}$ vérifiant $b_n \rightarrow \infty$ telle que

$$n\bar{F}(b_n z) \rightarrow z^{-\alpha}, z > 0, \quad (1.13)$$

3. Il existe une suite $(b_n)_{n \geq 0}$ vérifiant $b_n \rightarrow \infty$ telle que

$$\lambda_n(\cdot) = nP\left(\frac{Z}{b_n} \in \cdot\right) \xrightarrow{v} \nu_\alpha(\cdot) \quad (1.14)$$

dans $\mathcal{M}_+([0, \infty])$ où $\nu_\alpha \in \mathcal{M}_+([0, \infty])$ vérifie $\nu_\alpha([z, \infty]) = z^{-\alpha}, z > 0$.

Dans la proposition 1.1.1, il faut remarquer qu'on peut toujours choisir $b(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right)$ et poser $b_n = b(n)$. Pour plus d'information sur la théorie des fonctions à variation régulière, on renvoie le lecteur à [de Haan et Ferreira \(2006\)](#), [Bingham \(1987\)](#). Les notions décrites ci-dessus sont des éléments clés pour caractériser les domaine d'attraction des types extrêmes.

Définition 1.1.5. On dit qu'une variable aléatoire Z est dans le domaine d'attraction max-stable d'une distribution des valeurs extrêmes G_γ , s'il existe des suites $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ telles que : $F_Z^n(a_n z + b_n)$ converge vers G_γ , i.e.,

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) = F_Z^n(a_n z + b_n) \rightarrow G_\gamma, \quad n \rightarrow \infty, \quad (1.15)$$

et nous notons $Z \in MDA(G_\gamma)$.

Autrement dit, F_Z^n est de même type que G_γ , i.e que la classe des distributions max-stables coïncide avec la classe des trois types extrêmes : Gumbel, Fréchet, Weibull. En particulier le MDA est complètement caractérisé par la notion de distribution équivalente en queue dont nous définissons maintenant.

Définition 1.1.6. Les fonctions de distribution F_1 et F_2 sont dites de queues équivalentes si elles ont le même point terminal droit $z_{F_1} = z_{F_2} = z_F$ et si

$$\lim_{z \rightarrow z_F} \frac{1 - F_1(z)}{1 - F_2(z)} = c, \quad \text{avec } 0 < c < \infty, \quad (1.16)$$

1.1. EXTRÊMES UNIVARIÉS

En d'autres termes, chaque domaine d'attraction max-stable MDA est fermé sous l'équivalence en queue, i.e. si F_1 et F_2 sont équivalentes en queue, $F_1 \in MDA(G_\gamma)$ si et seulement $F_2 \in MDA(G_\gamma)$. Les résultats ci-dessus donnent une caractérisation unifiée des domaines d'attraction des trois types extrêmes (de Haan et Ferreira (2006)).

Théorème 1.1.5. Soit $\gamma \in \mathbb{R}$ et $U(z) = F^\leftarrow\left(1 - \frac{1}{z}\right)$ pour tout $z > 0$, la fonction quantile de queue. Les assertions suivantes sont équivalentes :

1. $F \in MDA(G_\gamma)$.
2. Il existe une fonction positive $a(\cdot)$ telle que pour tout $z > 0$,

$$\lim_{t \rightarrow \infty} \frac{U(tz) - U(t)}{a(t)} = D_\gamma(z) = \begin{cases} \frac{z^\gamma - 1}{\gamma}, & \text{si } \gamma \neq 0 \\ \log z, & \text{si } \gamma = 0 \end{cases} \quad (1.17)$$

3. Il existe une fonction positive $f(\cdot)$ telle que pour tout z vérifiant $1 + \gamma z > 0$,

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(t + zf(t))}{\bar{F}(t)} = \begin{cases} (1 + \gamma z)^{-1/\gamma}, & \text{si } \gamma \neq 0 \\ e^{-z}, & \text{si } \gamma = 0 \end{cases} ; \quad (1.18)$$

où f est une fonction qui peut être choisie telle que $f(t) = a(1/\bar{F}(t))$, où $a(\cdot)$ est la fonction utilisée dans (1.17).

En supposant que $U(\infty) > 0$, la condition (1.17) entraîne que :

$$\lim_{t \rightarrow \infty} \frac{U(tz) - U(t)}{a(t)/U(t)} = \begin{cases} \log z, & \text{si } \gamma \geq 0 \\ \frac{z^\gamma - 1}{\gamma}, & \text{si } \gamma < 0 \end{cases} . \quad (1.19)$$

Le résultat (1.19) est très intéressant car il est utilisé pour proposer l'estimateur des moments du paramètre de forme γ , pour plus de détails voir Dekkers *et al.* (1989). La preuve sur les résultats de caractérisation unifiée du domaine d'attraction max-stable peut être retrouvée dans Embrechts *et al.* (1997). Les caractérisations de chacun des trois domaines d'attraction de Fréchet, Gumbel et de Weibull ont été proposées dans (Leadbetter (1983), Resnick (1987), Embrechts *et al.* (2013), de Haan et Ferreira (2006)).

c. Inférence des maxima par blocs

En pratique pour décrire un modèle statistique associé aux données extrêmes par la méthode des maxima par blocs, on utilise le résultat unifié (1.5) en commençant par séparer ces données en k blocs de même taille, avec n observations. La figure (1.2) illustre une telle partition des données. En particulier n doit être fixé assez haut que possible pour garantir des blocs avec suffisamment d'observations afin d'assurer une convergence adéquate de la distribution limite des maxima sélectionnés. Néanmoins, un nombre suffisant de maxima indépendants est nécessaire pour garantir une bonne inférence, si bien que k doit être fixé aussi haut que possible, ce qui conduit à un compromis entre le biais et la variance. Il convient de noter

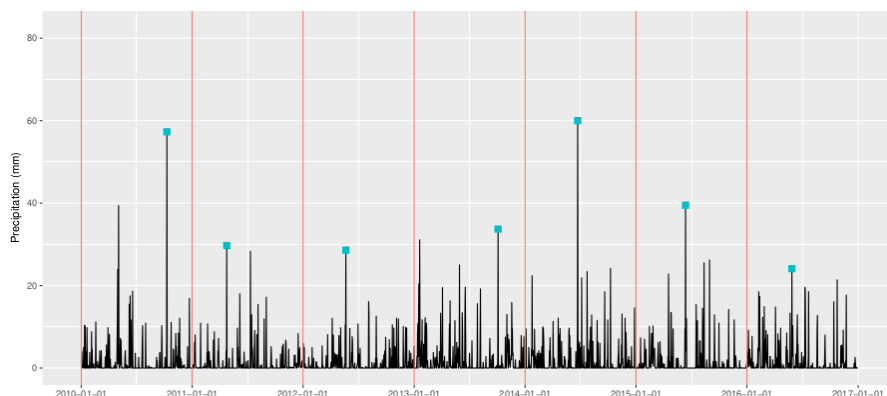


FIGURE 1.2 – Maxima annuel de précipitations (mm) enregistrés à la station synoptique de Ouagadougou.

que la taille du bloc est généralement donnée par le contexte dans de nombreuses applications pratiques, telles les observations quotidiennes, mensuelles ou annuelles dans le domaine du climat et de l'environnement. On peut alors inférer les paramètres (μ, σ, γ) du modèle statistique adapté aux données en utilisant les approches basées sur les moments, moments pondérés, le maximum de vraisemblance ou des méthodes bayésiennes. Cependant cette approche des maxima par blocs semble gaspiller l'information, car d'autres observations extrêmes comme les secondes valeurs les plus importantes dans les blocs devraient également fournir des informations pertinentes sur les extrêmes et sont parfois plus grandes que la plus grande valeur de certains blocs. De plus une modélisation détaillée des données extrêmes est très souvent nécessaire, car une succession de valeurs extrêmes peuvent engendrer plus de risque qu'une seule valeur extrême, mais l'analyse de maxima par bloc ne permet pas de modéliser des grappes d'événements rares. Pour pallier ce problème, une alternative possible dans la modélisation des événements extrêmes est de s'intéresser aux excès au-dessus d'un seuil (choisi en fonction du système et de la question posée). Cette stratégie assez complémentaire de l'analyse des maxima a l'avantage d'être plus efficace du point de vue statistique car elle offre un échantillon de taille plus grande que celui des maxima et prend notamment en compte les insuffisances relevées.

1.1.2. Approche des dépassements de seuils

a. Description et résultats fondamentaux

L'approche de dépassement de seuil (Peak Over Threshold (POT) en anglais) repose sur le comportement des valeurs observées au-delà d'un certain seuil donné. En d'autres termes, elle consiste non pas à observer les maxima ou les plus grandes valeurs mais à ne retenir que les observations dépassant un certain seuil fixé généralement assez grand. L'idée de base de cette approche consiste à choisir un seuil suffisamment élevé et à étudier les excès au-delà de ce seuil. Soient $Z_1, \dots, Z_n$ une suite de v.a *i.i.d* de distribution commune F et de point terminal droit $z_F = \sup \{z : F(z) < 1\}$. On définit un seuil $u \in \mathbb{R}$, $n_u = \text{card} \{t : t = 1, \dots, n, Z_t > u\}$

1.1. EXTRÊMES UNIVARIÉS

et $Y = \{Z - u \mid Z > u\}$ telle que $Y_t = Z_t - u$ pour $0 \leq t \leq n_u$, où n_u est le nombre de dépassements de seuil u et t est la date du j ème dépassement, $j = 1, \dots, n_u$. La figure (1.3) illustre une telle sélection des excès. On cherche alors à partir de la distribution F de Z à définir une loi conditionnelle F_u des excès Y de Z au-delà d'un seuil u . Plus précisément on appelle excès de la variable aléatoire Z au-delà

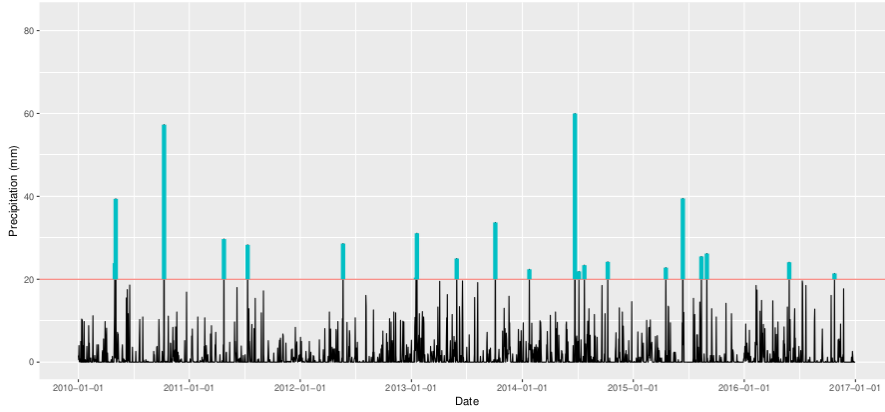


FIGURE 1.3 – Dépassements supérieurs à $u = 20\text{mm}$ de pluies enregistrés à la station synoptique de Ouagadougou.

d'un seuil $u < z_F$, la variable aléatoire Y qui prend ses valeurs dans $]0, z_F - u[$, définie par $Y = \{Z - u \mid Z > u\}$, $u < z_F$.

Définition 1.1.7. Soit Z une variable aléatoire de fonction de répartition F et de point terminal droit z_F . On appelle distribution des excès de la v.a Z au-delà d'un seuil $u < z_F$, la distribution de probabilité de la v.a conditionnelle $Y = \{Z - u \mid Z > u\}$ noté F_u et définie par : $F_u(y) = P(Z - u \leq y \mid Z > u)$, $y \in \mathbb{R}$ telle que

$$\text{Pour tout } y \in \mathbb{R}, F_u(y) = \begin{cases} 0 & \text{si } y \leq 0 \\ 1 - \frac{1 - F(u + y)}{1 - F(u)} & \text{si } 0 < y < z_F - u \\ 1 & \text{si } y \geq z_F - u \end{cases}, \quad (1.20)$$

Pickands (1975) et Balkema et De Haan (1974) établissent que seule la distribution de Pareto Généralisée modélise la variable excédentaire $(Y_1, \dots, Y_{n_u})$ sous certaines conditions de convergence. Ce résultat fondateur de l'analyse des dépassements de seuil est résumé par le théorème suivant.

Théorème 1.1.6. Soit F une distribution appartenant au domaine d'attraction d'une distribution de valeurs extrêmes G_γ . S'il existe une fonction σ_u strictement positive mesurable et $\gamma \in \mathbb{R}$ telles que :

$$\lim_{u \rightarrow z_F} \sup_{0 \leq y \leq z_F - u} |F_u(y) - H_\gamma(y \mid \sigma_u)| = 0, \quad (1.21)$$

où F_u est la fonction de répartition conditionnelle des dépassements de seuil, H_γ est appelé distribution de Pareto Généralisée (GPD) et définie par

$$H_\gamma(y \mid \sigma_u) = \begin{cases} 1 - (1 + \gamma y / \sigma_u)_+^{-1/\gamma}, & \gamma \neq 0 \\ 1 - \exp(-y / \sigma_u), & \gamma = 0 \end{cases}, \quad (1.22)$$

où $y \geq 0$, si $\gamma \geq 0$ et $0 \leq y \leq -\sigma_u / \gamma$ si $\gamma < 0$.

Le théorème (1.1.6) permet d'établir une relation étroite entre la loi *GPD* et la loi *GEV*. En effet si $F \in MDA(G_\gamma)$, alors la distribution de Pareto généralisée modélise le comportement des excès au dessus d'un seuil suffisamment grand. Le résultat suivant donne une formulation équivalente du théorème 1.1.6

Théorème 1.1.7. *Soit $Z_1, Z_2, \dots$ une suite de variables aléatoires i.i.d de distribution F et $(a_n)_{n \geq 1} > 0$, $(b_n)_{n \geq 1} \in \mathbb{R}$ des suites normalisantes telle que*

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \longrightarrow G_\gamma(z|\sigma, \mu), n \longrightarrow \infty,$$

où G est une distribution des valeurs extrêmes (*GEV*) non dégénérée. Alors

$$P(Z > u_n(u+z) \mid Z > u_n(u)) \longrightarrow 1 - H_\gamma(z \mid \sigma_u), n \longrightarrow \infty, \quad (1.23)$$

avec $u_n(z) = a_n z + b_n$ pour tout $z \in]0; \infty[$ et

$$H_\gamma(z \mid \sigma_u) = \begin{cases} 1 - (1 + \gamma z / \sigma_u)_+^{-1/\gamma}, & \gamma \neq 0 \\ 1 - \exp(-z / \sigma_u), & \gamma = 0 \end{cases}, \quad (1.24)$$

où $\sigma_u = \sigma + \gamma(u - \mu)$. La distribution limite est appelée *distribution de Pareto Généralisée GPD* de paramètre d'échelle σ_u et de forme γ .

Ainsi pour un seuil u suffisamment grand, la distribution des dépassements de seuil est approximée par une *GPD* dont les paramètres sont déterminés de manière unique à partir de ceux de la *GEV* donnée par l'équation (1.5). Le paramètre de forme γ est égal à celui de la *GEV* correspondante. Le paramètre d'échelle σ_u est étroitement lié au paramètre de forme et d'échelle de la loi *GEV* telle que $\sigma_u = \sigma + \gamma(u - \mu)$. Comme pour la loi *GEV*, le paramètre γ caractérise le comportement de la queue de la distribution. Si $\gamma < 0$, la distribution des excès a une limite supérieure égale à $u - \sigma_u / \gamma$. Si $\gamma > 0$, la distribution n'a pas limite supérieure. Par contre si $\gamma = 0$, la distribution n'a pas de limite supérieure finie et doit être interprétée comme la limite lorsque $\gamma \longrightarrow 0$, ce qui conduit à une distribution exponentielle de paramètre $1/\sigma_u$. La figure 1.4 illustre l'effet des paramètres sur la fonction de répartition et densité de la loi *GPD*.

A partir des équations (1.20) et (1.22) [Davison et Smith \(1990\)](#) proposent d'approximer la queue supérieure de la fonction de distribution F de la variable aléatoire Z par

$$F(z) = (1 - \phi_u) + \phi_u H_\gamma(z - u \mid \sigma_u), z > u \quad (1.25)$$

où $u > 0$ est un seuil suffisamment élevé, et $\phi_u = P(Z > u) = 1 - F(u)$ représente la fonction de survie ou la probabilité que Z dépasse le seuil u .

Une alternative assez récente dans analyse statistique des extrêmes par dépassement de seuil consiste à reformuler les résultats précédents en termes de mesures de convergence vague ([de Fondeville et Davison \(2020\)](#)) et basée sur la variation régulière (voir proposition 1.1.1). En effet si $F \in MDA(G_\gamma)$, alors grâce au développement limité de Taylor du logarithme de l'équation (1.15) nous avons le résultat équivalent suivant :

$$n[1 - F(a_n z + b_n)] \approx -\log \{G_\gamma(z)\}, n \longrightarrow \infty, \quad (1.26)$$

1.1. EXTRÊMES UNIVARIÉS

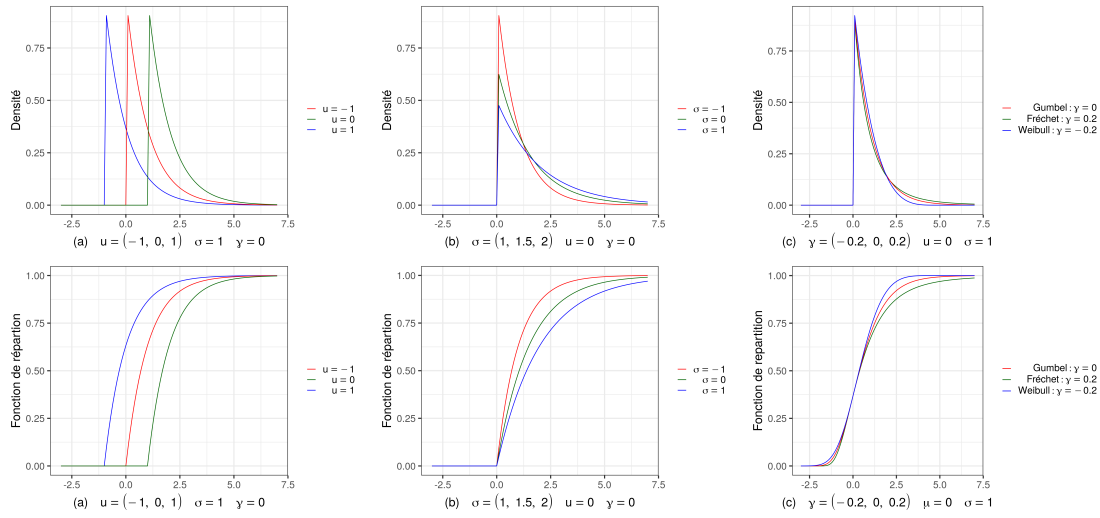


FIGURE 1.4 – Effet des paramètres sur la fonction de répartition (ligne2) et la densité de probabilité (ligne1) d’une distribution GPD. Dans chaque figure, deux paramètres sont maintenus constants, à la troisième, trois valeurs différentes sont attribuées. Dans (a) le paramètre de localisation est modifié, dans (b) l’échelle et dans (c) la forme.

pour tout $z \in \mathbb{R}$ et $1 + \gamma(z - b_n)/a_n > 0$. En reformulant alors l’équation (1.26) en termes de $\hat{w}$ -convergence vague de mesures, nous obtenons le résultat suivant.

Théorème 1.1.8. *Supposons qu’il existe $a_n > 0$, $b_n \in \mathbb{R}$, et $\gamma \in \mathbb{R}$ telles que*

$$\left. \begin{aligned} nP \left\{ \left(1 + \gamma \frac{Z - b_n}{a_n} \right)_+^{1/\gamma} > z \right\} \\ nP \left\{ \exp \left(\frac{Z - b_n}{a_n} \right)_+ > z \right\} \end{aligned} \right\} \xrightarrow{\hat{w}} \Lambda(z), \quad (1.27)$$

où $\{z\}_+ = \max(0, z)$. Alors soit Λ est dégénérée, soit $\Lambda([z, +\infty[) = z^{-1}$.

De façon analogue aux maxima par bloc, les conditions d’existence des suites normalisantes $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ telles que Λ ne soit pas dégénérée sont liées à la notion de variation régulière. En effet, si Z est à variation régulière d’indice $-\gamma/\sigma < 0$, alors il existe des suites $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ telles que

$$\lim_{n \rightarrow \infty} nP \left\{ \frac{Z - b_n}{a_n} > z \right\} = (1 + \gamma z/\sigma)_+^{-1/\gamma}, \quad n \rightarrow \infty, \quad (1.28)$$

dans $\mathcal{M}_+([0, +\infty[))$, où de manière similaire aux maxima par bloc, $a_n \rightarrow \infty$ lorsque $n \rightarrow \infty$ et $(b_n)_{n \geq 1}$ peuvent être choisis égaux à zéro. D’un point de vue pratique, le théorème (1.1.8) a l’avantage de fournir une base pour une description unifiée des queues de distribution en termes de dépassement de seuil. Les résultats précédents peuvent également être formulés en terme de processus ponctuels des dépassements de seuil suffisamment grand ; voir Leadbetter (1983), Falk *et al.* (1994) et Embrechts *et al.* (1997).

b. Processus ponctuels des dépassements

Dans l'approche des processus ponctuels pour les extrêmes, la notion d'événements rares se définit non pas en terme de maxima par bloc mais en terme de dépassements de seuil suffisamment grand comme dans l'approche POT. Les deux approches de la théorie des valeurs extrêmes univariées décrites dans les sections (1.1.1, 1.1.2) sont étroitement liées avec la théorie des processus ponctuels et les processus ponctuels de Poisson. Ici nous nous intéresserons aux processus ponctuels des excès qui comptent le nombre de dépassements d'une suite finie de variables aléatoires au-delà d'un certain seuil $u \in \mathbb{R}$ fixé.

Définition 1.1.8. *Un processus de Poisson de mesure d'intensité Λ est un processus ponctuel qui satisfait aux deux propriétés suivantes :*

1. Pour tout $k \geq 0$ et tout ensemble mesurable $A \subset \mathcal{E}$

$$P(N(A) = k) = \begin{cases} \frac{(\Lambda(A))^k}{k!} e^{-\Lambda(A)}, & \text{si } \Lambda(A) < \infty \\ 0, & \text{si } \Lambda(A) = \infty \end{cases}.$$

2. Pour tout $n \in \mathbb{N}$, si $A_1, \dots, A_n \in \mathcal{E}$ deux à deux disjoints, alors les variables aléatoires $N(A_1), \dots, N(A_n)$ sont mutuellement indépendantes.

Définition 1.1.9. Soit $(Z_n, n \geq 1)$ une suite de variables aléatoires iid et $u \in \mathbb{R}$ un seuil fixé. On appelle processus ponctuel des excès noté N_n associé à l'échantillon des variables aléatoires $\{Z_i, i = 1, \dots, n\}$ dépassant le seuil u , le processus à espace d'état $E =]0, 1]$ défini par :

$$N_n(A) = \sum_{i=1}^n \delta_{\frac{i}{n}}(A) \mathbb{1}_{[u, \infty[}(Z_i), \quad A \subset E, \quad n \in \mathbb{N}^*. \quad (1.29)$$

En particulier si $A = E =]0, 1]$, alors $N(A)$ représente le nombre de variables aléatoires de tout l'échantillon $\{Z_i, i = 1, \dots, n\}$ dépassant le seuil u . Le résultat clé donné par [Coles et al. \(2001\)](#) est résumé dans le théorème suivant :

Théorème 1.1.9. Soit $(Z_n, n \geq 1)$ une suite de variables aléatoires iid de fonction de répartition F et $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ des suites normalisantes réelles telles que pour tout $z \in \mathbb{R}$,

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \longrightarrow G_\gamma(z), \quad n \longrightarrow \infty, \quad (1.30)$$

où M_n est définie comme dans (1.1.1) et G_γ prend la forme définie dans (1.5). Alors le processus ponctuel N_n défini par

$$N_n = \left\{ \left(\frac{i}{n+1}, \frac{Z_i - b_n}{a_n} \right) : i = 1, \dots, n \right\}, \quad (1.31)$$

Converge dans des régions de la forme $]0, 1[\times [u, \infty[$ vers un processus Ponctuel de Poisson de mesure d'intensité $\Lambda(A_z) = (t_2 - t_1) \left[1 + \gamma \left(\frac{z - \mu}{\sigma} \right) \right]_+^{-1/\gamma}$, où $A_z = [t_1, t_2] \times [z, \infty[$, pour tout $z > u$ et $0 \leq t_1 < t_2 \leq 1$.

1.1. EXTRÊMES UNIVARIÉS

La preuve de ce théorème peut être retrouvé dans ([Embrechts et al. \(1997\)](#), [Coles et al. \(2001\)](#), [Dalhoumi \(2017\)](#)).

Le processus de Poisson est dit homogène si sa mesure d'intensité Λ est proportionnelle à la mesure de Lebesgue notée μ . Dans ce cas, il existe une constante κ telle que $\Lambda(A) = \int_A \kappa dx = \kappa \mu(A)$ (Processus de poisson non homogène). Le théorème suivant résume les équivalences entre les trois approches.

Théorème 1.1.10. *Soit $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ des suites normalisantes réelles et G_γ une loi GEV de support $[z_G, z^G]$ telle que $\bar{H}_\gamma(z) = -\log(G_\gamma(z))$, où z_G et z^G désignent respectivement la borne inférieure et supérieure de G_γ . Les assertions suivantes sont équivalentes :*

1. $F^n(a_n z + b_n) \longrightarrow G_\gamma(z)$, $n \rightarrow \infty$ pour tout $z < z^G$ (maxima),
2. $\exists f(t) > 0$ telle que $\frac{\bar{F}(u + f(t)z)}{\bar{F}(u)} \longrightarrow \bar{H}_\gamma(z)$, $u \rightarrow \infty$, pour tout $z_G < z < z^G$ (Survie des excès),
3. $N_n(\cdot) = \sum_{i=1}^n \mathbb{1}_{\{i/(n+1), (Z_i - b_n)a_n^{-1}\}}(\cdot) \longrightarrow N$, $n \rightarrow \infty$, où N est un processus de Poisson sur $[0, 1] \times [z_G, z^G]$ d'intensité $\Lambda([0, 1] \times [z, \infty]) = (t_2 - t_1) \bar{H}_\gamma(z)$ (Processus Ponctuels).

En pratique pour utiliser l'approche des processus ponctuels des excès, nous devons d'abord déterminer un seuil u au-dessus duquel l'approximation du processus de Poisson est appropriée; ceci peut être fait en utilisant des techniques similaires à celles décrites dans la section (1.1.2, c)). Ensuite, on définit $A_u =]0, 1[\times [u, \infty[$. Les points qui se trouvent à l'intérieur de A_u sont notés par : $\{(t_1, z_1), \dots, (t_{N(A_u)}, z_{N(A_u)})\}$ de sorte qu'on ait au total $N(A_u)$ points dans la région A_u . Les estimateurs du maximum de vraisemblance peuvent alors être trouvés en utilisant la vraisemblance

$$L(\mu, \sigma, \gamma, \mathbf{z}) = \exp \left\{ -k \left[1 + \gamma \left(\frac{u - \mu}{\sigma} \right) \right]_+^{-1/\gamma} \right\} \prod_{i=1}^{N(A_u)} \frac{\{1 + \gamma \left(\frac{z_i - \mu}{\sigma} \right)\}_+^{-1/\gamma - 1}}{\sigma} \quad (1.32)$$

Cette expression (1.32) est obtenue à partir de la vraisemblance du processus de Poisson, à savoir

$$L(\mu, \sigma, \gamma, \mathbf{z}) = \exp \{-\Lambda(A_u)\} \prod_{i=1}^{N(A_u)} \lambda(t_i, z_i); \quad (1.33)$$

où $\Lambda(A_u) = \int_{t_1}^{t_2} \lambda(t, z) dt$. L'approche des extrêmes par les processus ponctuels diffère légèrement de celle qui consiste à ajuster une GEV aux maxima des blocs, comme dans la section (1.1.1), puisque toutes les données supérieures au seuil u sont utilisées pour estimer les paramètres du modèle; cela devrait permettre une inférence plus précise.

c. Procédures de sélection des seuils

En pratique, le choix d'un seuil approprié constitue une difficulté. En fait, u doit être assez grand pour que l'approximation *GPD* soit valide, mais pas trop élevé de façon à maximiser le nombre de dépassements à utiliser pour obtenir une inférence plus précise et réduire autant que possible l'incertitude sur l'estimation des paramètres. Une multitude de techniques statistiques ont été proposées dans la littérature pour la sélection d'un seuil approprié, mais nous rappelons ici celles les plus utilisées.

Parmi les méthodes graphiques de diagnostic visuel, nous pouvons citer le graphe de la fonction "mean excess plot" et celui de la stabilité des paramètres. La méthode de stabilité des paramètres estimés consiste à ajuster une distribution *GPD* à une série de seuils croissants et à choisir le seuil le plus bas à partir duquel on a un comportement linéaire des paramètres. Les graphiques $\{(u, \hat{\sigma}_u) : u \leq z_{\max}\}$ et $\{(u, \hat{\gamma}_u) : u \leq z_{\max}\}$ des paramètres d'échelle et de forme estimés sont tracés en fonction de u et le seuil approprié correspond à la plus petite valeur de u à partir de laquelle les estimations de γ restent quasiment constantes et les estimations de σ_u linéaires (voir Figure 1.5).

L'approche "mean excess plot" consiste à trouver la moyenne des excès sur une série de seuil u et à choisir le seuil à partir duquel cette moyenne est approximativement linéaire. Sous l'hypothèse que les excès de seuil u suivent une distribution *GPD*, [Davison et Smith \(1990\)](#), propose d'utiliser la moyenne des excès donnée par :

$$E[Z - u \mid Z > u] = \frac{\sigma_u}{1 - \gamma}, \quad (1.34)$$

avec $\gamma < 1$, (si $\gamma \geq 1$ la moyenne est infinie) appelée fonction "mean excess plot" (ou de vie résiduelle moyenne) de Z . En suivant la propriété de stabilité de seuil de la distribution *GPD* évoquée précédemment, pour tout seuil $v > u$, l'espérance (1.34) prend la forme suivante :

$$E[Z - u \mid Z > u] = \frac{\sigma_v}{1 - \gamma} = \frac{\sigma_u + \gamma(v - u)}{1 - \gamma}, \text{ avec } \gamma < 1, \quad (1.35)$$

qui est linéaire en v avec une pente $\gamma/(1 - \gamma)$. Étant donné un échantillon indépendant et identiquement distribué $Z_1, \dots, Z_n$, l'estimateur empirique de la fonction "mean excess plot" est donné par :

$$\hat{e}(u) = \frac{\sum_{t=1}^{n_u} (Z_t - u) \mathbb{1}_{\{Z_t > u\}}}{\sum_{t=1}^{n_u} \mathbb{1}_{\{Z_t > u\}}}, \quad (1.36)$$

$\mathbb{1}_{\{Z_t > u\}} = 1$ si $Z > u$ et 0 sinon. Un seuil approprié est alors obtenu en traçant le graphe $\hat{e}(u)$ en fonction de u et en identifiant la plus petite valeur du seuil u au-dessus de laquelle $\hat{e}(u)$ est approximativement linéaire en v (voir Figure 1.5). Les graphiques des excès moyen et de stabilité du seuil peuvent être appliqués simultanément pour obtenir le seuil optimal. Les principaux inconvénients des approches

1.1. EXTRÊMES UNIVARIÉS

graphiques décrites ci-dessus en matière de sélection de seuils résident dans le fait qu'elles nécessitent l'expertise de l'analyste pour l'interprétation de ces diagnostics et qu'elles peuvent être assez subjectives (voir [Coles et al. \(2001\)](#) pour plus de détails).

Une autre approche alternative pour sélectionner le seuil optimal consiste à choisir u qui rend la distribution empirique des dépassements observés aussi proche que possible du modèle GPD ajusté. Il existe plusieurs méthodes qui visent à mesurer la distance entre deux distributions, mais pour les données non normales, la plus couramment utilisée est la statistique de Kolmogorov-Smirnov $D_u^{(1)}$, qui mesure la distance maximale entre la fonction de répartition empirique et la fonction de répartition ajustée :

$$D_u^{(1)} = \sup_y \left| F_{emp,u}(y) - H_{\hat{\gamma}_u}(y \mid \hat{\sigma}_u) \right|, \quad (1.37)$$

où $F_{emp,u}(y) = \frac{1}{k} \sum_{t=1}^k \mathbb{1}_{\{Y_t < y\}}$ est la distribution empirique des dépassements de seuil u et $H_{\hat{\gamma}_u}(y \mid \hat{\sigma}_u)$ la distribution GPD ajustée. L'objectif est alors de trouver le seuil u qui minimise $D_u^{(1)}$. Pour évaluer le seuil optimal basé sur cette procédure, on construit un graphique de "distance entre les distributions", représenté par $\{(u, D_u^{(1)}) : u \leq y_{\max}\}$. Il faut remarquer que la statistique de Kolmogorov-Smirnov, est très souvent élevée pour des valeurs de u très grand puis chute et se stabilise rapidement lorsque u diminue. Par conséquent le seuil donné par cette procédure est généralement très élevé, mais il peut servir à déterminer une limite supérieure de celui-ci.

Parmi les règles empiriques les plus couramment utilisées pour sélectionner la valeur du seuil, on trouve la méthode des percentiles ou méthode des quantiles. Dans ce cas, après avoir spécifié une valeur appropriée du percentile, le seuil est sélectionné de manière à ce qu'il corresponde au percentile de la série temporelle d'intérêt. La valeur du percentile choisie va dépendre de la nature des données traitées et de l'expertise de l'analyste à interpréter ces résultats. Malgré sa simplicité, le principal inconvénient de cette méthode est sa subjectivité qu'elle implique. Dans la littérature pertinente, une série de percentiles a été proposée. Pour plus de détails sur la méthode des quantiles voir [DuMouchel \(1983\)](#), [Grabemann et Weisse \(2008\)](#), [Arns et al. \(2013\)](#), [Eastoe et Tawn \(2012\)](#)

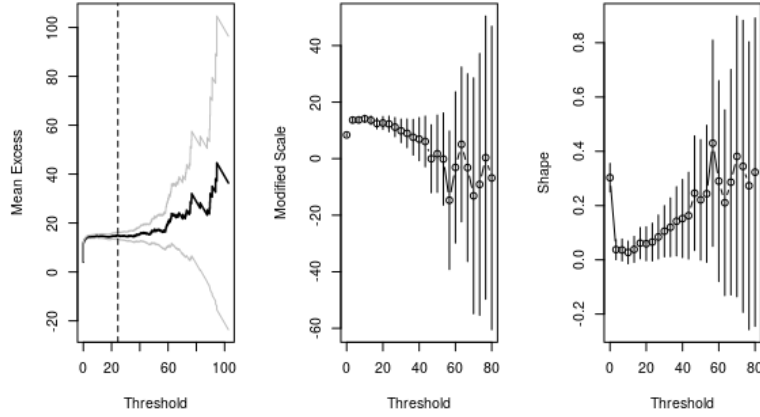


FIGURE 1.5 – Panneau de sélection des seuils de pluies enregistrés pour la station synoptique de Ouagadougou. Mean excess plot (gauche) et stabilité des paramètres (droite).

1.1.3. Niveau de retour

La gravité d'un évènement climatique extrême tel qu'une inondation est généralement communiquée via la probabilité de dépassement annuel ou la période de retour. Dans de nombreuses disciplines appliquées, la période et le niveau de retour sont utilisés pour décrire et quantifier le risque. Par exemple dans un climat stationnaire, un évènement extrême avec une période de retour $m = 100$ ans est en moyenne susceptible de se produire une fois tous les 100 ans, mais avec une probabilité de dépassement annuel de 1% de se produire au cours d'une année donnée. Dans ce cas on fait l'hypothèse que la distribution de probabilité sous-jacente reste inchangée au fil du temps (d'où la notion stationnarité). En d'autres termes dans un environnement stationnaire, la période de retour m , d'un évènement particulier z_m est l'inverse de la probabilité que l'évènement z_m soit dépassé au cours d'une année donnée. Plus formellement, pour une variable aléatoire Z *i.i.d* de fonction de répartition F_Z , cela se traduit par :

$$m = \frac{1}{F_Z(Z > z_m)}, \quad (1.38)$$

où z_m est le niveau de retour associé à la période de retour m . Autrement dit z_m est le quantile de $(1 - 1/m)$ associé à la distribution F_Z et peut être dérivé en fonction de l'approche des extrêmes que nous utilisons.

a. Maxima par blocs

Dans le cas où les données sont ajustées par une distribution $GEV(\mu, \sigma, \gamma)$, les informations obtenues peuvent servir à estimer les niveaux de retour ou quantiles extrêmes au delà des observations. Ainsi le niveau de retour z_m obtenu à partir de

1.2. EXTRÊMES MULTIVARIÉS

l'équation (1.38) est donné par :

$$z_m = \begin{cases} \mu - \frac{\sigma}{\gamma} \left[1 - \left\{ -\log \left(1 - \frac{1}{m} \right) \right\}^{-\gamma} \right], & \gamma \neq 0 \\ \mu - \sigma \log \left\{ -\log \left(1 - \frac{1}{m} \right) \right\}, & \gamma = 0 \end{cases}, \quad (1.39)$$

avec $m > 1$. Comme évoqué précédemment, différentes méthodes peuvent être utilisées pour obtenir des estimateurs des paramètres $(\hat{\sigma}, \hat{\mu}, \hat{\gamma})$ des valeurs extrêmes $GEV(\sigma, \mu, \gamma)$. Une approche courante consiste à utiliser le maximum de vraisemblance pour obtenir les estimateurs de ces paramètres. Ces estimateurs peuvent être substitués dans l'équation (1.39) pour obtenir les estimations du niveau de retour z_m .

b. Méthode POT

Par contre lorsque les données sont ajustées par un modèle GPD, le résultat obtenu est une distribution conditionnelle pour la queue supérieure et un déconditionnement est nécessaire pour obtenir un modèle marginal des observations de Z , i.e $P(Z > z) = P(Z \leq u + z \mid Z > u) P(Z > u)$. Dans ce cas le niveau de retour z_m obtenu en résolvant l'équation (1.38) est donné par :

$$z_m = \begin{cases} \mu + \frac{\sigma_u}{\gamma} [(n_z m \phi_u)^\gamma - 1], & \gamma \neq 0 \\ \mu + \sigma_u \log(n_z m \phi_u), & \gamma = 0 \end{cases}, \quad (1.40)$$

où $\phi_u = P(Z > u) = \frac{n_u}{N_{obs}}$ est le taux de dépassement de seuil u , n_z représente le nombre d'observations par an et N_{obs} le nombre total d'observations. Là encore, comme dans le cas des maxima par bloc, les paramètres σ_u et γ peuvent être remplacés dans l'équation (1.40) par leur estimateur $\hat{\sigma}_u$ et $\hat{\gamma}$ respectivement afin de calculer les niveaux de retour z_m .

1.2. Extrêmes multivariés

Dans le cadre univarié, les dépassements de seuil u d'une variable aléatoire correspondant aux valeurs élevées sont définis grâce à la relation d'ordre sur $\mathbb{R}$. Pour un vecteur $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{id}) \in \mathbb{R}^d, i = 1, \dots, n$, la relation d'ordre n'existe plus et dans ce contexte, un dépassement de seuil peut avoir plusieurs définitions. Les méthodes définissant les valeurs extrêmes multivariées sont proposées par [Rootzén et al. \(2006\)](#), [Rootzén et al. \(2018a\)](#) Barnett (1976). Par exemple pour un vecteur $\mathbf{Z}_i$, on peut supposer que toutes les composantes de $\mathbf{Z}_i$ dépassent un certain seuil, ou au moins une des composantes de $\mathbf{Z}_i$ dépasse un seuil fixé, ou encore l'image par une fonctionnelle dépasse un seuil. Cette dernière approche permet de généraliser les deux précédentes.

1. prendre le vecteur de maxima "composantes par composantes"

$$\mathbf{M}_n = \left(\max_{i=1}^n Z_{i,1}, \dots, \max_{i=1}^n Z_{i,d} \right);$$

2. définir les excès selon une fonction $\ell : \mathbb{R}^d \rightarrow \mathbb{R}$,
3. prendre les maxima concomitants : les observations dont l'une au moins des composantes est égale à son maximum observé.

Les notions présentées dans cette section se concentrent sur les deux premières approches. Dans toute la suite, nous noterons les vecteurs par des lettres gras, par exemple $\mathbf{z} = (z_1, \dots, z_d) \in \mathbb{R}^d$. En particulier nous noterons $\mathbf{0} = (0, \dots, 0)$, $\mathbf{1} = (1, \dots, 1)$ et $\infty = (\infty, \dots, \infty)$. De plus, dans toute la suite les opérations sont effectuées composante par composante.

1.2.1. Variation régulière multivariée

Soit $\mathcal{C} \in \bar{\mathbb{R}}^d$ ($d \geq 1$) telle que $\mathcal{C} \cap (\mathbb{R}^d \setminus \{\mathbf{0}\}) \neq \emptyset$ un cône convexe, i.e. que pour tout $\mathbf{z} \in \mathcal{C}$, $t\mathbf{z} \in \mathcal{C}$ avec $t > 0$.

Définition 1.2.1. Soit $\mathcal{C}$ un cône convexe de $\bar{\mathbb{R}}^d$ ($d \geq 1$) et $\mathbf{1} \in \mathcal{C}$. Une fonction $F : \mathcal{C} \rightarrow \mathbb{R}^+$ est à variation régulière d'indice α au voisinage de l'infini, s'il existe une fonction limite $h(\mathbf{z})$ telle que

$$\lim_{t \rightarrow \infty} \frac{F(t\mathbf{z})}{F(t\mathbf{1})} = h(\mathbf{z}), \quad \text{pour tout } \mathbf{z} \in \mathcal{C} \quad (1.41)$$

Pour tout $c > 0$, la fonction h vérifie la propriété suivante : $h(c\mathbf{z}) = c^\alpha h(\mathbf{z})$. En particulier comme dans le cas univarié, si $\alpha = 0$, h est dite à variation lente.

Définition 1.2.2. Soit $\|\cdot\|$ une norme sur $\mathbb{R}^d$ et $\mathcal{S} = \{\mathbf{z} \in \mathbb{R}^d, \|\mathbf{z}\| = 1\}$ la sphère unitaire associée à $\|\cdot\|$. On appelle transformation pseudo-polaire

$$T : \mathbb{R}^d \setminus \{\mathbf{0}\} \rightarrow]0, \infty[\times \mathcal{S} \quad \text{telle que} \quad T(\mathbf{z}) = \left(\|\mathbf{z}\|, \frac{\mathbf{z}}{\|\mathbf{z}\|} \right) = (r, \mathbf{w}), \quad (1.42)$$

où le couple $(r, \mathbf{w})$ est appelé coordonnées pseudo-polaires. $r > 0$ est le rayon et $\mathbf{w} \in \mathcal{S}$ est l'angle.

La transformation T est un homéomorphisme qui vérifie $T^{-1}(r, \mathbf{w}) = r \frac{\mathbf{w}}{\|\mathbf{w}\|}$. En particulier si Λ est une mesure homogène d'indice $-\alpha$ définie sur un cône convexe $\mathcal{C} \setminus \{\mathbf{0}\}$, pour tout borélien $A \subset \mathcal{C}$, Λ est homogène d'ordre $-\alpha$, i.e. $\Lambda(tA) = t^{-\alpha} \Lambda(A)$. En coordonnées pseudo-polaires, cette propriété d'homogénéité est traduite par :

$$\Lambda \circ T^{-1}(d(r, \mathbf{w})) = \alpha r^{-(\alpha+1)} dr \times \sigma(d\mathbf{w}), \quad (1.43)$$

dans $\mathcal{M}_+([0, \infty] \times \sigma)$ où σ est une mesure de Radon définie sur $\mathcal{S}_\mathcal{C} = \mathcal{S} \cap \mathcal{C}$ appelé mesure angulaire ou très souvent mesure spectrale. La variation régulière multivariée à l'infini est un outil souple et puissant permettant de caractériser la queue de distribution d'un vecteur aléatoire dans un cône convexe en terme de convergence vague de mesure (Resnick (1987)).

Théorème 1.2.1. Soit $\mathbf{Z} \geq 0$ un vecteur aléatoire de distribution F définie sur $[0, \infty]^d$ et $\mathcal{C} = [0, \infty]^d \setminus \{\mathbf{0}\}$. Les assertions suivantes sont équivalentes :

1.2. EXTRÊMES MULTIVARIÉS

1. Il existe une mesure de Radon Λ sur $\mathcal{C}$ telle que :

$$\lim_{t \rightarrow \infty} \frac{1 - F(t\mathbf{z})}{1 - F(t\mathbf{1})} = \lim_{t \rightarrow \infty} \frac{P(t^{-1}\mathbf{Z} \in [\mathbf{0}, \mathbf{z}]^c)}{P(t^{-1}\mathbf{Z} \in [\mathbf{0}, \mathbf{1}]^c)} = \Lambda([\mathbf{0}, \mathbf{z}]^c), \quad (1.44)$$

pour tout point de continuité $\mathbf{z} \in \mathcal{C}$ de la mesure limite $\Lambda([\mathbf{0}, \cdot]^c)$

2. Il existe $b(t) \rightarrow \infty$ et une mesure de Radon Λ sur $\mathcal{C}$ tels que :

$$\lim_{t \rightarrow \infty} P\left(\frac{\mathbf{Z}}{b(t)} \in \cdot\right) \xrightarrow{v} \Lambda(\cdot), \quad \text{dans } \mathcal{M}_+(\mathcal{C}), \quad (1.45)$$

où $\xrightarrow{v}$ désigne la convergence vague.

3. Il existe une mesure de probabilité $\sigma(\cdot)$ dans les boréliens $\mathcal{S}_{\mathcal{C}}$, une constante $c > 0$, et une fonction $b(t) \rightarrow \infty$ telle que pour tout $(r, \mathbf{w}) = \left(\|\mathbf{Z}\|, \frac{\mathbf{Z}}{\|\mathbf{Z}\|}\right)$, on ait

$$\lim_{t \rightarrow \infty} tP\left[\left(\frac{r}{b(t)}, \mathbf{w}\right) \in \cdot\right] \xrightarrow{v} c\Lambda_{\alpha} \times \sigma(\cdot), \quad \text{dans } \mathcal{M}_+(\mathcal{C}) \quad (1.46)$$

1.2.2. Extension GEV multivariée

Considérons un échantillon d'observations d-dimensionnelle $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{id})$, $i = 1, \dots, n$ et définissons $\mathbf{M}_n = (M_{n1}, \dots, M_{nd})$ un vecteur des maxima composante par composante (Beirlant *et al.* (2004)) dont les éléments sont définis par :

$$M_{nj} = \max_{1 \leq i \leq n} Z_{ij} \quad (j = 1, \dots, d).$$

Le vecteur $\mathbf{M}_n$ n'est pas nécessairement un point des données, car le maximum de chaque composante peut provenir des différentes observations. Etant donné une suite de vecteur aléatoires *i.i.d* $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ de $\mathbb{R}^d$ ($d \geq 2$) de distribution commune F , pour tout $\mathbf{z} \in \mathbb{R}^d$, on a :

$$P(\mathbf{M}_n \leq \mathbf{z}) = P(\mathbf{Z}_1 \leq \mathbf{z}, \dots, \mathbf{Z}_n \leq \mathbf{z}) = F^n(\mathbf{z}), \quad (1.47)$$

où les opérations sont effectuées composante par composante. De façon analogue au cadre univarié, on s'intéresse à la distribution limite du vecteur aléatoire $\mathbf{M}_n$ (Smith (1990)). Soient $\mathbf{a}_n = (a_{n1} > 0, \dots, a_{nd} > 0)$ et $\mathbf{b}_n = (b_{n1}, \dots, b_{nd})$ tels que :

$$\lim_{n \rightarrow \infty} P\left(\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{z}\right) = \lim_{n \rightarrow \infty} F^n(\mathbf{a}_n \mathbf{z} + \mathbf{b}_n) = G(\mathbf{z}), \quad (1.48)$$

où G est une distribution multivariée d-dimensionnelle dont les marges sont non-dégénérées. Chacune des marges G_j , $j = 1, \dots, d$ suit une distribution GEV univariée comme décrite dans la section 1.1.1. De plus il existe également une analogie max-stable multivariée.

Définition 1.2.3. Une distribution multivariée G est dite max-stable si pour tout $k \in \mathbb{N}$, il existe des vecteurs $\mathbf{a}_k > 0$ et $\mathbf{b}_k$ tels que :

$$G^k(\mathbf{z}) = G(\mathbf{a}_k \mathbf{z} + \mathbf{b}_k). \quad (1.49)$$

1.2. EXTRÊMES MULTIVARIÉES

Tout comme dans le cadre univarié, une distribution G est une distribution des valeurs extrêmes multivariées si et seulement si elle satisfait la propriété de max-stabilité et en plus les maxima par composante suivent une distribution des valeurs extrêmes multivariée. En pratique, il est courant que l'étude des extrêmes multivariés se déroule en deux étapes : Les distributions marginales sont d'abord estimées en utilisant une méthodologie univariée (en ajustant une GEV aux maxima ou une GPD aux dépassements de seuil), ensuite ces distributions marginales sont standardisées à l'échelle de Fréchet unitaire, afin de traiter la structure de dépendance. En effet on peut toujours s'y ramener moyennant une transformation des marges en posant : $t_j(.) = -\frac{1}{\log(G_j(.))}$, $j = 1, \dots, d$. Dans ce cas la classe des distributions limites dans (1.48) est donnée par :

$$G(z_1, \dots, z_d) = \exp(-V((z_1, \dots, z_d))), \quad (z_1, \dots, z_d > 0). \quad (1.50)$$

Ces distributions sont connues sous le nom de distribution des valeurs extrêmes multivariées (MEV : Multivariate Extremes Values). La fonction V dans (1.50) est appelée mesure exposante ou encore fonction de dépendance et est définie par :

$$V(z_1, \dots, z_d) = \int_{S_d} \max_{1 \leq j \leq d} \left(\frac{w_j}{z_j} \right) dH(w_1, \dots, w_d), \quad (z_1, \dots, z_d > 0), \quad (1.51)$$

où H est une mesure sur le simplexe S_d de dimension $d - 1$ tel que :

$$S_d = \{w \in [0, 1]^d : \sum_{j=1}^d w_j = 1\},$$

qui satisfait :

$$\int_{S_d} w_i dH(w_1, \dots, w_d) = 1, \quad (i = 1, \dots, d).$$

La mesure exposante V satisfait à deux propriétés suivantes :

1. Pour tout $z_k > 0$, $V(\infty, \dots, \infty, z_k, \infty, \dots, \infty) = \frac{1}{z_k}$
2. La fonction V est homogène d'ordre -1, i.e. pour tout $c > 0$,

$$V(cz_1, \dots, cz_d) = c^{-1}V(z_1, \dots, z_d).$$

La spécification d'un modèle paramétrique pour la mesure d'exposant $V(.)$ dans (1.50), ou de manière équivalente pour la mesure spectrale dH , revient à restreindre la dépendance à une structure particulière. Il est donc essentiel de construire des modèles de dépendance flexibles, mais parcimonieux, qui peuvent également être facilement interprétés. Cependant en haute dimension, les modèles paramétriques flexibles pour les extrêmes sont difficiles à construire. Quelques modèles ont été proposés dans la littérature (voir, par exemple, [Coles et Tawn \(1991\)](#), [Joe et al. \(1992\)](#), [Ballani et Schlather \(2011\)](#)), mais certains souffrent généralement d'un manque de flexibilité pour les grands dimension et entraînent des problèmes de calcul et d'inférence.

1.3. EXTRÊMES FONCTIONNELS

1.2.3. Extension POT multivariée

Des extensions multivariées de la *GPD* univariée présentée dans la section (1.1.2) ont été principalement étudiée par Rootzén *et al.* (2006), Rootzén *et al.* (2018a), Rootzén *et al.* (2018b). Dans cette section nous présentons brièvement les détails de cette distribution multivariée. Supposons que la limite (1.48) soit satisfaite, le vecteur aléatoire d -dimensionnel $\mathbf{Z}$ a la propriété suivante :

$$\left\{ \frac{\mathbf{Z} - \mathbf{b}_n}{\mathbf{a}_n} \right\} \mid \{\mathbf{Z} \not\leq \mathbf{b}_n\}. \quad (1.52)$$

converge en distribution vers un vecteur aléatoire Y , où Y suit une distribution multivariée de Pareto Généralisée (MGPD) noté par H et la notation $\not\leq$ signifie qu'il existe i tel que $Z_i > b_i$. Contrairement au modèle *MEV*, il faut noter que les distributions marginales H_j des $Y_j, j = 1, \dots, d$ ne sont pas nécessairement des distributions *GPD* univariées. Ceci ne se produit que si les marges sont conditionnées d'être positives de sorte que :

$$H_j^+(y) = P(Y_j > y \mid Y_j > 0) = 1 - (1 + \gamma_j y / \sigma_j)_+^{-1/\gamma_j}, \quad j = 1, \dots, d, \quad (1.53)$$

où $z_+ = \max(0, z)$.

1.3. Extrêmes fonctionnels

Dans cette section nous présentons les résultats essentiels existant sur l'analyse des extrêmes spatiaux afin de faciliter leur utilisation pour modéliser et détecter des tendances spatio-temporelles dans les extrêmes. Nous nous concentrerons en particulier sur l'extension fonctionnelle des dépassements de seuils construite dans le cadre de la variation régulière fonctionnelle. Nous commençons par rappeler brièvement le cadre des processus spatiaux max-stables avant de nous étendre un peu plus sur les processus ℓ -Pareto généralisés.

1.3.1. Processus spatiaux et spatio-temporels

a. Définitions, notations et propriétés

Soient $S \subset \mathbb{R}^d (d \in \mathbb{N}^*)$ un sous ensemble compact de $\mathbb{R}^d$ et $T \subset \mathbb{R}^+$ un sous ensemble compact de $\mathbb{R}^+$, représentant respectivement le domaine spatial et temporel. Nous désignons par $\mathcal{C}(S)$ l'ensemble des fonctions continues à valeurs réelles définies sur S muni de la norme uniforme $\|f\|_\infty = \sup_{s \in S} |f(s)|$ et $\mathcal{C}_+(S)$ sa restriction aux fonctions positives. Nous notons également par $\mathcal{B}\{\mathcal{C}(S)\}$ une σ -algèbre des Boréliens associée à $\mathcal{C}(S)$. De plus, sauf indication contraire explicite, nous utiliserons la notation suivantes :

- $s \in S$ pour désigner un emplacement générique dans le domaine spatial S ;
- d sera utilisée pour désigner la dimension du champ spatial ;
- m fera référence à la dimension de l'ensemble des points dans S .

1.3. EXTRÊMES FONCTIONNELS

En outre, nous utiliserons n pour désigner le nombre de réalisation du processus, et N pour le nombre total d'observations échantillonnées dans S . Dans le cadre spatio-temporel, les lettres m et n seront respectivement utilisées pour désigner le nombre de stations dans S et de dates dans T échantillonnés de sorte que $N = m \times n$.

Définition 1.3.1. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Un champ aléatoire à espace d'état $\mathcal{C}(S)$, est une famille $Z = \{Z(s), s \in S\}$ de variables aléatoires à valeurs réelles définies sur le même espace probabilisé commun $(\Omega, \mathcal{A}, P)$ et à valeurs dans $(\mathcal{C}(S), \mathcal{B}\{\mathcal{C}(S)\})$.

Autrement dit, Z est une fonction mesurable définie sur $(\Omega, \mathcal{A}, P)$ et à valeurs dans $(\mathcal{C}(S), \mathcal{B}\{\mathcal{C}(S)\})$ tels que :

$$\begin{aligned} Z : \Omega \times S &\longrightarrow \mathbb{R} \\ (\omega, s) &\longmapsto Z(\omega, s). \end{aligned} \quad (1.54)$$

En effet, pour tout $s_0 \in S$, $Z(\omega, s_0)$ est bien une variable aléatoire à valeurs réelles. Par contre pour chaque $\omega_0 \in \Omega$ fixé, l'application $Z(\omega_0, \cdot) \longrightarrow \mathbb{R}$ qui à tout $s \in S$ associe $Z(\omega_0, s)$ est appelée réalisation de Z ou encore variable régionalisée (voir [Adler \(1981\)](#)). Un champ aléatoire de dimension $d \geq 2$ est aussi appelé processus stochastique ou processus aléatoire. En particulier lorsque Z est un processus spatio-temporel, $S \times T \subset \mathbb{R}^d \times \mathbb{R}^+$ désigne le domaine spatio-temporel. La loi de Z (sa mesure de probabilité P) est complètement caractérisée par la connaissance de ses distributions fini-dimensionnelles. Ainsi, pour tout $m \geq 1$, et pour tout m -uplet $(s_1, \dots, s_m) \in S^m$, la loi de $(Z(s_1), \dots, Z(s_m))$ est l'image de la mesure de probabilité $P_{s_1, \dots, s_m}$ par l'application $\omega \longmapsto (Z(\omega, s_1), \dots, Z(\omega, s_m))$ telle pour tout $B_j \in \mathcal{A}, j = 1, \dots, m$,

$$P_{s_1, \dots, s_m}(B_1 \times \dots \times B_m) = P[Z(s_1) \in B_1, \dots, Z(s_m) \in B_m]. \quad (1.55)$$

Les lois fini-dimensionnelles doivent vérifier des conditions minimales de cohérence, généralement appelées conditions de Kolmogorov ([Kallenberg \(2002\)](#)). Si la famille des lois $F_{s_1, \dots, s_m}$ ne vérifie pas les conditions de Kolmogorov, alors il ne peut pas exister de champ aléatoire sur S ayant les lois $F_{s_1, \dots, s_m}$ comme lois fini-dimensionnelles. La famille de toutes les distributions finies-dimensionnelles de Z s'appelle la loi spatiale du processus, ou encore loi temporelle lorsque $S \subset \mathbb{R}$.

b. Fonction de covariance et semi-variogramme

Un champ aléatoire $\{Z(s), s \in S\}$, est généralement décrit à partir de ses deux premiers moments, à savoir l'espérance $\mu(s) = E[Z(s)], s \in S$, et la fonction de covariance définie pour tous $s_1, s_2 \in S$ par :

$$C(s_1, s_2) = E[\{Z(s_1) - \mu(s_1)\}\{Z(s_2) - \mu(s_2)\}]. \quad (1.56)$$

La variance est définie comme $Var(s) = C(s, s)$ et la fonction de corrélation est

$$\rho(s_1, s_2) = C(s_1, s_2) / \{Var(s_1)Var(s_2)\}^{-1/2}, s_1, s_2 \in S. \quad (1.57)$$

Par exemple une des classes importantes de champs aléatoires bien connue dans la littérature est la famille des processus Gaussiens, qui sont entièrement déterminés par la fonction moyenne $\mu(s)$ et la fonction de covariance $C(s_1, s_2)$.

1.3. EXTRÊMES FONCTIONNELS

Définition 1.3.2. Soit $\{Z(s), s \in S\}$ un processus stochastique. On dit Z est un processus Gaussien, si pour tout $m \in \mathbb{N}$, $s_1, \dots, s_m \in S$, la distribution conjointe du vecteur aléatoire $Z = (Z(s_1), \dots, Z(s_m))^T$ est une distribution gaussienne multivariée de dimension m , de vecteur moyen $\mu = (E[Z(s_i)])_{i=1, \dots, m}$, $s \in S$ et de matrice de covariance $\Sigma = \{C(s_k, s_j)\}_{k,j}$, pour $k, j = 1, \dots, m$.

Dans le cas où la matrice de covariance Σ est non singulière, Z a une fonction densité définie par :

$$f(z) = (2\pi)^{-m/2} (\det(\Sigma))^{-1/2} \exp\left\{-\frac{1}{2}(z - \mu)^T \Sigma^{-1}(z - \mu)\right\}. \quad (1.58)$$

Les processus Gaussiens ont été largement utilisés en géostatistique classique en raison de leurs propriétés théoriques attrayantes, de leur tractabilité en grande dimension, de leur distributions conditionnelles explicites et de leur facilité de simulation (voir [Gelfand et Schliep \(2016\)](#)).

Lors de la modélisation des phénomènes spatiaux, il est courant de supposer des hypothèses de stationnarité (homogénéité spatiale et temporelle) afin de réduire la complexité de l'étude. Parmi ces hypothèses nous distinguons alors trois "niveaux" de stationnarité énoncés dans les définitions suivantes :

Définition 1.3.3. Soit $\{Z(s), s \in S\}$ un processus stochastique défini sur S et à trajectoires dans $\mathcal{C}(S)$.

1. On dit que Z est strictement/fortement stationnaire, si pour tout $m \in \mathbb{N}^*$, $s_1, \dots, s_m \in S$ et $s_1 + h, \dots, s_m + h \in S$ avec $h \in S$,

$$P[Z(s_1) \leq z_1, \dots, Z(s_m) \leq z_m] = P[Z(s_1 + h) \leq z_1, \dots, Z(s_m + h) \leq z_m]. \quad (1.59)$$

2. On dit que Z est stationnaire au second ordre (ou faiblement stationnaire), si pour tout $s, h \in S$

$$E[Z(s)] = \mu(s) < \infty, \quad C[Z(s), Z(s + h)] = C(h) < \infty. \quad (1.60)$$

Dans le cas de la stationnarité stricte, les distributions fini-dimensionnelles de Z sont invariantes par une translation arbitraire des points par un vecteur h . Par contre sous l'hypothèse de stationnarité au second ordre, l'espérance (ponctuelle) de Z est constante sur S et la covariance entre deux points ne dépend que de la distance h qui les sépare. En particulier la variance prend une valeur constante $\text{Var}[Z(s)] = \sigma^2$ pour tout $s \in S$ et la fonction de corrélation est égale à $\rho(h) = C(h)/\sigma^2$.

Définition 1.3.4. Soit $\{Z(s), s \in S\}$ un processus stochastique défini sur S et à trajectoires dans $\mathcal{C}(S)$. On dit que Z est intrinsèquement stationnaire si, pour tout $h \in S$, le processus des incréments noté Δ_h défini, pour tout $s \in S$ par $\Delta_h = Z(s + h) - Z(s)$ est stationnaire au second ordre.

Les trois types de stationnarité ne sont pas équivalents, mais simplement liés par la propriété d'invariance par translation. Remarquons que, si $E[Z(s)]^2 < \infty$

1.3. EXTRÊMES FONCTIONNELS

pour tout $s \in S$, alors la stationnarité stricte implique une stationnarité au second ordre qui implique à son tour, une stationnarité intrinsèque. L'inverse n'est pas nécessairement vraie. Par contre dans le cas particulier des processus Gaussiens la stationnarité au second ordre est équivalente à la stationnarité stricte. Cependant, l'hypothèse de stationnarité du second ordre est parfois trop forte car $E[Z(s)]$ peut ne pas être constante et sa variance $Var[Z(s)]$ peut ne pas être finie pour tout $s \in S$. Pour contourner ce problème, nous faisons recours généralement à une fonction de dépendance croissante avec la distance appelée variogramme.

Définition 1.3.5. Soit $\{Z(s), s \in S\}$ un champ aléatoire intrinsèque défini sur S et à trajectoires dans $\mathcal{C}(S)$. Sa fonction semi-variogramme ν est définie pour tout $h \in S$ par :

$$\nu(h) = \frac{1}{2} Var[Z(s+h) - Z(s)], \quad 2\nu(h) \text{ est appelée variogramme.} \quad (1.61)$$

La fonction semi-variogramme est bien définie dès lors que le champ aléatoire Z est intrinsèquement stationnaire. De plus, si Z est stationnaire au second ordre, il est possible de déduire sa fonction de covariance à partir de l'égalité suivante $\nu(h) = C(0) - C(h)$ pour tout $h \in S$. Nous mentionnerons également que ν est conditionnellement semi-défini négatif. Plus précisément, pour tout ensemble de sites $s_1, \dots, s_m \subset S$ et pour toute suite de nombres réels $a_1, \dots, a_m$ tel que $\sum_{j=1}^m a_j = 0$

$$\sum_{k=1}^m \sum_{j=1}^m a_k a_j \nu(s_k - s_j) \leq 0. \quad (1.62)$$

Sous les hypothèses de stationnarité intrinsèque et de second ordre, le semi-variogramme devient $\nu(h) = \frac{1}{2} E[Z(h) - Z(0)]^2$, pour tout $h \in S$. En pratique, on utilise généralement le semi-variogramme empirique ou expérimental pour estimer la structure de la variabilité spatiale sous-jacent du phénomène d'intérêt. [Matheron \(1962\)](#) a proposé l'estimateur empirique non biaisé suivant du variogramme, appelé variogramme expérimental :

$$\hat{\nu}(h) = \frac{1}{2|\mathcal{N}_h|} \sum_{(s_k, s_j) \in \mathcal{N}_h} (Z(s_k) - Z(s_j))^2, \quad (1.63)$$

pour tout $h \in S$, où $\mathcal{N}_h = \{(s_k, s_j) : s_k - s_j = h, k, j = 1, \dots, m\}$ et $|\mathcal{N}_h|$ est le nombre de paires de points distinctes distantes de h dans $\mathcal{N}_h$. Le tableau 1.1 présente quelques exemples de fonction de corrélation et de semi-variogramme. La figure 1.6 montre l'allure des fonctions de corrélation.

Pour résumer, les champs aléatoires sont des outils pertinents pour la modélisation stochastique des événements spatiaux. Ils sont mieux adaptés pour étudier et décrire la dépendance des extrêmes spatio-temporels. En pratique, il est souvent très intéressant d'évaluer la probabilité que Z dépasse un certain seuil critique $z^* \in \mathbb{R}^+$ en au moins un emplacement $s \in S$, i.e.

$$P\left(\max_{s \in S} Z(s) > z^*\right) \iff 1 - P(Z(s_1) < z^*, \dots, Z(s_m) \leq z^*) \quad (1.64)$$

1.3. EXTRÊMES FONCTIONNELS

Famille	Fonction de corrélation	Domaine de validité
Whittle-Matern	$\rho(h) = \frac{2^{1-v}}{\Gamma(v)} \left(\frac{\ h\ }{a}\right)^v \kappa_v\left(\frac{\ h\ }{a}\right)$	$a > 0, v > 0$
Cauchy	$\rho(h) = \left[1 + \left(\frac{\ h\ }{a}\right)^2\right]^{-v}$	$a > 0, v > 0$
Stable	$\rho(h) = \exp\left(-\left(\frac{\ h\ }{a}\right)^v\right)$	$a > 0, 0 < v < 2$
Puissance exponentiel	$\rho(h) = \exp\left(-\left(\frac{\ h\ }{a}\right)^v\right)$	$a > 0, 0 < v < 2$
Famille	Fonction semi-variogramme	Domaine de validité
Schather et Moreva	$\nu(h) = \frac{(1 + \ h/\tau\ ^\alpha)^{\beta/\alpha}}{2^{\beta/\alpha}}$	$0 < \alpha \leq 2, \tau > 0, \beta \leq 2$
Puissance	$\nu(h) = \sigma^2 \left(\frac{\ h\ }{\tau}\right)^\kappa$	$0 < \kappa \leq 2, \tau, \sigma^2 > 0$
Puissance Exponentiel	$\nu(h) = \sigma^2 \left[1 - \exp\left(-\left(\frac{\ h\ }{\tau}\right)^\kappa\right)\right]$	$0 < \kappa \leq 2, \sigma^2, \tau > 0$
Bernstein	$\nu(h) = \nu_{matern}(h) + \frac{\ h\ }{(1 + \ h\ ^2)^2}$	$\alpha > 0$

TABLE 1.1 – Familles paramétriques de fonctions de corrélation et de semi-variogramme isotrope, où K_v est la fonction de Bessel d'ordre v , $\Gamma(v)$ est la fonction Gamma.

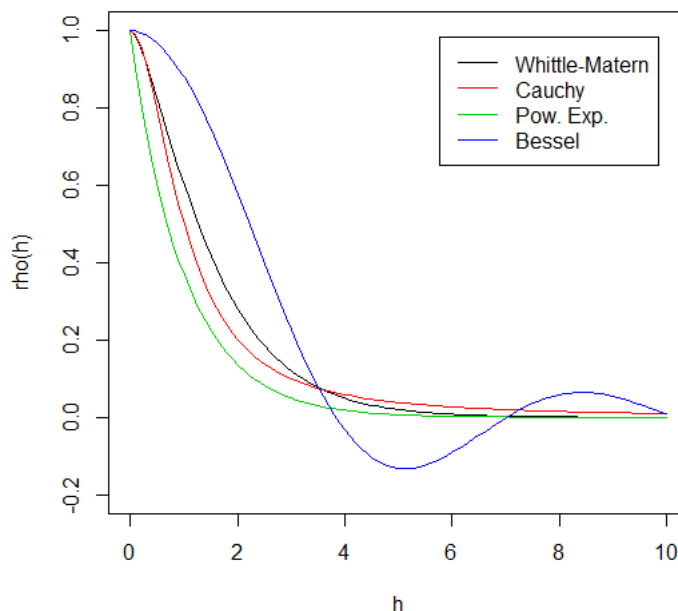


FIGURE 1.6 – Courbe de quelques fonctions de corrélation isotrope.

1.3. EXTRÊMES FONCTIONNELS

De même, il peut être également intéressant d'évaluer la probabilité qu'une fonction Z sur un sous ensemble mesurable $S' \subset S$ dépasse un certain seuil critique $z^* \in \mathbb{R}^+$ i.e.

$$P \left[\int_{S'} Z(s) ds > z^* \right], \quad \text{ou} \quad P \left[\sup_{s \in S'} Z(s) > z^* \right]. \quad (1.65)$$

Cependant les sous-ensembles $\{f \in \mathcal{C}(S) : \int_{S'} f(s) ds > z^*\}$ et $\{f \in \mathcal{C}(S) : \sup_{s \in S'} f(s) > z^*\}$, peuvent ne pas être dans $\mathcal{B}\{\mathcal{C}(S)\}$ et par conséquent les probabilités (1.65) ne sont pas nécessairement bien définies. Ces difficultés peuvent être contournées en introduisant les notions de mesurabilité conjointe et de séparabilité [Adler \(1981\)](#) et [Billingsley \(1995\)](#). Par contre lorsque les probabilités (1.64, 1.65) sont bien définies, leur estimation dépasse le cadre de la théorie des dépassements de seuil univariés.

1.3.2. Processus max-stables

Les processus max-stables ([De Haan et al. \(1984\)](#)) sont des outils populaires et utiles intervenant dans la modélisation des extrêmes dans les applications environnementales, financières et d'ingénierie. C'est une extension des méthodes de distribution de valeurs extrêmes multivariées aux données fonctionnelles. Soit S un sous-ensemble compact de $\mathbb{R}^d$ et $\mathcal{C}(S)$ l'ensemble des fonctions continues à valeurs réelles sur S , muni de la norme uniforme $\|f\|_\infty = \sup_{s \in S} |f(s)|$ pour $f \in \mathcal{C}(S)$. Supposons que $Z_t(s)$, $t = 1, \dots, n$, sur $s \in S$, est une suite de n répétitions indépendantes d'un certain processus stochastique continu stationnaire $\{Z(s), s \in S\}$. On dit que le processus stochastique Z appartient au domaine d'attraction max-stable d'un certain processus $\{Y(s), s \in S\}$, s'il existe des suites de fonctions de normalisation appropriées $a_n : S \rightarrow]0, \infty[$, $b_n : S \rightarrow \mathbb{R}$, toutes continues sur S , telles que $\max(Z - b_n)/a_n$ converge en distribution vers Y dans $\mathcal{C}(S)$ lorsque $n \rightarrow \infty$, c'est-à-dire,

$$\lim_{n \rightarrow \infty} \frac{\max_t Z_t(s) - b_n(s)}{a_n(s)} \xrightarrow{v} \{Y(s), s \in S\}, \quad (1.66)$$

où $\xrightarrow{v}$ désigne la convergence vague dans $\mathcal{C}(S)$. En particulier si le processus limite Y existe et possède des distributions marginales non dégénérées pour tout $s \in S$, alors $\{Y(s), s \in S\}$ est un processus stationnaire max-stable, où les marges suivent une distribution $GEV(\mu(s), \sigma(s), \gamma(s))$ (1.5). [Lin et de Haan \(2001\)](#) donne une caractérisation complète des conditions de max-domaine d'attraction pour le cas $S = [0, 1]$. Les paramètres de renormalisation peuvent varier dans l'espace, mais aussi l'indice de queue γ rendant l'analyse assez complexe. Pour réduire la complexité du problème, il est généralement pratique de transformer le processus Z pour avoir des marginales de Fréchet unitaires ([de Haan et Ferreira \(2006\)](#), de chapitre 9) ou Pareto unité ([Kluppelberg et Resnick \(2008\)](#)). Par exemple pour une Fréchet unitaire, les suites normalisatrices sont connues : $a_n(s) = n$ et nous pouvons choisir $b_n(s) = 0$. [de Haan et Ferreira \(2006\)](#) (Theorem 9.2.1) montrent que la convergence de M_n vers un processus max-stable est équivalente dans un premier

1.3. EXTRÊMES FONCTIONNELS

temps, à la convergence de $Z(s)$ vers une distribution de valeurs extrêmes généralisée pour tout $s \in S$ et deuxièmement, à la convergence du processus normalisé Z^* :

$$\{Z^*(s) = \frac{1}{1 - F_{Z(s)}(Z(s))}, s \in S\} \xrightarrow{v} \{Y^*(s), s \in S\}, n \rightarrow \infty, \quad (1.67)$$

où $F_{Z(s)}$ est la fonction de distribution de Z pour tout $s \in S$ et $Y^*(.)$, est un processus max-stable simple (ou standard) dont les marges suivent une distribution de Fréchet unitaire, i.e, $P(Y^*(s) \leq y) = \exp(-1/y)$. En pratique, la normalisation du processus nécessite la connaissance de la fonction de distribution $F_{Z(s)}$, qui est en général inconnue. Pour contourner cette limitation, des modèles semi-paramétriques des distributions marginales $F_{Z(s)}$ sont proposés dans [Palacios-Rodríguez et al. \(2020\)](#), [Richards et al. \(2021\)](#). D'autres propositions de transformations des marges peuvent également être trouvées dans [Coles et al. \(1999\)](#), [Coles et Tawn \(1991\)](#). Le prétraitement des marges de Z est une pratique courante dans la théorie des valeurs extrêmes multivariées et dans modélisation par les copules ([Nelsen \(2006\)](#)) car l'inférence statistique conjointe du comportement de la queue marginale et la dépendance extrême est très souvent difficile. Sans perte de généralité dans tout le reste de cette section nous travaillerons avec des marges de Fréchet unitaires et nous présenterons une description des propriétés de quelques modèles paramétriques des processus max-stables simples.

a. Représentation spectrale

Soit Z^* un processus stochastique appartenant au domaine d'attraction d'un processus max-stable Y^* , issue de la normalisation de Z en marge de Fréchet ou Pareto unitaire. [de Haan et Ferreira \(2006\)](#) montrent que pour tout suite de fonction continue $a_n : S \rightarrow]0, \infty[$, le processus renormalisé Z^* converge dans $\mathcal{C}_+(S)$, vers une mesure limite Λ telle que :

$$nP(a_n^{-1}Z^* \in A) \rightarrow \Lambda(A), n \rightarrow \infty, \quad A \subset \mathcal{C}_+(S), \quad (1.68)$$

où $\mathcal{C}_+(S) = \{z \in A \subset \mathcal{C}(S) : z(s) \geq 0, s \in S\}$. La mesure limite Λ caractérise la distribution du processus max-stable Y^* , c'est à dire que pour tout $u > 0$,

$$P[Y^* \in \{z \in \mathcal{C}_+(S) : \|z/u\|_\infty \leq 1\}] = \exp[-\Lambda\{\mathcal{A}_\infty(u)\}], \quad (1.69)$$

où $\mathcal{A}_\infty(u) = \{z \in \mathcal{C}_+(S) : \|z/u\|_\infty \geq 1\}$ est la mesure exposante et permet de caractériser la dépendance extrême. De plus, la mesure limite Λ est homogène d'ordre -1 , c'est-à-dire pour tout $t \in \mathbb{R}^+$, $\Lambda(tA) = t^{-1}\Lambda(A)$, $A \subset \mathcal{C}_+(S)$. Cette propriété caractéristique des processus extrêmes permet d'extrapoler les probabilités de queue au-delà des niveaux d'intensité observés. En effet si A est un ensemble fermé contenant au moins une observation de l'échantillon $Y_t^*, t = 1, \dots, n$, alors pour tout $t \in \mathbb{R}^+$ suffisamment grand, l'ensemble $tA = \{tz : z \in A\}$ ne contient aucune observation, mais sa probabilité $P(tA)$ peut être directement déduite à partir de $P(A)$ en utilisant la propriété d'homogénéité. La représentation de la mesure limite Λ n'est pas unique et la propriété d'homogénéité est la clé des représentations alternatives. [Giné et al. \(1990\)](#), propose le résultat suivant qui constitue une base pour la construction spectrale des processus max-stables simples.

Théorème 1.3.1. Soit Z^* un processus appartenant au domaine d'attraction d'un processus max-stable Y^* de mesure limite Λ . Alors il existe une mesure $\sigma_{\|\cdot\|}$ définie sur le simplexe unité $\mathcal{C}_{\|\cdot\|}(S) = \{z \in \mathcal{C}_+(S) : \|z\| = 1\}$ telle que :

$$\int_{\mathcal{C}_{\|\cdot\|}(S)} z(s) d\sigma_{\|\cdot\|}(z) = 1, \quad s \in S, \quad (1.70)$$

et

$$\Lambda \left(\left\{ z \in \mathcal{A}_{\|\cdot\|}(r) : \frac{z}{\|z\|} \in \mathcal{A} \right\} \right) = r^{-1} \sigma_{\|\cdot\|}(\mathcal{A}), \quad (1.71)$$

où $\mathcal{A}_{\|\cdot\|}(r) = \{z \in \mathcal{C}_+(S) : \|z\| \geq r\}$, $r > 0$. Dans la décomposition (1.71), r^{-1} mesure l'intensité du processus Y^* , tandis que $\sigma_{\|\cdot\|}$ est la mesure angulaire (ou mesure spectrale) et permet de caractériser sa structure de dépendance. Dans le théorème 1.3.1, toute mesure $\sigma_{\|\cdot\|}$ sur $\mathcal{C}_{\|\cdot\|}(S)$ satisfaisant la condition (1.70) donne une mesure limite valide Λ et donc un processus max-stable simple sur S . Le choix de différentes normes dans (1.71) conduit à des décompositions différentes, ce qui entraîne également différentes représentations spectrales possibles pour un processus max-stable simple de mesure limite Λ . Pour des raisons de commodité pratique d'inférence, le choix de la norme L_1 est souvent effectuée, car dans ce cas, le calcul de la mesure $\Lambda(\mathcal{A}_{\|\cdot\|_1}(r))$ est simple. Le résultat suivant donne une des représentations spectrales introduite par [De Haan et al. \(1984\)](#), [Giné et al. \(1990\)](#), [Penrose \(1992\)](#)

Théorème 1.3.2. Soit Y^* un processus max-stable simple avec des trajectoires sur $\mathcal{C}_+(S)$. Alors Y^* peut s'écrire :

$$Y^*(s) = \max_{i \in \mathbb{N}} U_i Q_i(s), \quad s \in S, \quad (1.72)$$

où $\{U_i, i \in \mathbb{N}\}$ sont les points d'un processus ponctuel de Poisson sur $]0; +\infty[$ de mesure d'intensité $u^{-2} du$ et les fonctions spectrales $Q_i, i \in \mathbb{N}$ sont des copies indépendantes d'un processus stochastique non-négatif $\{Q(s), s \in S\}$ vérifiant $E[Q(s)] = 1$ pour tout $s \in S$ et $E[\sup_{s \in S} Q(s)] < \infty$.

L'idée principale dans cette formulation est de considérer qu'en chaque point $s \in S$, le processus max-stable est obtenu en prenant le maximum du produit entre U_i et Q_i . La structure de dépendance spatiale est caractérisée par le processus spectral (ou angulaire) $\{Q(s), s \in S\}$, alors que l'intensité des événements extrêmes, est engendrée par des grandes valeurs du processus de Poisson. Dans le domaine du climat et de l'environnement, cette représentation spectrale (1.72) a été comparée aux tempêtes de pluies ([Smith \(1990\)](#)), où chaque couple $(U_i, Q_i), i = 1, \dots, n$ représente une tempête contribuant aux champs des maxima dont l'intensité est mesurée par U_i et la structure de dépendance spatiale est caractérisée par le processus spectral Q . Dans la représentation spectrale donnée par le théorème 1.3.2, différents choix du processus angulaire $\{Q(s), s \in S\}$ dans la modélisation conduisent à des modèles plus ou moins souples pour les maxima spatiaux. Nous présentons maintenant quelques exemples de modèles paramétriques max-stables simples bien connus, où différentes formes de $Q_i, i \in \mathbb{N}$ sont considérées. De plus comme notre intérêt porte principalement sur les processus limites des dépassements de seuil nous présenterons uniquement ici les modèles susceptibles d'être utilisés dans l'inférence des processus ℓ -Pareto généralisés.

1.3. EXTRÊMES FONCTIONNELS

b. Modèle Brown–Resnick

Parfois appelé modèle Gaussien géométrique (Brown et Resnick (1977), Kabluchko *et al.* (2009)), le modèle Brown-Resnick est construit en supposant que chaque $Q_i, i \in \mathbb{N}$ dans la représentation spectrale du théorème (1.3.2) du processus max-stable limite est de la forme

$$Q_{\max,i}(s) = \exp(G_i(s) - \nu(s)), s \in S \quad (1.73)$$

où G_i sont des copies indépendantes d'un processus Gaussien de moyenne nulle, à incréments stationnaires et de fonction de dépendance $\nu(\cdot)$ appelé semi-variogramme. Le processus max-stable résultant est un processus de Brown-Resnick max-stable stationnaire avec des marges de Fréchet standard. En d'autres termes, la fonction spectrale (ou mesure exposante) associée aux processus Brown-Resnick est un processus log-gaussien de moyenne unité avec des incréments stationnaires, et dont les distributions en dimension finie ne dépendent que du semi-variogramme ν . Dans ce cas Huser et Davison (2013) montre que la mesure exposante en dimension m du processus max-stable correspondant aux sites $s_1, \dots, s_m$ est donnée par

$$\Lambda_{BR}\{\mathcal{A}_\infty(u)\} = \sum_{j=1}^m \frac{1}{u_j} \Phi(\eta_j(u), R_j), \quad (1.74)$$

où $\mathcal{A}_\infty(u) = \{z \in \mathbb{R}^m : \max(z/u) \geq 1\}$, $\nu_{j,i}$ désigne $\nu(s_j, s_i)$, $s_j, s_i \in S$, $\Phi(\cdot, R_j)$ est la distribution normale multivariée de moyenne nulle, de matrice covariance R_j ayant (i, k) entrées $(\nu_{j,i} + \nu_{j,k} - \nu_{i,k}) / \{2(\nu_{j,i}\nu_{j,k})^{1/2}\}$, $i, k \in \{1, \dots, j-1, j+1, \dots, m\}$, et η_j est un vecteur de dimension $(m-1)$ de i ème composante

$$\eta_{j,i} = (\nu_{j,i}/2)^{1/2} + \log(u_i/u_j) / (2\nu_{j,i})^{1/2}. \quad (1.75)$$

Engelke *et al.* (2015) montre qu'une forme explicite de la fonction d'intensité du modèle Brown Resnick notée λ_{BR} , peut être trouvée en prenant les dérivées partielles de la mesure exposante $\Lambda_{BR}\{\mathcal{A}_\infty(u)\}$ par rapport à $u_1, \dots, u_m$

$$\lambda_{BR}(z) = \frac{|\Sigma|^{-1/2}}{z_1^2 z_2, \dots, z_m (2\pi)^{(m-1)/2}} \exp\left(-\frac{1}{2} \bar{z}^T \Sigma^{-1} \bar{z}\right), \quad z \in \mathbb{R}_+^m, \quad (1.76)$$

où $\bar{z}$ est un vecteur de dimension $(m-1)$ de composante $\log(z_i/z_1) + \nu_{i,1}$, $i = 2, \dots, m$ et Σ est une matrice de taille $(m-1) \times (m-1)$ de coefficients $\Sigma_{i,k} = \nu_{i,1} + \nu_{k,1} - \nu_{i,k}$, $i, k \in \{2, \dots, m\}$. Le modèle Brown Resnick a un coefficient extrémal bivarié

$$\theta(s, s+h) = \theta(h) = 2\Phi\left(\left\{\frac{\nu(h)}{2}\right\}^{1/2}\right), \quad (1.77)$$

$\theta(h) \rightarrow 2$ lorsque $h \rightarrow \infty$ si et seulement si $\nu(h) \rightarrow \infty$. Ainsi, l'indépendance asymptotique pour une distance finie entre les sites n'est pas possible dans le cas où le processus Gaussien sous-jacent est stationnaire. Car dans ce cas le semi-variogramme est borné, c'est-à-dire $\nu(h) \rightarrow \sigma^2$ lorsque $h \rightarrow \infty$ et le coefficient extrémal a une borne supérieure $2\{1 - \Phi(\sigma/2)\}$. Le comportement à long terme du variogramme détermine la dépendance extrême à de grandes distances, tandis que son comportement à l'origine détermine la régularité du processus.

c. Modèle t-Extremal

Le modèle t-extremal (Opitz (2013)) se présente comme la limite des maxima par composante des fonctions aléatoires elliptiques et son processus spectral est caractérisé par :

$$Q_{\max,i}(s) = \pi^{1/2} 2^{-(v-2)/2} \max\{G_i(s), 0\}^v, s \in S, \quad (1.78)$$

où $v > 0$ et G_i sont des copies indépendantes d'un processus Gaussien stationnaire, de variance unité et de fonction de corrélation ρ . Lorsque $v \rightarrow \infty$, nous obtenons le modèle de Brown Resnick (Kablichko *et al.* (2009)). Par contre pour $v = 1$, on tombe sur le modèle de Schlather qui ne fera pas l'objet de présentation ici dont les détails peuvent être trouvés dans Schlather (2002), ce qui fait du modèle t-extremal un modèle plus général. La mesure exposante du processus max-stable simple Y^* résultant défini aux emplacements s_1 et s_2 est donnée par (Ribatet et Sedki (2013)) :

$$\Lambda_{\text{tExtremal}}\{\mathcal{A}_\infty(u)\} = \frac{1}{u_1} t_{v+1} \left\{ -\frac{\rho(h)}{b} + \frac{1}{b} \left(\frac{u_2}{u_1} \right)^{1/v} \right\} + \frac{1}{u_2} t_{v+1} \left\{ -\frac{\rho(h)}{b} + \frac{1}{b} \left(\frac{u_1}{u_2} \right)^{1/v} \right\}, u_1, u_2 > 0 \quad (1.79)$$

où $\Lambda_{\text{tExtremal}}\{\mathcal{A}_\infty(u)\} = \{z \in \mathbb{R}^2 : \max(z_1/u_1, z_2/u_2) \geq 1\}$, t_{v+1} est la distribution standard de Student à $v+1$ degré de liberté et $b^2 = \{1 + \rho(h)^2\}/(v+1)$. Le coefficient extrémal bivarié résultant est donné par :

$$\theta(h) = 2t_{v+1} \left[(v+1)^{1/2} \left\{ \frac{1 - \rho(h)}{1 + \rho(h)} \right\}^{1/2} \right], \quad (1.80)$$

avec $\theta(h) \leq 2t_{v+1}\{(v+1)^{1/2}\}$. Le processus max-stable résultant est asymptotiquement dépendant. En pratique une limite similaire est obtenue pour le modèle de Schlather dans les grandes régions lorsque v est très petit. Nikoloulopoulos *et al.* (2009) propose une forme fermée de la mesure exposante en dimension m associée aux sites $s_1, \dots, s_m$ par :

$$\Lambda_{\text{tExtremal}}\{\mathcal{A}_\infty(u)\} = \sum_{j=1}^m u_j^{-1} t_{v+1} \left\{ \frac{u_{-j}}{u_j} - \Sigma_{-j,j}, (v+1)^{-1} (\Sigma_{-j,-j} - \Sigma_{j,j} \Sigma_{-j,j}^T) \right\}, \quad (1.81)$$

où $\mathcal{A}_\infty^m(u) = \{z \in \mathbb{R}^m : \max(z/u) \leq 1\}$, $\Sigma = \{\rho(s_j, s_k)\}_{j,k=1,\dots,m}$ est la matrice de corrélation du processus Gaussien sous-jacent. Tandis que $t_{v+1}(\cdot, \Sigma)$ est la distribution de student multivariée à $v+1$ degré de liberté, de moyenne nulle de matrice de corrélation Σ . Thibaud et Opitz (2015) propose une forme fermée de la fonction d'intensité correspondant que nous notons ici $\lambda_{\text{tExtremal}}$ en dérivant la mesure exposante $\Lambda_{\text{tExtremal}}$

$$\lambda_{\text{tExtremal}}(y) = \frac{v^{1-m} \pi^{(1-m)/2} \Gamma\{(v+m)/2\}}{|\Sigma|^{-1/2} \Gamma\{(v+1)/2\}} \left(\prod_{j=1}^m (y_j) \right)^{1/v-1} \left\{ T_{1/v}(y)^T \Sigma^{-1} T_{1/v}(y) \right\}^{-(v+d)/2}, \quad (1.82)$$

1.3. EXTRÊMES FONCTIONNELS

avec $y \in \mathbb{R}^m$ et $T_{1/2} = \text{sign}(y)|y|^{1/v}$. Elle a servi à ces auteurs de développer une procédure d'inférence de vraisemblance censurée similaire à celle proposée par [Wadsworth et Tawn \(2014\)](#). Le processus t-extremal est l'un des processus max-stables les plus connus et les plus utilisés, avec le processus Brown-Resnick ([Brown et Resnick \(1977\)](#), [Kabluchko et al. \(2009\)](#)), le processus gaussien des valeurs extrêmes ([Smith \(1990\)](#)) et les processus gaussiens extremal ([Schlather \(2002\)](#)) peuvent être considérés comme des cas particuliers. Les processus de Brown-Resnick et t-extremal sont respectivement des processus extrémaux limites des processus gaussiens et de Student strictement stationnaires. Cependant, en pratique les données peuvent être non stationnaires et présenter des distributions asymétriques dans de nombreuses applications.

d. Modèle skew t-Extremal

Contrairement aux modèles Brown Resnick et t-extremal, le modèle skew t-extremal ([Beranger et al. \(2017\)](#)) est un processus max-stable non-stationnaire limite des fonctions aléatoires skew-normal. Cette classe de processus max-stables a des propriétés très intéressantes en pratique pour étudier les non-stationnarités spatio-temporelles des extrêmes dans le cadre du changement climatique. Ce modèle est un sous-ensemble de processus non-stationnaires qui inclut le processus t-extremal stationnaire et capable de décrire une gamme de structure de dépendance non-stationnaire isotrope ou anisotrope. La fonction spectrale issue de la présentation spectrale (1.72) est caractérisée par :

$$Q_{\max,i} = \{G_i(s)\}_+ / \{\mu_+(s)\}^{1/v}, \quad (1.83)$$

où $\{G(s), s \in S\}$ est un processus stochastique skew-normal non-stationnaire à trajectoires continues sur S dont la distribution finie dimensionnelle est notée $\mathcal{SN}_m(\Sigma, \alpha, \tau)$ de matrice de corrélation Σ et de paramètres d'inclinaison et d'extension respectifs $\alpha \in \mathbb{R}^m$ et $\tau \in \mathbb{R}$ tels que $E[\sup_{s \in S} G(s)] < \infty$ et $\mu_+(s) = E[\{G(s)\}_+^v] \in]0, \infty[$, $s \in S, v > 0$ où $\{G(s)\}_+ = \max\{G(s), 0\}$ désigne la partie positive de $G(s)$.

Dans ce cas les $\{U_i, i \in \mathbb{N}\}$ sont maintenant des points d'un processus de Poisson inhomogène défini sur $]0, \infty[$, d'intensité $vu^{-(v+1)}$, $v > 0$ et indépendants des $G_i, i \in \mathbb{N}$. Le processus max-stable non-stationnaire résultant a des marges v -Fréchet dont la structure de dépendance est non-stationnaire et dépend de l'écart entre les sites. En particulier si les marges sont Fréchet unitaire, c'est-à-dire $v = 1$, alors le processus est dit processus max-stable standard. La mesure exposant de dimension m associée aux sites $s = s_1, \dots, s_m$ est donnée par :

$$\Lambda_{\text{Skew-t}}\{\mathcal{A}_\infty(y)\} = \sum_{j=1}^m y_j^{-v} \Psi_{m-1} \left[\left(\sqrt{\frac{v+1}{1-\rho_{k,j}^2}} \left(\frac{y_k^\circ}{y_j^\circ} - \rho_{k,j} \right), k \in I_j \right)^T ; \Sigma_j^\circ, \alpha_j^\circ, \tau_j^\circ, \kappa_j^\circ, v+1 \right], \quad (1.84)$$

où $y \in \mathbb{R}^m$ de composante $y_j = y(s_j)$, $y_j^\circ = (y_j \mu_+(s_j))^{1/v}$. En outre Ψ_{m-1} est une distribution asymétrique non-centré étendue de Student multivariée de degré de liberté $v+1$, de matrice de corrélation Σ_j° , de paramètres de forme, d'extension et

de non centralité respectives $\alpha_j^\circ \in \mathbb{R}^{m-1}$, $\tau_j^\circ \in \mathbb{R}$ et $\kappa_j^\circ \in \mathbb{R}$ où $I = \{1, \dots, m\}$, $I_j = I \setminus \{j\}$ et $\rho_{k,j}$ est le (k, j) -ème élément de Σ . En particulier il est facile de remarquer qu'en fixant $\alpha_s \equiv 0$ un vecteur nul et $\tau = 0$, on retrouve le processus t-extrémal (Opitz (2013)). Par contre en fixant ensuite $v = 1$, on obtient le modèle de Schlather (Schlather (2002)). En prenant les dérivées partielles de la mesure exposante Λ_{Skew} par rapport à $y = y_1, \dots, y_m$, Beranger *et al.* (2020) propose une forme explicite de la fonction d'intensité noté λ_{Skew} et définie par :

$$\lambda_{\text{Skew}}(y) = \frac{2^{(v-2)/2} v^{1-m} \Gamma\left(\frac{m+v}{2}\right) \Psi\left(\tilde{\alpha}_s \sqrt{m+v}; -\tau_s, m+v\right) \prod_{j=1}^m \left(\mu_j + y_j^{1-v}\right)^{1/v}}{\pi^{m/2} |\Sigma_s|^{1/2} Q_{\Sigma_s}(y^\circ)^{(m+v)/2} \Phi\left(\tau_s \{1 + Q_{\Sigma_s^{-1}}(\alpha_s)\}^{-1/2}\right)}, \quad (1.85)$$

où $y^\circ = (y\mu_+(s_j))^{1/v} \in \mathbb{R}^m$, $\tilde{\alpha}_s = \alpha_s^T y^\circ Q_{\Sigma_s}(y^\circ)^{-1/2} \in \mathbb{R}$, $\alpha_s \in \mathbb{R}^d$ et $Q_{\Sigma_s}(\alpha_s) = \alpha_s^T \Sigma_s \alpha_s$.

1.3.3. Processus ℓ -Pareto et extensions

Dans cette section nous introduisons, le cadre de modélisation des processus spatiaux de dépassements de seuil et sa généralisation. Nous rappelons quelques propriétés importantes sur la construction de ces processus. Le lien de ces processus avec la variation régulière, la mesure exposante ainsi que les processus max-stables sont également rappelés. Un accent particulier est accordé à quelques méthodes d'inférences de ces processus. Ces outils seront la base de la modélisation de l'approche POT spatio-temporelle présentée dans le chapitre 3 à 5.

a. Variation régulière fonctionnelle

Dans ce qui suit nous rappelons le lien étroit entre la variation régulière fonctionnelle et la loi limite des dépassements de seuils. Nous avons besoin pour cela d'introduire quelques notations et les objets limites qui apparaîtront dans la théorie fonctionnelle des valeurs extrêmes. Pour un contexte général sur la théorie fonctionnelle des extrêmes, les détails sont fournis dans de Haan et Ferreira (2006), chapitre 9, Dombry et Ribatet (2015), de Fondeville et Davison (2018); de Fondeville et Davison (2020). Rappelons qu'un sous-ensemble fermé mesurable $\mathcal{C} \in \mathcal{C}_+(S)$ est un cône si $tf \in \mathcal{C}$, pour tout $f \in \mathcal{C}$ et $t > 0$. Dans l'étude des extrêmes le cône $\mathcal{C} = \mathcal{C}_a \cup \mathcal{C}_b$ avec $\mathcal{C}_a = \{0\}$ et $\mathcal{C}_b = \{f \in \mathcal{C}_+(S) : \inf_{s \in S} f(s) = 0\}$ sont souvent exclus de $\mathcal{C}_+(S)$ afin d'éviter des mesures limites avec des masses infinies à l'origine ou sur les axes des coordonnées. Nous notons par $\mathcal{F} = \mathcal{C}_+(S) \setminus \mathcal{C}$, $\|\cdot\|_{\text{ang}}$ la norme angulaire sur $\mathcal{F}$ qui définit les angles, $\mathcal{S}_{\text{ang}} = \{f \in \mathcal{F} : \|f\|_{\text{ang}} = 1\}$ la sphère unitaire par rapport à la norme angulaire $\|\cdot\|_{\text{ang}}$ et $\mathcal{M}(\mathcal{F})$ la classe de mesure sur $\mathcal{B}(\mathcal{F})$.

Définition 1.3.6. (Hult et Lindskog (2005)) *Un processus stochastique Z avec des trajectoires sur $\mathcal{F}$ est à variation régulière s'il existe une suite de nombre réels positifs $a_1, \dots, a_n$ avec $\lim_{n \rightarrow \infty} a_n = \infty$ et une mesure unique $\Lambda \in \mathcal{M}(\mathcal{F})$ tel que :*

$$nP\left(a_n^{-1}Z \in A\right) \longrightarrow \Lambda(A), \quad n \rightarrow \infty, \quad A \in \mathcal{B}(\mathcal{F}). \quad (1.86)$$

1.3. EXTRÊMES FONCTIONNELS

On note $Z \in RV(\mathcal{F}, a_n, \Lambda)$. La mesure limite Λ est homogène d'ordre α , i.e., $\Lambda(tA) = t^{-\alpha}\Lambda(A)$, $t > 0$, $A \in \mathcal{B}(\mathcal{F})$. En particulier si Z est convenablement normalisé par exemple, étant donné les marges unitaires de Fréchet ou de Pareto, l'équation (1.86) est équivalente à la convergence des maxima ponctuels renormalisés $n^{-1} \max_{j=1}^m Z_j$ vers un processus max-stable non dégénéré, avec des marges de Fréchet unitaires et de mesure exposante Λ (Lin et de Haan (2001)). Plus généralement, on dit que Z est à variation régulière fonctionnelle généralisée (Ferreira et al. (2014), notée $Z \in GRV(\gamma, a_n, b_n, \Lambda)$, i.e qu'il existe des fonctions $a_n : S \rightarrow \mathbb{R}_+$, $b_n : S \rightarrow \mathbb{R}$ et un indice de queue $\gamma \in \mathbb{R}$ tel que

$$nP \left[(1 + \gamma((Z - b_n)/a_n))^{1/\gamma} \in A \right] \rightarrow \Lambda(A), n \rightarrow +\infty, \quad (1.87)$$

où Λ est une mesure non-nulle sur $\mathcal{M}\{\mathcal{F}\}$. La mesure limite satisfait $\Lambda(tA) = t^{-1}\Lambda(A)$ pour tout $t > 0$ et $A \subset \mathcal{F}$, i.e. Λ est homogène d'ordre -1 (Lindskog et al. (2014)). La généralisation de l'analyse des dépassements de seuil aux fonctions est possible dans le cadre de la variation régulière fonctionnelle. Nous présentons ici deux approches de cette généralisation. L'une des approches construite sous l'hypothèse que Z appartient au domaine d'attraction de Fréchet : les processus ℓ -Pareto standard (Dombry et Ribatet (2015)) et l'autre assez générale couvre les différents régimes de queue lourde : les processus ℓ -Pareto généralisés (de Fondeville et Davison (2020)).

b. Fonction de risque et notion de ℓ -excès

Soient $\{Z(s), s \in S\}$ un processus stochastique défini sur $\mathcal{C}_+(S)$ admettant un attracteur max-stable à marges non dégénérées, $(Z(s_1), \dots, Z(s_m))$ le vecteur aléatoire associé à la restriction du processus aux différentes positions $s_1, \dots, s_m$ et $u = (u(s_1), \dots, u(s_m)) \in \mathbb{R}^m$ un vecteur seuil. Introduit par Buishand et al. (2008), Aulbach et Falk (2012) et Ferreira et al. (2014) l'approche *POT* fonctionnelle, est construite en définissant par exemple un dépassement de seuil spatial lorsque au moins une des composantes du vecteur aléatoire observé $(z(s_1), \dots, z(s_m))$ est plus grande que son seuil marginal, i.e $z(s_j) > u(s_j)$. Plus précisément, un dépassement de seuil élevé d'un processus stochastique $\{Z(s), s \in S\}$ peut être défini comme suit

Définition 1.3.7. Soient $\{Z(s), s \in S\}$ un processus stochastique défini sur $\mathcal{C}_+(S)$ et $a_n(s) > 0$ et $b_n(s)$ des fonctions de normalisation. On appelle dépassement de seuil spatial l'évènement :

$$\left\{ \sup_{s \in S} \frac{Z(s) - b_n(s)}{a_n(s)} > 0 \right\} \text{ tq. } \lim_{n \uparrow \infty} P \left(\sup_{s \in S} \frac{Z(s) - b_n(s)}{a_n(s)} > 0 \right) \rightarrow 0. \quad (1.88)$$

Dombry et Ribatet (2015) généralisent la notion de dépassement de seuil spatial introduit par Ferreira et al. (2014) à toute fonction de risque ℓ non négative continue et homogène. Comme dans Ferreira et al. (2014), il s'intéresse à la loi conditionnelle $u_n^{-1}Z \mid \{\ell(Z) > u_n\}$, $n \rightarrow \infty$ en particulier dans le cas où le processus Z appartient au domaine d'attraction de Fréchet.

Définition 1.3.8. Soit $Z = \{Z(s), s \in S\}$ un processus stochastique défini sur $\mathcal{C}_+(S)$. Un dépassement de seuil peut être défini comme un évènement $\{\ell(Z) > u_n\}$

tel que $\lim_{n \uparrow \infty} P(\ell(Z) > u_n) \rightarrow 0$ où ℓ est une fonction continue non négative, homogène d'ordre $\alpha > 0$: $\ell(tZ) = t^\alpha \ell(Z)$, $t > 0$ et u_n une suite de seuils.

Par exemple, si Z représente la force du vent dans une région donnée, alors $\ell_1(Z) = \int_S |Z(s)|^2 ds$ est une fonctionnelle qui représente l'énergie du système. De même si Z représente le niveau de la mer et $\ell_2(Z) = \inf_{s \in S} Z(s)$, alors $\{\ell_2(Z) > u_n\}$ représente l'évènement d'observer simultanément des hauts niveau de la mer à tous les endroits. Si nous supposons que Z est la température moyenne journalière au Burkina Faso et $\ell_3(Z) = \sup_{s \in S} Z(s)$, alors $\{\ell_3(Z) > u_n\}$ définit l'évènement d'observer une température extrême au moins dans un emplacement. C'est d'ailleurs cette fonction de risque qui été utilisée par [Ferreira et al. \(2014\)](#) pour jeter les bases des processus spatiaux de Pareto. De plus si Z désigne les précipitations d'une région et $\ell_4(Z) = \int_{S'} |Z(s)| ds$ alors $\{\ell_4(Z) > u_n\}$ définit les dépassements du cumul régional de précipitations observées sur les sous-régions $S' \subset S$. À des fins théoriques et sans perte de généralité, nous exigeons que ℓ satisfasse la propriété d'homogénéité suivante :

$$\ell(tZ) = t\ell(Z), \quad t > 0, \quad Z \in \mathcal{C}_+(S). \quad (1.89)$$

En effet puis que la fonctionnelle ℓ est non négative, la fonctionnelle $\ell' = \ell^{1/\alpha}$ est clairement homogène d'ordre $\alpha = 1$ et satisfait

$$\{\ell(Z) > u_n\} = \{\ell'(Z) > u_n^{1/\alpha}\}. \quad (1.90)$$

Les observations issues de la définition (1.3.8), pour une fonction de risque ℓ donnée sont alors désignées sous le nom de ℓ -excès (ou ℓ -dépassement). Etant donné une fonction de risque ℓ , une formulation naturelle des résultats des dépassements de seuils spatiaux, exploite la variation régulière fonctionnelle et utilise une décomposition pseudo-polaire.

Définition 1.3.9. Soit $\ell : \mathcal{C}_+(S) \rightarrow [0, +\infty[$ un risque fonctionnel mesurable, positif et $\|\cdot\|_{ang}$ une norme angulaire sur $\mathcal{F}$. On appelle transformation pseudo-polaire T_ℓ une fonction définie par :

$$T_\ell : \mathcal{F} \rightarrow [0, +\infty[\times \mathcal{S}_{ang}, \quad T_\ell(f) = \left\{ r = \ell(f), w = \frac{f}{\|f\|_{ang}} \right\}, \quad (1.91)$$

où $\mathcal{S}_{ang} = \{f \in \mathcal{F} : \|f\|_{ang} = 1\}$ est la sphère unité et $\|\cdot\|_{ang}$ une norme sur $\mathcal{F}$. Les coordonnées pseudo-polaires (r, w) sont composées du rayon $r > 0$ et de l'angle $w \in \mathcal{S}_{ang}$.

En particulier si ℓ est continue et la transformation pseudo-polaire T_ℓ est restreinte à $\{f \in \mathcal{F} : \ell(f) > 0\}$, alors T_ℓ est un homéomorphisme tel que :

$$T_\ell^{-1}(r, w) = r \frac{w}{\ell(w)}. \quad (1.92)$$

La définition suivante met l'accent entre la variation régulière, la mesure ponctuelle de Poisson et les maxima d'échantillon. Ces notions nous apparaissent comme des outils essentiels pour introduire le lien entre la variation régulière et les dépassements de seuil dans un cadre fonctionnel.

1.3. EXTRÊMES FONCTIONNELS

Définition 1.3.10. Pour tout $\alpha > 0$ et σ une mesure de probabilité sur $\mathcal{S}_{ang}$, nous définissons :

i) $\Lambda_{\alpha,\sigma}$ la mesure unique sur $]0, +\infty[\times \mathcal{S}_{ang}$ telle que :

$$\Lambda_{\alpha,\sigma} \{[r, +\infty[\times B\} = r^{-\alpha} \sigma(B), \quad r > 0, \quad B \in \mathcal{B}(\mathcal{S}_{ang}). \quad (1.93)$$

ii) $M_{\alpha,\sigma}$ un processus max-stable continu sur $\mathcal{C}_+(S)$ de mesure exposante $\Lambda_{\alpha,\sigma}$.
 iii) $\Pi_{\alpha,\sigma}$ une mesure ponctuelle de Poisson sur $\mathcal{F}$ et d'intensité $\Lambda_{\alpha,\sigma}$.

Le résultat suivant énoncé dans le théorème 2.1 de [Lindskog et al. \(2014\)](#) qui fournit une formulation intéressante du lien entre la variation régulière et les ℓ -excès en utilisant la transformation pseudo-polaire T_ℓ définie ci-dessus.

Propriété 1.3.1. Pour tout $Z \in RV(\mathcal{F}, a_n, \Lambda)$ et tout risque fonctionnel ℓ uniformément continu ne variant pas Λ -presque partout, il existe $\alpha > 0$ et une mesure σ_ℓ sur $\mathcal{B}(\mathcal{S}_{ang})$ tels que :

$$nP\left(T_\ell^{-1}\left(a_n^{-1}r, w\right) \in \cdot\right) \longrightarrow \Lambda \circ T_\ell^{-1}(\cdot) = \Lambda^\alpha \times \sigma_\ell(\cdot), \quad n \longrightarrow \infty, \quad (1.94)$$

où $\Lambda^\alpha[r, \infty[= r^{-\alpha}$ et $\sigma_\ell(\cdot) = \Lambda\{f \in \mathcal{F} : \ell(f) \geq 1, \|f\|_{ang} = 1, f/\|f\|_{ang} \in (\cdot)\}$ est la mesure angulaire sur $\mathcal{S}_{ang}^\ell$.

Bien que le choix d'une fonction de risque, permet de se concentrer sur des événements extrêmes particuliers, le choix de la norme angulaire $\|\cdot\|_{ang}$ n'a pas d'effet et est généralement effectué pour des raisons de commodité. Dans toute la suite nous assimilerons la norme angulaire $\|\cdot\|_{ang}$ à la norme L_1 pour des raisons pratiques dans l'inférence statistique.

c. Processus ℓ -Pareto Standard

Sous l'hypothèse de la variation régulière fonctionnelle et certaines conditions sur la fonction de risque les ℓ -excès du processus Z convenablement normalisés, convergent faiblement dans $\mathcal{C}_+(S)$ vers un processus stochastique appartenant à la famille des processus ℓ -Pareto dont nous présenterons maintenant quelques propriétés caractéristiques et des méthodes d'inférences. Les résultats ci-dessus découlent de la stabilité des dépassements de seuil des processus de Pareto. En d'autres termes, la structure de dépendance des processus ℓ -Pareto est invariante par rapport au niveau d'extrême de l'événement. Cependant, en pratique, ce n'est pas le cas, car pour des niveaux finis, le seuil u ne peut pas être pris arbitrairement élevé et les événements extrêmes sélectionnés dépendent de ℓ . Néanmoins, différents choix de fonction du risque peuvent améliorer le comportement sous-asymptotique des processus ℓ -Pareto standard. Une définition constructive des processus de Pareto standard a d'abord été proposée par [Ferreira et al. \(2014\)](#) dans le cadre des dépassements du supremum avant d'être reformuler par [Dombry et Ribatet \(2015\)](#) pour les ℓ -dépassements.

Définition 1.3.11. ([Dombry et Ribatet \(2015\)](#)) Soit $W_\ell^* = \{W_\ell^*(s), s \in S\}$ un processus stochastique défini sur $\mathcal{F}$. On dit que W_ℓ^* est un Processus ℓ -Pareto spatial standard, s'il peut se décomposer de la manière suivante :

$$W_\ell^*(s) = RQ(s), \text{ pour tout } s \in S. \quad (1.95)$$

où

- i) Q est processus stochastique borné satisfaisant $\ell(Q(s)) \equiv 1$,
- ii) R est une variable aléatoire de Pareto standard de paramètres d'échelle 1 et de forme γ_R , i.e $P(R > r) = r^{-\gamma_R}, r > 1$,
- iii) R et Q sont indépendants.

Dans cette construction nous avons nécessairement $R = \ell(W_\ell^*)$ et $Q = W_\ell^* / \ell(W_\ell^*)$. Grâce à la décomposition en coordonnées pseudo-polaires, [Dombry et Ribatet \(2015\)](#) propose de construire les processus ℓ -Pareto de la manière suivante :

$$W_{\alpha, \sigma_\ell}^* = R \frac{Q}{\ell(Q)}, \quad (1.96)$$

où R est une v.a univarié de Pareto, i.e $P(R > r) = r^{-\alpha} (r \geq 1)$ et $\{Q(s), s \in S\}$ un processus aléatoire avec des trajectoires dans $\mathcal{S}_{\text{ang}}^\ell = \{f \in \mathcal{F} : \ell(f) \geq 1, \|f\|_{\text{ang}} = 1\}$ et de mesure de probabilité $\sigma_{\text{ang}} = \sigma_\ell(\cdot) / \sigma_\ell(\mathcal{S}_{\text{ang}})$, où σ_ℓ est la mesure de $Q(s) / \ell(Q(s))$ et $\|\cdot\|_{\text{ang}}$ est la mesure angulaire sur $\mathcal{F}$.

Propriété 1.3.2. Soient une fonctionnelle continue $\ell : \mathcal{C}_+(S) \rightarrow \mathbb{R}_+$ et $Z \in \mathcal{C}_+(S)$, alors on a la convergence faible dans $\mathcal{C}_+(S)$:

$$P(u^{-1}Z \in A \mid \ell(Z) > u) \rightarrow P(W_\ell^* \in A), \quad u \rightarrow +\infty \quad (1.97)$$

où $A \subset \{f \in \mathcal{C}_+(S) : \ell(f) \geq 1\}$, $W_\ell^* \in \mathcal{C}_+(S)$ est un processus ℓ -Pareto standard avec $P(\ell(W_\ell^*) = 1) = 1$.

Une propriété importante des processus ℓ -Pareto, est la stabilité par seuillage. Pour tout $A \in \mathcal{B}(\mathcal{F})$ et $P(\ell(W_\ell^*) > u) > 0$ alors nous avons :

$$P(u^{-1}W_\ell^* \in A \mid \ell(W_\ell^*) > u) = P(W_\ell^* \in A). \quad (1.98)$$

Définition 1.3.12. Soient σ_ℓ une mesure de probabilité sur $\mathcal{S}_{\text{ang}}^\ell$ et ℓ une fonction de risque continue, non négative et homogène d'ordre 1. Le processus max-stable M associé au processus ℓ -Pareto (eq. 1.96) est défini par :

$$M(s) = \max_{n \in \mathbb{N}} R_n \frac{Q_n(s)}{\ell(Q_n)}, \quad s \in S, \quad (1.99)$$

où $\{R_n, n \in \mathbb{N}\}$ est un processus de Poisson sur $]0, \infty[$ d'intensité $r^{-2}dr$ et $Q_1, \dots, Q_n$ sont des répétitions du processus Q de mesure de probabilité σ_ℓ .

La mesure exposante Λ_θ de paramètres θ du processus max-stable M vaut $\Lambda_\theta(A_{\max}(f)) = \Lambda^1 \times \sigma_\ell\{A_{\max}(f)\}$, avec $A_{\max}(f) = \{f' \in \mathcal{F} : \sup_{s \in S} f'(s) \geq f(s)\}$. La relation (1.99) relie un processus ℓ -Pareto à son équivalente max-stable.

d. Processus ℓ -Pareto généralisés

Initialement développés pour des processus stochastiques appartenant au domaine d'attraction de Fréchet, les processus ℓ -Pareto ont été étendus par [Fondeville et Davison \(2020\)](#), afin de prendre en compte les différents régimes de queue

1.3. EXTRÊMES FONCTIONNELS

lourde et garantir une bonne interprétabilité du risque. Nous supposons qu'il existe une fonction réelle continue et positive $A_Z : S \rightarrow]0, \infty[$, $Z \in GRV(\gamma, a_n, b_n, \Lambda)$ un processus stochastique vérifiant (1.87) et ℓ une fonction de risque telle que la suite de fonction a_n satisfait la décomposition asymptotique suivante (Ferreira *et al.* (2012), Engelke *et al.* (2019), de Fondeville et Davison (2020)) :

$$\lim_{n \rightarrow \infty} \sup_{s \in S} \left| \frac{a_n(s)}{\ell(a_n)} - A_Z(s) \right| = 0, \quad (1.100)$$

Les distributions marginales de Z sont supposées appartenir à une famille d'échelle de localisation, ce qui permet en particulier de garantir un paramètre de forme $\gamma(s) = \gamma \in \mathbb{R}$ constant sur tout $s \in S$, c'est-à-dire qu'il existe une fonction de distribution F telle que

$$P(Z(s) \leq z(s)) = F \left[\frac{z(s) - B_Z(s)}{A_Z(s)} \right], \quad (1.101)$$

où $A_Z : S \rightarrow]0, \infty[$ vérifie la décomposition asymptotique (1.100) et $B_Z : S \rightarrow \mathbb{R}$ sont des fonctions continues sur S . En particulier F appartient au domaine d'attraction d'une distribution GEV d'indice de queue $\gamma \in \mathbb{R}$, i.e il existe des suites $\tilde{a}_n > 0$ et $\tilde{b}_n \in \mathbb{R}$ telles que : $\lim_{n \rightarrow \infty} F^n(\tilde{a}_n z + \tilde{b}_n) = G_\gamma(z)$. Dans ces conditions les fonctions normalisations $a_n(s)$ et $b_n(s)$ satisfont :

$$a_n(s) = \tilde{a}_n A_Z(s), \quad b_n(s) = B_Z(s) + A_Z(s) \tilde{b}_n, \quad s \in S, \quad (1.102)$$

Les fonctions $A_Z(s)$ et $B_Z(s)$ ainsi que la structure de dépendance spatiale de Z peuvent appartenir à des familles de fonctions paramétriques, i.e, $\{A_{Z, \theta_A} : \theta_A \in \Theta_A\}$, $\{B_{Z, \theta_B} : \theta_B \in \Theta_B\}$ et $\{\Lambda_{\theta_\Lambda} : \theta_\Lambda \in \Theta_\Lambda\}$. Sous des hypothèses minimales sur la fonction de risque et les hypothèses (1.87, 1.100) de Z , de Fondeville et Davison (2020) montre que la distribution conditionnelle des ℓ -dépassements pour un certain seuil $u \geq 0$ du processus normalisé $(Z - b_n)/\ell(a_n)$ peut être approximée par un processus de ℓ -Pareto généralisé, pour tout n suffisamment grand.

Théorème 1.3.3. *Soit Z un processus stochastique à trajectoire continue sur S . Si $Z \in GRV(\gamma, a_n, b_n, \Lambda)$, $u \geq 0$ et ℓ une fonction de risque valide pour Z , alors*

$$P \left\{ \left\lfloor \frac{Z - b_n}{\ell(a_n)} \right\rfloor \in A \mid \ell \left(\frac{Z - b_n}{\ell(a_n)} \right) \geq u \right\} \rightarrow P \{W_\ell \in A\}, n \rightarrow \infty, \quad (1.103)$$

où $\lfloor z \rfloor = \max(z, A_Z \gamma^{-1})$ si $\gamma > 0$ et $\lfloor z \rfloor = z$ si $\gamma \leq 0$. W_ℓ est un processus stochastique non dégénéré sur S et appartient à la famille des processus de ℓ -Pareto généralisés d'indice de queue γ , de fonction d'échelle A_Z , de localisation zéro, et une mesure limite Λ .

Plus précisément, un processus ℓ -Pareto généralisé W_ℓ associé à la mesure limite Λ et d'indice de queue $\gamma \in \mathbb{R}$ est un processus stochastique prenant des valeurs dans $\{z \in \mathcal{C}_+(S) : \ell \{(z - b)/\ell(a)\} \geq 0\}$ et défini par :

$$W_\ell = \begin{cases} a(Y_\ell^\gamma - 1)/\gamma + b, & \gamma \neq 0 \\ a \log Y_\ell + b, & \gamma = 0, \end{cases} \quad (1.104)$$

1.3. EXTREMES FONCTIONNELS

où $a \equiv a_n(s)$ avec $a = \ell(a)A_Z > 0$ et b sont des fonctions continues sur S respectivement des fonctions d'échelle et de localisation et Y_ℓ est un processus stochastique dont la mesure de probabilité est complètement déterminée par la mesure limite Λ . Pour la modélisation de la structure de dépendance spatiale du processus latent, nous utilisons Y_ℓ dont les marges sont dans le domaine d'attraction de Fréchet avec un indice de queue $\gamma = 1$, comme processus de référence. Une décomposition pseudo-polaire de Y_ℓ dans (1.104) conduit à la formulation suivante (Dombry et Ribatet (2015), de Fondeville et Davison (2018))

$$Y_\ell = RQ \quad (1.105)$$

où R est une variable aléatoire unitaire de Pareto d'indice γ_R représentant l'intensité du processus, et Q est un processus stochastique noté composante angulaire à espace d'état S et prenant des valeurs dans $\mathcal{S} = \{y \in \mathcal{C}_+(S) : \|y\|_1 = 1\}$ dont la mesure de probabilité est caractérisée par la mesure limite Λ . On trouvera plus de détails sur les processus généralisés de ℓ -Pareto dans (Ferreira *et al.* (2014), Dombry et Ribatet (2015), de Fondeville et Davison (2018); de Fondeville et Davison (2020)).

Toutes les définitions et résultats précédents sont valables pour des dimensions finies, c'est-à-dire pour des vecteurs aléatoires de dimensions m , en remplaçant la convergence faible par la convergence vague (Resnick, 2007, section 3.3.5) sur $\mathcal{M}(\mathbb{R}_+^m \setminus \{0\})$, la classe des mesures de Borel sur $\mathcal{B}(\mathbb{R}_+^m \setminus \{0\})$ munie de la norme $\|\cdot\|$; voir la thèse de Thomas Opitz mentionnée ci-dessus.

e. Inférence des processus ℓ -Pareto

Dans cette section nous fournissons une revue de la littérature des méthodes d'inférence des processus ℓ -Pareto résumé par de Fondeville et Davison (2018). En pratique, Z n'est observé que dans un ensemble fini m de sites, mais les résultats de Lindskog *et al.* (2014) impliquent alors qu'un résultat similaire s'applique au processus dérivé $Z_{\{s_1, \dots, s_m\}}$.

e.1. Vraisemblance spectrale. Pour un vecteur seuil $u = (u_1, \dots, u_m) \in \mathbb{R}_+^m$, et étant donné un modèle paramétrique approprié, la fonction densité de $Z_{\{s_1, \dots, s_m\}}$ conditionnellement à $\ell(Z_{\{s_1, \dots, s_m\}}) > u$ (d'un processus ℓ -Pareto), s'obtient en re-normalisation la fonction d'intensité λ_θ par $\Lambda_\theta(A_\ell(u))$:

$$\lambda_{\theta, u}^\ell(z) = \frac{\lambda_\theta(z)}{\Lambda_\theta\{A_\ell(u)\}}, \quad z \in A_\ell(u), \quad (1.106)$$

où

$$\Lambda_\theta\{A_\ell(u)\} = \int_{A_\ell(u)} \lambda_\theta(z) dz, \quad (1.107)$$

λ_θ est la fonction d'intensité que l'on retrouve en prenant les dérivées partielles de la mesure exposante Λ_θ , θ est un élément de l'espace des paramètres $\Theta \subset \mathbb{R}^p$ et $A_\ell(u) =$

1.3. EXTRÊMES FONCTIONNELS

$\{z \in \mathbb{R}_+^m : \ell(z/u) \geq 1\}$, est la région des dépassements. La log-vraisemblance associée à (1.106) est donnée par :

$$\ell(\theta, z^1, \dots, z^N) = \sum_{t=1}^N \mathbb{1} \left\{ \ell \left(\frac{z^t}{u} \right) \geq 1 \right\} \log \left[\frac{\lambda_\theta(z^t)}{\Lambda_\theta \{A_\ell(u)\}} \right], \text{ (vrais spectrale)} \quad (1.108)$$

où $N = \text{card}(A_\ell(u))$ et $\{z^t, t = 1, \dots, N\}$ sont des répétitions indépendantes du vecteur aléatoire ℓ -Pareto tel que $z^t \in A_\ell(u)$. L'évaluation numérique de l'intégrale $\Lambda_\theta \{A_\ell(u)\}$ est généralement coûteuse en terme de temps de calcul et devient très difficile pour m assez grand, bien qu'elle soit simplifiée pour certains risques fonctionnels. Par exemple pour $\ell_1(z) = \max_j z_j$, l'intégrale $\Lambda_\theta \{A_{\ell_1}(u)\}$ est une somme de fonctions de probabilité multivariées (Davis *et al.* (2013)). De même si $\ell_2(z) = m^{-1} \sum_j z_j$ (Coles et Tawn (1991)), l'intégrale $\Lambda_\theta \{A_{\ell_2}(u)\}$, évaluée dans la région des dépassements est une constante qui ne dépend pas de θ .

$$\Lambda_\theta \{A_{\ell_2}(u)\} = \sum_{i=1}^m \frac{1}{u_j}, \quad u = (u_1, \dots, u_m). \quad (1.109)$$

L'estimateur du maximum de vraisemblance basée sur cette fonction de risque est appelé estimateur spectral et a la forme suivante :

$$\hat{\theta}_{\ell_2}(z^t, t = 1, \dots, N) = \arg \max_{\theta \in \Theta} \sum_{t=1}^N \mathbb{1} \left\{ \sum_{j=1}^m \frac{z_j^t}{u_j} > 1 \right\} \log \lambda_{\theta, u}^{\ell_2}(z^t). \quad (1.110)$$

Cependant, pour un seuil u fini et des composantes en dessous du seuil u dans $z^t \in A_\ell(u)$ peuvent conduire à une estimation biaisée qui croît avec la dimension m . Le biais peut alors être réduit par une vraisemblance censurée proposée dans le cadre multivarié par Joe *et al.* (1992) et utilisé pour le modèle Brown-Resnick par Wadsworth et Tawn (2014) et le processus t-extrémal par Thibaud et Opitz (2015). Cette méthode fonctionne bien dans la pratique, mais nécessite des calculs de probabilités multivariées qui deviennent fastidieux avec la dimension.

e.2. Vraisemblance censurée. La vraisemblance censurée pour les processus des valeurs extrêmes associés aux fonctions aléatoires log-gaussiennes a été développée par Wadsworth et Tawn (2014) et est basée sur (1.108) avec un risque fonctionnel $\ell(z) = \max_{j=1}^m \{z_j/u_j\}$. Ce choix de la fonction de risque permet de censurer toutes les composantes situées sous le vecteur seuil $(u_1, \dots, u_m) > 0$ par $z_j^c = \max(z_j, u)$ et de réduire le biais lié aux composantes faibles. L'estimateur correspondant a une variance plus élevée mais un biais plus faible que l'estimateur spectral. La fonction de densité de vraisemblance censurée associée au processus de Brown-Resnick est énoncée comme suit : (Asadi *et al.* (2015))

$$\lambda_{\theta, u}^{\text{cens}}(z) = \frac{1}{\Lambda_\theta \{A_\ell(u)\}} \frac{1}{z_1^2 z_2 \dots z_k} \phi_{k-1}(\bar{z}_{2:k}; \Sigma_{2:k}) \Phi_{m-k} \{\mu_{\text{cens}}(z_{1:k}), \Sigma_{\text{cens}}(z_{1:k})\} \quad (1.111)$$

où k est le nombre de dépassement de seuil, $\bar{z}_{2:k}$ et $\Sigma_{2:k}$ sont des sous-ensemble de $\bar{z}$ et Σ_θ dans (1.76) et ϕ_{k-1} et Φ_{m-k} sont les fonctions de densité et de distribution

gaussiennes multivariées de moyenne nulle. L'argument et la matrice de covariance pour Φ_{m-k} sont respectivement

$$\mu_{\text{cens}}(z_{1:k}) = \{\log(u_j/z_1) + \gamma_{j,1}\}_{j=k+1,\dots,m} - \Sigma_{(k+1):m,2:k} \Sigma_{2:k,2:k}^{-1} \bar{z}_{2:k},$$

$$\Sigma_{\text{cens}}(z_{1:k}) = \Sigma_{(k+1):m,(k+1):m} - \Sigma_{(k+1):m,2:k} \Sigma_{2:k,2:k}^{-1} \Sigma_{2:k,(k+1):m}.$$

L'estimateur de la vraisemblance censurée est donnée par :

$$\hat{\theta}_{\text{cens}}(z^t : t = 1 \cdots N) = \arg \max_{\theta \in \Theta} \sum_{t=1}^N \mathbb{1} \left\{ \max_{j=1,\dots,m} \left(\frac{z_j^t}{u_j} \right) \geq 1 \right\} \log \lambda_{\theta,u}^{\text{cens}}(z^t) \quad (1.112)$$

où $z^t \in A_\ell(u)$, $t = 1, \dots, N$ et $\lambda_{\theta,u}^{\text{cens}}(z^t : t = 1, \dots, N)$ est la fonction de densité de vraisemblance censurée proposée par [Engelke et al. \(2015\)](#). Cet estimateur est également consistant et asymptotiquement normal. De plus, pour les seuils finis, il est plus robuste en ce qui concerne les composants faibles. Cependant, il est très coûteux en calcul pour le modèle Brown-Resnick et les processus analogues associés, en raison du grand nombre d'intégrales normales multivariées impliquées dans le calcul de $\lambda_{\theta,u}^{\text{cens}}(z^t : t = 1, \dots, N)$.

e.3. Méthode du score de gradient. L'inférence de la vraisemblance classique nécessite l'évaluation de la constante de normalisation $\Lambda_\theta(A_\ell(u))$ dont la complexité augmente avec la dimension. Il s'agit de trouver une alternative qui ne nécessite pas son calcul.

Définition 1.3.13. [Gneiting et Raftery \(2007\)](#) Soient $\mathcal{A} \subset \mathbb{R}_+^m$, un espace d'échantillon, $\mathcal{P}$ une classe convexe de mesures de probabilité sur $\mathcal{A}$ et Z une v.a de distribution $F \in \mathcal{A}$. Un score approprié (ou de scoring) est une fonction

$$\delta : \mathcal{P} \times \mathcal{A} \longrightarrow \mathbb{R} / \Delta_\delta(G, F) = E_Z \{ \delta(G, Z) \} - E_Z \{ \delta(F, Z) \} \geq 0, \quad G \in \mathcal{P} \quad (1.113)$$

Si $\Delta_\delta(G, F) = 0 \Rightarrow G = F$, le score est dit strictement positif et sous cette hypothèse Δ_δ , définit une mesure de divergence sur $\mathcal{P}$ ([Thorarinsdottir et al. \(2013\)](#)).

Proposition 1.3.1. Soient δ un score propre strictement positif, $\{F_\theta, \theta \in \Theta\} \subset \mathcal{A}$ une famille paramétrique de distribution et $Z^1, \dots, Z^n$ des observations indépendantes provenant de F_{θ_0} . Le premier terme de la divergence $\Delta_\delta(F_\theta, F_{\theta_0})$ peut être estimée par :

$$\delta(\theta) = \frac{1}{n} \sum_{t=1}^n \delta(F_\theta, Z^t). \quad (1.114)$$

La minimisation de cet estimateur sous les conditions de régularité de [Dawid et al. \(2016\)](#), donne un estimateur sans biais et asymptotiquement normal de θ_0 . Pour un risque fonctionnel ℓ , l'estimateur résultant est :

$$\hat{\theta}_{\delta,u}^\ell(z^t : t = 1, \dots, N) = \arg \min_{\theta \in \Theta} \sum_{t=1}^N \mathbb{1} \left\{ \ell \left(\frac{z^t}{u} \right) > 1 \right\} \delta(\lambda_{\theta,u}^\ell, z^t/u), \quad (1.115)$$

où $\delta(\lambda_{\theta,u}^\ell, z^t/u)$ est un score propre. Les propriétés asymptotiques de cet estimateur peuvent être prolongées à des échantillons provenant d'un vecteur aléatoire à variation régulière avec des marges normalisées.

1.3. EXTRÊMES FONCTIONNELS

Proposition 1.3.2. Soient $Y^1, \dots, Y^N$ des répliques indépendantes d'un vecteur aléatoire Y^* à variation régulière, avec des marges normalisées et de mesure limite Λ_{θ_0} . Soit δ un score propre satisfaisant les conditions du Théorème 4.1 de [Dawid et al. \(2016\)](#). Si $N \rightarrow \infty$ et $N_u \rightarrow \infty$ de sorte à ce que $N_u/N \rightarrow 0$ quand $N \rightarrow \infty$ alors

$$N_u^{1/2} \left\{ \hat{\theta}_{\delta, N/N_u}^\ell(Y^1, \dots, Y^N) - \theta_0 \right\} \rightarrow \mathcal{N}(O, K^{-1}J(K^{-1})^T) \quad (1.116)$$

$$\text{où, } J = E \left\{ \frac{\partial \delta(\theta)}{\partial \theta} \frac{\partial \delta(\theta)}{\partial \theta^T} \right\} \Big|_{\theta=\theta_0} ; \quad K = E \left\{ \frac{\partial^2 \delta(\theta)}{\partial \theta \partial \theta^T} \right\} \Big|_{\theta=\theta_0}$$

Bien qu'il soit efficace, il n'est pas robuste, ce qui peut être gênant lors de l'ajustement des modèles asymptotiques comme les processus de Pareto, de plus le coefficient de normalisation $\Lambda_\theta \{A_\ell(u)\}$ n'est disponible que dans des cas particuliers.

La règle de calcul du gradient ([Hyvärinen et Dayan \(2005\)](#)) utilise uniquement les $\nabla_z \log \lambda_{\theta, u}^\ell$ et ne nécessite pas le calcul de $\Lambda_\theta \{A_\ell(u)\}$. [Hyvärinen \(2007\)](#) adapte ce score propre aux variables strictement positives. Dans ce travail, elle est étendue à tout domaine de la forme $A_\ell(u) = \{z \in \mathbb{R}_+^m : \ell(z/u) \geq 1\}$ en utilisant la mesure de divergence.

$$\Delta_{grad}(\theta, \theta_0) = \int_{A_\ell(u)} \|\nabla_z \log \lambda_\theta(z) \otimes w(z) - \nabla_z \log \lambda_{\theta_0}(z) \otimes w(z)\|_2^2 \lambda_{\theta_0}(z) dz, \quad (1.117)$$

λ_{θ_0} est différentiable pour tout $\theta \in \Theta$ sur $A_\ell(u) \setminus \partial A_\ell(u)$, ∇_z est l'opérateur de gradient, $w : A_\ell(u) \rightarrow \mathbb{R}_+^m$ est une fonction de pondération positive et $\otimes$ désigne le produit d'Hadarnard. Si w est différentiable sur $A_\ell(u)$ et satisfaisant à certains conditions de limites, alors un score approprié est donnée par [de Fondeville et Davison \(2018\)](#) :

$$\delta_w(\lambda_\theta, z) = \sum_{j=1}^m \left(2w_j(z) \frac{\partial w_j(z)}{\partial z_j} \frac{\partial \log \lambda_\theta(z)}{\partial z_j} + w_j(z)^2 \left[\frac{\partial^2 \log \lambda_\theta(z)}{\partial z_j^2} + \frac{1}{2} \left\{ \frac{\partial \log \lambda_\theta(z)}{\partial z_j} \right\}^2 \right] \right), \quad (1.118)$$

où $z \in A_\ell(u)$. Le score du gradient d'un processus de Pareto log-Gaussien satisfait aux conditions de régularité du théorème 4.1 de [Dawid et al. \(2016\)](#) de sorte que l'estimateur $\hat{\theta}_{\delta, u}^\ell(z^t : t = 1, \dots, N)$ soit asymptotiquement normal. En particulier, deux fonctions de poids sont possibles pour le modèle Brown-Resnick :

$$w_j^1(z) = z_j [1 - \exp \{1 - \ell(z/u)\}] \quad (j = 1, \dots, m) \quad (1.119)$$

$$w_j^2(z) = [1 - \exp \{-3(z_j - u_j)/u_j\}] [1 - \exp \{1 - \ell(z/u)\}] \quad (j = 1, \dots, m) \quad (1.120)$$

où ℓ est un risque fonctionnel différentiable sur $\mathbb{R}_+^m$, $u \in \mathbb{R}_+^m$ est le vecteur seuil. Les poids w^1 sont dérivés de [Hyvärinen \(2007\)](#), tandis que w^2 est conçu pour approximer l'effet de la censure, en appliquant des poids faibles aux valeurs z près du seuil.

1.4. CONCLUSION

Il faut noter que pour ces fonctions de pondération, un vecteur $z \in A_\ell(u)$ est pénalisé si $\ell(z/u)$ est proche de 1 et des composantes faibles de z induisent des poids faibles associés aux dérivées partielles, donc z ne contribue pas au problème d'optimisation (1.115). Cela fournit une raison de s'attendre à ce que l'inférence utilisant le score de gradient soit plus robuste en ce qui concerne les composantes faibles que la vraisemblance spectrale. L'estimateur $\hat{\theta}_{\delta,u}^\ell(z^t : t = 1, \dots, N)$ est beaucoup moins coûteux à calculer que $\hat{\theta}_{cens}$ et peut être obtenu pour tout risque fonctionnel différentiable sur $\mathbb{R}_+^m$.

1.4. Conclusion

Ce chapitre a introduit les notions nécessaires à l'étude des événements extrêmes spatio-temporels. Nous avons dans un premier temps décrit les distributions asymptotiques univariées des maxima de blocs et des dépassements de seuils. Pour les applications dans le domaine de l'environnement et du climat, les lois asymptotiques univariées sont limitées car elles ne rendent pas compte de la nature spatio-temporelle des processus physiques. Les processus max-stables se présentent comme une généralisation fonctionnelle de la distribution des valeurs extrêmes généralisées. Nous avons ainsi passé en revue les principales propriétés de cette famille de processus, qui a été appliquée avec succès dans les études environnementales.

Enfin, nous avons présenté le processus ℓ -Pareto qui généralise l'analyse des dépassements de seuils aux fonctions et qui nous serviront comme base pour monter les modèles à même de prendre en compte des non-stationnarités temporelles et spatiales.

Approches POT ponctuelles non-stationnaires et applications

Dans ce chapitre nous montrons comment l'approche des dépassements de seuil peut être utilisée pour analyser les tendances temporelles des extrêmes des paramètres météorologiques. Une synthèse des résultats de ce chapitre a fait l'objet d'une communication au JDS2019.

Résumé : Dans cette étude, les tendances des températures et des précipitations extrêmes au Burkina Faso sont évaluées en utilisant la théorie statistique des valeurs extrêmes, en particulier l'approche des dépassements de seuil dans un contexte non-stationnaire. Une méthodologie basée sur des tendances paramétriques est utilisée pour décrire l'évolution dans le temps des paramètres de la distribution des excès au dessus d'un seuil fixé et l'intensité du processus de Poisson non-stationnaire. A partir de l'extrapolation des tendances identifiées, les niveaux de retour non-stationnaires et leur intervalle de confiance sont également calculés.

Mots-clefs : Modèle POT non-stationnaire, Extrêmes Values, Changement Climatique

2.1. Introduction

Les événements climatiques extrêmes causent chaque année à travers le monde la perte en vies humaines ainsi que de nombreux dégâts matériels. Ils ne sont certes pas des phénomènes nouveaux, mais ils ont pris une ampleur particulièrement importante au cours de ces trente dernières années ([Sall et al. \(2011\)](#), [Puget et al. \(2010\)](#)).

En Afrique sub-sahélienne, la fin de la décennie 2000 et la première moitié de celle de 2010, ont été marquées, par ces événements extrêmes aux forts impacts socio-économiques et environnementaux, jamais enregistrés depuis plus de 30 ans (Centre Régional AGRHYMET, 2010). En 2012 par exemple, 363 personnes sont mortes au Nigéria dans la pire des inondations qu'a connues le pays. Il y' a eu environ 2 millions de déplacés et 618 000 maisons détruites. Au Niger, les inondations ont provoqué également des dégâts considérables : 65 décès et 125 000 personnes sans abri. Des dégâts ont été aussi enregistrés dans d'autres pays d'Afrique de l' Ouest, notamment au Sénégal, au Mali et au Burkina Faso.

Le Burkina Faso, à l'instar de tous ces pays de la région Ouest africaine a connu une modification de son régime pluviométrique depuis la fin des années 1960. Cette modification s'est traduite par une baisse des cumuls pluviométriques et des

pluies journalières. Selon les dernières publications de l'Agence Nationale de la Météorologie, la pluviométrie a amorcé un retour vers des périodes humides depuis la fin des années 1980 et les décennies 1990 et 2000. Cette reprise est plus prononcée dans la zone sahélienne que dans les espaces soudanien et soudano-sahélien. Le retour des pluies est plutôt lié à une fréquence élevée des événements de forte intensité pluviométrique qu'à une augmentation des jours pluvieux. Le pays est donc confronté à une récurrence des phénomènes d'inondations dont les impacts sont très préjudiciables aux populations, sur toute l'étendue du territoire et en particulier dans la capitale Ouagadougou et ses périphéries (PANA, 2007). Selon le Centre de Recherche sur l'Epidémiologie des Désastres (CRED, 2016), le Burkina Faso a connu douze (12) grandes inondations entre 2000 et 2015. Les souvenirs les plus marquants remontent au 1er Septembre 2009 où les inondations ont causé des dommages considérables, tant sur le plan humain que matériel à Ouagadougou et dans certaines localités environnantes (principalement dans les régions du Centre et du Plateau Central). En effet, selon le rapport d'activité 2009 du CONASUR, près de 150 000 personnes ont été affectées par ces inondations, plus de 24 000 maisons se sont écroulées et 9 personnes ont trouvé la mort. Selon les autorités, une telle quantité d'eau (261,3 mm) n'est jamais tombée sur le Burkina Faso depuis 1919.

Cette situation a suscité un regain d'intérêt des chercheurs à s'investir dans la modélisation et l'analyse de tendances des variables météorologiques en Afrique sub-sahélienne. De nombreuses études ont été menées sur la question à l'aide des techniques d'analyse statistiques variées, visant à caractériser les éventuels changements dans les paramètres climatiques et de proposer des adaptations appropriées (Alhassane *et al.* (2013), Vischel *et al.* (2015), Salack *et al.* (2016), Sultan *et al.* (2017), PAS-PNA (2021), Faye (2018)). La plus part de ces études concluent à une tendance globale à la hausse de certains paramètres climatiques comme les températures et des précipitations. D'autres approches statistiques plus focalisée sur les événements climatiques extrêmes ont permis de montrer que les extrêmes de températures et de précipitations sont devenus de plus en plus fréquents dans la région (Panthou *et al.* (2014), Taylor *et al.* (2017), Wilcox *et al.* (2018) Chagnaud *et al.* (2022)). Les simulations climatiques en scénarios socio-économiques futurs (Deme *et al.* (2015), Sylla *et al.* (2015), Salack *et al.* (2016), Panthou *et al.* (2018)) confirment certains résultats ci-dessus et suggèrent l'émergence de nouvelles conditions hydroclimatiques sur la région dans les prochaines décennies c'est à dire que les nouvelles caractéristiques de la pluviométrie du sahel incluent des faux début et fin précoces des saisons pluvieuses, une augmentation de la fréquence et l'intensité de pluies journalières intenses et une augmentation du nombre de nuits chaudes et de jours chauds. Au Burkina Faso en particulier, les projections de la répartition des précipitations saisonnières en s'appuyant les données Cordex des scénarii RCP4.5 et RCP8.5 révèlent toutes que pour les périodes de 2021-2050, le pays enregistrera des poches de sécheresses de plus en plus sévères ayant une plus forte variabilité en début et fin de saison (Ibrahim *et al.* (2014), PAS-PNA (2021)). Des études plus récentes (Panthou *et al.* (2018), Lejeune et Saeed (2019), Salack *et al.* (2015)), confirment une évolution incertaine des futures précipitations annuelles moyennes mais aussi, une augmentation de température et une plus grande fréquence des événements extrêmes. Toutefois il faut rappeler que la plupart des

2.2. PRÉSENTATION DES DONNÉES

modèles proposés dans ces études permettent entre autres de décrire les jeux de données et de faire des projections climatiques le plus souvent développés dans un cadre stationnaire, la dimension temporelle tenant lieu de répétitions indépendantes. Or ce n'est généralement pas le cas surtout dans un contexte de changement climatique, les paramètres climatiques évoluent de manière continue.

Dans cette étude on cherche à détecter et analyser les tendances récentes des températures et des précipitations extrêmes au Burkina Faso. Pour cela, nous utilisons les méthodes développées par [Parey *et al.* \(2007, 2010\)](#) et [Acero *et al.* \(2014\)](#), dans le cadre des méthodes POT (Peak Over Threshold) dans un contexte non stationnaire. Ces méthodes permettent de détecter des tendances polynomiales (on se restreindra aux ordres 1 et 2) dans le paramètre d'intensité des occurrences de dépassement et le paramètre d'échelle de la loi de Pareto Généralisée. Ces techniques sont mises en œuvre sur les observations des deux paramètres météorologiques collectés sur la période 1957-2016.

2.2. Présentation des données

Le Burkina Faso a un climat de type tropical à dominance soudano-sahélienne, caractérisé par l'alternance entre une courte saison de pluies (Juin à Octobre) et une longue saison sèche dont une période fraîche et sèche (Novembre à Février) et un temps chaud (Mars à Mai). On distingue trois zones climatiques qui sont : la zone sahélienne au Nord avec une pluviométrie moyenne annuelle inférieure à 600 mm ; la zone Nord soudanienne au centre avec une pluviométrie moyenne annuelle comprise entre 600 et 900 mm et la zone Sud soudanienne au Sud avec une pluviométrie moyenne annuelle supérieure à 900 mm. Les températures et les précipitations connaissent une forte variabilité spatio-temporelle. Les températures moyennes journalières dépassent rarement 35°C et les extrêmes durant ces trente dernières années se rencontrent au Nord avec une valeur minimale de 21°C et une valeur maximale absolue de 44.2°C enregistrées à la station de Dori. L'analyse des tendances climatiques montre une augmentation de la température moyenne et une baisse des cumuls pluviométriques et des pluies journalières avec une migration des isothermes et des isohyètes du Nord vers le Sud (PANA, 2007).

Les données météorologiques traitées sont issues de la base de données climatologiques de l'Agence National de la Météorologie (ANAM) du Burkina Faso qui est assez bien fournie avec des observations remontant à 1902 pour certains paramètres météorologiques. Nous disposons des mesures journalières de précipitations et de températures (minimales ($T_{\min}$), moyennes (T_{moy}) et maximales ($T_{\max}$)) effectuées de façon continue de 1957 à 2016 sur dix stations synoptiques (Ouaga, Dori, Ouahigouya, Dédougou, Bogandé, Fada N'Gourma, Bobo Dioulasso, Boromo, Pô et Gaoua). Les caractéristiques des différentes stations sont consignées dans le tableau 2.1. Ces stations sont choisies de sorte à garantir non seulement une bonne représentativité des zones climatiques mais aussi une bonne couverture spatiale du territoire avec au moins trois stations par zone climatique dont trois pour la zone sahélienne au Nord (Ouahigouya, Dori et Bogandé), quatre pour la zone sub-sahélienne au centre

2.2. PRÉSENTATION DES DONNÉES

Stations	Lat	Long	Alt (m)	Période d'observation	Données NA (%)		
					Pluies	Tmax	Tmin
Ouaga	12.35	-1.52	303	1957-2016	0.09	0.23	0.25
Dori	14.03	-0.033	288	1957-2016	0.09	1.84	1.86
Ouahigouya	13,58	-2.43	329	1957-2016	0.09	0.16	0.16
Dédougou	12.47	-3.47	308	1957-2016	0.55	1.75	2.2
Bogandé	12.59	0.80	250	1997-2016	2.21	0.00	0.00
Fada	12.03	0.37	308	1957-2016	0.21	0.21	0.23
Bobo	11.17	-4.32	459	1957-2016	0.09	0.52	0.21
Boromo	11.75	-2.93	270	1957-2016	0.09	0.25	0.25
Pô	11.17	-1.15	326	1978-2016	0.27	0.00	0.00
Gaoua	10.33	-3.18	333	1957-2016	0.31	0.3	0.61

TABLE 2.1 – Caractéristiques des stations météorologiques.

(Ouaga, Dédougou, Boromo et Fada) et trois pour la zone nord soudanienne au Sud (Bobo, Pô et Gaoua). La figure 2.1 donne la répartition spatiale des différentes stations. Afin de limiter les problèmes liés au cycle saisonnier des précipitations et

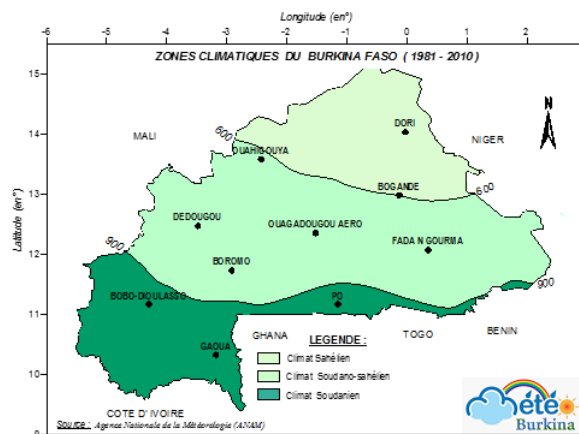


FIGURE 2.1 – Emplacement des différentes stations synoptiques

des températures, sur chaque station nous avons travaillé respectivement à partir de la sous-série correspondant aux jours les plus pluvieux et les jours les plus chauds. Les périodes du 15 Mars au 15 Juin et de Juin à Octobre ont été retenues respectivement pour les séries temporelles de températures et de précipitations. Ainsi une sous-série de 60 fois 184 jours et 60 fois 93 jours est extraite respectivement de la série d'observations des précipitations et des températures pour constituer la série des extrêmes journaliers entre 1957 et 2016. Sur ces séries d'extrêmes, nous appliquons une procédure de déclusterisation (runs declustering en anglais) à pas journalier ($r=1$ jour) pour identifier les groupes d'observations extrêmes approximativement indépendants au sein de l'échantillon afin d'éviter les dépendances à court terme dans les séries chronologiques pour toute la période 1957-2016.

2.3. Modèle POT non stationnaire

2.3.1. Distribution de Paréto Généralisée

Dans l'approche POT stationnaire classique, les observations d'une variable aléatoire donnée, doivent au préalable valider les hypothèses d'indépendance et de stationnarité. Cependant la présence de certains types de variabilité dans les extrêmes des variables météorologiques (tendance, effets saisonniers, régimes climatiques, dépendance avec d'autres covariables) peut remettre en cause l'hypothèse de stationnarité surtout dans le contexte actuel de changement climatique. Ainsi une bonne analyse des tendances des extrêmes des variables météorologiques ne peut se faire sans la prise en compte de cette question de la non-stationnarité des processus environnementaux. Généralement dans l'approche POT ponctuelle, la non-stationnarité est prise en compte en supposant que les paramètres de la loi $GPD(\sigma_u, \gamma)$ évoluent en fonction du temps ou d'autres covariables. La dépendance entre les covariables et les paramètres peut prendre différentes structures de dépendance, polynomiale, semi-paramétrique, linéaire par morceaux ou non-linéaire (Coles *et al.* (2001), Parey *et al.* (2007), Cooley (2013)). Pour tenir compte des caractéristiques de la zone d'étude et des récents développements au niveau des structures de dépendance utilisées en analyse non-stationnaire, les modèles de tendance polynomiale sont retenus pour traduire la relation entre les paramètres de la loi $GPD(\sigma_u(s), \gamma(s))$ (non-stationary generalized Pareto distribution) :

$$\left. \begin{aligned} \sigma_u(s) &= \exp \left(\sigma_0(s) + \sum_{j=1}^{n_{\sigma}(s)} \sigma_j(s) X_j^{(\sigma)}(s) \right) \\ u(s) &= u_0(s) + \sum_{j=1}^{n_u(s)} u_j(s) X_j^{(u)}(s) \\ \gamma(s) &= \gamma_0(s) + \sum_{j=1}^{n_{\gamma}(s)} \gamma_j(s) X_j^{(\gamma)}(s) \end{aligned} \right\} s \in S \subset \mathbb{R}^2, \quad (2.1)$$

où $X_1^{(u)}(\cdot), \dots, X_{n_u(s)}^{(u)}(\cdot)$; $X_1^{(\sigma)}(\cdot), \dots, X_{n_{\sigma}(s)}^{(\sigma)}(\cdot)$ et $X_1^{(\gamma)}(\cdot), \dots, X_{n_{\gamma}(s)}^{(\gamma)}(\cdot)$ sont des covariables appropriées. Les $u_0(s), \dots, u_{n_u(s)}(s)$; $\sigma_0(s), \dots, \sigma_{n_{\sigma}(s)}(s)$ et $\gamma_0(s), \dots, \gamma_{n_{\gamma}(s)}(s)$ sont des réels et S l'ensemble spatial de sites de mesure. Dans ce cadre statistique, $Z(s) \sim GPD(\sigma_u(s), \gamma(s))$ dont la distribution est donnée par l'équation (2.2) :

$$GPD(z \mid \sigma_u(s), \gamma(s)) = \begin{cases} 1 - (1 + \gamma(s)z/\sigma_u(s))^{-\frac{1}{\gamma(s)}}, & \text{si } \gamma(s) \neq 0 \\ 1 - \exp(-z/\sigma_u(s)), & \text{si } \gamma(s) = 0 \end{cases}, \quad (2.2)$$

où $\sigma_u(s)$ et $\gamma(s)$ sont respectivement les paramètres d'échelle et de forme au site s donnés par la relation 2.1.

2.3.2. Processus de Poisson non-homogène

Les dates des dépassements décrits dans la section 2.3.1 sont généralement approchées par un processus ponctuel de Poisson stationnaire d'intensité λ . Le processus de Poisson non homogène est une extension du processus de Poisson homogène au contexte non-stationnaire. Dans ce cas, l'intensité du processus évolue

en fonction du temps et d'autres covariables. Les hypothèses d'indépendance et de non-multiplicité des événements sont maintenues tandis que l'hypothèse de stationnarité ne l'est plus. Le nombre d'événements $N(t)$ est maintenant distribué selon une loi de Poisson de paramètre : $m(t) = \int_0^t \lambda_\alpha(u) du$, où $m(t)$ est le nombre moyen d'événements arrivant dans l'intervalle $[0, t]$. Ainsi pour tout $I = [t_1, t_2] \subset [0, T]$, un processus de Poisson non-stationnaire est complètement caractérisé par la relation 2.3 suivante :

$$P(N(I) = k) = \exp(-\Lambda(I)) \frac{\Lambda(I)^k}{k!}, \quad \text{avec } \Lambda(I) = \int_{t_1}^{t_2} \lambda(t) dt \quad (2.3)$$

Dans toute la suite nous optons également pour un modèle de tendance polynomiale pour décrire la fonction d'intensité des occurrences aussi bien pour les précipitations que pour les températures :

$$\lambda(s) = \lambda_0(s) + \sum_{j=1}^{n_\lambda(s)} \lambda_j(s) X_j^{(\lambda)}(s), \quad s \in S \subset \mathbb{R}^2, \quad (2.4)$$

où $X_1^{(\lambda)}(\cdot), \dots, X_{n_\lambda(s)}^{(\lambda)}(\cdot)$ sont des covariables appropriés et les $\lambda_0(s), \dots, \lambda_{n_\lambda(s)}$ sont des réels.

2.3.3. Niveaux de retour non stationnaires

Dans le contexte non-stationnaire, l'approche la plus simple pour décrire les risques consiste à donner une estimation du risque via des probabilités ou la période de retour. Ainsi la probabilité de dépassement d'un niveau donné, peut évoluer d'un instant à un autre, et doit être prise en compte dans la définition du niveau de retour non-stationnaire. Par conséquent la relation donnée par l'équation 1.38 devient caduque. Ainsi, la notion de période et de niveau de retour devrait être révisée et réadaptée pour prendre en compte la non-stationnarité généralement présente dans les données spatio-temporelles.

Récemment, de nombreuses études ont proposé des approches afin de prendre en compte la non-stationnarité dans l'estimation des quantiles extrêmes ([Olsen et al. \(1998\)](#), [Parey et al. \(2007, 2010\)](#), [Salas et Obeysekera \(2014\)](#), [Cooley \(2013\)](#), [Yan et al. \(2017\)](#)). Parmi ces approches, nous avons deux méthodes basées sur la période de retour m , à savoir le temps d'attente moyen (Expected Waiting Time : EWT, [Olsen et al. \(1998\)](#), [Cooley \(2013\)](#)) et le nombre moyen de dépassements (Expected Number of Exceedances : ENE, [Parey et al. \(2007, 2010\)](#)). Dans le domaine de la fiabilité, des auteurs pour prendre en compte les impacts de la durée de vie des équipements hydrologiques ont introduit de nouvelles notions ([Read et Vogel \(2015\)](#), [Rootzén et Katz \(2013\)](#)). Parmi ces méthodes alternatives introduites, nous avons encore celles basées sur le concept de période de retour : la fiabilité équivalente (Equivalent Reliability : ER) ([Hu et al. \(2018\)](#)) et le niveau de vie de conception moyen (Average Design Life Level : ADLL) ([Yan et al. \(2017\)](#)).

2.3. MODÈLE POT NON STATIONNAIRE

a. Expected Waiting Time (EWT)

La méthode EWT a été proposée par [Olsen et al. \(1998\)](#), puis étendue par [Salas et Obeysekera \(2014\)](#) en utilisant une distribution géométrique avec des paramètres variant dans le temps. Dans des conditions non stationnaires, la distribution géométrique décrivant le temps d'attente T avant la première occurrence d'un événement dépassant le quantile z est

$$\begin{aligned} f(t) = P(T = t) &= P(Z_1 \leq z, Z_2 \leq z, \dots, Z_{t-1} \leq z, Z_t > z) \\ &= P(Z_1 \leq z) P(Z_2 \leq z) \times \dots \times P(Z_{t-1} \leq z) P(Z_t > z) \\ &= \prod_{i=1}^{t-1} F_{Z,i}(z) (1 - F_{Z,t}(z)) = (1 - p_t) \prod_{i=1}^{t-1} p_i, \quad t = 1, 2, \dots, t_{\max} \end{aligned}$$

Le temps d'attente moyen $E[T]$ est donc calculé comme suit :

$$E[T] = \sum_{t=1}^{t_{\max}} t f(t) = \sum_{t=1}^{t_{\max}} t P(T = t) \quad (2.5)$$

Ce qui donne en développant puis en regroupant les termes de $E[T]$:

$$\begin{aligned} E[T] &= P(T = 1) + \sum_{t=2}^{t_{\max}-1} t P(T = t) + t_{\max} P(T = t_{\max}) \\ &= 1 - p_1 + \sum_{t=2}^{t_{\max}-1} t (1 - p_t) \prod_{i=1}^{t-1} p_i + t_{\max} \prod_{i=1}^{t_{\max}-1} p_i; \\ &= 1 - p_1 + \sum_{t=2}^{t_{\max}-1} t \prod_{i=1}^{t-1} p_i - \sum_{t=3}^{t_{\max}} (t-1) \prod_{i=1}^{t-1} p_i + t_{\max} \prod_{i=1}^{t_{\max}-1} p_i; \\ &= 1 - p_1 + \sum_{t=2}^{t_{\max}-1} t \prod_{i=1}^{t-1} p_i - \sum_{t=3}^{t_{\max}} t \prod_{i=1}^{t-1} p_i + \sum_{t=3}^{t_{\max}} \prod_{i=1}^{t-1} p_i + t_{\max} \prod_{i=1}^{t_{\max}-1} p_i; \\ &= 1 + \sum_{t=1}^{t_{\max}-1} \prod_{i=1}^t p_i. \end{aligned} \quad (2.6)$$

Le niveau de retour z_m associé à la période de retour m est la valeur pour laquelle le temps d'attente moyen du dépassement est égale à m . En posant $E[T] = m$, le niveau de retour associé noté z_m^{EWT} est alors solution de l'équation suivante :

$$m = E[T] = 1 + \sum_{i=1}^{t_{\max}-1} \prod_{t=1}^i F_{Z,t}(z_m^{EWT}). \quad (2.7)$$

b. Expected Number of Exceedances (ENE)

La méthode ENE a été proposée pour la première fois par [Parey et al. \(2007, 2010\)](#). Soit N le nombre de dépassements du niveau z_m de la variable aléatoire (Z_t) sur la période m avec $N = \sum_{t=1}^m \mathbb{1}_{\{Z_t > z_m\}}$. Le nombre moyen de dépassements $E[N]$ est :

$$E[N] = \sum_{t=1}^m E[\mathbb{1}_{\{Z_t > z_m\}}] = \sum_{t=1}^m P(Z_t > z_m) = \sum_{t=1}^m \{1 - F_{Z,t}(z_m)\}. \quad (2.8)$$

Dans la méthode ENE, le niveau de retour associé à une période de retour m noté z_m^{ENE} , est la valeur qui est dépassée en moyen une fois pendant la période m . Ainsi z_m^{ENE} est la solution de l'équation suivante :

$$1 = \sum_{t=1}^m \left\{ 1 - F_{Z,t} \left(z_m^{ENE} \right) \right\}. \quad (2.9)$$

En particulier dans une approche POT-PP non-stationnaire qui combine une loi POT non-stationnaire avec une distribution de Poisson non-stationnaire de fonction d'intensité $\lambda(t)$ (Leadbetter (1983)), la probabilité de dépassement

$$P(Z_t > z) = P(Z_t > z \mid Z_t > u(t)) P(Z_t > u(t)). \quad (2.10)$$

D'après les équations 2.2, 2.10 et 2.9, le niveau de retour z_m en approche POT-PP non-stationnaire est tel que :

$$\sum_{t=1}^m \left(1 + \frac{\gamma(t)}{\sigma(t)} \left(z_m^{ENE} - u(t) \right) \right)^{-\frac{1}{\gamma(t)}} \lambda(t) = 1, \quad (2.11)$$

où $\lambda(t) = E(N)$ à chaque instant t . La définition du niveau de retour dans le contexte stationnaire est un cas particulier de la définition donnée ci-dessus.

c. Equivalent Reliability (ER)

La méthode ER a été proposée par Hu *et al.* (2018). Dans un cadre stationnaire, la fiabilité pendant la durée de vie de conception $T_2 - T_1$ d'un équipement associé à une période de retour m , est notée $RE_{T_2-T_1}^s$ et calculée par :

$$P(Z_{T_1} < z_m, \dots, Z_{T_2} < z_m) = RE_{T_2-T_1}^s = \left(1 - \frac{1}{m} \right)^{T_2-T_1+1}. \quad (2.12)$$

Dans un contexte non stationnaire, la fiabilité équivalente de conception $RE_{T_2-T_1}^{ns}$ associée à une valeur de conception z_m pendant une durée de vie de conception $T_2 - T_1$ est donnée par :

$$RE_{T_2-T_1}^{ns} = \prod_{t=T_1}^{T_2} F_{Z,t}(z_m). \quad (2.13)$$

En posant $RE_{T_2-T_1}^s = RE_{T_2-T_1}^{ns}$, la valeur de conception $z_{T_2-T_1,m}^{ER}$ basée sur la méthode ER peut être calculée en résolvant l'équation suivante :

$$RE_{T_2-T_1}^{ns} = \prod_{t=T_1}^{T_2} F_{Z,t}(z_{T_2-T_1,m}^{ER}) = \left(1 - \frac{1}{m} \right)^{T_2-T_1+1}. \quad (2.14)$$

d. Average Design Life Level (ADLL)

La fiabilité moyenne annuelle dans un cadre non stationnaire est définie comme (Read et Vogel (2015), Yan *et al.* (2017))

$$RE_{T_2-T_1}^{ave} = \frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} p_t = \frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} F_{Z,t}(z_m). \quad (2.15)$$

2.3. MODÈLE POT NON STATIONNAIRE

La méthode ADLL suppose que pour un équipement dont la durée de vie de conception va de T_1 à T_2 , la fiabilité moyenne annuelle associée à une valeur de conception z_m doit être identique à la fiabilité annuelle $1 - \frac{1}{m}$, i.e $RE_{T_2-T_1}^{ave} = 1 - \frac{1}{m}$. Ainsi, la valeur de conception de m-année $z_{T_2-T_1,m}^{ADLL}$ basée sur la méthode ADLL peut être dérivée de l'équation suivante :

$$\frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} F_{Z,t} \left(z_{T_2-T_1,m}^{ADLL} \right) = 1 - \frac{1}{m}. \quad (2.16)$$

2.3.4. Inférence statistique

Les paramètres de la fonction d'intensité du processus ponctuel de Poisson non-homogène et de la fonction d'échelle de la loi GPD sont estimés par maximum de vraisemblance. Comme dans d'autres études (Coles *et al.* (2001), Pujol *et al.* (2007), Parey *et al.* (2007)), nous maintenons constant le paramètre de forme γ et le seuil u afin de ne pas augmenter les incertitudes d'estimation. On sait que les dates des dépassements convergent vers un processus de Poisson non-stationnaire dont la loi de probabilité est donnée par l'équation 2.3. La vraisemblance associée alors à ce modèle a la forme suivante :

$$L(t_1, \dots, t_{n_u} | \lambda_\alpha(t)) = \exp \left(- \int_0^{AL} \lambda_\alpha(t) dt \right) \prod_{i=1}^{n_u} \lambda_\alpha(t_i), \quad (2.17)$$

où A représente en années la taille de la période d'enquête, L en jours le nombre d'observations prise en compte dans l'année et α le vecteur des paramètres de la fonction d'intensité λ_α . On en déduit la log-vraisemblance :

$$\ell(t_1, \dots, t_{n_u} | \lambda_\alpha(t)) = - \int_0^{AL} \lambda_\alpha(t) dt + \sum_{i=1}^{n_u} \log [\lambda_\alpha(t_i)]$$

Ce qui donne

$$\ell(t_1, \dots, t_{n_u} | \lambda_\alpha(t)) = -\Lambda_\alpha(AL) + \sum_{i=1}^{n_u} \log(\lambda_\alpha(t_i)),$$

où Λ_α est la primitive de la fonction d'intensité λ_α . L'établissement des équations de vraisemblance associées aux paramètres α_i conduit généralement à un système d'équations dont la solution n'est pas toujours explicite. Pour cela des algorithmes d'optimisation numérique, sont utilisés pour obtenir les estimateurs du maximum de vraisemblance.

La série des dépassements $(z_1, \dots, z_{n_u}) \sim GPD(\exp(\sigma_\beta(t)), \gamma)$ de fonction d'échelle $\exp(\sigma_\beta(t))$, sa distribution est donnée par l'équation 2.2. Sa fonction densité de probabilité est définie pour tout $z \in \mathbb{R}$ par :

$$h(z_i | \sigma_\beta(t)) = \begin{cases} \frac{1}{\exp(\sigma_\beta(t))} \left(1 + \gamma \frac{z}{\exp(\sigma_\beta(t))} \right)^{-(1+\frac{1}{\gamma})}, & si \ \gamma \neq 0 \\ \frac{1}{\exp(\sigma_\beta(t))} \exp \left(-\frac{z}{\exp(\sigma_\beta(t))} \right), & si \ \gamma = 0 \end{cases} \quad (2.18)$$

où β représente le vecteur de paramètres de la fonction d'échelle. La vraisemblance associée est donnée par :

$$L(z_i | \sigma_\beta) = \prod_{i=1}^{n_u} h(z_i | \sigma_\beta(t_i)) \quad (2.19)$$

Par suite la log-vraisemblance se déduit comme suit :

$$\ell(z_i | \sigma_\beta) = \begin{cases} -\sum_{i=1}^{n_u} \sigma_\beta(t_i) - \left(1 + \frac{1}{\gamma}\right) \sum_{i=1}^{n_u} \log(1 + \gamma z_i \exp(-\sigma_\beta(t_i))), & \gamma \neq 0 \\ -\sum_{i=1}^{n_u} \sigma_\beta(t_i) - \sum_{i=1}^{n_u} z_i \exp(-\sigma_\beta(t_i)), & \gamma = 0 \end{cases} \quad (2.20)$$

Là aussi, on établit facilement les équations de vraisemblance associées aux β_i qui peuvent se résoudre comme précédemment par des méthodes d'optimisation numériques.

2.4. Résultats

Dans cette section nous présentons les principaux résultats issus de la modélisation des données de températures (Tmax, Tmin) et des précipitations extrêmes au cours de ces 60 dernières années au Burkina Faso. Une extrapolation des tendances identifiées a permis de calculer des niveaux de retour et leur intervalle de confiance à 95%. La méthode ENE étant la moins coûteuse en temps de calcul, c'est elle que nous avons choisie pour le calcul des quantiles extrêmes.

2.4.1. Application aux données de températures

a. Estimation des paramètres des modèles GPD et NHPP

Nous avons couplé la méthode des quantiles et la méthode Mean Excess Plot (MEF), voir section 1.1.2 c) du Chapitre 1, pour choisir le seuil u . Nous avons pris d'abord le quantile à 95% et après nous avons tracé la fonction MEF afin de vérifier que le seuil sélectionné se situe bien dans la partie où la fonction MEF est approximativement linéaire. Ainsi les seuils optimaux identifiés sur l'ensemble de nos stations synoptiques sont consignés dans le tableau 2.2. La Figure 2.2 nous montre les graphes des points $(u, e(u))$ de six stations pour les températures maximales, où $e(u)$ est la moyenne des excès au delà du seuil u . Les paramètres estimés par la méthode du maximum de vraisemblance des variables Tmin et Tmax sont consignés dans le tableau 2.3.

b. Tendances identifiées

Les tendances sont identifiées dans les paramètres des distributions des valeurs extrêmes et des occurrences comme décrites ci-dessus (paragraphe 2.3). Le paramètre de forme γ est supposé constant afin de réduire l'incertitude dans les estimations. Par contre nous optons pour des tendances polynomiales restreintes aux

2.4. RÉSULTATS

Variables	Ouaga	Dori	Ouahigouya	Dédougou	Bogandé	Fada	Bobo	Boromo	Pô	Gaoua
Tmin	29.6	31	30.3	29.4	30.6	29	27	28.9	28	27.5
Tmax	41.9	41.6	43	42.4	38.4	42	39.5	41.8	40.9	39.6

TABLE 2.2 – Seuils optimaux u (en ° C) estimés pour Tmin et Tmax.

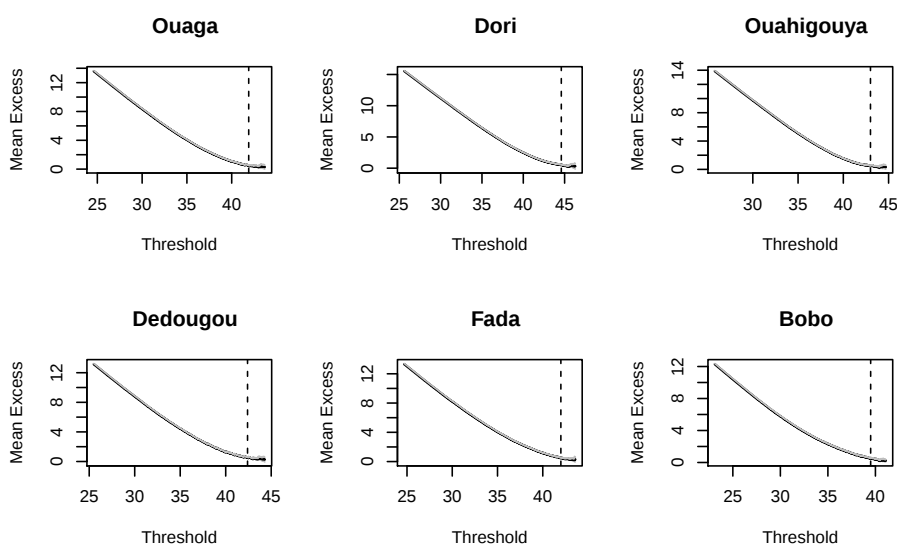


FIGURE 2.2 – Choix du seuil u pour Tmax par la méthode "Mean Excess Plot".

ordres un et deux pour décrire la fonction d'intensité du processus de Poisson non-homogène et d'échelle de la loi des excès. Les résultats des meilleurs tendances ajustant les données pour les deux paramètres aussi bien pour Tmin, que pour Tmax sont obtenus en utilisant le test du rapport du maximum de vraisemblance au risque de 5%. Les degrés des fonctions polynômes sélectionnées par le test sont rangés dans le tableau 2.4. Le tableau 2.5 résume les p-values et les déviations statistiques de ces tests. Les tendances optimales de l'intensité des occurrences des températures extrêmes en considérant à la fois les températures minimales et maximales montrent une augmentation importante des occurrences dans certaines localités. Pour Tmax, 2 stations montrent une tendance linéaire, 5 quadratiques. Tandis que l'augmentation semble plus importante pour les températures minimales avec 4 stations montrant des tendances linéaires et 4 quadratiques. La figure 2.3 illustre ces tendances dans quelques localités du pays.

2.4. RÉSULTATS

Variables	Paramètres	Ouaga	Dori	Ouahigouya	Dédougou	Fada	Bobo	Boromo	Gaoua
Tmin	α_0	1.064	-4.569	-0.327	-0.007	0.095	0.065	0.313	0.059
	α_1	-0.075	0.390	0.124	0.109	0.037	0.091	0.100	0.009
	α_2	0.004	-0.004	0.000	0.000	0.001	0.000	0.000	0.002
	β_0	-0.856	0.127	-0.201	-0.487	-0.534	0.037	-0.252	-0.495
	β_1	0.017	0.000	0.000	0.000	0.011	0.000	0.000	0.008
	γ	-0.191	-0.257	-0.120	0.126	-0.296	-0.328	-0.138	-0.261
Tmax	α_0	1.504	0.918	2.383	2.169	0.538	1.708	1.652	1.816
	α_1	-0.044	0.054	0.000	-0.141	0.008	-0.106	-0.079	0.024
	α_2	0.002	0.000	0.000	0.004	0.002	0.003	0.003	0.000
	β_0	-0.902	-0.645	-0.290	-0.778	-0.256	-0.295	-0.325	0.045
	β_1	0.011	0.012	0.000	0.009	0.000	0.005	0.000	0.000
	γ	-0.162	-0.331	-0.246	-0.234	-0.273	-0.461	-0.261	-0.409

TABLE 2.3 – Paramètres estimés par la méthode du maximum de vraisemblance pour Tmin et Tmax.

Variables	Fonctions	Ouaga	Dori	Ouahigouya	Dédougou	Bogandé	Fada	Bobo	Boromo	Pô	Gaoua
Tmin	σ_β	1	0	0	0	0	1	0	1	0	1
	λ_α	2	2	1	1	2	2	1	1	0	2
Tmax	σ_β	1	1	0	1	0	0	1	0	0	0
	λ_α	2	1	0	2	0	2	2	2	0	1

TABLE 2.4 – Degré de la tendance optimale du paramètre d'échelle ($\sigma_{\beta,u}(t)$) et l'intensité du processus de Poisson non stationnaire ($\lambda_\alpha(t)$).

c. Niveaux de retour non-stationnaire

Les niveaux de retour non-stationnaires en degré celsius ($^{\circ}\text{C}$) associés aux périodes de retour $m = 20, 30, 50, 100$ et leur intervalle de confiance à 95% sont calculés et présentés dans les tableaux 2.7 et 2.8 respectivement pour Tmax et Tmin. Dans ces calculs nous avons pris en compte à la fois les tendances identifiées dans l'intensité des occurrences des événements extrêmes et le paramètre d'échelle de la loi GPD en utilisant la méthode ENE 2.9. Globalement, nous notons une tendance à

2.4. RÉSULTATS

Modèles	Indicateurs	Ouaga	Dori	Ouahigouya	Dédougou	Fada	Bobo	Boromo	Gaoua
$\alpha_0 + \alpha_1 t$	Déviances	37.96	22.78	3.20	53.14	61.98	30.34	37.84	4.51
	p-values	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.03
$\alpha_0 + \alpha_1 t + \alpha_2 t^2$	Déviance	9.40	0.01	0.04	33.05	6.27	26.20	14.96	0.53
	p-value	0.00	0.90	0.84	0.00	0.01	0.00	0.00	0.47
$\exp(\beta_0 + \beta_1 t)$	Déviances	5.81	8.54	0.13	7.59	0.37	5.31	2.71	1.44
	p-values	0.02	0.00	0.72	0.01	0.54	0.02	0.10	0.23

TABLE 2.5 – p-value du test du rapport de vraisemblance des tendances de la fonction d'intensité λ_α et d'échelle σ_β pour Tmax.

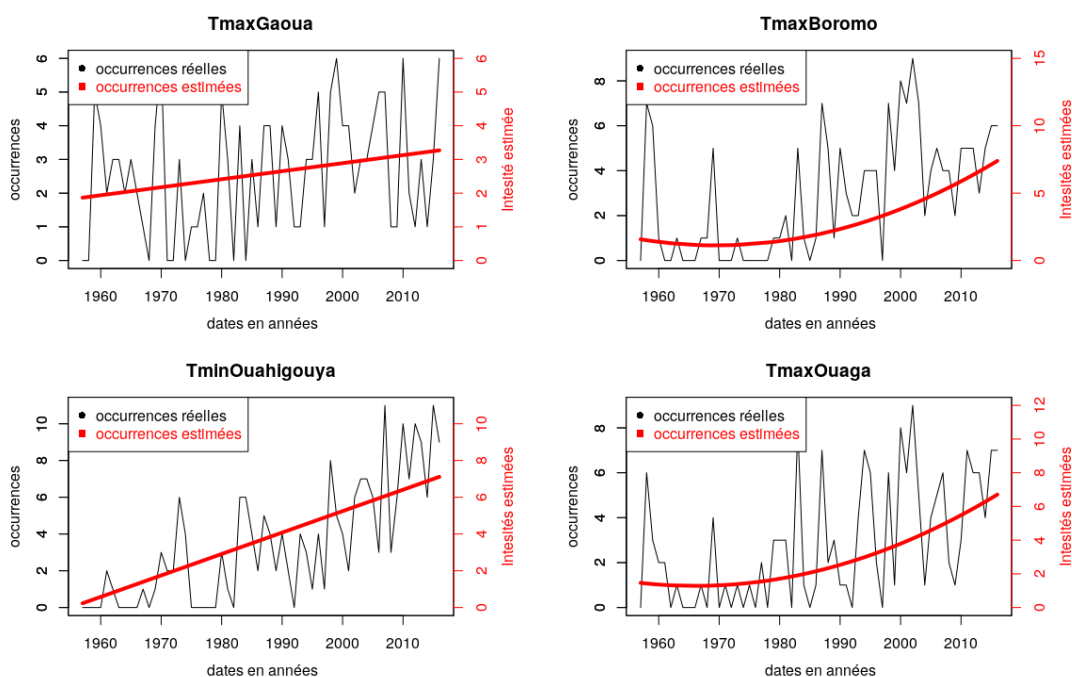


FIGURE 2.3 – Tendence des occurrences des températures minimales et maximales.

2.4. RÉSULTATS

Modèles	Indicateurs	Ouaga	Dori	Ouahigouya	Dédougou	Fada	Bobo	Boromo	Gaoua
$\alpha_0 + \alpha_1 t$	Déviances	97.93	161.33	109.64	78.34	95.13	69.94	70.73	108.13
	p-values	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\alpha_0 + \alpha_1 t + \alpha_2 t^2$	Déviances	22.89	8.49	3.40	0.17	6.16	1.05	2.18	9.94
	P-values	0.00	0.00	0.07	0.68	0.01	0.31	0.14	0.00
$\exp(\beta_0 + \beta_1 t)$	Déviances	11.01	1.29	0.94	0.13	6.18	1.37	1.00	4.20
	p-values	0.00	0.26	0.33	0.72	0.01	0.24	0.32	0.04

TABLE 2.6 – p-value du test du rapport de vraisemblance des tendances de la fonction d'intensité λ_α et d'échelle σ_β pour Tmin.

la hausse des niveaux de retour non-stationnaires de l'ordre de 0.1°C en moyenne aussi bien pour Tmin que pour Tmax sur les prochaines années. La tendance à la hausse est beaucoup plus significative pour les températures minimales que pour les maxima. Ils sont beaucoup plus élevés au nord qu'au sud du pays, ce qui est en parfait accord avec les tendances identifiées dans la section précédente et les caractéristiques climatiques de ces régions. La figure 2.5 illustre ces différentes variations spatiales.

Stations	20RL[95% CI]	30RL[95% CI]	50RL[95% CI]	100RL[95% CI]
Ouaga	44.1 [42.9, 48.2]	44.4 [43.1, 48.2]	44.4 [43.1, 48.2]	44.4 [43.1, 48.4]
Dori	46.2 [45.6, 46.7]	46.2 [45.6, 46.7]	46.2 [45.6, 46.7]	46.2 [45.6, 46.7]
Ouahigouya	45.8 [45.0, 46.7]	45.8 [45.0, 46.7]	45.8 [45.0, 46.8]	45.9 [45.0, 47.0]
Dédougou	44.3 [43.5, 45.7]	44.4 [43.5, 45.7]	44.4 [43.5, 45.7]	44.4 [43.5, 45.7]
Fada	44.6 [44.0, 45.8]	44.6 [44.0, 46.0]	44.7 [44.0, 46.0]	44.7 [44.0, 46.1]
Bobo	41.1 [40.6, 41.4]	41.1 [40.6, 41.4]	41.1 [40.6, 41.4]	41.1 [40.6, 41.4]
Boromo	44.3 [43.8, 45.8]	44.3 [43.8, 45.8]	44.4 [43.8, 45.8]	44.5 [43.8, 45.8]
Gaoua	42.1 [41.7, 42.4]	42.1 [41.7, 42.4]	42.1 [41.7, 42.5]	42.1 [41.7, 42.5]

TABLE 2.7 – Niveau de retour non- stationnaire(en $[\circ]\text{C}$) et intervalle de confiance à 95% pour Tmax calculés par la méthode de bootstrap.

2.4. RÉSULTATS

Stations	20RL [95% CI]	30RL [95% CI]	50RL [95% CI]	100RL [95% CI]
Ouaga	31.8 [30.8, 33.7]	31.8 [30.8, 33.8]	31.8 [30.8, 33.8]	31.8 [30.8, 33.8]
Dori	35.4 [34.0, 36.7]	34.8 [34.1, 36.8]	35.1 [34.1, 36.9]	35.2 [34.1, 37.0]
Ouahigouya	34.8 [33.3, 36.4]	35.0 [33.4, 37.0]	35.3 [33.5, 37.7]	35.6 [33.5, 38.7]
Dédougou	39.6 [34.3, 50.0]	41.2 [34.7, 50.0]	43.5 [35.2, 50.0]	47.1 [35.9, 50.0]
Fada	31.0 [30.1, 31.8]	31.0 [30.1, 31.9]	31.0 [30.1, 31.9]	31.0 [30.1, 31.9]
Bobo	30.0 [29.4, 30.4]	30.0 [29.4, 30.5]	30.1 [29.4, 30.5]	30.1 [29.4, 30.6]
Boromo	32.9 [31.8, 34.5]	33.1 [31.8, 34.9]	33.3 [31.9, 35.4]	33.5 [31.9, 36.1]
Gaoua	29.8 [28.7, 31.0]	29.8 [28.7, 31.2]	29.8 [28.7, 31.3]	29.8 [28.7, 31.2]

TABLE 2.8 – Niveau de retour non- stationnaire (en $^{\circ}$ C) et intervalle de confiance à 95% pour Tmin calculés par la méthode de bootstrap.

2.4.2. Application aux données pluviométriques

a. Estimation des paramètres des modèles GPD et NHPP

Comme pour les températures, nous avons estimé ici aussi les paramètres par la méthode du maximum de vraisemblance. Ces paramètres sont résumés dans le tableau 2.9

Variable	Paramètres	Ouaga	Dori	Ouahigouya	Dédougou	Bogandé	Fada	Bobo	Boromo	Pô	Gaoua
Pluie	α_0	8.733	8.850	6.968	8.833	8.600	8.700	8.650	8.833	7.214	8.750
	α_1	0.000	0.000	0.058	0.000	0.000	0.000	0.000	0.000	0.052	0.000
	β_0	2.665	2.735	2.658	2.683	2.688	2.803	2.754	2.636	2.669	2.689
	γ	0.059	-0.049	0.023	-0.032	0.019	-0.021	-0.020	0.028	0.027	0.037

TABLE 2.9 – Paramètres estimés des fonctions d'intensité λ_{α} et d'échelle σ_{β} pour les précipitations

b. Tendances identifiées

Nous notons que contrairement aux températures, les résultats des tendances optimales estimées pour les précipitations extrêmes résumés dans le tableau 2.10 montrent que l'intensité des occurrences des pluies extrêmes est pratiquement constante pour la même période dans plusieurs localités du pays. Par contre certaines régions comme le Nord et le Sud du pays pourraient enregistrer une forte intensification des précipitations extrêmes à tendance linéaire. Ce qui est en parfait accord avec le retour à des périodes humides marquées par une fréquence élevée des événements de forte intensité pluviométrique (PANA, 2007). Les p-values des différents tests de détection de tendance sont consultables dans le tableau 2.11. La figure 2.4 montre les tendances identifiées dans le Nord et le Sud-ouest du pays.

2.4. RÉSULTATS

Variable	Fonctions	Ouaga	Dori	Ouahigouya	Dédougou	Bogandé	Fada	Bobo	Boromo	Pô	Gaoua
Pluie	σ_β	0	0	0	0	0	0	0	0	0	0
	λ_α	0	0	1	0	0	0	0	0	1	0

TABLE 2.10 – Degrés de la tendance optimale du paramètre d'échelle ($\sigma_\beta(t)$) et l'intensité du processus de Poisson ($\lambda_\alpha(t)$) pour les précipitations.

Modèles	Indicateurs	Ouaga	Dori	Ouahigouya	Dédougou	Bogandé	Fada	Bobo	Boromo	Pô	Gaoua
$\alpha_0 + \alpha_1 t$	Déviances	1.38	0.54	8.24	0.59	0.08	0.86	0.03	1.53	6.11	2.15
	p-values	0.24	0.46	0.00	0.44	0.78	0.35	0.87	0.22	0.01	0.14
$\alpha_0 + \alpha_1 t + \alpha_2 t^2$	Déviances	4.82	10.36	18.59	9.97	1.84	1.75	3.21	1.28	4.66	0.45
	p-values	0.03	0.00	0.00	0.00	0.18	0.19	0.07	0.26	0.03	0.50
$\exp(\beta_0 + \beta_1 t)$	Déviances	0.08	0.01	0.00	1.14	0.39	0.56	0.11	1.19	2.56	0.22
	p-values	0.78	0.94	0.99	0.28	0.53	0.46	0.74	0.28	0.11	0.64

TABLE 2.11 – p-value des tests du rapport de vraisemblance pour les précipitations au niveau de confiance de 95%.

c. Niveaux de retour non-stationnaire

Les résultats des niveaux de retour des précipitations sont également calculés en prenant en compte les tendances identifiées en utilisant la méthode ENE 2.9. Ces résultats attestent que des événements à forte intensité pluviométrique à l'image de celui du premier septembre 2009 pourraient survenir dans certaines localités du Pays. Le niveau de retour centennal au Centre et au Sud-ouest du pays est respectivement estimé à 295,56 mm et 265,63 mm de pluies. Il faut ajouter que pour la même période de retour, d'autres régions du pays pourront être le théâtre

2.4. RÉSULTATS

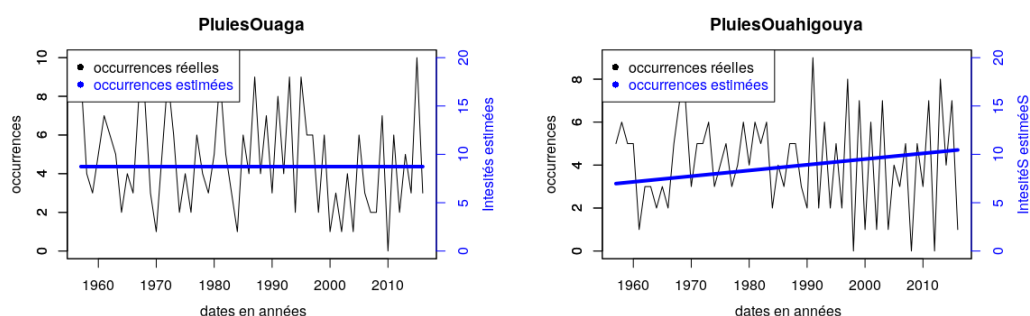


FIGURE 2.4 – Tendence des occurrences des précipitations extrêmes pour les stations de Ouaga au Centre et Ouahigouya au Nord du pays.

des événements pluvieux mais avec des intensités assez variables. C'est notamment le cas de la région du Nord (235,56 mm, station de Ouahigouya), la région de l'Est (233, 09 mm, station de Bogandé) et celle de la Boucle du Mouhoun (238,14 mm, station de Boromo). Le tableau 2.12 résume l'essentiel de ces résultats exprimés en mm de pluies.

Stations	20RL [95% CI]	50RL [95% CI]	100RL [95% CI]
Ouaga	249.1 [169.4, 410.3]	275.2 [179.3, 477.6]	295.9 [186.6, 500]
Dori	150 [129.9, 500]	158.1 [135.1, 500]	163.9 [138.8, 500]
Ouahigouya	200.3 [141.3, 286.8]	219.4 [149.6, 326.3]	235.7 [155.7, 365.6]
Dedougou	162.9 [133.8, 500]	172.2 [139, 500]	179 [142.5, 500]
Bogandé	203.3 [147.4, 330.9]	220.2 [154.7, 377.3]	233.1 [159.6, 416.8]
Fada	189 [146.2, 500]	200.9 [151.3, 500]	209.7 [155, 500]
Bobo	185.1 [144.2, 500]	196.5 [149.4, 500]	205.1 [153.2, 500]
Boromo	206.9 [145.4, 330.6]	224.5 [152.4, 377.2]	238.1 [157.5, 416.4]
Pô	212 [149.5, 331]	231.9 [156.7, 384.5]	248.7 [162, 432.9]
Gaoua	229 [156.6, 376.1]	249.6 [164.3, 418.7]	265.6 [169.8, 465]

TABLE 2.12 – Niveau de retour non- stationnaire (en mm de pluies) et intervalle de confiance à 95% pour les précipitations calculés par la méthode de bootstrap.

2.4.3. Krigeage des niveaux de retour

La production de cartes de niveaux de retour est l'un des objectifs courants lors d'une étude des extrêmes spatiaux. En utilisant les conclusions de cette étude nous fournissons maintenant une carte de krigeage des niveaux de retour des températures et de précipitations extrêmes associés aux périodes de retour $m = 20$ et $m = 100$ ans. Pour ce faire nous définissons une grille régulière de positions assez fine qui couvre le domaine d'étude. La structure spatiale est modélisée grâce à un semi-variogramme exponentiel. La figure 2.5 montre les cartes des niveaux de retour non-stationnaires de Tmax et Tmin sur 20 ans (resp. 100 ans). Les températures maximales susceptibles d'être observées une fois au moins tous les 20 ans (resp. 100 ans) atteignent $44 \pm 1.1^\circ\text{C}$ à $46 \pm 1.1^\circ\text{C}$ en particulier dans la zone sahélienne.

La région du centre pourrait connaître des périodes de chaleur assez remarquables. Pour les mêmes périodes de retour, les températures minimales attendues tournent entre $30 \pm 1.2^\circ\text{C}$ et $32 \pm 1.2^\circ\text{C}$ dans la zone soudanienne, $32 \pm 2.9^\circ\text{C}$ et $34 \pm 2.9^\circ\text{C}$ dans la zone soudano-sahélienne. Les plus fortes minima sont attendues dans la zone sahélienne. Il faut ajouter que les valeurs prédites des niveaux de retour non-

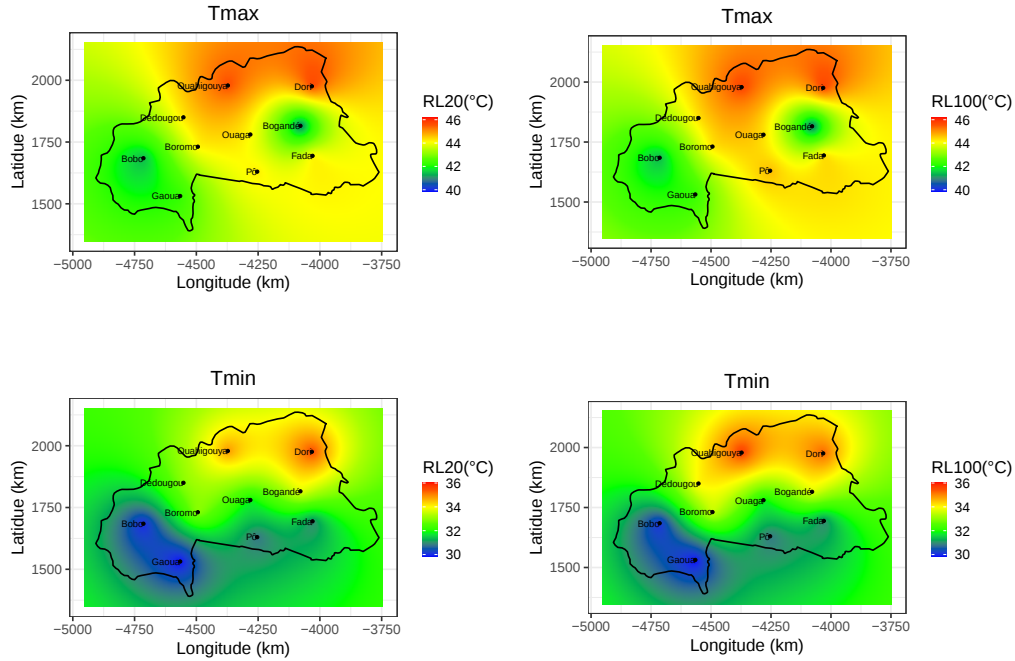


FIGURE 2.5 – Krigeage des niveaux de retour non-stationnaires de Tmax et Tmin pour des périodes de retour $m = 20$ et $m = 100$ ans

stationnaires des températures maximales ne bougent que très peu d'une période à l'autre, parce que nous avons identifié un paramètre de forme négatif sur la plus part des stations. Ce qui signifie que la distribution de queue des températures maximales est bornée et de ce fait ça bouge assez lentement dans la queue. Par contre nous notons une différence significative entre les niveaux de retour non-stationnaires à 20 ans et à 100 ans pour les températures minimales. Nous notons une légère tendance à la hausse de ces niveaux de retour non-stationnaires d'une période à l'autre. Et comme conséquence il faut s'attendre à des nuits de plus en plus chaudes dans les prochaines années. Cela pourrait avoir un impact négatif sur la santé des personnes fragiles. Notons que l'approche du krigeage spatiale ajuste assez bien la structure spatiale des températures minimales et maximales. L'erreur de krigeage est presque nul sur les stations et varie entre 1°C et 2°C au delà des stations. La figure 2.6 donne les détails de ces calculs. Les cartes des niveaux de retour non-stationnaires des températures minimales traduisent bien l'évolution de ces températures dans les différentes zones climatiques. Les extrêmes des pluies journalières susceptibles de tomber une fois au moins tous les 100 ans varient entre 200 mm et 290 mm selon les endroits. Les régions du Sud-Ouest, du Centre et de l'Est seront les plus touchées par les précipitations extrêmes très localisées. Nous n'avons pas détecté de tendance

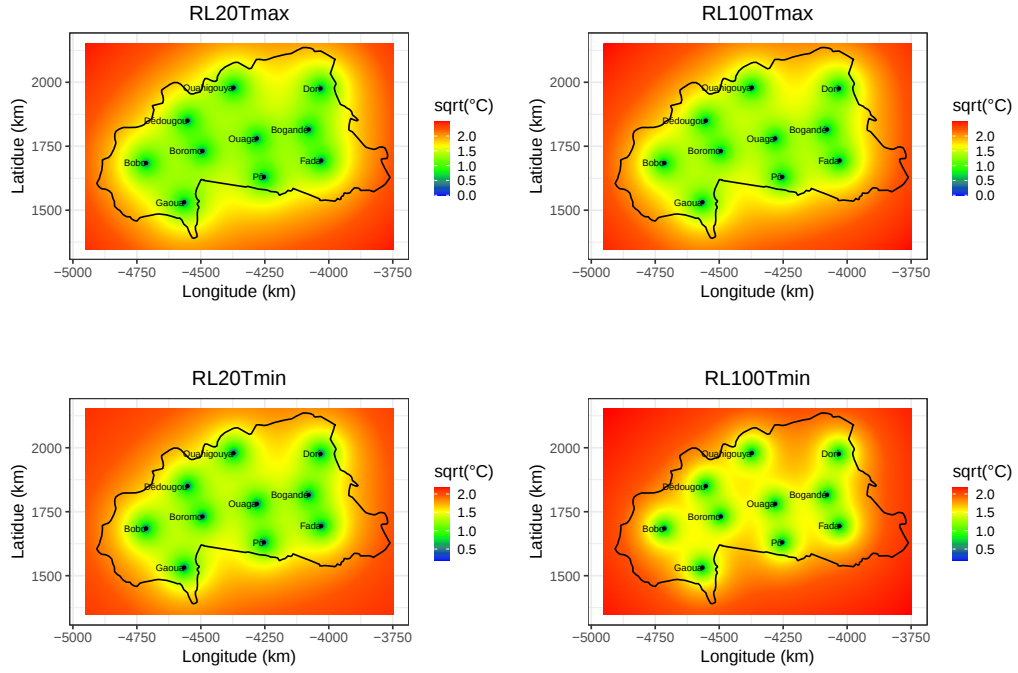


FIGURE 2.6 – Ecart-type de krigeage des niveaux de retour non-stationnaires de Tmax et Tmin pour des périodes de retour $m = 20$ et $m = 100$ ans

pour les précipitations extrêmes ni de structure spatiale cohérente dans l'analyse ponctuelle. Le krigeage spatial, n'ajuste pas très bien la structure spatiale des précipitations non-stationnaires. L'ajustement est meilleur aux stations, cependant en dehors des stations, cette approche n'arrive pas à capturer la structure spatiale sous-jacente de la non-stationnarité des précipitations. Une étude plus approfondie des précipitations extrêmes dans un contexte de changement temporel et spatial est fournie dans les chapitres 4 et 5.

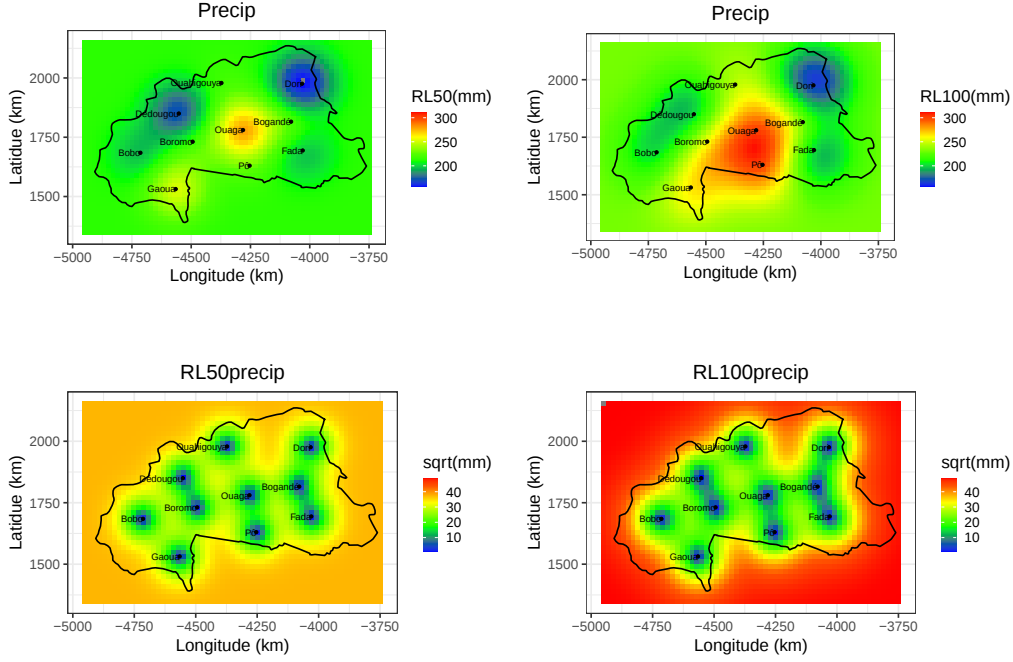


FIGURE 2.7 – Krigeage des niveaux de retour non-stationnaires des précipitations extrêmes pour des périodes de retour $m = 50$ et $m = 100$ ans

2.5. Discussion et perspectives

L'étude présentée ici utilise des séries chronologiques de températures et de précipitations journalières entre 1957 et 2016 de dix stations synoptiques extraites de la base de données climatologiques du Burkina Faso. Ces stations ont été choisies de sorte à garantir une bonne uniformité spatiale, une bonne représentativité des différents régimes climatiques et la qualité des données. Nous avons utilisé la théorie statistique de valeurs extrêmes, en particulier l'approche des dépassements de seuil pour décrire la non stationnarité des températures et des précipitations extrêmes. Pour chaque série temporelle, les dépassements au dessus d'un seuil (95 ème percentile) ont été soumis à une procédure de declustérisation à pas journalier ($r = 1$) afin de garantir l'hypothèse d'indépendance dans les séries des excès. Les séries des excès résultantes ont ensuite été ajustées par une loi *GPD* flexible et les dates des dépassements à un processus de Poisson non stationnaire afin d'identifier les éventuelles tendances et calculer les niveaux de retour. Les tendances optimales pour le paramètre d'échelle de la loi *GPD* et l'intensité du processus de Poisson ont été identifiées en supposant le paramètre de forme constant. Les résultats obtenus pour Tmin et Tmax montrent une augmentation assez significative de la fréquence de fortes chaleurs car la moitié des stations présentent une tendance quadratique pour l'intensité de la loi *NHPP* et la plupart des autres linéaires.

Contrairement aux températures, les tendances optimales identifiées pour les précipitations montrent une constance des évènements de forte intensité pluvio-

métrique sur toute la région, mais néanmoins n'excluent pas de pluies extrêmes localisées, en attestent les stations de Ouaga, Pô et Ouahigouya. Ce constat serait sans doute lié à la forte variabilité spatiale et temporelle des précipitations extrêmes, qui complique davantage la détection de tendances significatives des extrêmes pluviométriques dans la région. A partir de l'extrapolation des tendances des paramètres ci-dessus mentionnés (σ_β et λ_α) les niveaux de retour dans un contexte non stationnaire ont été calculés ainsi que leur intervalle de confiance. Les valeurs plus élevées pour les 50RLs par exemple sont atteintes au Nord et au Sahel dépassant $45.8^\circ C$ pour Tmax et $35.3^\circ C$ pour Tmin. Pour la même période de retour, les précipitations les plus élevées sont atteintes par contre au centre (275.2 mm) et au Sud-ouest (249.6 mm) de pays.

En somme nous constatons qu'il n'y a pas de différences significatives entre les niveaux de retour non-stationnaires à 20 ans et à 100 ans pour les températures maximales. Pour la simple bonne raison que nous avons identifié un paramètre de forme négatif ($\gamma_u < 0$) quasiment sur toutes les stations. Dans ce cas il faut noter que la distribution de queue des températures maximales est bornées, ce qui justifie le fait que ça bouge assez très lentement dans la queue de la distribution. Par contre pour les températures minimales, nous notons une différence significative entre les niveaux de retour non-stationnaires à 20 ans et à 100 ans. Ces niveaux ont une légère tendances à la hausse sur pratiquement les différences régions climatiques et en conséquence il faut s'attendre à des nuits de plus en plus chaudes dans les prochaines années. Pour les précipitations, il faut souligner que nous avons pas détecté de tendances significatives dans la fonction d'échelle et d'intensité ni de structures spatiales. Cela mets en évidence la difficulté des approches ponctuelles à prendre en compte les non-stationnarités des précipitations, d'où le développement des approches spatio-temporelles entreprises dans la suite de la thèse. Nos résultats sont en accord avec ceux trouvés par [Panthou \(2013\)](#); [Panthou et al. \(2014\)](#), [Wilcox et al. \(2018\)](#) dans la région Ouest Africaine. Dans ces travaux à partir d'un modèle *GEV* régional et ponctuel dépendant du temps ces auteurs ont montré la non-stationnarité des extrêmes des certains paramètres climatiques dans le sahel ouest-africain. Ils concluent que les températures et les précipitations extrêmes sont devenues de plus en plus fréquentes au cours de ces trente dernières années. Il faut ajouter que nous avons considéré ici que des tendances polynomiales limitées aux ordres un et deux pour caractériser l'évolution de la fonction d'échelle σ_β et d'intensité λ_α . Mais on pourrait considérer des tendances polynomiales d'ordre supérieure à deux assez raisonnable. Du point de vue numérique, il n'y a pas de difficultés dans l'estimation des paramètres. Cependant il faut rappeler que l'exploitation des tendances identifiées au déla des données dans ce cas, pourrait surtout surestimer les prédictions des quantiles extrêmes non-stationnaires ([Parey et al. \(2007\)](#), [Parey et al. \(2010\)](#)). Pour contourner ces difficultés, nous pouvons envisager également d'autres modèles plus flexibles permettant de prendre en compte la particularité du terrain d'étude, mais aussi le type de données traitées. Ces modèles de tendances peuvent être constant, constant par morceaux, semi-paramétrique où non linéaire. Elles peuvent intégrer des covariables climatiques qui dépendent eux même du temps. Ces types de tendances apporteront sans doute des informations signi-

ficatives dans l'estimation des niveaux de retour non-stationnaires. Pour comparer rigoureusement ces modèles candidats et tester la significativité des évolutions, dans la pratique nous pouvons diviser les données en deux périodes climatiques de 30 ans. La première période servira de données d'ajustement tandis que la seconde période des données va jouer de rôle de données de validation. Après on ajuste les modèles de tendances candidates aux données expérimentales, et ensuite on extrapole les tendances identifiées au delà des données expérimentales puis on compare les prédictions aux données de validation. Pour cela on pourrait alors calculer et comparer quelques indicateurs tels que l'erreur moyenne absolue, l'erreur moyenne quadratique où encore la cross-validation.

A spatial approach of the POT method in trend detection of extreme temperatures and precipitations in Burkina Faso

Béwentaoré Sawadogo, Liliane Bel, Diakarya Barro

Abstract. In this study, trends in extreme temperatures and rainfall are evaluated using the statistical theory of extreme values, in a non-stationary context. The temporal evolution of the extremes is handled pointwise by the peak over threshold approach. First evolution of the parameters of the distribution of exceedances and the intensity of extreme event occurrences for several meteorological variables are described as linear or quadratic functions of time. Then, the spatial dependence structure is integrated using generalized ℓ -Pareto processes in a non-stationary framework. This allows us to produce a map of return levels with the extrapolation of trends identified in the marginal distributions and the estimated dependence structure. Finally, our approach is applied to temperature and rainfall data in Burkina Faso.

Keywords : Non-stationary functional POT, Space-time Modeling, Spatial Dependence, Trends detection, Climate Change.

3.1. Introduction

Climate change related impacts, such as natural disasters, droughts, floods, sea level rise, and heat waves, have become particularly significant over the last thirty years, with increasing economic losses and a major influence on social stability, economic development, and people's livelihoods. According to the latest report of the Intergovernmental Panel on Climate Change (IPCC, 2018, 2019), average air temperature is currently increasing by $0.2 \pm 0.1^\circ\text{C}$ every ten years and warming could exceed 1.5°C on the horizon 2050. These same reports (IPCC, 2007, 2018, 2019) suggest a trend increasing of the probability of occurrence and the intensity of precipitation extremes. Thus the needs for modeling and accurate prediction of the intensity and spatial extent of future extreme events have multiplied and remain today a research issue of great interest in many areas such as the environment or risk management.

Extreme Value Theory (EVT) is extended to observations that are non-stationary

and not independent and provides rigorous mathematical tools for modeling and detecting recent trends in the extremes of climate variables and extrapolating these events beyond observed data in a spatial and temporal framework. Indeed, climate extremes have been widely studied in several papers, some of which present methods to deal with the non-stationarity generally present in the extremes of environmental processes, e.g. [Dacunha-Castelle et al. \(2015\)](#), [Mentaschi et al. \(2016\)](#) and [Cabral et al. \(2020\)](#) have transformed a non-stationary time series into a stationary series for which classical EVT can be applied. A frequently used alternative for addressing non-stationarity in EVT models is to include covariates, such as space and time, in the model parameters ([Davison et Smith \(1990\)](#), [Eastoe et Tawn \(2009\)](#), [Parey et al. \(2010\)](#), [Cooley \(2013\)](#), [Salas et Obeysekera \(2014\)](#), [Yan et al. \(2017\)](#), [Hu et al. \(2018\)](#)). These approaches are built under certain important assumptions that the underlying process that generates the data is locally stationary and that the observations generated can be considered as independent approximations and generally do not take into account spatial dependence. Recently some forms of non-stationary dependence structures, i.e. integrated in spatial max-stable processes have been studied by [Huser et Davison \(2014\)](#), [Huser et Genton \(2016a\)](#) and [Nicolet et al. \(2018\)](#), [Chevalier et al. \(2020\)](#), but here again these models do not take into account marginal non-stationarities. Although attractive, these spatial models are very computationally expensive for large dimensions and are not adapted to model threshold exceedances. The functional Peak Overs Threshold (POT) approach was introduced in a stationary framework by [Ferreira et al. \(2014\)](#), and generalized by [Dombry et Ribatet \(2015\)](#) and [de Fondeville et Davison \(2020\)](#) which have defined the family of generalized ℓ -Pareto process. Neglecting the non-stationarity of extremes, whether in marginal distributions or dependence structures, would not only give a poor description of the data but would also have dramatic consequences on the estimation of return levels in a climate change context. Although it is relatively simple to construct non-stationary models in the uni-variate framework, for example by letting the parameters of the marginals depend on time and other covariates ([Coles et al. \(2001\)](#), [Engelke et al. \(2019\)](#) and [Parey et al. \(2007\)](#)), it is more difficult to model spatial and temporal trends in the dependence structure. In addition, even if a parametric family can be identified, it can be very difficult to make an inference if the data set is not spatially rich enough to be used for the inference.

This paper proposes a different methodology from the previous ones by coupling two parametric approaches. One of these approaches captures non-stationarity in the marginal distributions and intensity of event occurrences ([Coles et al. \(2001\)](#), [Parey et al. \(2007, 2010\)](#)) while the other focuses on the dependence structure using a functional POT model([de Fondeville et Davison \(2018\)](#); [de Fondeville et Davison \(2020\)](#)). This coupling allows us to obtain a new spatio-temporal threshold exceedance method for assessing the influence of climate change.

This approach has proven to be flexible in dealing with non-stationary data, which is generally difficult when working with environmental variables, such as precipitation or wind, due to seasonal effects or trends in space and time. The way in which it deals with non-stationary data is first to study the temporal evolu-

3.2. POINTWISE PEAKS-OVER-THRESHOLDS

tion of the extremes treated at pointwise in time by describing the parameters of the distribution of exceedances and the intensity of extreme event occurrences of meteorological variables as linear or quadratic functions of time. Then, the spatial dimension is integrated using generalized ℓ -Pareto processes in a non-stationary framework, following the inference procedure developed by [de Fondeville et Davison \(2018\)](#). We use the method of threshold exceedances in a context of temporal and spatial change to analyze future changes in rainfall and temperature in sub-sahelian Africa, particularly in Burkina Faso. We want to determine if the parameters of the pointwise POT distributions show significant trends. Then we show how the trends identified in the marginal distributions fit into the dependence structure of the space-time process underlying rainfall and temperature, in order to evaluate the non-stationary return levels.

The paper is organized as follows : section 2 provides a brief overview on pointwise POT approach in a stationary and non-stationary framework. Section 3 details the space-time functional POT model and statistical inference. In section 4, we present our main results obtained by applying our approach to temperature and precipitation real data.

3.2. Pointwise Peaks-Over-Thresholds

3.2.1. Stationary threshold exceedances modeling

Extreme value theory (EVT) is a branch of statistics that describes the behavior of the largest (or smallest) observations in a data set. It has a very long history and has its origins in the asymptotic convergence theory results of Fisher-Tippett-Gnedenko and Pickands-Belkema-de Haan. The literature associated with this theory essentially proposes two modeling approaches, the block maxima approach and Peaks Over Threshold (POT). In block maxima analysis, the maximum or minimum statistic is investigated by establishing that a generalized parametric family of continuous probability distributions summarizes the asymptotic behavior of the suitably normalized maximum distribution of an independent and identically distributed sample $Z_t, t = 1, \dots, n$. Let $M_n = \max_{1 \leq t \leq n} Z_t$, if there are suitable sequences $a_n > 0$ and $b_n \in \mathbb{R}$ and a non-degenerated G distribution function such as $P\{a_n^{-1}(M_n - b_n) \leq z\} \rightarrow G(z), n \rightarrow \infty$, then the results established by [Fisher et Tippett \(1928\)](#) and [Gnedenko \(1943\)](#) indicate that G is necessarily a Stationary Generalized Extreme Values distribution denoted by $\text{SGEV}(\mu, \sigma, \xi)$, and defined on $\{z \in \mathbb{R} : 1 + \xi(z - \mu)/\sigma > 0\}$, such as

$$\text{SGEV}(z \mid \mu, \sigma, \xi) = \begin{cases} \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, & \text{if } \xi \neq 0 \\ \exp\left\{-\exp\left(-\frac{z-\mu}{\sigma}\right)_+\right\}, & \text{if } \xi = 0 \end{cases} \quad (3.1)$$

Beyond the study of maxima, it is also possible to look at excesses above a threshold (chosen according to the system and the question asked). This strategy, which is quite complementary to the maxima analysis, has the advantage of being more efficient from a statistical point of view because it offers a larger sample size than

that of the maxima. The Peak Over Threshold (POT) approach consists of retaining only those observations that exceed a large threshold. If a random variable Z has distribution function H and there exist sequences of constants $a_n > 0$ and b_n such that

$$nP(a_n^{-1}(Z - b_n) > z) \longrightarrow -\log G(z), n \longrightarrow \infty, \quad (3.2)$$

where G is a non-degenerate distribution function, then Z is said to belong to the max-domain of attraction of G (Resnick (1987)). For a large enough threshold $u < \inf\{z : H(z) = 1\}$, the tail behaviour of Z can be described using a stationary generalized Pareto (SGP) distribution (Balkema et De Haan (1974) and Pickands (1975)),

$$P(Z > z + u \mid Z > u) \approx \bar{H}_{\xi, \sigma} = \begin{cases} (1 + \xi z / \sigma)_+^{-1/\xi}, & \xi \neq 0 \\ \exp(-z / \sigma), & \xi = 0 \end{cases} \quad (3.3)$$

where $\sigma = \sigma(u) > 0$ and here below, $a_+ = \max(a, 0)$ for real a . the shape parameter ξ is also called the tail index. According to the sign of ξ , we end up with three types of distributions of extreme values known as Gumbel ($\xi \rightarrow 0$), Frechet ($\xi > 0$) and Weibull ($\xi < 0$). The tail of the distribution is fully characterized by these two parameters and the probability $\phi_u = P(Z > u)$ of exceeding the threshold u . The modeling of exceedances has been extended to non-stationary settings (Davison et Smith (1990), Eastoe et Tawn (2009), Parey et al. (2010, 2007)).

3.2.2. Non-stationary pointwise POT-PP modeling

In the classical stationary POT approach, observations of a given random variable must first validate the hypotheses of independence and stationarity. However, the presence of different types of variability in the extremes of the weather series (trends, seasonal effects, weather patterns, dependence on other covariates) may call into question the assumption of stationarity, especially in the current context of climate change. In this context, the trend analysis of weather extremes cannot be performed without taking into account the question of the non-stationarity of environmental processes. In the literature, we find two approaches for modeling threshold exceedances of a non-stationary time series. Most of the time, the non-stationarity of environmental processes is dealt with either by a stationarization procedure (Dacunha-Castelle et al. (2015), Ferreira et al. (2017), Cabral et al. (2020)) that allows the classical approach to be applied to the resulting stationary random variable. Another alternative is the treaty by a trend parametric approach in which the non-stationarity is incorporated in the parameters of the tail distribution of threshold exceedances. In this case, the dependence between the parameters and the relevant covariates can take different dependence structures polynomial, semi parametric, or non-linear for detail see (Davison et Smith (1990), Eastoe et Tawn (2009), Coles et al. (2001), Parey et al. (2007, 2010), Cooley (2013), or Salas et Obeysekera (2014)). In order to take into account the characteristics of the study area and recent developments in dependence structures, we decide to treat non-stationarity in this work by using the second modeling alternative, i.e, we describe the parameters of the distribution of exceedances and the intensity of extreme

3.3. NON-STATIONARY SPATIAL ℓ -POT MODELING

events as linear or quadratic functions of time (Coles *et al.* (2001), Parey *et al.* (2007, 2010)) such as,

$$\left. \begin{aligned} \log(\sigma_u(s, t)) &= \sigma_0(s) + \sum_{j=1}^p \sigma_j(s) t^j \\ u(s, t) &= u_0(s) + \sum_{j=1}^p u_j(s) t^j \\ \xi_u(s, t) &= \xi_0(s) + \sum_{j=1}^p \xi_j(s) t^j \end{aligned} \right\} \text{ for all } s \in S \subset \mathbb{R}^d, \quad (3.4)$$

where the parameters are polynomials function of time; $u_j(s)$, $\sigma_j(s)$, $\xi_j(s)$ and $\lambda_j(s)$ are a real and S the spatial set of sites in the region concerned by the study. In this non-stationary environment $Z - u \mid Z > u \sim H_{SN}(\sigma_u(s, t), \xi_u(s, t))$ such that

$$H_{SN}(z) = \begin{cases} 1 - (1 + \xi_u(s, t)z/\sigma_u(s, t))^{-1/\xi_u(s, t)}, & \text{if } \xi_u(s, t) \neq 0 \\ 1 - \exp(-z/\sigma_u(s, t)), & \text{if } \xi_u(s, t) = 0 \end{cases} \quad (3.5)$$

The distribution of exceedance occurrences is modeled by a non-stationary Poisson process whose intensity λ is a time varying function, i.e,

$$P(N(I) = k) = \exp(-\Lambda(I)) \frac{\Lambda(I)^k}{k!}, \quad (3.6)$$

where $\Lambda(I) = \int_{t_1}^{t_2} \lambda(s, t) dt$ and N is the number of occurrences for the whole time interval $I = [t_1, t_2]$ and $\lambda(s, t) = \lambda_0(s) + \sum_{j=1}^p \lambda_j(s) t^j$, $s \in S$. In this case the exceedance probability for a non-stationary POT-PP model is approximated by :

$$\begin{aligned} P(Z(s, t) \geq z) &= P(Z(s, t) - u(s, t) \geq z \mid Z(s, t) \geq u(s, t)) P(Z(s, t) \geq u(s, t)) \\ &\simeq \left(1 + \xi_u(s, t) \frac{z}{\sigma_u(s, t)} \right)^{-1/\xi_u(s, t)} \lambda(s, t) \end{aligned} \quad (3.7)$$

We estimate the parameters of the intensity of the non-homogeneous Poisson process and the distribution of NSGP using the maximum likelihood method under the assumption that exceedances series of $(Z_i(s, t) - u(s, t))_{i=1, \dots, n_u}$ and the dates of the exceedances $(t_i)_{i=1, \dots, n_u}$ are independent. The best model is selected using the likelihood ratio test statistic denoted D_n and defined by :

$$D_n = 2(\ell_n(M_1) - \ell_n(M_0)), \quad (3.8)$$

where M_0 and M_1 are respectively models of degree d_0 and d_1 such that M_0 is nested in M_1 and for n large and under H_0 , D_n follows the distribution of χ_d^2 at $d = d_1 - d_0$ degree of freedom. We extend these non-stationarity pointwise modeling techniques to function data by assuming that the parameters of the marginal distributions are functions of time.

3.3. Non-stationary Spatial ℓ -POT modeling

Let S and T be a two compact subsets in respectively $\mathbb{R}^d$ (the space) and T (the time). We denote by $C(S \times T)$ the set of continuous real functions on $S \times T$ provided with the uniform norm $\|\cdot\|_\infty$ and $\mathcal{C}_+(S \times T)$ is the restriction of $\mathcal{C}(S \times T)$ to non-negative functions. We designate by $\mathcal{C}_0(S \times T)$ the subset of $\mathcal{C}_+(S \times T)$ containing

only non-negative functions, i.e. $\mathcal{C}_0(S \times T) = \mathcal{C}_+(S \times T) \setminus \{0\}$, thus avoiding the appearance of degenerated limit probability measures. Let $Z = \{Z(s, t)\}_{s \in S, t \in T}$ be a space-time stochastic process defined on $\mathcal{C}_+(S \times T)$. The peak-over threshold modeling has been extended to continuous processes (Ferreira *et al.* (2014), Dombry et Ribatet (2015), de Fondeville et Davison (2018); de Fondeville et Davison (2020)) as part of regular functional variation. A stochastic process Z is regularly varying, if there exist a sequence $a_n : \mathcal{C}_+(S \times T) \rightarrow \mathbb{R}^+$ of continuous functions such that

$$nP(a_n^{-1}(s, t)Z(s, t) \in A) \longrightarrow \Lambda(A), \quad n \longrightarrow +\infty, \quad (3.9)$$

then we write $Z \in RV\{\mathcal{C}_0(S \times T), a_n(s, t), \Lambda\}$, where Λ is a measure on $\mathcal{C}_0(S \times T)$ such that $\Lambda(aA) = a^{-1}\Lambda(A)$ for any $a > 0$ and any borelian $A \in \mathcal{B}(\mathcal{C}_0(S \times T))$.

In the multivariate and spatial framework a threshold exceedances for a random function $Z = \{Z(s, t)\}_{s \in S, t \in T}$ is defined by Dombry et Ribatet (2015) to be an event of the form $\{\ell(Z) > u\}$ for some $u \geq 0$, i.e., an event for which the variable $\ell(Z)$ exceeds a threshold u , where $\ell : \mathcal{C}(S \times T) \rightarrow \mathbb{R}^+$ is a continuous and homogeneous non-negative risk function, i.e., there exists $\alpha > 0$ such that $\ell(ay) = a^\alpha \ell(y)$ when $y \in \mathcal{C}_+(S \times T)$ and $\lambda > 0$. The risk function ℓ determines the type of extreme events of interest. For example, such a function can be the maximum, minimum, average or value at a specific point $(s_0, t_0) \in S \times T$. As in Dombry et Ribatet (2015), we transform the marginal distributions of Z into unit Pareto using the following relationship :

$$Z^*(s, t) = \frac{1}{1 - F_{Z(s, t)}(Z(s, t))}, \text{ for all } s \in S \text{ and } t \in T \quad (3.10)$$

For given a ℓ risk function, the ℓ -excess of Z^* converge weakly in $\mathcal{C}_+(S \times T)$ to a process $W_\ell^*(s, t)$ called standard ℓ -Pareto process and constructed by Ferreira *et al.* (2014) as follows :

$$W_\ell^*(s, t) = YV(s, t), \quad (3.11)$$

where Y is a standard Pareto random variable, i.e. $P(Y > y) = y^{-\gamma_Y}, y > 1$ and independent of $\{V(s, t), s \in S, t \in T\}$ a bounded stochastic process, such as $\ell(V) \equiv 1$. In this construction (3.11), we have necessarily $Y = \ell(W_\ell^*)$ and $V = W_\ell^*/\ell(W_\ell^*)$. A pseudo-polar stochastic representation of standard ℓ -Pareto process W_ℓ^* leads to the following formulation (Dombry et Ribatet (2015), de Fondeville et Davison (2018))

$$W_\ell^* = R \frac{Q}{\ell(Q)}; \quad (3.12)$$

where R is a unit Pareto random variable of index γ_R representing the intensity of process, and Q being a stochastic process denoted the angular component with state space S and taking values in $\mathcal{S}_\ell = \{f \in \mathcal{C}_0(S \times T) : \ell(f) \geq 1, \|f\|_1 = 1\}$ whose distribution is determined by Λ and models the dependence structure of W_ℓ^* . In all the following we have opted that Q belongs to the family of Brown-Resnick processes. But other spectral max-stable models can be chosen(Thibaud

et Opitz (2015), Belzile *et al.* (2022)). In addition to the homogeneity of the risk function, one of the limitations of standard ℓ -Pareto process approach is that it is constructed under the assumption that the process Z belongs to the Fréchet max-domain of attraction. In practice this constraint leads to transforming the marginals of Z into unit Fréchet or standard Pareto. de Fondeville et Davison (2020) generalize this approach to any risk function. They study the conditional ℓ -exceedances limit distribution of the normalized process $(Z - b_n(s, t)) / a_n(s, t)$, where $a_n(s, t)$ is a scaling function and $b_n(s, t)$ is a location function. For given a standard ℓ -Pareto process W_ℓ^* , the generalized ℓ -Pareto process taking values in $\{z \in \mathcal{C}_0(S \times T) : \ell\{(z - b_n)/\ell(a_n)\} \geq 0\}$ is defined by :

$$W_\ell(s, t) = \begin{cases} a_n(s, t) (W_\ell^*(s, t)^{\xi_u(s, t)} - 1) / \xi_u(s, t) + b_n(s, t), & \text{if } \xi_u(s, t) \neq 0 \\ a_n(s, t) \log W_\ell^*(s, t) + b_n(s, t), & \text{if } \xi_u(s, t) = 0 \end{cases} \quad (3.13)$$

For some common max-stable models (e.g. Brown-Resnick) we are able to simulate W_ℓ^* on a number of locations $s \in S$ and derive W_ℓ (de Fondeville et Davison (2018), de Fondeville et Davison (2020)).

3.4. Application to climate data in Burkina Faso

3.4.1. Data set

Burkina Faso has a tropical climate with sudano-sahelian climate, characterized by the alternation between a short rainy season (from June to October) and a long dry season with a cool and dry period (from November to February) and hot weather (from March to May). There are three climatic zones : the sahelian zone in the north with an average annual rainfall of less than 600 mm ; the northern sudanian zone in the center with an average annual rainfall of between 600 and 900 mm and the southern sudanian zone in the south with an average annual rainfall of more than 900 mm. Temperatures and precipitation are highly variable in space and time. Average daily temperatures rarely exceed 35°C and extremes over the last 30 years are found in the North with a minimum value of 21° C and an absolute maximum value of 44.2°C recorded at station of Dori. The analysis of climatic trends shows an increase in the average temperature and a decrease in cumulative rainfall and daily rainfall with a migration of isotherms and isohyets from North to South (PANA, 2007). The meteorological data processed are taken from the climatological database of Burkina Faso which is quite well supplied with observations dating back to 1902 for some meteorological parameters. We have daily measurements of rainfall and temperatures (minimum (Tmin), average (Tmean) and maximum (Tmax)) sampled continuously from 1957 to 2016 on ten synoptic stations (Ouaga, Dori, Ouahigouya, Dedougou, Bogande, Fada N’Gourma, Bobo Dioulasso, Boromo, Po and Gaoua) whose data quality is no longer to be demonstrated. These stations are chosen so as to guarantee not only a good representation of the climatic zones but also a good spatial coverage of the country, with at least three stations per climatic zone, including three for the sahelian zone in the north (Ouahigouya, Dori and Bogande), four for the sub-sahelian zone in the center (Ouaga, Dedougou,

Boromo and Fada) and three for the north-Sudanian zone in the south (Bobo, Po and Gaoua). Figure 3.1 gives the spatial distribution of the different stations. In

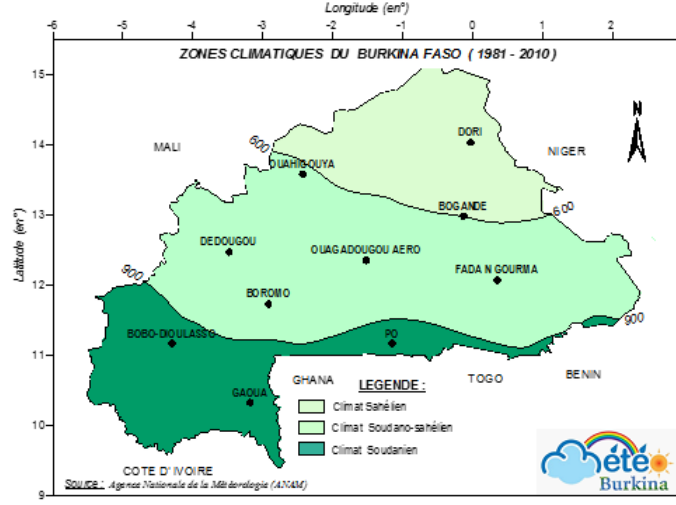


FIGURE 3.1 – Spatial distribution of synoptic sampling stations associated with climatic regimes, source : National Agency Meteorology of Burkina Faso.

order to limit the problem related to the seasonal cycle of rainfall and temperatures, on each station we worked respectively from the sub-series corresponding to the wettest and hottest days. The periods from 15 March to 15 June and from June to October were retained respectively for the temperature and precipitation time series. Thus, a series of 60 by 184 days and 60 by 93 days is extracted respectively from the series of precipitation and temperature observations to constitute the series of daily extremes between 1957 and 2016. On these series of extremes, we apply a runs declustering procedure with daily steps ($r=1$ day) to identify the groups of approximately independent extreme observations within the sample in order to avoid short-term dependencies in the time series for the entire 1957-2016 survey period.

3.4.2. Non-stationary marginal modeling

In this study, we consider that daily temperature and precipitation were generated by a stochastic process $Z = \{Z(s, t)\}_{s \in S, t \in T}$ taking its values in the space of continuous functions $\mathcal{C}_+(S \times T)$, where $S \times T$ denotes a subset of $\mathbb{R}^2 \times \mathbb{R}^+$. First we studied the local tail behavior of the daily precipitation distribution by fitting a non-stationary generalized Pareto distribution for each station $s \in S$, i.e, for any $z \geq 0$

$$P(Z(s, t) - u(s, t) > z \mid Z(s, t) > u(s, t)) = \left(1 + \xi_u(s, t) \frac{z}{\sigma_u(s, t)}\right)_+^{-\frac{1}{\xi_u(s, t)}}.$$

Trends are identified by incorporating non-stationarity into the parameters of the distribution of excesses and the intensity of the occurrences of extreme events for

3.4. APPLICATION TO CLIMATE DATA IN BURKINA FASO

several meteorological variables are described as linear or quadratic functions of time, i.e, for any $s \in S \subset \mathbb{R}^2$

$$\begin{cases} \sigma_u(s, t) = \exp(\sigma_0 + \sigma_1 t + \sigma_2 t^2) \\ \lambda(s, t) = \lambda_0 + \lambda_1 t + \lambda_2 t^2 \\ \xi_u(s, t) = \xi_0 \end{cases}. \quad (3.14)$$

The shape parameter and threshold are kept constant so as not to increase the uncertainty in the estimation of parameters and extreme quantiles. In the pointwise POT approach the results of the best trends adjusting the data for climates variables are obtained using the maximum likelihood ratio test at risk 5%. The degrees of the optimal polynomial trends retained for the different variables are summarized in the table 3.1. The optimal trends identified for minimum and maximum temperatures

Variables	Functions	Ouaga	Dori	Ouahigouya	Dédougou	Bogandé	Fada	Bobo	Boromo	Pô	Gaoua
Tmin	σ	1	0	0	0	0	1	0	1	0	1
	λ	2	2	1	1	2	2	1	1	0	2
Tmax	σ	1	1	0	1	0	0	1	0	0	0
	λ	2	1	0	2	0	2	2	2	0	1
Precip	σ	0	0	0	0	0	0	0	0	0	0
	λ	0	0	1	0	0	0	0	0	1	0

TABLE 3.1 – Degree of the optimal trend of the scale parameter ($\sigma_u(s, t)$) and the intensity of the non-stationary Poisson process ($\lambda(s, t)$).

and precipitation show an increasing frequency of heat waves in recent years (see Figure 3.2 and Figure 3.3).

The non-stationary return level $z_T(s)$ associated with a period of interest return T in years are evaluated from the trends identified using the equation 3.15 by a numerical algorithm, for any $s \in S$

$$\sum_{t=t_0}^{t_0+mn_z} \left[\left(1 + \frac{\xi_u(s, t)}{\sigma_u(s, t)} (z_T(s) - u(s, t)) \right)^{-1/\xi_u(s, t)} \lambda(s, t) \right] = 1, \quad (3.15)$$

where n_z is the number of days in the year and t_0 the initial date of the observed data period (Parey *et al.* (2007, 2010)). The results of the local return levels and their confidence interval calculated for $T = 50$ years are shown in the following table 3.2. Maximum temperatures likely to be observed on average at least once every 50 years reach 45° C in the central sub-Saharan zone. For the same return period, maximum expected temperatures range between 40° C (resp. 45° C) in the Sahelian zone in the north and North Sudanian zone in the south. The 50-year return period levels for daily rainfall vary between 196 mm and 275 mm depending on localities. Extreme rainfall is most expected on average once every 50 years in Fada, Bogande,

3.4. APPLICATION TO CLIMATE DATA IN BURKINA FASO

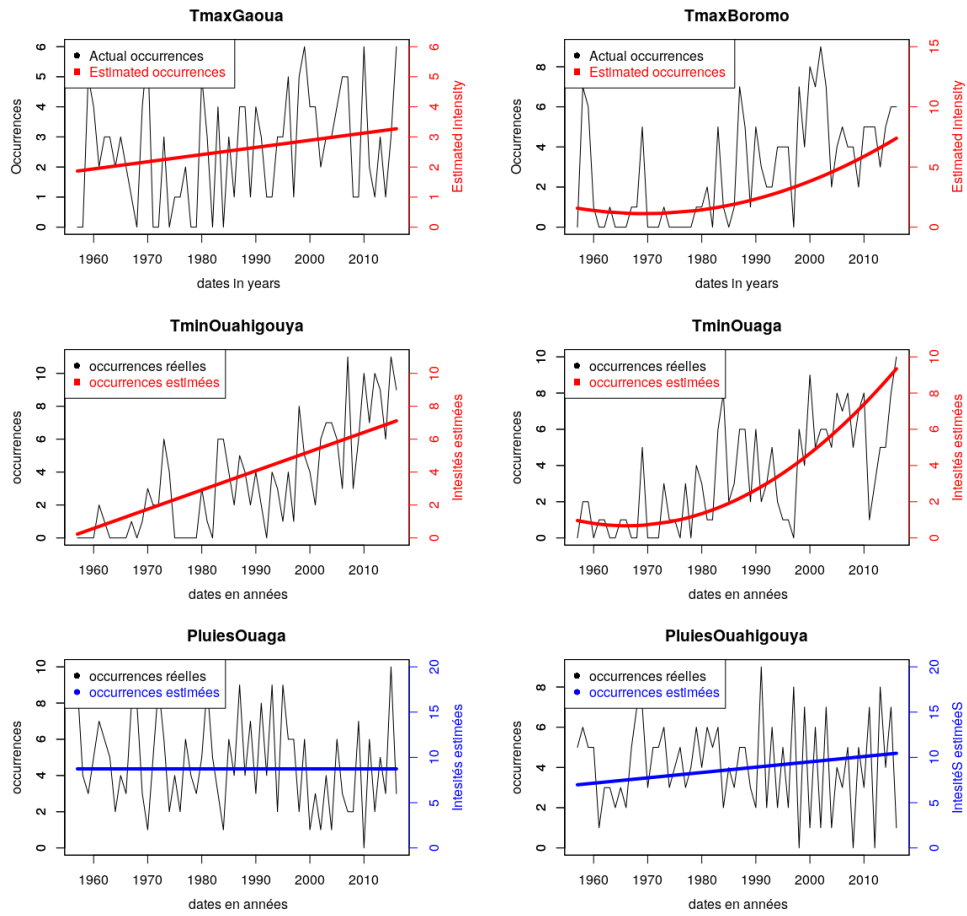


FIGURE 3.2 – Optimal trends of the occurrences of Tmax (line 1), Tmin (line 2) and Precip (line 3) of some reference stations.

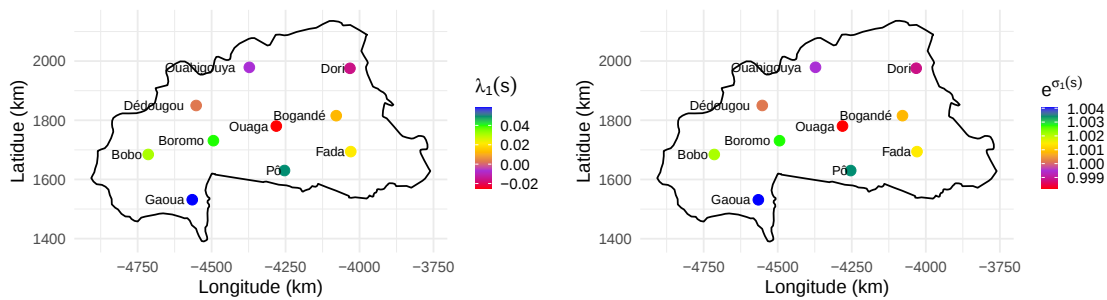


FIGURE 3.3 – Evaluation of significant trend coefficients for the variability and intensity of extreme.

Gaoua, Ouagadougou. In order to produce risk maps in a context of temporal and spatial change, we consider the spatial dependence structure and trends from the marginal distributions discussed in the following section 3.4.2.

3.4. APPLICATION TO CLIMATE DATA IN BURKINA FASO

Stations	Tmin	Tmax	Precip
Ouaga	31.8 [30.8, 33.8]	44.4 [43.1, 48.2]	275.2 [179.3, 477.6]
Dori	35.1 [34.1, 36.9]	46.2 [45.6, 46.7]	158.1 [135.1, 283.5]
Ouahigouya	35.3 [33.5, 37.7]	45.8 [45.0, 46.8]	219.4 [149.6, 326.3]
Dédougou	35.5 [35.2, 50.0]	44.4 [43.5, 45.7]	172.2 [139, 328.6]
Bogandé	33.7 [32.4, 36.5]	39.9 [39.4, 40.6]	220.2 [154.7, 377.3]
Fada	31.0 [30.1, 31.9]	44.7 [44.0, 46.0]	200.9 [151.3, 437.2]
Bobo	30.1 [29.4, 30.5]	41.1 [40.6, 41.4]	196.5 [149.4, 500]
Boromo	33.3 [31.9, 35.4]	44.4 [43.8, 45.8]	224.5 [152.4, 377.2]
Pô	30.8 [30.1, 31.8]	44.4 [43.0, 47.9]	231.9 [156.7, 384.5]
Gaoua	29.8 [28.7, 31.3]	42.1 [41.7, 42.5]	249.6 [164.3, 418.7]

TABLE 3.2 – Pointwise return level associated with the return period $m = 50$ years for the different synoptic stations.

3.4.3. Spatial and temporal dependence

a. Estimation

Spatial and temporal non-stationarity is taken into account in our model by letting the parameters of the distribution of threshold exceedances evolve as a function of time or other covariates, while modeling the dependence structure using generalized ℓ -Pareto processes. The marginal trends are first removed by transforming the data to Pareto standard or Frechet unit using the relationship 3.10. To reduce the complexity of this problem, we have kept the threshold and the shape parameter constant. The variability parameter σ_u is a log-linear function in time of order one or two. These standardized data are then used to estimate the parameters of dependence structure. In order to model the dependence structure of ℓ -Pareto processes, we choose to model its angular component in (3.12) using a Brown-Resnick model for which the formula of intensity function λ_{BR} is available. The d -dimensional intensity function of the Brown-Resnick model is given by [Engelke et al. \(2015\)](#)

$$\lambda_{BR}(z) = \frac{|\Sigma|^{-1/2}}{z_1^2 z_2 \cdots z_d (2\pi)^{(d-1)/2}} \exp\left(-\frac{1}{2} \bar{z}^T \Sigma^{-1} \bar{z}\right), \quad z \in \mathbb{R}_+^d \quad (3.16)$$

where $\bar{z}_i = \log \frac{z_i}{z_1} + \nu(s_i - s_1)$, Σ the covariance matrix and ν is the selected semi-variogram function. We could also choose other families of max-stable spectral processes ([Belzile et al. \(2022\)](#)). In addition, to better take into account the dependence structure of precipitations and temperatures data, we choose a flexible isotropic parametric semi-variogram ν belonging to the bounded power-exponential family models such that $\nu(h) = c_0 [1 - \exp\{-(\|h\|/\tau)^\kappa\}]$, with $\tau > 0$, $c_0 > 0$, $h \in \mathbb{R}^2$, and $\kappa \in]0, 2]$. But we could also choose other semi-variogram models and also take into account the anisotropy according to the studied data. We check the adequacy of our dependence model fitting the data using the extremogram function. Under the Brown-Resnick model, the extremogram function is defined by $\pi(h) = 2 \left(1 - \Phi \left[\{\nu(h)/2\}^{1/2}\right]\right)$, where $h = s_j - s_{j'}$ and Φ denotes the standard normal cumulative distribution function. The different parameters of parametric semi-variogram are estimated using the gradient scoring rule method ([de Fondeville](#)

et Davison (2018)) or censored likelihood (Wadsworth et Tawn (2014)). The estimated variogram parameters for the Tmin and Tmax are reported in table 3.3. The fit of the semi-variogram for Tmin and Tmax is shown in figure 3.4. Figures (3.5 Tmin)

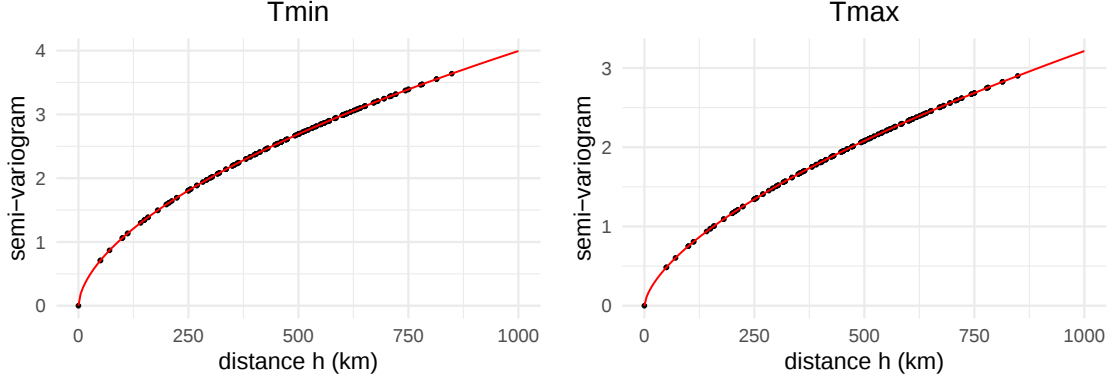


FIGURE 3.4 – Adjustment of the semi-variogram calculated for each pair of sites, for minimum and maximum temperatures.

Variables	κ	τ	c_0
Tmin	0.59(0.53 ; 0.65)	95288.73(95288.66 ; 95288.74)	59.73(59.71 ; 59.79)
Tmax	0.64(0.57 ; 0.66)	158563.14(158563.1 ; 158563.2)	82.83(82.79 ; 82.89)

TABLE 3.3 – Dependence structure parameters estimated by gradient score rule method and their confidence interval obtained by bootstrap

and (3.5 Tmax) show respectively the good quality of the fit of the dependence measures such as the extremogram. Points in gray represent the calculated empirical extremogram and the blue curve is the fitted empirical extremogram, while the red curve represents our fitted dependence model. From these adjustments, we note that there is a strong dependence in the extremes (Tmin and Tmax) of temperatures in Burkina Faso for distances $h \leq 250km$ with an exceedance probability $\pi(h) \geq 0.25$. But, for distance $h > 250km$ the exceedance probability is $0.125 < \pi(h) \leq 0.25$.

b. Simulation and Non-stationary Return Level

We aim at simulating the underlying process from the identified trends and the estimated dependence model at grid points where we have no observations. We discretize the space into a fine grid $\mathcal{G} \subset S$. For each grid point $s_0 \in \mathcal{G}$, we assume that the spatial dependence model is locally stationary in small neighborhoods $\mathcal{N}_{s_0} \subset S$. In what follows, these regional neighborhoods will be determined by a small number D_0 , of nearest stations from the reference site $s_0 \in \mathcal{G}$, thus we will write $\mathcal{N}_{s_0} = \mathcal{N}_{s_0, D_0}$. Obviously, the choice of neighbourhood is important ; the assumed stationary marginal trend model (3.14) could be a poor approximation for large neighbourhoods (i.e., for large D_0), while the simulation of the process could be cumbersome for small neighbourhoods characterized by a small D_0 number of stations. In principle, the choice of D_0 should be such that the spatial dependence

3.4. APPLICATION TO CLIMATE DATA IN BURKINA FASO

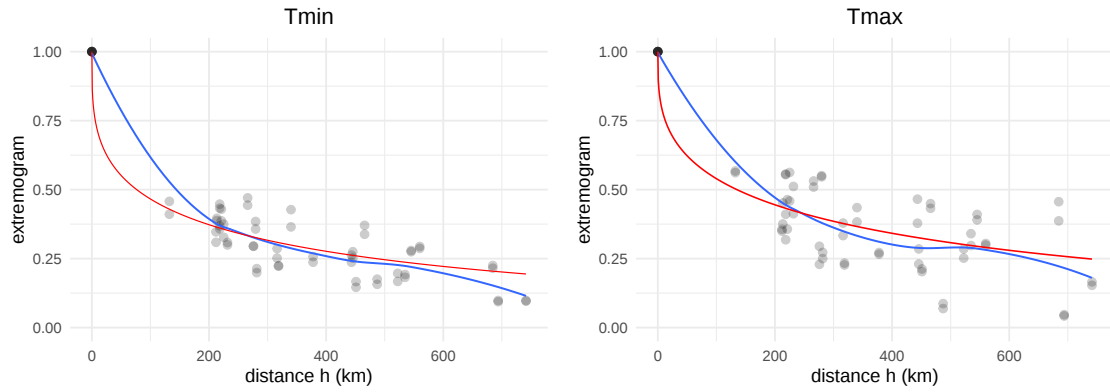


FIGURE 3.5 – Adjustment of the extremogram calculated for each pair of sites, for minimum and maximum temperatures.

structure of trends is approximately stationary within each selected neighborhood $\mathcal{N}_{s_0, D_0}$ around $s_0 \in \mathcal{G}$. We obtain the relatively homogeneous, non-overlapping sub-regions using the k Nearest Neighbors (k -NN) method with $k = 1$ considering our reference stations. This method is extensively used because it is computationally simple and produces accurate results, compared to other more complex clustering methods. The longitude and latitude of the grid points were used as input variables in the k -NN algorithm to form the ten clusters centered on the reference stations. The figure 3.6 shows the spatial structure of obtained clusters. We simulate the

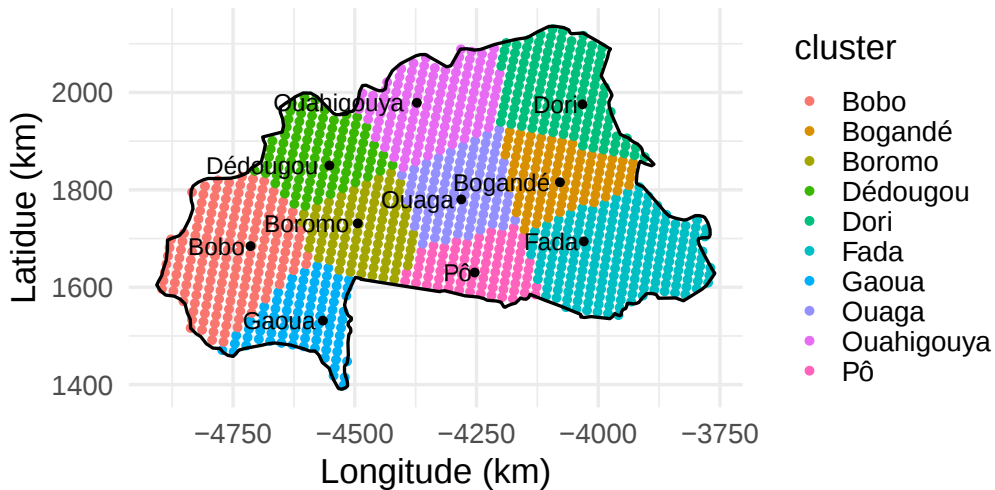


FIGURE 3.6 – Repartition of the grid points in ten clusters.

stationary ℓ -Pareto process W_ℓ^* resulting from modeling the ℓ -excesses over the

entire country using an adapted acceptance-rejection method (Dombry et Ribatet (2015)) or without acceptance-rejection method of mvPot package (de Fondeville et Davison (2018)). For this we exploit the estimated parameters of the variogram and the marginal trends identified on the different clusters obtained (cf. Figure 3.6). The simulation approach consists of first generating a sample of events with size $N = n * LJ * T$ from the ℓ -Pareto process $W_t^{*,i}(s), t = 1, \dots, T$ and $i = 1, \dots, N$ on the unit Pareto margins, where n is the number of repetitions for each path, T is the return period in years and LJ is the number of days in a year. These simulations are then transformed into the scale of the data using the inverse transformation of the relationship 3.10 and trends margins parameters 3.14. This technique consists in computing the simulated sample at the uniform scale using the relation $U_t^i(s) = GPD_{(0,1,1)}(W_t^{*,i}(s))$. Then for each cluster $c_j, j = 1, \dots, 10$, if $U_t^i(s_{c_j}) < 1 - \alpha$, $\tilde{Z}_t^i(s_{c_j}) = \hat{F}_{Z(c)}^{-1}(U_t^i(s))$ otherwise $\tilde{Z}_t^i(s_{c_j}) = GPD_{(u_c, \sigma_{t,c}, \gamma)}^{-1}(U_t^i(s))$ with $\alpha \in]0, 1[$, where $\hat{F}_{Z(c)}$ is the empirical distribution of Z . Since the ℓ -Pareto process is stationary in time, we can simulate the non-stationary process for each time step t . Figure 3.7 gives maps of simulated processes at time $t = 100$ and $t = 200$ years respectively. From these simulations for each period $1, \dots, T$ we are

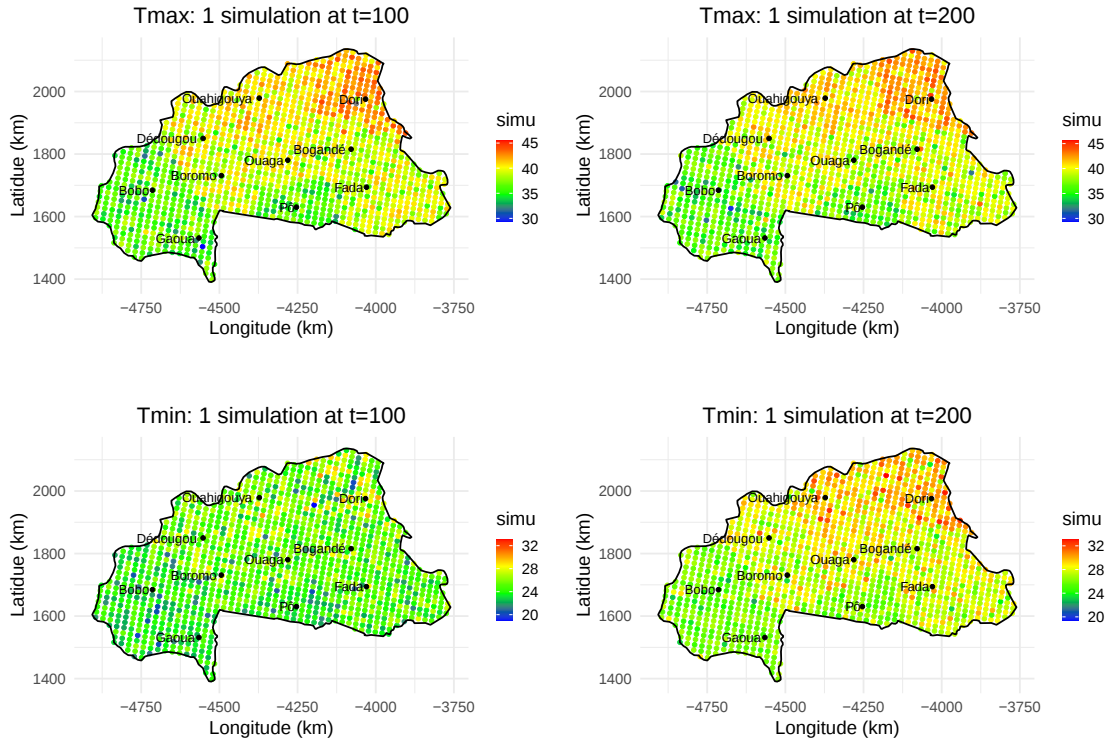


FIGURE 3.7 – Simulate Spatiotemporal processes for Tmax(Top) and Tmin (bottom) by standard ℓ -Pareto processes approach for each step t .

able to deduce a non-stationary return level. Several definitions extend the concept of point-wise return level to non-stationary processes (Parey et al. (2010), Cooley (2013), Salas et Obeysekera (2014)). Among these definitions, we will use the one

3.4. APPLICATION TO CLIMATE DATA IN BURKINA FASO

that is the least expensive in terms of computation time, namely the Expected Number of Exceedances (ENE) method (Parey *et al.* (2010)). The ENE is the value $z_T(s)$, $s \in \mathcal{G}$ for which Expected Number of events exceeding $z_T(s)$ during the period $1, \dots, T$ equals to one, it is the solution of the equation :

$$1 = \sum_{t=1}^T P(Z_t(s) > z_T(s)), \quad s \in \mathcal{G}. \quad (3.17)$$

The probability in the relationship 3.17 can be approximated empirically from simulate process, which leads to the following equation :

$$1 = \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \mathbb{1}_{\{\tilde{Z}_{i,t}(s_j) > z_T(s_j)\}}, \quad j = 1, \dots, nGrid. \quad (3.18)$$

where $\tilde{Z}_{i,t}$ represents the simulation of the process Z for each time step $t = 1, \dots, T$, N is the number of simulation and $nGrid$ is the number of points of grid $\mathcal{G}$ and T is the return period in days. From these simulations we are able to deduce a non-stationnary return levels. Figure 3.8 gives the map of the non-stationnary return levels of Tmax and Tmin for $T = 2$ and $T = 10$ years.

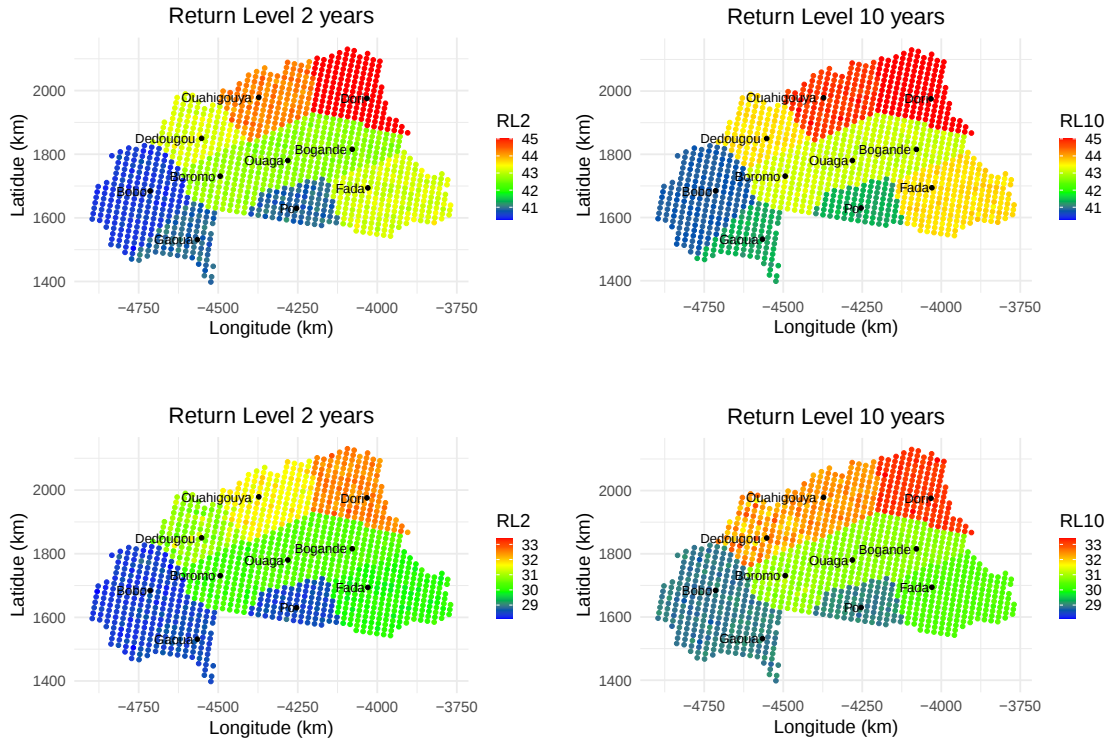


FIGURE 3.8 – non-stationary return levels for Tmax (Top) and Tmin (bottom) by standard ℓ -Pareto processes approach.

3.5. Discussion and perspectives

In this study we have modeled recent temperature and precipitation trends in Burkina Faso using non-stationary ℓ -Pareto processes. Non-stationarity is taken into account in the parameters of the marginal distributions and in the intensity function of the occurrences. The spatial dependence is evaluated by a semi-variogram belonging to the exponential family. The marginal parameters are estimated by the maximum likelihood method while those of dependence structure are obtained by the gradient score rule or the censored likelihood method (de Fondeville et Davison (2018), Wadsworth et Tawn (2014)). Over the period from 1957 to 2016, we note an increasing trend in the intensity of extreme temperature occurrences and the scale function of GPD. We simulate a standard ℓ -Pareto process with unit Pareto margins by acceptance-rejection method (Dombry et Ribatet (2015)) or without acceptance-rejection method (de Fondeville et Davison (2018)) on each cluster using the identified trends and dependence structure. Then we transform them to the scale of data for each future time step via the inverse of equation 3.10 and we deduce the non-stationary return levels. There is a significant increase of non-stationary return levels from South to North as a function of latitude. The Sudano-Sahelian and Sahelian regions are likely to be affected by high minimal and maximal temperatures in the coming years. These extreme maximum temperatures could reach an average of 41 to 45°C for a period $T = 10$ years. In particular we note an increasing trend in non-stationary return levels of minimum temperatures. The non-stationary return levels of minimum temperatures obtained have an increasing trend. As a consequence, we can expect more and more warm nights in the coming years over the region. This could have quite a negative impact on the health of fragile people. On the other hand, it should be noted that we did not detect any significant difference between the non-stationary return levels of the maximum temperatures at 2 years and at 10 years. For the same reasons mentioned in chapter 2, i.e. the tail distribution of maximum temperatures is bounded. Furthermore, it should be added that for $T_{\min}$ and $T_{\max}$, the marginal non-stationary return levels estimated by taking into account the spatial dependence structure are similar or even identical to the return levels estimated at the stations. But the associated uncertainties are not at all the same outside the stations. In other words, our model better quantifies space-time incertitudes and estimates reliably the non-stationary extreme quantiles unlike the spatial kriging implemented in chapter 2.

Note that we have considered in our approach only polynomial trends limited to order one and two to characterize the evolution of the scale σ_β and intensity λ_α function. But polynomial trends of order greater than two can be considered quite reasonable. From a numerical point of view, there are no major difficulties in the estimation of the parameters. However, it should be remembered that exploiting the trends identified beyond the observed data in this case, could especially overestimate the predictions of non-stationary extreme quantiles (Parey *et al.* (2010)). To work around these difficulties, we can also consider other more flexible trend models that can take into account not only the particularity of the study field, but also the nature of the data processed. These flexible models can take different forms : constant, piecewise constant, semi-parametric or non-linear that can integrate climatic covariates

3.5. DISCUSSION AND PERSPECTIVES

that themselves depend on time. These types of trends will undoubtedly provide significant information in the estimation of future non-stationary extreme quantiles. Moreover, we have only integrated stationary dependence structures into our modeling in order to not increase the number of parameters to be estimated, thus reducing the uncertainties in the estimates. The choice fell on the Brown Resnick model to adjust the spatial dependence structure because of its flexibility and its ease of simulation. Moreover, this model makes it possible to achieve asymptotic independence when the variogram is not bounded. But one could choose other spectral models associated with the max-stable process (De Haan *et al.* (1984), Belzile *et al.* (2022)) and integrate forms of non-stationary dependence structures as in Nicolet *et al.* (2016, 2018) in order to better take into account spatial non-stationarities. This is a subject that will be developed in our next work.

Modélisation hétéroscédastique et application à la recherche de surfaces de retour

4.1. Introduction

La modélisation spatio-temporelle des événements climatiques extrêmes est une tâche essentielle pour évaluer les risques dans de nombreuses disciplines appliquées. Dans le domaine de l'environnement et du climat une façon d'évaluer le risque est la période ou le niveau de retour. Dans un contexte stationnaire et pour une variable aléatoire Z , le niveau de retour z_τ associé à la période de retour d'intérêt τ est défini de manière unique par :

$$P(Z > z_\tau) = \frac{1}{\tau} \Leftrightarrow F_Z(z_\tau) = 1 - \frac{1}{\tau}, \quad (4.1)$$

où F_Z est la fonction de répartition de Z . Contrairement au cadre univarié, la notion du niveau de retour spatial ou courbe de retour dans le cadre multivarié n'est pas définie de manière unique à cause de la notion de relation d'ordre dans $\mathbb{R}^m$. La probabilité de dépassement peut se définir de plusieurs façons : le cas "OU" au moins une des variables dépasse son seuil, le cas "ET" toutes les variables dépassent leur seuil et plus généralement l'image par une fonctionnelle dépasse un certain seuil. Ainsi pour une fonction aléatoire $Z = \{Z(s), s \in S\}$ et en supposant que toutes les composantes excèdent un seuil, la courbe de retour ou isoline associée à la période de retour τ est définie par :

$$\left\{ z_\tau = (z(s_1), \dots, z(s_m)) \in \mathbb{R}^m : F_Z(z_\tau) = 1 - \frac{1}{\tau} \right\}, \quad (4.2)$$

où F_Z représente maintenant la fonction de répartition conjointe ([Salvadori et al. \(2011\)](#), [Jiang et al. \(2019\)](#), [Richards et Wadsworth \(2021\)](#)). Le véritable problème est que d'une part nous n'avons pas accès le plus souvent à la loi fini-dimensionnelle de Z . D'autre part même si elle existe son évaluation s'avère très difficile en grande dimension d'un point de vue numérique.

Cependant lorsque l'on traite des données climatiques, notamment dans le cadre du changement climatique, un problème majeur est la non-stationnarité dans l'espace et dans le temps. Une façon habituelle de traiter les non-stationnarités dans l'espace est de transformer les marges via une fonction des coordonnées et de travailler ponctuellement dans le cadre stationnaire ([Eastoe et Tawn \(2009\)](#), [Mentacchi et al. \(2016\)](#)). De même, la non-stationnarité dans le temps peut être traitée en modélisant les paramètres marginaux en fonction du temps, par exemple par des

polynômes ou des splines (Davison et Smith (1990), Parey et al. (2010), Hao et al. (2018)). Travailler de manière ponctuelle ne nous permet pas de prendre en compte la structure spatiale des données et des modèles spatiaux sont nécessaires pour décrire la dépendance spatiale. Pour traiter la nature spatiale des données climatiques en se concentrant sur les valeurs extrêmes, nous utilisons les modèles ℓ -Pareto introduits par Dombry et Ribatet (2015) qui sont une extension de la distribution de Pareto généralisée pour les dépassements spatiaux, donnée par une fonction de risque ℓ . Les non-stationnarités dans l'espace et le temps sont traitées comme dans (Einmahl et al. (2016), Mefleh et al. (2020)) via une fonction de scédasticité introduite dans le cadre des "extrêmes hétéroscédastiques" qui établit une relation entre la distribution de queue du processus spatio-temporel non-stationnaire et un processus spatial latent.

Dans ce travail, nous proposons de définir la notion de surface de retour de période τ , comme une fonction paramétrique f pour tout $s \in S$. Soit $X = \{X_t(s), s \in S, t \in T\}$ un processus spatio-temporel non-stationnaire. La fonction f vérifie sur $S \subset \mathbb{R}^2$

$$f : S \longrightarrow \mathbb{R}, \operatorname{argmin}_{f \in \mathcal{C}_+(S)} \left(1 - \sum_{t=1}^{\tau} P\{X_t(s) > f(s)\} \right)^2, \quad (4.3)$$

où $\mathcal{C}_+(S)$ est l'ensemble des fonctions positive et continues sur S . La probabilité est évaluée empiriquement à l'aide d'une approximation par un processus de ℓ -Pareto qu'il est possible de simuler. La méthodologie proposée est mise en oeuvre sur des données réelles de températures maximales quotidiennes du Burkina Faso.

4.2. Modèles hétéroscédastiques

4.2.1. Hétéroscédasticité de queue et pseudo observations

Soit $X = \{X_t(s), s \in S, t \in T\}$ un processus spatio-temporel défini sur une région $S \times T \subset \mathbb{R}^d \times \mathbb{R}^+$. Soient $F_{t,s}$ les marginales univariées de point terminal x_F et $Z = \{Z(s), s \in S\}$ un processus spatial latent non-observé de fonction de répartition marginale $F_{Z(s)}$ tels qu'on ait la condition de queue proportionnelle suivante : (de Haan et al. (2015), Ferreira et al. (2017), Cabral et al. (2020))

$$\lim_{x \rightarrow x_F} \frac{P(X_t(s) > x)}{P(Z(s) > x)} = c_\theta(t/n, s), \text{ avec } \int_0^1 c_\theta(u, s_j) du = 1, \quad u \in [0, 1], \quad (4.4)$$

où c_θ est appelée fonction de tendance ou fonction de scédasticité de paramètres θ . Les dates d'observations sont ramenées à $[0, 1]$. Elle décrit la fréquence des événements extrêmes conjointement dans l'espace et dans le temps. Les distributions marginales du processus latent Z sont supposées appartenir au domaine d'attraction d'une distribution des valeurs extrêmes généralisées d'indice de queue $\gamma \in \mathbb{R}$, i.e. $F_{Z(s)} \in \mathcal{D}(G_\gamma)$. Autrement dit il existe des constantes $a_n > 0$ et $b_n \in \mathbb{R}$ de sorte que les maxima convenablement renormalisés d'un échantillon de $F_{Z(s)}$ converge vers G_γ tel que :

$$\lim_{n \rightarrow \infty} F_{Z(s)}^n(b_n + xa_n) = G_\gamma(x) \quad (4.5)$$

4.2. MODÈLES HÉTÉROSCÉDASTIQUES

La condition 4.5 est en réalité une condition sur la queue de $1 - F_{Z(s)}$. En prenant le logarithme de cette relation et en utilisant un développement limité d'ordre un et en passant en continu, nous obtenons :

$$\lim_{y \rightarrow \infty} y \{1 - F_{Z(s)}(b_Z(y) + xa_Z(y))\} = (1 + \gamma x)^{-1/\gamma}, \quad x > 0, \quad (1 + \gamma x) > 0 \quad (4.6)$$

où $a_n \rightarrow a_Z(y)$, $b_n \rightarrow b_Z(y)$. De même l'hypothèse $F_{Z(s)} \in \mathcal{D}(G_\gamma)$ dans la relation 4.4, reste valable pour tout $F_{t,s}$ avec la même distribution limite que dans la relation 4.6, i.e. :

$$\lim_{y \rightarrow \infty} y \{1 - F_{t,s}(b_{t,s}(y) + xa_{t,s}(y))\} = (1 + \gamma x)^{-1/\gamma}, \quad x > 0, \quad (1 + \gamma x) > 0 \quad (4.7)$$

où $a_{t,s} > 0$ et $b_{t,s} \in \mathbb{R}$ sont des fonctions continues sur $\mathcal{C}(S \times T)$ et choisies avec soin. En passant maintenant aux fonctions quantiles (de Haan *et al.* (2015)), la relation 4.4 est équivalente à l'équation 4.8 suivante :

$$\lim_{p \rightarrow 1} \frac{F_{t,s}^\leftarrow(p) - F_{Z(s)}^\leftarrow(p)}{a_Z\left(\frac{1}{1-p}\right)} = \frac{c_\theta(t/n, s)^\gamma - 1}{\gamma} \quad (4.8)$$

En changeant de notation et en posant $b_{t,s}(y) = F_{t,s}^\leftarrow(1 - 1/y) = (1/(1 - F_{t,s}))^\leftarrow(y)$ et $b_Z(y) = F_{Z(s)}^\leftarrow(1 - 1/y) = (1/(1 - F_{Z(s)}))^\leftarrow(y)$, $y > 1$, l'équation 4.6 est équivalente à (de Haan *et al.* (2015), Cabral *et al.* (2020)) :

$$\begin{cases} \lim_{y \rightarrow \infty} \frac{b_{t,s}(yc_\theta(t/n, s)) - b_Z(y)}{a_Z(y)} = \frac{c_\theta(t/n, s)^\gamma - 1}{\gamma}, \\ \lim_{y \rightarrow \infty} \frac{a_{t,s}(yc_\theta(t/n, s))}{a_Z(y)} = c_\theta(t/n, s)^\gamma \end{cases}, \quad y > 0 \quad (4.9)$$

Les fonctions quantiles de queue $b_{t,s}$ et $a_{t,s}$ du processus non-stationnaire sont donc reliées aux fonctions quantiles de queue b_Z et a_Z du processus spatial latent Z . Nous pouvons déduire une relation entre les paramètres du processus X_t et Z . En injectant la relation 4.9 dans 4.7 nous obtenons :

$$\lim_{y \rightarrow \infty} y [1 - F_{t,s}\{b_{t,s}(yc_\theta(t/n, s)) + xa_{t,s}(yc_\theta(t/n, s))\}] = (1 + \gamma x)^{-1/\gamma} \quad (4.10)$$

Par suite nous avons,

$$\lim_{y \rightarrow \infty} y \left[1 - F_{t,s} \left\{ b_Z(y) + a_Z(y) \frac{c_\theta(t/n, s)^\gamma - 1}{\gamma} (1 + \mathcal{O}(1)) + xa_Z(y) c_\theta(t/n, s)^\gamma (1 + \mathcal{O}(1)) \right\} \right] = (1 + \gamma x)^{-1/\gamma}$$

En factorisant et en calculant les limites à l'intérieur des accolades, nous avons

$$\lim_{y \rightarrow \infty} y \left[1 - F_{t,s} \left\{ b_Z(y) + a_Z(y) \left[\frac{c_\theta(t/n, s)^\gamma - 1}{\gamma} + xc_\theta(t/n, s)^\gamma \right] \right\} \right] = (1 + \gamma x)^{-1/\gamma} \quad (4.11)$$

Finalement lorsque $y \rightarrow \infty$ nous déduisons les approximations suivantes :

$$\begin{cases} b_{t,s} \approx b_{Z(s)} + a_{Z(s)} \frac{c_\theta(t/n, s)^\gamma - 1}{\gamma} \\ a_{t,s} \approx a_{Z(s)} c_\theta(t/n, s)^\gamma \end{cases} \quad (4.12)$$

Les pseudo-observations extrêmes du processus spatial latent $\{Z_t(s_j)\}_{t,j}$ sont alors déduites à partir de celles du processus non-stationnaire $\{X_t(s_j)\}_{t,j}$ de sorte que

$$Z_t(s_j) = \{c_\theta(t/n, s_j)\}^{-\gamma} \left[X_t(s_j) - \frac{\{c_\theta(t/n, s_j)\}^\gamma - 1}{\gamma} (a_{Z(s_j)} - \gamma b_{Z(s_j)}) \right], \quad (4.13)$$

$$j = 1, \dots, m; \quad t = 1, \dots, n$$

où $\gamma, a_{Z(s_j)}, b_{Z(s_j)}$ et θ sont des paramètres à estimer.

4.2.2. Estimation de la dépendance spatiale et marginale

a. Fonction de tendance

Le paramètre θ de la fonction de tendance c_θ peut être estimé en utilisant la méthode du maximum de vraisemblance en prenant par exemple des modèles paramétriques en raison de leur flexibilité dans la détection de tendance. Ils sont également flexibles dans l'extrapolation de la tendance au-delà des données observées. On suppose que les dates des occurrences des dépassements de seuil $x(s) \in \mathbb{R}, s \in S$ sont asymptotiquement indépendantes et identiquement distribuées de fonction densité c_θ . Comme dans [Mefleh et al. \(2020\)](#) nous nous intéresserons aux modèles monotones de tendance log-linéaires et linéaires simples suivant :

$$\begin{cases} c_\theta^1(\frac{t}{n}, s) = \frac{\theta(s)}{\exp\{\theta(s)\} - 1} \exp\{\theta(s)\frac{t}{n}\}, \quad \theta(s) \in \mathbb{R}; \quad \frac{t}{n} \in [0, 1] \\ c_\theta^2(\frac{t}{n}, s) = \theta(s) \left\{ 2\frac{t}{n} - 1 \right\} + 1, \quad \theta(s) \in]-1; 1[; \quad \frac{t}{n} \in [0, 1], \end{cases} \quad (4.14)$$

où $\theta(s) > 0$ signifie que les événements extrêmes sont devenus plus fréquents tandis que $\theta(s) < 0$ indique une tendance à la baisse des fréquences des événements extrêmes. La figure 4.1 illustre la fonction de tendance log-linéaire et linéaire pour différentes valeurs de θ .

Une autre alternative pour estimer c_θ consiste à l'approcher par un estimateur à densité de type noyaux appliqué aux dépassements. Nous prenons un seuil dans l'espace et dans le temps $\hat{b}_{t,s} = q_{1-\alpha}(X_t(s)), 0 < \alpha < 1$.

$$\hat{c}_\theta(u, s_j) = \frac{1}{kh} \sum_{t=1}^n \mathbb{1}\{X_t(s_j) > \hat{b}_{t,s}\} G\left(\frac{u - t/n}{h}\right), \quad j = 1, \dots, m, \quad u \in [0, 1]. \quad (4.15)$$

où k est le nombre de dépassements et G est une fonction continue de noyau symétrique sur $[-1, 1]$ telle que $\int_{-1}^1 G(s) ds = 1$, $G(s) = 0$ pour $|s| > 1$. La fonction intégrée $C_j(u)$ est définie par :

$$C_j(u) = \frac{1}{m} \int_0^u c_\theta(t, s_j) dt, \quad u \in [0, 1], \quad (4.16)$$

4.2. MODÈLES HÉTÉROSCÉDASTIQUES

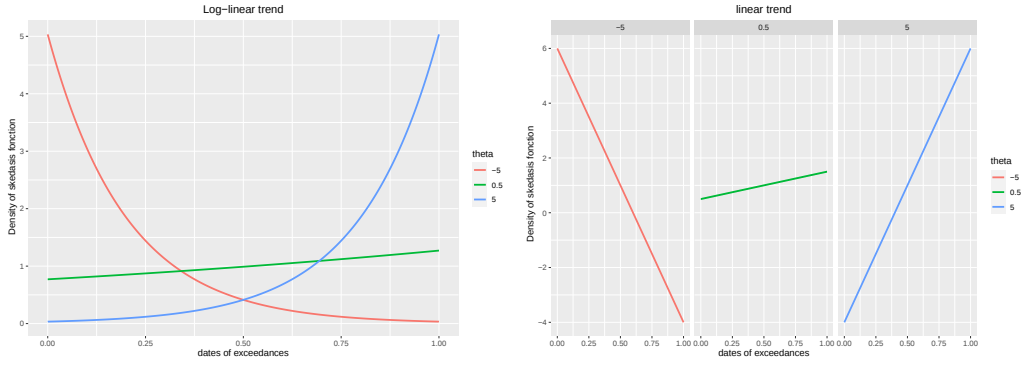


FIGURE 4.1 – Allure de la fonction de tendance Log-linéaire à droite et celle de la fonction de tendance Linéaire simple à gauche pour différentes valeur de θ .

peut être estimée par :

$$\hat{C}_j(u) = \frac{1}{k} \sum_{t=1}^{n_u} \mathbb{1}\{X_t(s_j) > \hat{b}_{t,s}\}, j = 1, \dots, m, \quad u \in [0, 1] \left(\sum_{j=1}^m \hat{C}_j(1) = 1 \right) \quad (4.17)$$

Elle est utilisée dans le test statistique de détection de tendance $H_0 : \theta = 0$ contre $H_1 : \theta \neq 0$. Cependant l'approche non paramétrique est un peu limitée car elle ne permet pas une extrapolation des tendances identifiées au-delà des données observées et mérite d'être réadaptée pour prendre en compte cette limite dans la détection de tendance.

b. Paramètres marginaux

Pour estimer les paramètres a_Z , b_Z et γ plusieurs options se présentent. Dans [Mefleh et al. \(2020\)](#), a_Z et b_Z sont considérés tous constants dans l'espace, b_Z est un quantile d'ordre $1 - \alpha$ de l'ensemble des distributions $\{X_{t,s}\}$, a_Z et γ sont estimés par la méthode de Hill généralisée ([Amenah et al. \(2021\)](#)) :

$$\hat{\gamma} = M_n^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)^{-1} \quad (4.18)$$

$$\hat{a}_Z = \hat{b}_Z M_n^{(1)} [1 - (\hat{\gamma} - M_n^{(1)})] = \frac{1}{2} \hat{b}_Z M_n^{(1)} \left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)^{-1} \quad (4.19)$$

où

$$M_N^\ell = \frac{1}{k} \sum_{t=1}^n \sum_{j=1}^m \left\{ \log X_t(s_j) - \log \hat{b}_Z \right\}^\ell \mathbb{1}\{X_t(s_j) > \hat{b}_Z\}, \ell = 1, 2. \quad (4.20)$$

Une autre option est de tirer profit des relations 4.12 et de les injecter dans l'expression de la vraisemblance indépendantes de $X_t(s_j)$.

$$\mathcal{L}_{indep}(\gamma, a_{t,s_j}) = \sum_{t=1}^n \sum_{j=1}^m \mathbb{1}_{\{x_t(s_j) \geq b_{t,s_j}\}} \log \left[\frac{1}{a_{t,s_j}} \left\{ 1 + \gamma \frac{x_t(s_j) - b_{t,s_j}}{a_{t,s_j}} \right\}^{-1/\gamma-1} \right] \quad (4.21)$$

Dans ce cas $a_{Z(s)}$ et $b_{Z(s)}$ varient dans l'espace avec s . Ainsi à partir des estimateurs de $a_{Z(s)}$, $b_{Z(s)}$ et γ , nous pouvons déduire pour chaque pas de temps t les paramètres des distributions $F_{t,s}$.

c. Excès de Z , ℓ -Pareto

Après la suppression de la non-stationnarité présente dans X_t la modélisation est maintenant concentrée sur l'évaluation de la structure de dépendance spatiale restante dans Z . En supposant que Z est à variation régulière, soit ℓ une fonction de risque et $u > 0$ un seuil, alors les ℓ -dépassements de Z peuvent être approchés par un processus limite stationnaire ℓ -Pareto W_ℓ i.e. :

$$P(u^{-1}Z \in \cdot \mid \ell(Z) > u) \longrightarrow P(W_\ell \in \cdot). \quad (4.22)$$

Les marges de W_ℓ sont des distributions *GPD*. Le processus W_ℓ peut être écrit $W_\ell(s) = RQ(s)$ où R est un processus de Pareto standard d'indice de queue $\gamma \in \mathbb{R}$ et $Q = \{Q(s), s \in S\}$ un processus stochastique continu à trajectoires dans $\mathcal{S} = \{f \in \mathcal{C}_+(S) : \ell(f) \geq 1, \|f\|_{ang} = 1\}$ et de mesure angulaire ou mesure spectrale σ_ℓ (voir chapitre 1, section 1.3.3). Le choix du processus Q avec une distribution qui se situe dans une famille paramétrique dont les paramètres peuvent être déduits conduit à une estimation de la structure de dépendance spatiale. Un choix standard est un processus log-gaussien avec une fonction de variogramme paramétrique, pour lequel le processus max-stable correspondant est le processus bien connu de Brown-Resnick ([de Fondeville et Davison \(2018\)](#)). D'autres modèles de processus max-stables sont également des candidats potentiels ([Belzile et al. \(2022\)](#)). Des algorithmes d'estimation et de simulation du processus ℓ -Pareto sont développés et des logiciels sont disponibles. Nous estimons les paramètres de la structure de dépendance en utilisant la méthode du score de gradient ([de Fondeville et Davison \(2018\)](#)) ou celle de la vraisemblance censurée ([Wadsworth et Tawn \(2014\)](#)).

4.2.3. Surface de retour

Soit $\{X_t(s) : s \in S, t \in T\}$ un processus spatio-temporel. Plusieurs définitions sont proposées pour étendre la notion du niveau de retour au contexte non stationnaire (voir chapitre 2, section 2.3.3). Par exemple le nombre moyen de dépassements (ou Expected Number of Exceedances (ENE)) est la valeur pour laquelle le nombre moyen attendu d'événements dépassant le niveau $x_\tau(s)$, $s \in S$ pendant la période $1, \dots, \tau$ est égale à un ([Parey et al. \(2010\)](#)). Le niveau $x_\tau(s)$ est alors la solution de l'équation :

$$1 = \sum_{t=1}^{\tau} P(X_t(s) > x_\tau(s)) = \sum_{t=1}^{\tau} \{1 - F_{t,s}(x_\tau(s))\} \quad (4.23)$$

La définition 4.23 permet d'estimer seulement le niveau de retour ponctuellement et ne prend pas en compte la structure de dépendance spatiale. Nous proposons d'étendre cette définition aux processus stochastiques spatio-temporels.

4.2. MODÈLES HÉTÉROSCÉDASTIQUES

Définition 4.2.1. Soit $\tau \in \mathbb{N}$ fixé. On appelle niveau de retour spatial de période τ une fonction $x_\tau : S \rightarrow \mathbb{R}$ qui vérifie

$$x_\tau = \operatorname{argmin}_{f \in \mathcal{C}_+(S)} \left(1 - \sum_{t=1}^{\tau} \mathbb{P}(X_t > f) \right)^2, \quad (4.24)$$

où f est une fonction continue de $\mathcal{C}_+(S)$.

Remarquons que,

$$\begin{aligned} \mathbb{P}\{X_t > f\} &= \mathbb{P}\left\{Z_t > c_\theta^{-\gamma}(t/n, \cdot) \left[f - \frac{c_\theta^\gamma(t/n, \cdot) - 1}{\gamma}(a_Z - \gamma b_Z) \right]\right\} \\ &= \mathbb{P}\left\{\frac{Z_t - b_Z}{\ell(a_Z)} > \left(c_\theta^{-\gamma}(t/n, \cdot) \left[f - \frac{c_\theta^\gamma(t/n, \cdot) - 1}{\gamma}(a_Z - \gamma b_Z) \right] - b_Z \right) / \ell(a_Z) \mid \ell\left(\frac{Z - b_Z}{\ell(a_Z)}\right) > u\right\} \\ &\quad \times \mathbb{P}\left\{\ell\left(\frac{Z - b_Z}{\ell(a_Z)}\right) > u\right\} \\ &\approx \mathbb{P}\left\{W_\ell > \left(c_\theta^{-\gamma}(t/n, \cdot) \left[f - \frac{c_\theta^\gamma(t/n, \cdot) - 1}{\gamma}(a_Z - \gamma b_Z) \right] - b_Z \right) / \ell(a_Z)\right\} \\ &\quad \times \mathbb{P}\left\{\ell\left(\frac{Z - b_Z}{\ell(a_Z)}\right) > u\right\}, \end{aligned}$$

où W_ℓ est un processus ℓ -Pareto.

Posons $L_t(f) := \left(c_\theta^{-\gamma}(t/n, \cdot) \left[f - \frac{c_\theta^\gamma(t/n, \cdot) - 1}{\gamma}(a_Z - \gamma b_Z) \right] - b_Z \right) / \ell(a_Z)$. On peut alors remplacer ces dernières probabilités par leurs expressions empiriques :

$$\hat{\mathbb{P}}(X_t > f) = \frac{1}{k} \sum_{i=1}^k \mathbb{1}\{W_\ell^i > L_t(f)\} \quad \hat{\mathbb{P}}\left(\ell\left(\frac{Z - b_Z}{\ell(a_Z)}\right) > u\right),$$

où $W_\ell^1, \dots, W_\ell^k$ sont des simulations *i.i.d* de W_ℓ . De cette manière, nous obtenons un estimateur empirique du niveau de retour spatial :

$$\hat{x}_T = \operatorname{argmin}_{f \in \mathcal{C}(S)} \left(1 - \hat{\mathbb{P}}\left(\ell\left(\frac{Z - b_Z}{\ell(a_Z)}\right) > u\right) \frac{1}{k} \sum_{t=1}^{\tau} \sum_{i=1}^k \mathbb{1}\{W_\ell^i > L_t(f)\} \right)^2, \quad (4.25)$$

Si f appartient à une famille paramétrique (par exemple un polynôme de degré fixé des coordonnées) on est ramené à un problème d'optimisation : trouver les points critiques de la fonction :

$$Q_n(\tilde{\beta}) = \left(1 - \hat{\mathbb{P}}\left(\ell\left(\frac{Z - b_Z}{\ell(a_Z)}\right) > u\right) \frac{1}{k} \sum_{t=1}^{\tau} \sum_{i=1}^k \mathbb{1}\{W_\ell^i > L_t(f_{\tilde{\beta}})\} \right)^2.$$

La fonction Q_n n'étant pas dérivable sur les points $\tilde{\theta}$ tels que $L_t(f_{\tilde{\beta}}) = W_\ell^i$, on utilisera une version dérivable de Q_n en remplaçant la fonction indicatrice $\mathbb{1}_{\{X > x\}}$ par son approximation dérivable : $\phi_X(x) = \frac{1}{1 + \exp(-\alpha(X - x))}$, $\alpha > 0$

4.3. Application aux données de température

Nous illustrons maintenant notre méthodologie sur des données réelles de températures maximales quotidiennes du Burkina Faso à la recherche de surface de retour. Dans toute la suite on se limitera dans notre application à des surfaces polynomiales simples d'ordre inférieur ou égale à deux telles que $f_\beta(x, y) = \beta_0 + \beta_1x + \beta_2y + \beta_3xy + \beta_4x^2 + \beta_5y^2$, où x et y désignent respectivement la longitude et la latitude pour tout point $s \in S$. On examinera les cas où f_β est une fonction constante, linéaire et quadratique. Mais nous pouvons aussi envisager une fonction constante par morceaux pour faire référence aux trois régions climatiques de notre zone d'étude.

$$f_\beta(x, y) = \begin{cases} \beta_{\text{sahel}}, & \text{si } y \geq 1021 \\ \beta_{\text{centre}}, & \text{si } 334 \leq y \leq 1020 \\ \beta_{\text{sud}}, & \text{si } 0 \leq y \leq 333 \end{cases} \quad (4.26)$$

Aussi sur ces données de températures nous avons opté de modéliser le processus angulaire Q issu de la décomposition pseudo-polaire de W_ℓ par un modèle Brown-Resnick de semi-variogramme appartenant à la famille puissance-exponentielle. Nous commençons d'abord par estimer le paramètre θ de la fonction de scédasticité en utilisant la méthode du maximum de vraisemblance. Ensuite les paramètres marginaux a_Z , et γ , sont estimés en tirant profit des relations 4.12 et en les injectant dans la log-vraisemblance indépendante des $X_t(s_j)$, $j = 1, \dots, 10$, tandis que b_Z est fixé à un quantile d'ordre $1 - \alpha$ de tous les ℓ -excès. L'estimation du paramètre d'échelle a_Z montre une forte variabilité spatiale des températures maximales dans les climats sahélien et soudanien. Elles varient relativement moins dans le climat soudano-sahélien. Les paramètres $b_{t,s}$ et $a_{t,s}$ peuvent être déduits pour chaque pas de temps t . La figure 4.2 donne les paramètres marginaux a_Z et θ et une extrapolation des paramètres $b_{t,s}$ et $a_{t,s}$ pour une date future $t = 10$ ans. Les niveaux $b_{t,s}$ de dépassement évoluent de façon décroissante du nord vers le Sud. Nous déduisons les pseudo-observations de Z en utilisant les paramètres marginaux estimés et la relation 4.13. Les paramètres de la structure de dépendance sont estimés en prenant une fonction de semi-variogramme puissance-exponentielle en utilisant la méthode du score de gradient. La figure 4.3 illustre la qualité de l'ajustement de la fonction d'extrémogramme utilisé ici pour évaluer la structure de dépendance spatiale. La courbe bleue représente l'ajustement de l'extrémogramme empirique tandis que celle en rouge est le modèle ajusté. Enfin nous simulons un échantillon du processus ℓ -Pareto obtenu sur les ℓ -excès de Z de taille $N = \tau * LJ * nRep$ à l'échelle de Pareto unité sur une grille fine, où τ est la période de retour en année, $LJ = 93$ le nombre de jour par ans et $nRep = 100$ le nombre de répétitions sur chaque point $s \in S$. Les paramètres $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ de la fonction de retour f_β sont alors estimés en optimisant la relation 4.25 à l'aide des simulations précédentes pour chaque période de retour τ fixée. Nous déduisons les surfaces de retour non-stationnaires du processus $X_{t,s}$ à l'aide des paramètres de la fonction f_β . La figure 4.4 présente les surfaces de retour non-stationnaires calculées pour différentes périodes de retour $\tau = 2, 10$ ans. La première colonne de la figure 4.4 correspond aux surfaces de retour linéaires tandis que la seconde colonne donne les cartes correspondant aux surfaces

4.3. APPLICATION AUX DONNÉES DE TEMPÉRATURE

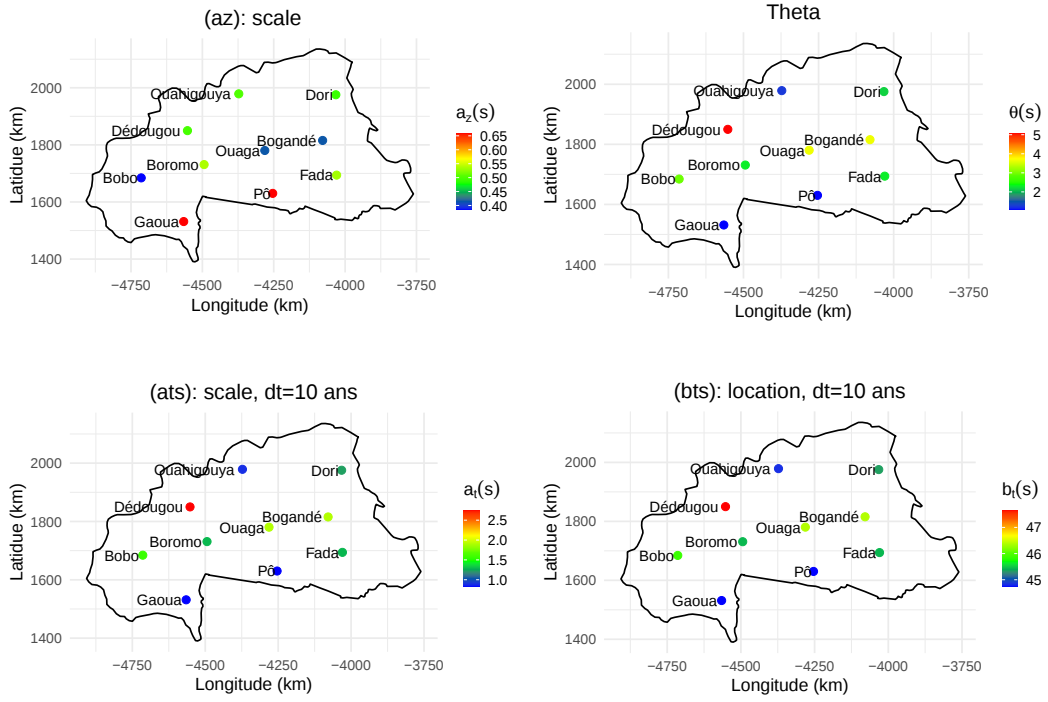


FIGURE 4.2 – Paramètres marginaux a_Z et θ avec une extrapolation des paramètres $a_{t,s}$ et $b_{t,s}$ à une date futur $t = 10$ ans.

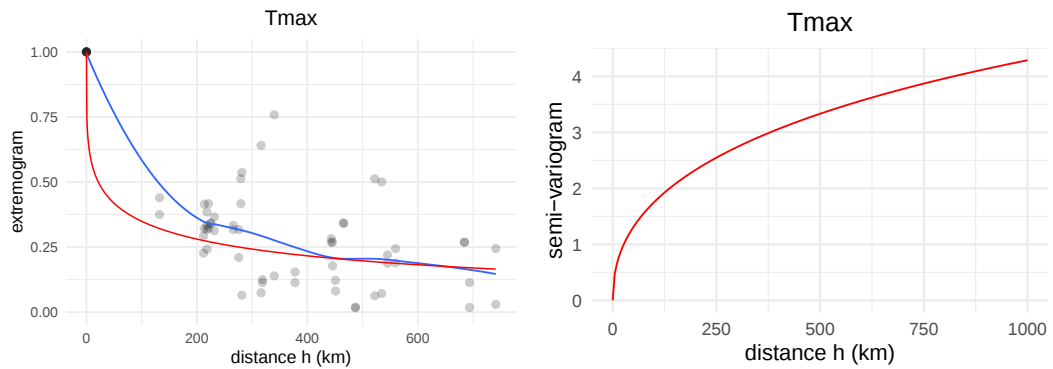


FIGURE 4.3 – Ajustement de la fonction d'extrémogramme à gauche et la fonction semi-variogramme à droite appliquée aux pseudo observations.

de retour quadratique.

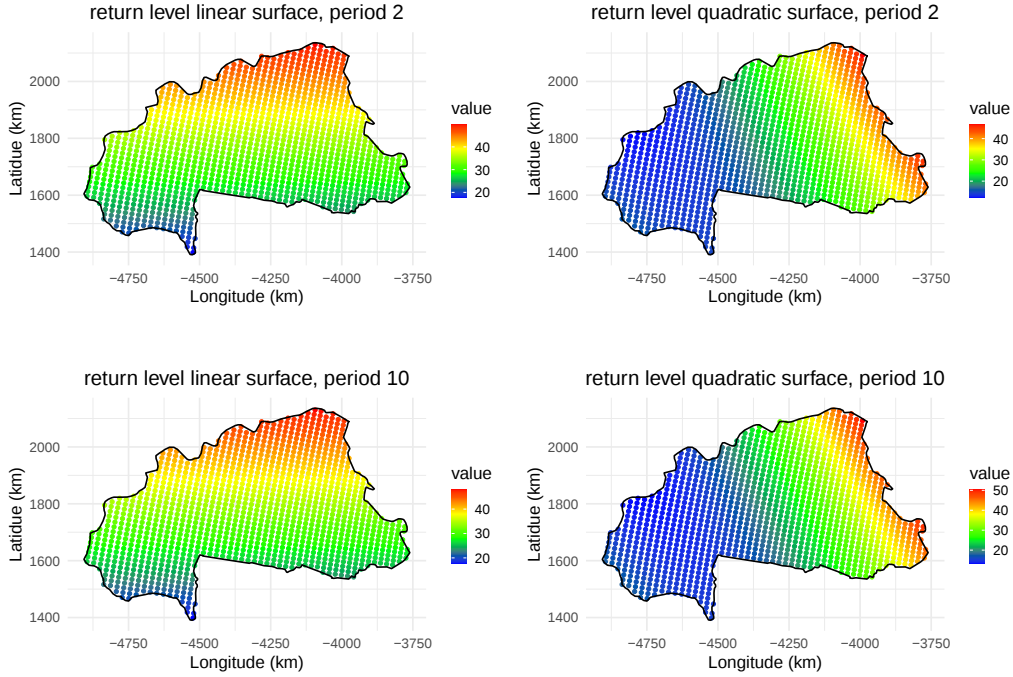


FIGURE 4.4 – Surface de retour des températures maximales associées aux fonctions linéaire et quadratique dans le cas d’un ajustement Brown-Resnick.

4.4. Discussion et perspectives

Dans ce chapitre, nous avons proposé des modèles d’analyse de tendances d’extrêmes dans un cadre spatio-temporel par l’approche fonctionnelle des dépassements de seuil. Nous avons étendu la définition du niveau de retour non-stationnaire univarié (Parey *et al.* (2010)) à des processus stochastiques spatio-temporels. Les surfaces de retour non-stationnaires sont déduites à partir des simulations pour une période de retour τ fixée, en utilisant la définition proposée 4.24. Ces surfaces recherchées dépendent de la forme paramétrique qu’on se donne à f_τ et elles sont susceptibles d’être dépassées par le processus sur toute la région d’étude au moins une fois sur une période de temps τ donnée. Nous avons illustré notre nouvelle méthodologie sur des données réelles de températures maximales quotidiennes du Burkina Faso. Notre nouvelle approche permet de prendre en compte non seulement une structure de dépendance spatiale, mais aussi de mieux quantifier les incertitudes spatio-temporelles dans l’estimation des quantiles extrêmes non-stationnaire contrairement aux approches plus simples notamment sans la prise en compte de la structure de dépendance spatiale. C’est une approche très intéressante qui s’avère d’un intérêt crucial pour estimer les surfaces de retour non-stationnaires des processus spatio-temporels dans le contexte actuel de changement climatique. Pour des raisons de temps de calculs, nous n’avons considéré ici que des périodes de retour relativement

4.4. DISCUSSION ET PERSPECTIVES

courte. Mais nous sommes en train de travailler sur notre algorithme de calculs afin de l'optimiser et nous permettre de prendre en compte des périodes de retour plus ou moins longue. C'est vrai que nous n'avons considéré dans notre analyse que des formes de tendance polynomiale d'ordre un et deux des latitudes et longitudes pour f_τ . Mais nous pouvons envisager d'autres formes de tendances plus flexibles dépendant par exemple des covariables climatiques qui elles-mêmes dépendent du temps. Elles peuvent prendre les formes constantes par morceaux, semi-paramétrique, ou non-linéaire. En perspectives nous comptons étendre d'autres définitions du niveau de retour ponctuel aux processus spatio-temporels. Un accent particulier sera aussi accordé à l'aspect numérique des différents programmes afin de réduire le temps de calculs.

Space-time trend detection and dependence modelling in extreme events approaches by functional Peaks-Over-Thresholds. Application to precipitation in Burkina Faso

Ce chapitre introduit une approche POT spatio-temporelle permettant de modéliser et d'analyser les tendances dans les extrêmes. C'est un article publié dans "International Journal of Mathematics and Mathematical Sciences".

Béwentaoré Sawadogo, Diakarya Barro

Abstract. In this paper, we propose a new method for estimating trends in extreme spatio-temporal processes using both information from marginal distributions and dependence structure. We combine two statistical approaches of extreme value theory : the temporal and spatial non-stationarities are handled via a tail trend function in the marginal distributions. The spatial dependence structure is modeled by a latent spatial process using generalized ℓ -Pareto processes. This methodology for trend analysis of extreme events is applied to precipitation data from Burkina Faso. We show that a significant increasing trend for the 50 and 100 years return levels in some parts of the country. We also show that extreme precipitation is spatially correlated with distance for radius of approximately 200 km.

Keywords : Non-stationary POT, Generalized ℓ -Pareto process, Space-time Extremes, Dependence Modeling, Trends detection, Climate Change.

5.1. Introduction

In the framework of climate change, the modelling and accurate prediction of the magnitude and extent of extreme events that occur in space and time of climate variables is a particularly difficult task, since it implies taking into account spatial and temporal non-stationarities. Nowadays there is a general consensus in the scientific community that climate change has accelerated in recent decades and that the climate will continue to change in the coming decades, mainly due to natural and anthropogenic changes (IPCC 2007, 2018, 2019). This change is manifested in most

regions of the world by a resurgence of heavy rainfall, heat waves and pollution peaks with very significant economic and social consequences.

In sub-Saharan Africa, works on this topic have shown an increasing trend in the occurrence of extremes in meteorological parameters [Blanchet *et al.* \(2018\)](#); [Panthou *et al.* \(2018, 2014\)](#); [Wilcox *et al.* \(2018\)](#). In Burkina Faso, a study conducted in the framework of the National Action Program for Adaptation to climate change (PANA, 2007) showed that precipitation is highly variable in space and time. There is a decreasing trend in cumulative and daily rainfall. According to recent publications from the National Meteorology Agency, precipitation has returned to humid periods since the late 1980s and decades of 1990 and 2000. The return of rains is more related to a high frequency of high intensity rainfall events than to an increase in the number of rainy days. For example, we can cite the floods of 1st September 2009 that affected Ouagadougou and its outskirts, with a record of 261.3mm of rainfall not registered since 1919. The floods have affected more than 150000 peoples, damaged several bridges and flooded more than 9300 hectares of crops throughout the country. From a statistical point of view, this raises the question of trend detection in extreme events.

The classical extreme value theory extended both to non-stationary and non independent observations provides a rigorous mathematical framework to deal with this question of trend detection in extremes [Cabral *et al.* \(2020\)](#); [Einmahl *et al.* \(2016\)](#); [Ferreira *et al.* \(2017\)](#); [Mefleh *et al.* \(2020\)](#). This issue was first generally studied in extreme value literature by parametric pointwise approaches in which an extreme value model (GEV or GPD) is fitted to the data at each site in turn, leaving the parameters of the marginal distributions to evolve with time or other significant covariates [Cooley \(2013\)](#); [Davison et Smith \(1990\)](#); [Olsen *et al.* \(1998\)](#); [Parey *et al.* \(2010\)](#); [Salas et Obeysekera \(2014\)](#). Although it is relatively simple to construct non-stationary models in the univariate framework, it is more difficult to account for spatial and temporal trends in these univariate models. The spatial and temporal dimension of extreme events was first developed for block maxima in a stationary framework and modeled by max-stable processes [Diakarya *et al.* \(2012\)](#); [Bechler *et al.* \(2015\)](#); [De Haan *et al.* \(1984\)](#); [Dossou-Gbete *et al.* \(2009\)](#); [de Haan et Ferreira \(2006\)](#). These spatial models have been readapted to handle non-stationarities induced by global warming [Chevalier *et al.* \(2020\)](#); [Huser et Genton \(2016b\)](#). Although attractive, these models are expensive in computing time when tending towards large dimensions and are not adapted for modeling threshold exceedances.

The threshold exceedance approach introduced by [Balkema et De Haan \(1974\)](#) and [Pickands \(1975\)](#) was first extended to multivariate environments [Rootzén *et al.* \(2018b\)](#) before being generalized to functional data [de Fondeville et Davison \(2018\)](#); [de Fondeville et Davison \(2020\)](#); [Dombry et Ribatet \(2015\)](#); [Ferreira *et al.* \(2014\)](#) to give birth to the family of generalized ℓ -Pareto processes. This family of processes are good candidates for modelling the spatial dependence structure of exceedances data. However, these approaches do not sufficiently take into account marginal non-stationarities and spatial dependence between margins. In this paper we propose

a new method to capture non-stationarities in marginal distributions, while taking into account the spatial dependence structure. To reach this goal, we combine two statistical approaches to extreme value theory. First temporal and spatial non-stationarities are controlled via a tail trend function in marginal distributions [Cabral et al. \(2020\)](#); [Einmahl et al. \(2016\)](#); [Ferreira et al. \(2017\)](#). Subsequently the spatial dependence structure is taken into account by a hidden auxiliary spatial process using generalized ℓ -Pareto processes [de Fondeville et Davison \(2020\)](#); [Dombry et Ribatet \(2015\)](#). Finally we use the model to simulate and predict future extreme events.

The article is organized as follows : section 2 details the methodology implemented and the methods used to estimate the parameters. In section 3, we present our main results from the data analysis, and in particular we calculate the non-stationary return levels from the developed approaches. Section 4 concludes the study.

5.2. Methodology

5.2.1. Space-time trends modelling

Let $X = \{X_t(s), s \in S, t \in T\}$ be a continuous non-stationary space-time stochastic process with sample paths in the family of continuous functions $\mathcal{C}(S \times T)$ equipped with the uniform norm $\|\cdot\|_\infty$, where $S \times T \subset \mathbb{R}^d \times \mathbb{R}^+$ and $\mathcal{C}_+(S \times T)$ its restriction to non-negative functions deprived of the null function. In practice X is observed at each stations $s_1, \dots, s_m$ and at given dates $t = 1, \dots, n$. Let $F_{t,s}$ be the continuous univariate marginal distribution with a common right endpoint x_F and $Z = \{Z(s), s \in S\}$ an unobserved latent spatial stochastic process with sample paths in $\mathcal{C}(S \times T)$ satisfying the proportional tail condition such that

$$\lim_{x \rightarrow x_F} \frac{P(X_t(s_j) > x)}{P(Z(s_j) > x)} = c_\theta\left(\frac{t}{n}, s_j\right), \text{ with, } \frac{1}{m} \sum_{j=1}^m \int_0^1 c_\theta(u, s_j) du = 1, \quad u \in [0, 1], \quad (5.1)$$

where $c_\theta : [0, 1] \times S \rightarrow (0, \infty)$ is assumed to be a continuous and positive function depending on a parameter vector $\theta \in \Theta \subset \mathbb{R}$, called tail trend function or skedasis function [Cabral et al. \(2020\)](#); [Ferreira et al. \(2017\)](#). The skedasis function describes the evolution of extreme events jointly in space and time. In the framework of the model (5.1), Mefleh et al. [Mefleh et al. \(2020\)](#) shows that the empirical point measure converges in distribution in the space of point measure $\mathcal{M}_p = \mathcal{M}_p([0, 1] \times (0, \infty])$ to a Poisson point process with intensity measure $c(u)du\gamma z^{-(\gamma+1)}dz$ on $[0, 1] \times (0, \infty)$. In this case the times of exceedances for high threshold x and the value of exceedances are asymptotically independent with distributions respectively equal to the trend density function $c_\theta(u, s)$, $s \in S$ and the Pareto distribution of tail index γ .

Moreover we assume that the continuous marginal distributions F_Z of the latent process is in the maximum domain of attraction condition for some constant $\gamma \in \mathbb{R}$ and appropriate real normalization constants $\tilde{a}_n > 0$, $\tilde{b}_n \in \mathbb{R}$. Thanks to the equation (5.1) and the convergence of Z exceedances to a GPD distribution we deduce a

sample of $\{Z_t(s), s \in S\}$ as mentioned in [Ferreira et al. \(2017\)](#) and [Cabral et al. \(2020\)](#) in the following manner :

$$Z_t(s_j) = \left\{ c_\theta \left(\frac{t}{n}, s_j \right) \right\}^{-\gamma} \left[X_t(s_j) - \frac{\left\{ c_\theta \left(\frac{t}{n}, s_j \right) \right\}^\gamma - 1}{\gamma} (\tilde{a}_n - \gamma \tilde{b}_n) \right], \quad (5.2)$$

$$j = 1, \dots, m; \quad t = 1, \dots, n,$$

where $\gamma, \tilde{a}_n, \tilde{b}_n$ and c_θ are parameters to be estimated. Parameter inference techniques will be discussed in the section (a).

5.2.2. Non-stationary return period and return level

The concept of return period and level becomes very ambiguous when we leave the stationary context to the non-stationary framework. For example, in the stationary framework, an extreme event with a 100-year return period is likely to occur on average once every 100 years, but with an annual exceedance probability of 1% in a given year. In this case it is assumed that the underlying probability distribution remains unchanged over time. More formally, for a stationary random variable Z with distribution function F_Z , the m -year return period is expressed as $m = 1/F_Z(Z > z_m)$, where z_m is the return level associated with the return period. However, in the non-stationary context, the underlying probability distribution is no longer constant, but is supposed to evolve with time. In this paper, we have chosen to follow return period based approaches, i.e., Expected Number of Exceedances (ENE) [Parey et al. \(2010\)](#) and Expected Waiting Time (EWT) [Cooley \(2013\)](#); [Olsen et al. \(1998\)](#). In the ENE approach, the number of times the variable $X_t(s), s \in S$ exceeds the return level value x_m in m years is defined by $N_m = \sum_{t=1}^{n_x m} \mathbb{1}_{\{X_t(s) > x_m, s \in S\}}$ under non stationary context. The return level x_m can be defined as the value for which the expected number of events exceeding x_m in m years equals to one, i.e., the return level x_m is the solution of the following equation (3.17) :

$$1 = \sum_{t=1}^{n_x m} \{1 - F_{t,s}(x_m | \theta_t(s))\}, \quad s \in S, \quad (5.3)$$

where n_x is the number of days in the year and $\theta_t(s)$ the vector of time-dependent marginal parameters or other covariates. Parey et al. [Parey et al. \(2010\)](#) uses the ENE method in a pointwise POT model where the parameters of the distribution of exceedances and the intensity of extreme event occurrences are described as polynomial functions of time.

The EWT method was first proposed by Olsen et al. [Olsen et al. \(1998\)](#), and then derived by Salas and Obeysekera [Salas et Obeysekera \(2014\)](#) using a geometric distribution with time-varying parameters. Under non stationary conditions, the distribution describing waiting time Y before the first occurrence of an event exceeding the return level x_m is

$$f_s(y) = P(Y(s) = y) = p_{y,s} \prod_{i=1}^{y-1} (1 - p_{i,s}), \quad y = 1, 2, \dots, y_{\max} \quad \text{and } s \in S, \quad (5.4)$$

5.2. METHODOLOGY

where variable Y is the day of the first occurrence of an event exceeding the quantile x_m , $p_{t,s} = 1 - F_{t,s}(x_m | \theta_t(s))$ is daily exceedance probability varying with time step t . $y_{\max}$ is the time when the daily exceedance probability $p_{t,s}$ is equal to 1 for an increasing-trend series or is equal to 0 for a decreasing-trend series. Reused and simplified by Cooley (2013), the EWT approach defines the m -year return level x_m , as the value for which the expected waiting time until an exceedance of this level is m years, i.e, x_m is the solution of the equation :

$$n_x m = E[Y] = 1 + \sum_{y=1}^{\infty} \prod_{t=1}^y F_{t,s}(x_m | \theta_t(s)), s \in S. \quad (5.5)$$

Using the relationship (5.1), $P(X_t(s) > x_m)$ can be rewritten $c_{\theta}(\frac{t}{n}, s)P(Z(s) > u)P(Z(s) > x_m | Z(s) > u)$ for $u < x_m$ and we obtain the following results

Proposition 5.2.1. *Let $\{X_t(s), s \in S, t \in T\}$ be a non-stationary stochastic process defined on a region $S \subset \mathbb{R}^d$ and $\{Z(s), s \in S\}$ a latent spatial process satisfying the equation (5.1). Given a return period m and threshold $u < x_m$, the return level x_m for all $s \in S$ is a solution of the following two equations :*

i) *Return period as expected number of events*

$$1 = \sum_{t=t_0}^{t_0+n_x m} \left\{ c_{\theta} \left(\frac{t}{n}, s \right) \phi_u(s) \bar{F}_{Z(s)|\ell(Z)>u}(x_m - u) \right\}, \quad (5.6)$$

ii) *Return period as expected waiting time*

$$n_x m = 1 + \sum_{i=1}^{\infty} \prod_{t=1}^i \left\{ 1 - c_{\theta} \left(\frac{t}{n}, s \right) \phi_u(s) \bar{F}_{Z(s)|\ell(Z)>u}(x_m - u) \right\}, \quad (5.7)$$

where $\bar{F}_{Z(s)|\ell(Z)>u}$ is a survival of generalized Pareto distribution of ℓ -exceedances at position $s \in S$; $\phi_u(s) = P(Z(s) > u)$ is the probability of exceedances, c_{θ} is the tail trend function and n_x is the number of days in the year.

The non-stationary return levels obtained from equations (5.6 and 5.7) are evaluated by numerical algorithms taking into account information from the extrapolation of the trend function c_{θ} . To derive the non-stationary return level at points where we have not observations, we use spatial marginal model (5.15) described in section (a), to extrapolate the parameters at these points. An alternative approach is to calculate the non-stationary return levels from the following result.

Proposition 5.2.2. *Let $\{X_t(s), s \in S, t \in T\}$ be a non-stationary spatio-temporal stochastic process defined on a region $S \subset \mathbb{R}^d$ and $\{Z(s), s \in S\}$ a latent spatial process. The non-stationary return level $x_m(s), s \in S$ of the non-stationary process X is deduced from the return level z_m of the latent spatial process Z such that :*

$$x_m(s) = z_m(s) \hat{c}_{\theta}(t_m, s)^{\hat{\gamma}} + \frac{\hat{c}_{\theta}(t_m, s)^{\hat{\gamma}} - 1}{\hat{\gamma}} \left(\hat{a}_n - \hat{\gamma} \hat{b}_n \right), \quad s \in S, \quad (5.8)$$

where $\hat{a}_n, \hat{b}_n, \hat{c}_{\theta}$ and $\hat{\gamma}$ are the respective estimators of $\tilde{a}_n, \tilde{b}_n, c_{\theta}$ and γ . $t_m = 1 + \frac{n_x m}{n}$ with n_x the number of days in the year and n the size of the sample observed.

This result is a consequence of the equation (5.2). In practice, the return level z_m of the latent spatial process is first computed at grid points where we have not observations using equation (5.15). Then the non-stationary return level x_m is deduced using the proposition (5.2.2)

5.2.3. Spatial POT modelling

After removing the non-stationarity thanks to the equation (5.2), the modelling is then focused on the evaluation of the extreme spatial dependence structure in Z using functional POT [de Fondeville et Davison \(2018\)](#); [de Fondeville et Davison \(2020\)](#); [Dombry et Ribatet \(2015\)](#). In the multivariate and spatial framework a threshold exceedances for a random function $Z = \{Z(s), s \in S\}$ is defined by [Dombry et Ribatet \(2015\)](#) to be an event of the form $\{\ell(Z) > u\}$ for some $u \geq 0$, where $\ell : \mathcal{C}(S) \rightarrow \mathbb{R}^+$ is a continuous and homogeneous non-negative risk function, i.e, there exists $\alpha > 0$ such that $\ell(\lambda y) = \lambda^\alpha \ell(y)$ when $y \in \mathcal{C}_+(S)$ and $\lambda > 0$. The risk function ℓ determines the type of extreme events of interest. For example, such a function can be the maximum, minimum, average or value at a specific point $s_0 \in S$. Under minimal assumptions on the risk function, the conditional distribution of ℓ -exceedance for some threshold $u \geq 0$ of the process $(Z - b_n)/a_n$ can be approximated by a generalized ℓ -Pareto process, for n large enough [de Fondeville et Davison \(2020\)](#)

$$P \left\{ \left\lfloor \frac{Z - b_n}{\ell(a_n)} \right\rfloor \in A \mid \ell \left(\frac{Z - b_n}{\ell(a_n)} \right) \geq u \right\} \rightarrow P \{W_\ell \in A\}, n \rightarrow \infty, \quad (5.9)$$

where W_ℓ is a non-degenerate stochastic process over S and belongs to the family of generalized ℓ -Pareto processes with tail index γ (see appendix 1.B for more details). Specifically, W_ℓ is a stochastic process taking values in $\{z \in \mathcal{C}_+(S) : \ell\{(z - b)/\ell(a)\} \geq 0\}$ and defined by :

$$W_\ell = \begin{cases} a(Y_\ell^\gamma - 1)/\gamma + b, & \gamma \neq 0 \\ a \log Y_\ell + b, & \gamma = 0 \end{cases}, \quad (5.10)$$

where $a > 0$ and b are continuous functions on S respectively scale and location functions and Y_ℓ is a stochastic process whose probability measure is completely determined by the limit measure Λ . More details on generalized ℓ -Pareto processes can be found in [de Fondeville et Davison \(2018\)](#); [de Fondeville et Davison \(2020\)](#)

For the modelling of spatial dependence structure of latent process, we use Y_ℓ whose margins are in the Frechet domain of attraction with tail index $\gamma = 1$, as the process of reference. A pseudo-polar decomposition of Y_ℓ in (5.11) leads to the following formulation [de Fondeville et Davison \(2018\)](#); [Dombry et Ribatet \(2015\)](#)

$$Y_\ell = RQ \quad (5.11)$$

where R is a unit Pareto random variable of index γ_R representing the intensity of process, and Q is stochastic process denoted the angular component with state space S and taking values in $\mathcal{S} = \{y \in \mathcal{C}_+(S) : \|y\|_1 = 1\}$ whose probability measure is characterized by limit measure Λ .

5.2.4. Statistical inference

a. Marginal model of spatial latent process

Marginal parameters a_n , b_n and γ of latent process can be estimated into fixing $\tilde{b}_n$ at a high quantile of $\ell(Z)$, i.e, $\tilde{b}_n = q_{1-\alpha}\{\ell(Z)\}$. In general, a parametric model may be necessary for a_n and b_n , as in Engelke et al. [Engelke et al. \(2019\)](#). However, some forms of parameterization can lead to problems of parameter identifiability and inference. For these reasons and simplicity purpose, we consider in this work that $a_n(s_j) = a_j > 0$ and $b_n(s_j) = b_j \in \mathbb{R}$ for any $j = 1, \dots, m$. The threshold stability of the generalized Pareto distributions does not allow us to identify the b_n function without additional assumptions, so we set

$$b_n(s) = u_{q'}\{Z(s)\} - \tilde{b}_n, \quad s \in S, \quad (5.12)$$

where $u_{q'}\{Z(s)\}$ is an empirical quantile of the order q' of the ℓ -exceedances at each location $s \in S$, where q' is chosen such that $\ell(b_n) = \tilde{b}_n$ in order to impose the identifiability of the parameters. Thus, the tail index $\gamma \in \mathbb{R}$ and the scale parameters $a_n(s) > 0$ are estimated by maximizing the independent log-likelihood; that is

$$\ell_{indep}(\gamma, a_n(s_j)) = \sum_{t=1}^n \sum_{j=1}^m \mathbb{1}_{\{x_t(s_j) \geq b_n(s_j)\}} \log \left[\frac{1}{a_n(s_j)} \left\{ 1 + \gamma \frac{x_t(s_j) - b_n(s_j)}{a_n(s_j)} \right\}_+^{-1/\gamma-1} \right] \quad (5.13)$$

The assumptions of parameter identifiability (see appendix 1.B) implies that $\ell(a_n) = \tilde{a}_n$.

The trend function parameter c_θ is estimated using maximum likelihood method under the assumption that the dates of exceeding the marginal thresholds $\tilde{b}_n$ are asymptotically independent and identically distributed, from the density function $c_\theta(\frac{t}{n}, s)$ [Mefleh et al. \(2020\)](#). Several models are eligible to model the function c_θ . But in this study we opt for parametric models because of their flexibility in trend detection and in order to make extrapolations of the trend beyond the observed data. To facilitate inference by the maximum likelihood method, we are interested in monotonic log-linear and simple linear trend models of c_θ such that

$$\begin{cases} c_\theta^1(\frac{t}{n}, s) = \frac{\theta(s)}{\exp\{\theta(s)\} - 1} \exp\{\theta(s)\frac{t}{n}\}, & \theta(s) \in \mathbb{R}; \frac{t}{n} \in [0, 1] \\ c_\theta^2(\frac{t}{n}, s) = \theta(s) \left\{ 2\frac{t}{n} - 1 \right\} + 1, & \theta(s) \in]-1; 1[; \frac{t}{n} \in [0, 1], \end{cases} \quad (5.14)$$

where $\theta(s), s \in S$ is characteristic parameter of trend function. $\theta(s) > 0$ suggests that extreme are getting more frequent, while $\theta(s) < 0$ indicates that extremes are getting less frequent.

b. Spatial extremal dependence

The spatial dependence structure is taken into account in the marginal parameters of spatial process Z by letting these parameters evolve as a function of space covariates. More precisely, we use a spatial model of the marginal parameters as

functions of the significant space covariates using a generalized linear regression model [Cooley \(2013\)](#); [Parey et al. \(2010\)](#); [Salas et Obeysekera \(2014\)](#).

$$\begin{cases} a_n(s) = g_{a_n}(v(s)) > 0 \\ b_n(s) = g_{b_n}(v(s)) \in \mathbb{R} \\ \theta(s) = g_\theta(v(s)) \in \mathbb{R} \\ \gamma(s) = \gamma_0 \in \mathbb{R} \end{cases}, \quad s \in S \quad (5.15)$$

where v is a vector of covariates which can contain for example the space coordinates, or any other spatialized data explaining significantly the weather variable studied.

Furthermore, after removing the non-stationarity, modeling is focused on the evaluation of the remaining spatial dependence structure in Z . Thus we approach the limit distribution of ℓ -exceedances of spatial latent process by a generalized ℓ -Pareto process (5.9). To estimate the parameters of spatial dependence structure, we have chosen to model its angular component Q (5.11) by a log-Gaussian process with stationary increments and within the framework of the Brown-Resnick model. In this case, we have access to the limit measure Λ . In order to better capture this dependence structure, we use a flexible parametric semi-variogram belonging to the class of power semi-variograms. The parameter of the dependence structure is estimated using the gradient score method or the censored likelihood method [de Fondeville et Davison \(2018\)](#); [Wadsworth et Tawn \(2014\)](#).

5.3. Application to extreme precipitation in Burkina Faso

5.3.1. Data set analysis

This study uses time series of daily precipitation measurements from 1957 to 2016 provided by ten synoptic stations extracted from the Burkina Faso climatological database. These stations have been selected to ensure good spatial uniformity and representativeness of different climatic regimes and data quality. The Figure 5.1 gives the spatial distribution of selected stations in our study area. In order to limit the problems related to seasonal rainfall cycles on each station, we worked from the sub-series corresponding to rainy days. The period from May to October was chosen because during this period that the most rainfall is recorded in the country. Thus, a series of 60 years (1957-2016) by 184 days (May to October) is extracted to constitute the time series of daily rainfall. On these time series, we apply a run declustering procedure with a daily step ($r=1$ day) to identify the groups of approximately independent extreme observations within the sample in order to avoid short-term dependencies in the time series.

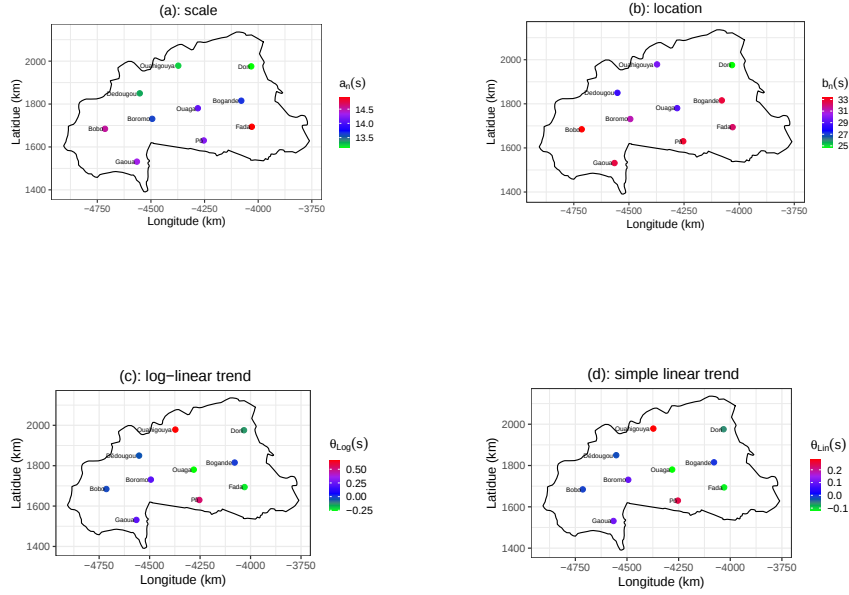


FIGURE 5.2 – Estimated functions a_n (5.2a) and b_n (5.2b) of the generalized ℓ -Pareto process for modelling extreme precipitation, and θ for Log-linear (5.2c) and simple linear (5.2d) trends. Estimates are obtained by shifting the local empirical quantiles $u_{0.80} \{\ell(Z(s))\}$ by $\hat{b}_n = 33.4mm$.

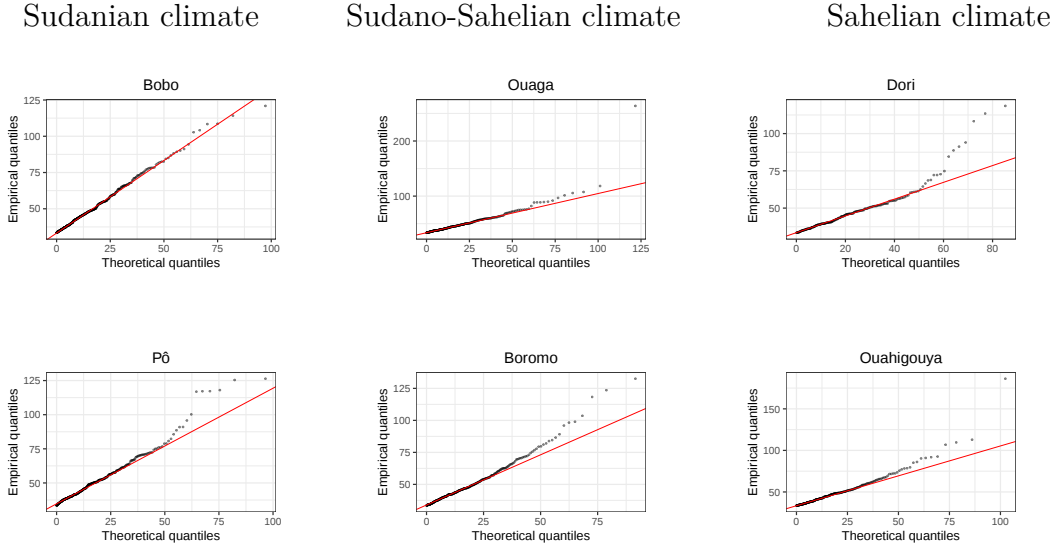


FIGURE 5.3 – q-q plot of the distributions of the local tails of six synoptic positions including two stations for each climate region. left column : Sudanian climate; middle column : Sudano-Sahelian climate; right column : Sahelian climate. The threshold is taken equal to $b_n(s)$ for each station.

5.3. APPLICATION TO EXTREME PRECIPITATION IN BURKINA FASO

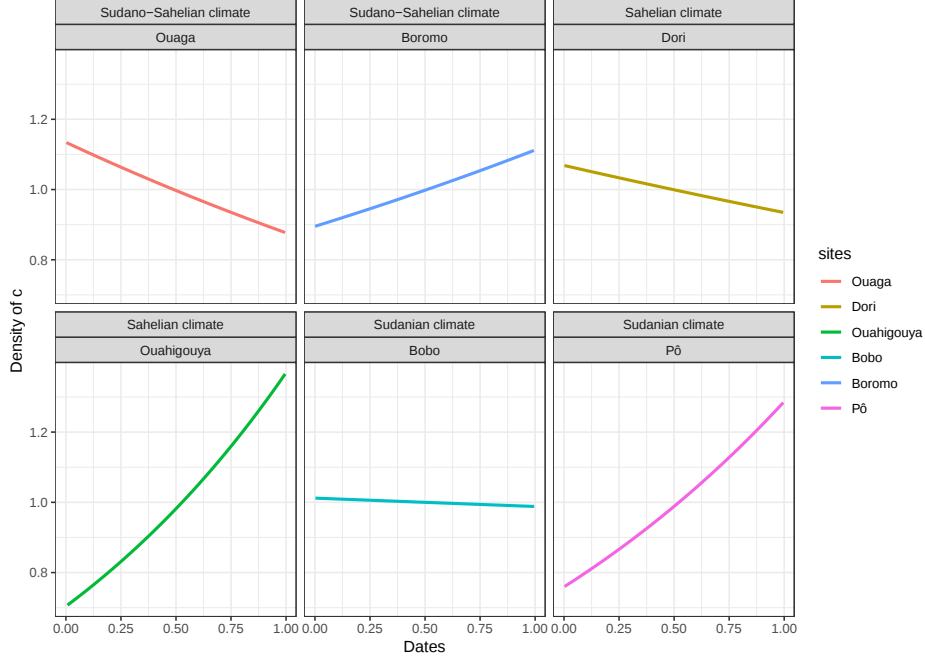


FIGURE 5.4 – Local fitting of the log-linear trend model c_θ^1 for some of the selected sites with two stations for each climate region

5.3.3. Estimated spatial dependence model

We use a spatial marginal model (5.15) to estimate the marginal parameters at the different grid points. This is to prepare the ground for the calculation of non-stationary return levels at locations where we have no observations.

$$\begin{cases} a_n(s) = a_0 + a_1 \text{Long} \times \text{Lat}(s) \\ b_n(s) = b_0 + b_1 \text{Lat}^2(s) + b_2 \text{Long} \times \text{Lat}(s) \\ \theta(s) = \theta_0 + \theta_1 \text{Long}(s) + \theta_2 \text{Lat}(s) \\ \gamma(s) = \gamma_0 \end{cases}, \quad s \in S, \quad (5.17)$$

An alternative to the spatial model of marginal parameters is to consider small subsets $\mathcal{N}_{s_0} \subset S$, where $\mathcal{N}_{s_0}$ is a small neighborhood around s_0 . In what follows, these regional neighborhoods will be determined by a small number D_0 of nearest stations from the site s_0 , thus we will write $\mathcal{N}_{s_0} = \mathcal{N}_{s_0, D_0}$. In principle, the choice of D_0 should be such that the spatial dependence structure and marginal parameters is approximately stationary within each selected neighborhood $\mathcal{N}_{s_0, D_0}$ around $s_0 \in S$. Obviously, the choice of neighbourhood is important; the assumed stationary marginal parameters could be a poor approximation for large neighbourhoods (i.e., for large D_0), while the simulation of the process could be cumbersome for small

neighbourhoods characterized by a small D_0 number of stations. We obtain the relatively homogeneous, non-overlapping sub-regions using the k-means clustering method centered on the reference stations [Bernard *et al.* \(2013\)](#); [Janßen et Wan \(2020\)](#). This method is extensively used because it is computationally simple and produces accurate results, compared to other more complex clustering methods. The longitude and latitude of the grid points were used as input variables in the k-means clustering algorithm to form the ten clusters centered on the reference stations. In addition, to better take into account the dependence structure of precipitation data and capture the possible isotropic, we choose a flexible parametric semi-variogram ν belonging to the power class models such that $\nu(h) = (\|h\|/\tau)^\kappa$, with $\tau > 0$, $h \in \mathbb{R}^2$, $\kappa \in]0, 2]$. We check the adequacy of the fit of the model to the data using the extremogram and the variogram. Figures 5.5a and 5.5b show respectively the good quality of the fit of the dependence measures such as the semi-variogram and extremogram. In Figure 5.5a, the points in gray represent the calculated empirical extremogram and the blue curve is the fitted empirical extremogram, while the red curve represents our fitted dependence model. The red curve in Figure 5.5b is the variogram model fitted to the data as a function of distance. It is noted that

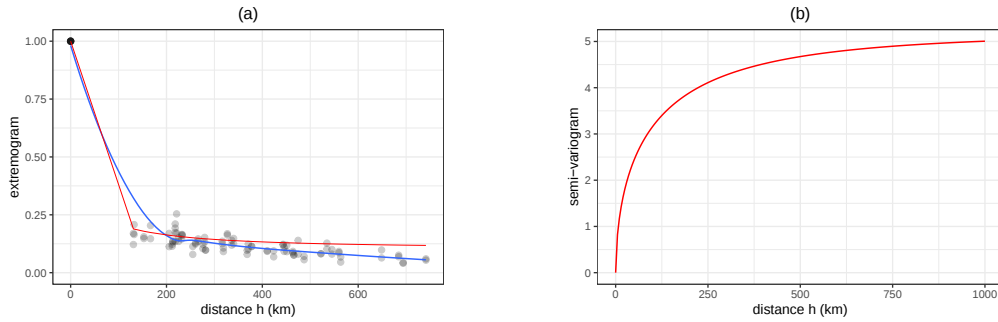


FIGURE 5.5 – Isotropic extremogram $\pi(h) = \lim_{q \rightarrow 1} P(Z(s+h) > u_q(s+h) \mid \{Z(s) > u_q(s), \ell(Z) > u\})$ (Figure 5.5a) and variogram (Figure 5.5b) estimated for risk functional $\ell(Z) = \max_{s \in S} Z(s)$ of distance between locations s and $s+h$. Blue line : empirical extremogram and red line : fitted dependence model.

extreme precipitation is spatially correlated for distances of the order of 200km. In other words, an extreme precipitation event is likely to affect an area within a 200km radius. This spatial dependence decreases gradually as the distance increases before stabilizing for example for a value of $\pi(h) = 0.0625$. We find the very localized nature of extreme precipitation for long distances. This is generally a property present in rainfall data [Tawn *et al.* \(2018\)](#)

5.3.4. Non-stationary return level results

After estimating the marginal parameters and those of the dependence structure, we now evaluate the non-stationary return levels. The return levels x_{50} and

5.3. APPLICATION TO EXTREME PRECIPITATION IN BURKINA FASO

x_{100} are first estimated (see Figure 5.7) using the equations (5.6 and 5.7). Then they are spatially interpolated at points where we have no observations using the spatially extrapolated trend function and dependence structure estimated from the observed data. For a given return period m , we first compute the return levels z_m

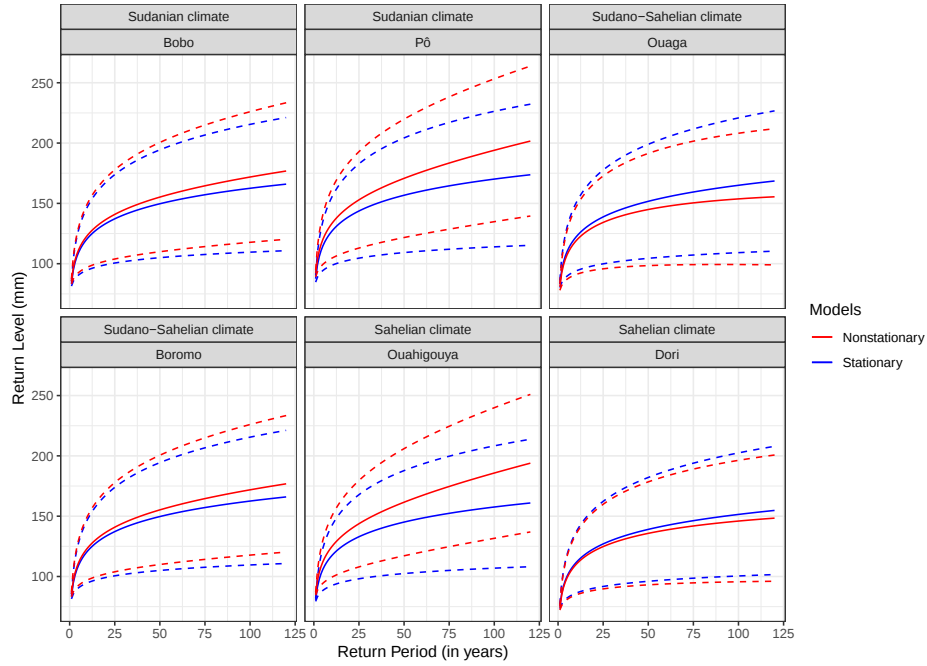


FIGURE 5.6 – Trends of stationary and non-stationary return levels at reference stations computed with a log-linear tail trend function.

on each point of grid from the latent spatial process Z using the estimated spatial model. The non-stationary return levels x_m is deducted using the equation 5.8 of proposition 5.2.2. We repeat this operation for different values of the return period m and we deduce on each point of the grid, the associated return level (Figure 5.6). We can then produce maps of the return levels obtained from the dependence structure and the extrapolation of the trend function for a future period m . Thus, the extreme precipitation likely to be observed on average at least once every 50 years (resp. 100 years), will be particularly intense in the Sudanian and Sudano-Sahelian zone and less intense in the Sahelian zone, with a potentially quite strong spatial dependence within a radius of 200 km. The southwest and eastern regions of the country will be most affected by extreme precipitations. The results of the return level for a return period $m = 50$ and 100 years are displayed in Figures 5.7.

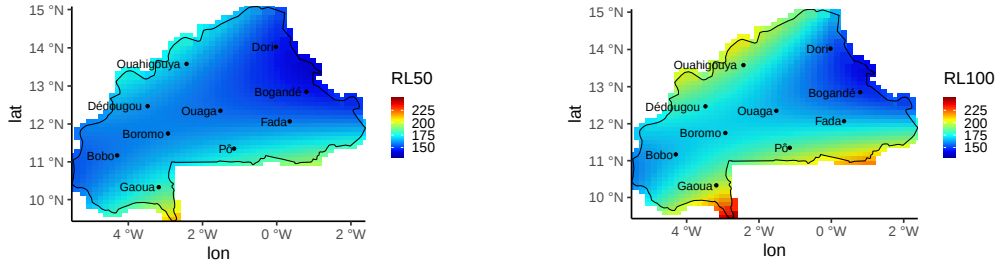


FIGURE 5.7 – Maps of the non-stationary 50-years return levels(left) and 100-years return level (right) obtained by extrapolating the information from the log-linear tail trend function and the dependence structure.

5.4. Discussion and conclusion

In summary, modeling the spatio-temporal trends of extreme precipitation in sub-Saharan Africa is a challenge and a necessity. Climate models do not often agree on this issue. Advanced statistical methods remain a fairly powerful alternative to obtain pertinent information on possible spatio-temporal changes in the underlying precipitation distribution. A frequently used approach to dealing with non-stationarity in trend analysis of extreme precipitation is to include covariates, such as space and time, in the parameters of the underlying distribution [Cooley \(2013\)](#); [Davison et Smith \(1990\)](#); [Parey et al. \(2010\)](#); [Salas et Obeysekera \(2014\)](#). These approaches are built under certain important assumptions that the underlying process that generates the data is locally stationary and that the observations generated can be considered as independent approximations and generally do not take into account spatial dependence.

In this study, we proposed a new flexible methodology for trend detection in extremes of spatio-temporal processes. This is capable of capturing marginal non-stationarities and the dependence structure between margins. Marginal non-stationarity is taken into account by a trend function called skedacis function which models the frequency of extreme events jointly in space and time. While spatial dependence structure is governed by a latent spatial process and modeled using generalized ℓ -Pareto processes. We calculated the non-stationary return levels of precipitation in Burkina Faso and exhibit some regions in which there may be a significant increasing or decreasing of extreme precipitation. We showed that these extreme rainfall events are spatially correlated around a radius of 200km. This spatial dependence decreases progressively as the distance increases. In sum, we set up a non-stationary stochastic generator of extreme rainfall in Burkina Faso.

Previous studies conducted in the sub-Saharan African region have shown trends in extreme precipitation similar to those we obtained using Burkina Faso data

5.4. DISCUSSION AND CONCLUSION

set. According to the latest publications of the National Meteorological Agency of Burkina Faso, rainfall has started to move back towards a wet period since the end of the 1980s and the decade 1990-2000. This return of rain is more related to an increase in the frequency of heavy rainfall events than to an increase in the number of rainy days. Our results are also in agreement with the results found by [Panthou *et al.* \(2018, 2014\)](#); [Wilcox *et al.* \(2018\)](#) in the region which indicate a similar intensification of rainfall events. Based on a point *GEV* model with time-dependent parameters, these authors show that intense precipitation events will become more frequently occurring in the region. Nkrumah *et al.* [Nkrumah *et al.* \(2019\)](#) and Mouhamed *et al.* [Mouhamed *et al.* \(2013\)](#) observe in the same region that the average annual precipitation has a decreasing trend due to less frequent but more intense precipitation during the last decade.

Our method generates extreme events consistent enough with the data already observed. The return levels computed can be improved by introducing also a non-stationarity in the structure of spatial dependence.

Acknowledgement

We thank the National Meteorology Agency of Burkina Faso for providing us with the rainfall data. We also thank the Cooperation and Cultural Action Service of the French Embassy in Burkina Faso and UMR 518 MIA-Paris-Saclay, AgroParisTech, INRAe for their financial support.

Conflict of interest

The authors declare that there is no conflict of interest for the publication of this paper.

1.A. Proof of proposition 5.2.1

i) *Démonstration.* Let N_m be a random variable describing the number of exceedances for a period of time m ; it comes that

$$N_m = \sum_{t=t_0}^{t_0+n_x m} \mathbb{1}_{\{X_t(s) > x_m(s)\}} \Rightarrow E[N_m] = \sum_{t=t_0}^{t_0+n_x m} E\left(\mathbb{1}_{\{X_t(s) > x_m(s)\}}\right).$$

And according to the equation (5.1), we have

$$E[N_m] = \sum_{t=t_0}^{t_0+n_x m} P(X_t(s) > x_m(s)) = \sum_{t=t_0}^{t_0+n_x m} c_\theta \left(\frac{t}{n}, s\right) P(Z_t(s) > x_m(s)).$$

Moreover, given a sufficiently large $u(s)$ threshold with $s \in S$,

$$\begin{aligned} P[Z_t(s) > x_m(s)] &= P[Z_t(s) > u(s)] * P[Z_t(s) - u(s) > x_m(s) \mid Z_t(s) > u(s)] \\ &\approx \phi_u(s) * \bar{F}_Z[x_m(s) - u(s), \sigma(s), \gamma(s)], x_m(s) > u(s) \end{aligned}$$

Furthermore, it follows that

$$E[N_m] = 1 \Rightarrow \sum_{t=t_0}^{t_0+n_x m} \left\{ c_\theta \left(\frac{t}{n}, s\right) \phi_u(s) * F_Z(x_m(s) - u(s), \sigma(s), \gamma(s)) \right\} = 1$$

Which gives

$$\sum_{t=t_0}^{t_0+n_x m} \left\{ \phi_u(s) c_\theta \left(\frac{t}{n}, s\right) \left(1 + \frac{\gamma(s)}{\sigma(s)} \{x_m(s) - u(s)\}\right)^{-1/\gamma(s)} \right\} = 1$$

with $\sigma(s) = a_n(s) + \gamma(s)(u(s) - b_n(s))$

ii) The result (5.7) is shown in a similar way □

1.B. Overview on generalized ℓ -Pareto process

Let S be a compact subset of $\mathbb{R}^d$ denoting the spatial domain of the study region. We note by $\mathcal{C}(S)$ the set of continuous real functions on S equipped with the uniform norm $\|\cdot\|_\infty$; $\mathcal{C}_+(S)$ its restriction to non-negative functions deprived of the null function and $\mathcal{M}\{\mathcal{C}_+(S)\}$ the class of measures associated with $\mathcal{C}_+(S)$.

Let $\{Z(s), s \in S\}$ be a latent spatial stochastic process indexed by $s \in S \subset \mathbb{R}^d$ with sample paths in $\mathcal{C}(S)$ of continuous marginal distribution F_Z and common right endpoint z_F . As mentioned in [de Fondeville et Davison \(2020\)](#); [Ferreira et al. \(2014\)](#), we assume that the spatial process Z is a general functional regular variation (GRV), i.e that there exists suitable sequences of continuous functions $a_n : S \rightarrow \mathbb{R}^+$, $b_n : S \rightarrow \mathbb{R}$ and $\gamma \in \mathbb{R}$ such that

$$nP \left[(1 + \gamma((Z - b_n)/a_n))_+^{1/\gamma} \in \cdot \right] \rightarrow \Lambda(\cdot), n \rightarrow +\infty, \quad (18)$$

1.B. OVERVIEW ON GENERALIZED ℓ -PARETO PROCESS

and noted by $Z \in GRV(\gamma, a_n, b_n, \Lambda)$, where Λ is a non-zero measure in $\mathcal{M}\{\mathcal{C}_+(S)\}$ and homogeneous of order -1 , with $\Lambda(tA) = t^{-1}\Lambda(A)$ for any positive real $t > 0$ and Borel set $A \subset \mathcal{C}_+(S)$. The shape parameter $\gamma \in \mathbb{R}$ is also called the tail index, determines the strength of the tail and its support. According to the sign of γ , we end up with three types of distributions of extreme values known as Gumbel ($\gamma \rightarrow 0$), Frechet ($\gamma > 0$) and Weibull ($\gamma < 0$).

In the multivariate and spatial framework a threshold exceedances for a random function $Z = \{Z(s), s \in S\}$ is defined by Dombry and Ribatet [Dombry et Ribatet \(2015\)](#) to be an event of the form $\{\ell(Z) > u\}$ for some $u \geq 0$, where $\ell : \mathcal{C}(S) \rightarrow \mathbb{R}^+$ is a continuous and homogeneous non-negative risk function, i.e, there exists $\alpha > 0$ such that $\ell(\lambda y) = \lambda^\alpha \ell(y)$ when $y \in \mathcal{C}_+(S)$ and $\lambda > 0$. The risk function ℓ determines the type of extreme events of interest. For example, such a function can be the maximum, minimum, average or value at a specific point $s_0 \in S$. Moreover, as in [de Fondeville et Davison \(2020\)](#); [Engelke et al. \(2019\)](#) we assume that there exists a continuous and positive real function A_Z such that the sequence a_n and the risk function ℓ satisfy the following asymptotic decomposition

$$\lim_{n \rightarrow \infty} \sup_{s \in S} \left| \frac{a_n(s)}{\ell(a_n)} - A_Z(s) \right| = 0, \quad \text{i.e. } a_n(s) \approx \ell(a_n)A_Z(s), \quad n \rightarrow \infty \quad (19)$$

Furthermore, the marginal distributions of Z are supposed to belong to a location-scale family, thus ensuring a constant $\gamma \in \mathbb{R}$ shape parameter over any S , i.e. that there exist a H distribution function such that

$$P(Z(s) \leq z) = H[(z(s) - B_Z(s))/A_Z(s)], \quad (20)$$

where $A_Z : S \rightarrow (0; \infty[$ verifies asymptotic decomposition (19) and $B_Z : S \rightarrow \mathbb{R}$ are continuous functions. In particular H must belong to the domain of attraction of a generalized extreme value (GEV) distribution of tail index $\gamma \in \mathbb{R}$, i.e., $F_Z \in \mathcal{D}(G_\gamma)$. In other words, there are appropriate normalization real sequences $\tilde{a}_n > 0$, $\tilde{b}_n \in \mathbb{R}$ such that $\lim_{n \rightarrow \infty} H^n(\tilde{a}_n z + \tilde{b}_n) = G_\gamma(z)$. The condition of the max-domain of attraction is equivalent to

$$\lim_{n \rightarrow +\infty} n \left[1 - H(\tilde{a}_n z + \tilde{b}_n) \right] = -\log G_\gamma(z), \quad z > 0, \quad (21)$$

Assuming these conditions and general functional regular variation, the appropriate sequence of continuous functions $a_n(s)$ and $b_n(s)$, satisfy :

$$a_n(s) = \tilde{a}_n A_Z(s), \quad b_n(s) = B_Z(s) + A_Z(s)\tilde{b}_n, \quad s \in S. \quad (22)$$

Functions $A_Z(s)$ and $B_Z(s)$ as well as the spatial dependence structure of Z are assumed to belong to a family of parametric functions $\{A_{Z,\theta_A} : \theta_A \in \Theta_A\}$; $\{B_{Z,\theta_B} : \theta_B \in \Theta_B\}$ and $\{\Lambda_{\theta_\Lambda} : \theta_\Lambda \in \Theta_\Lambda\}$. Under minimal assumptions on the risk function and assumptions (18, 19) of Z , the conditional distribution of ℓ -exceedance for some threshold $u \geq 0$ of the process $(Z - b_n)/\ell(a_n)$ can be approximated by a generalized ℓ -Pareto process, for n large enough [de Fondeville et Davison \(2020\)](#)

$$P \left\{ \left\lfloor \frac{Z - b_n}{\ell(a_n)} \right\rfloor \in A \mid \ell \left(\frac{Z - b_n}{\ell(a_n)} \right) \geq u \right\} \rightarrow P \{W_\ell \in A\}, \quad n \rightarrow \infty, \quad (23)$$

where $\lfloor z \rfloor = \max(z, A_Z \gamma^{-1})$ if $\gamma > 0$ and $\lfloor z \rfloor = z$ if $\gamma \leq 0$. W_ℓ is a non-degenerate stochastic process over S and belongs to the family of generalized ℓ -Pareto processes with tail index γ , zero location, scaling function A_Z and limit measure Λ . Specifically, a generalized ℓ -Pareto process W_ℓ associated to the limit measure Λ and tail index $\gamma \in \mathbb{R}$ is a stochastic process taking values in $\{z \in \mathcal{C}_+(S) : \ell\{(z - b)/\ell(a)\} \geq 0\}$ and defined by :

$$W_\ell = \begin{cases} a(Y_\ell^\gamma - 1)/\gamma + b, & \gamma \neq 0 \\ a \log Y_\ell + b, & \gamma = 0, \end{cases} \quad (24)$$

where $a = \ell(a)A_Z > 0$ and b are continuous functions on S respectively scale and location functions. Y_ℓ is a stochastic process whose probability measure is completely determined by the limit measure Λ .

1.C. Asymptotic distribution of ℓ -Pareto Process

For a threshold vector $u = (u_1, \dots, u_d) \in \mathbb{R}_+^d$ and for any regularly varying stochastic process $Z \in GRV\{\gamma, a_n, b_n, \Lambda\}$, the density function of the ℓ -excess of a ℓ -Pareto process is obtained by renormalizing suitably the intensity function λ found by taking the partial derivatives of the previously defined measure Λ (18) by $\Lambda(A_\ell(u))$:

$$f_{\ell,u}(z) = \frac{\lambda(z)}{\Lambda\{A_\ell(u)\}}, \quad z \in A_\ell(u), \quad (25)$$

where

$$\Lambda\{A_\ell(u)\} = \int_{A_\ell(u)} \lambda(z) dz,$$

while λ is the intensity function and $A_\ell(u) = \{z \in \mathbb{R}_+^d : \ell(z/u) \geq 1\}$ the region of exceedances. In order to make inferences and to model the dependence structure of ℓ -Pareto processes, we focus on the Brown-Resnick model for which the formulas of Λ and λ are available. The d -dimensional intensity function of the Brown-Resnick model is given :

$$\lambda_{BR}(z) = \frac{|\Sigma|^{-1/2}}{z_1^2 z_2 \dots z_d (2\pi)^{(d-1)/2}} \exp\left(-\frac{1}{2} \bar{z}^T \Sigma^{-1} \bar{z}\right), \quad z \in \mathbb{R}_+^d$$

where $\bar{z}_i = \log \frac{z_i}{z_1} + \gamma(s_i - s_1)$ and Σ the covariance matrix. To better capture the possible dependence structure we use anisotropic semi-variograms whose parameters change with time and other covariates. The different parameters are estimated using the gradient scoring rule method [de Fondeville et Davison \(2018\)](#). For any $z \in A_\ell(u)$, the log-density function is given by :

$$\delta_w(\lambda_\theta, z) = \sum_{j=1}^d \left(2w_j(z) \frac{\partial w_j(z)}{\partial z_j} \frac{\partial \log \lambda_\theta(z)}{\partial z_j} + w_j(z)^2 \left[\frac{\partial^2 \log \lambda_\theta(z)}{\partial z_j^2} + \frac{1}{2} \left\{ \frac{\partial \log \lambda_\theta(z)}{\partial z_j} \right\}^2 \right] \right) \quad (26)$$

where w is a differentiable weighting function that vanishing on the boundaries of $A_\ell(u)$ and θ is an element in the $\Theta \subset \mathbb{R}^p$ parameter space.

Conclusions et perspectives

Dans cette thèse nous nous sommes penchés sur l'analyse et la détection de tendances des événements climatiques extrêmes. Les outils statistiques essentiels des extrêmes spatiaux et spatio-temporels sont rappelés dans le chapitre 1. En particulier nous avons accordé une grande importance aux derniers développements des processus spatiaux de dépassements de seuil rangés dans la famille des processus ℓ -Pareto. Dans le chapitre 2, nous avons analysé les tendances des températures et des précipitations extrêmes au Burkina Faso à l'aide d'une approche POT ponctuelle paramétrique non-stationnaire. Dans cette approche nous avons laissé les paramètres des distributions des dépassements marginaux et l'intensité des occurrences des événements extrêmes évoluer en fonction du temps ou avec d'autres covariables. En effet on considère $X = \{X_t(s)\}_{s \in S, t \in T}$ un processus spatio-temporel non-stationnaires défini sur $S \times T \subset \mathbb{R}^d \times \mathbb{R}^+$. X peut être un paramètre météorologique comme la température, la pluie, la vitesse du vent ou le niveau de pollution pour ne citer que ces exemples. Etant donné un seuil $u(s) > 0$ suffisamment grand pour tout $s \in S$, la loi conditionnelle de $Z_t(s) = \{X_t(s) - u(s) \mid X_t(s) > u(s)\}$ converge asymptotiquement vers une distribution des dépassements de seuil non-stationnaire non dégénérée, i.e $Z_t(s) \sim GPD(\sigma_\beta(s, t), \gamma_u(s))$ tandis que la loi des occurrences des événements extrêmes $N(s, t)$ est elle donnée par un processus de Poisson non-homogène, i.e $N(s, t) \sim PPNH(\lambda_\alpha(s, t))$. Les fonctions d'échelle σ_β et d'intensité λ_α sont limitées à des polynômes d'ordre un et deux telles que :

$$\left. \begin{aligned} \sigma_\beta(s, t) &= \exp(\beta_0 + \beta_1 t + \beta_2 t^2) \\ \lambda_\alpha(s, t) &= \alpha_0 + \alpha_1 t + \alpha_2 t^2 \\ \gamma_u(s, t) &= \gamma_0 \end{aligned} \right\} s \in S \subset \mathbb{R}^2.$$

Le paramètre de forme $\gamma_u(s)$ et le seuil $u(s)$ sont pris constants afin de ne pas augmenter les incertitudes dans les estimations, mais cela n'est pas restrictif, car nous pouvons les considérer comme des fonctions du temps. Les vecteurs paramètres β , α et la paramètre de forme $\gamma_u(s)$ sont estimés en utilisant la méthode du maximum de vraisemblance grâce aux lois *POT* et *PP* non-stationnaires. Le seuil $u(s)$ est estimé en couplant la méthode des quantiles et la méthode Mean Excess Plot. Les niveaux de retour non-stationnaires ponctuels sont alors évalués en utilisant la méthode *ENE* (Expected Number of Events) dans une approche *POT-PPNH* qui combine la loi des dépassements de seuil et celle des occurrences des événements extrêmes, i.e que le niveau de retour $z_\tau(s)$ est solution de cette équation :

$$\sum_{t=1}^{\tau} \left[\left(1 + \frac{\gamma_u(s)}{\sigma_\beta(s, t)} (z_\tau(s) - u(s)) \right)^{-1/\gamma_u(s)} \lambda_\alpha(s, t) \right] = 1$$

Une extrapolation des meilleures tendances identifiées via le test du rapport de vraisemblance, nous ont permis de calculer des niveaux de retour non stationnaires en résolvant cette équation implicite à l'aide d'un algorithme numérique. La structure spatiale de ces niveaux de retour non-stationnaires calculés aux stations est évaluée en utilisant la méthode du krigeage spatial. Les résultats ont montré qu'il

n'y a pas de différence significative entre les niveaux de retour non-stationnaires des températures maximales d'une période à l'autre, tout simplement parce que nous avons estimé un paramètre de forme négatif sur la plupart de nos stations d'observations. En pareille situation, cela signifie que la queue de distribution des températures maximales est bornée, du coup ça bouge assez lentement dans les extrêmes. Par contre nous avons décelé une différence significative entre les niveaux de retour non-stationnaires des températures minimales d'une période de temps à l'autre, et il faudrait s'attendre à des nuits de plus en plus chaudes sur la région dans les prochaines années. En ce qui concerne les précipitations, nous n'avons pas détecté de tendance significative ni de structure spatiale cohérente. Les niveaux de retour estimés des précipitations sont alors plus ou moins entachés d'une certaine incertitude au regard des intervalles de confiance assez larges. Cela met en évidence la difficulté de certaines approches ponctuelles à prendre en compte les non-stationnarités des précipitations, d'où le développement des approches spatio-temporelles entreprises dans la suite de la thèse.

Le chapitre 3 propose un modèle couplant une approche POT ponctuelle paramétrique non-stationnaire aux dépassements de seuils spatiaux. Dans ce modèle nous adoptons les techniques de modélisation décrites dans le chapitre 2 tout en modélisant la structure de dépendance spatiale par un processus ℓ -Pareto standard au lieu de l'évaluer en utilisant la méthode du krigeage spatial. On considère $Z = \{Z(s, t)\}_{s \in S, t \in T}$, un processus stochastique spatio-temporel non-stationnaire défini sur $\mathcal{C}_+(S \times T)$, la classe des fonctions continues et positives sur $S \times T \subset \mathbb{R}^d \times \mathbb{R}^+$ muni de la norme uniforme $\|\cdot\|_\infty$, sous l'hypothèse que Z est à variation régulière, de sorte que :

$$nP \left(a_n^{-1}(s, t) Z(s, t) \in A \right) \longrightarrow \Lambda(A), \quad n \longrightarrow +\infty,$$

où $a_n : \mathcal{C}_+(S \times T) \longrightarrow \mathbb{R}^+$ est une suite de fonctions continues et Λ est une mesure sur $\mathcal{C}_+(S \times T)$ et homogène d'ordre -1 , i.e $\Lambda(tA) = t^{-1}\Lambda(A)$, pour tout $t > 0$. Pratiquement, la non-stationnarité des données observées est retirée en transformant les marginales en Pareto unité ou en Frechet unité via la relation suivante :

$$Z^*(s, t) = \frac{1}{1 - F_{Z(s, t)}(Z(s, t))}, \text{ pour tout } s \in S \text{ and } t \in T$$

Etant donné un seuil $u > 0$ suffisamment grand et une fonction de risque $\ell : \mathcal{C}_+(S \times T) \longrightarrow [0, +\infty[$ continue positive et homogène, i.e $\ell(tZ) = t\ell(Z)$, $t > 0$, la loi conditionnelle des ℓ -dépassements de Z^* converge asymptotique dans $\mathcal{C}_+(S \times T)$ vers un processus stochastique non dégénéré $W_{\ell, \gamma, \Lambda}^*$ appelé processus ℓ -Pareto standard, i.e :

$$P \left(u^{-1} Z^* \in \cdot \mid \{\ell(Z^*) > u\} \right) \longrightarrow P \left(W_{\ell, \gamma, \Lambda}^* \in \cdot \right), \quad u \longrightarrow +\infty.$$

La décomposition pseudo-polaire de $W_{\ell, \gamma, \Lambda}^*$ est donnée par : $W_{\ell, \gamma, \Lambda}^* = R_\gamma \frac{Q_\Lambda}{\ell(Q_\Lambda)}$ où R_γ est une variable aléatoire de Pareto unité d'indice de queue γ et Q_Λ un processus stochastique de mesure spectral Λ défini sur S et à valeurs dans le simplexe unité $\mathcal{S}_\ell = \{f \in \mathcal{C}_+(S \times T) : \ell(f) \geq 1, \|f\|_1 = 1\}$. C'est donc sur le processus stochastiques Q_Λ que nous drainons la structure de dépendance spatiale et devrait être

1.C. ASYMPTOTIC DISTRIBUTION OF ℓ -PARETO PROCESS

choisi avec soin. Ainsi le choix du processus stochastique Q_Λ avec une distribution qui se situe dans une famille paramétrique dont les paramètres peuvent être déduits, conduit à l'estimation des paramètres de la structure de dépendance spatiale. Parmi les modèles admissibles nous pouvons choisir les processus log-gaussien de semi-variogramme ν , le processus t -extrémal de fonction de covariance ρ et de degré de liberté $df = \nu$, ou tout autre modèle spectral associé aux processus max-stables. Dans cette étude, nous nous sommes placés dans le cadre du modèle Brown Resnick avec $Q_\Lambda(s) = \exp(G(s) - \nu(s))$, où G est un processus Gaussien à incrément stationnaire de semi-variogramme ν . La dépendance spatiale est ajustée par une fonction d'extrémogramme $\pi(h) = \lim_{q \rightarrow 1} P[Z^*(s+h) > u_q \mid \{Z^*(s) > u_q, \ell(Z^*) \geq u_q\}]$. Les paramètres de la structure de dépendance spatiale sont alors estimés sur ces données standardisées Z^* , en utilisant la méthode du score de gradient ou la vraisemblance censurée. Les résultats d'ajustements de la dépendance spatiale, montrent une forte dépendance dans les extrêmes des températures minimales et maximales lorsque la distance $h \leq 250km$ avec une probabilité $\pi(h) \geq 0.25$. Par contre lorsque la distance $h > 250km$, la dépendance spatiale diminue de façon progressive avec une probabilité $0.125 \leq \pi(h) < 0.25$. Nous sommes en mesure de simuler le modèle ℓ -Pareto obtenu sur une grille fine couvrant le domaine d'étude et ramener ces simulations à l'échelle des données non-stationnaires grâce à la fonction inverse de la stationnarisation et les tendances marginales identifiées pour une horizon de temps t donné. Pour cela nous commençons par subdiviser notre région d'étude en des sous-régions homogènes grâce à la méthode des k plus proches voisins, avec $k = 1$ et nous faisons l'hypothèse que la tendance identifiée sur une station $s \in S$ appartenant une sous-région homogène est constante sur cette région. Ainsi à partir des données simulées, nous déduisons maintenant les niveaux de retour non-stationnaires en utilisant la méthode ENE , i.e z_τ est une valeur unique de sorte que le nombre moyen de dépassements soit égal à un :

$$1 = \sum_{t=1}^{\tau} P(Z_t(s) > z_\tau(s)), \quad s \in \mathcal{G}.$$

où $P(Z_t(s) > z_\tau(s))$ est approximé de manière empirique à partir simulations du processus Z . Ainsi sur chaque point $s_j \in \mathcal{G}$, $j = 1, \dots, nGrid$, le niveau de retour $z_\tau(s_j)$ est déterminé par approximation et est solution de l'équation suivante :

$$1 = \frac{1}{N} \sum_{t=1}^{\tau} \sum_{i=1}^N \mathbb{1}_{\{\tilde{Z}_{i,t}(s_j) > z_\tau(s_j)\}}, \quad j = 1, \dots, nGrid.$$

où $\tilde{Z}_{i,t}$ représente les simulations du processus spatio-temporel $Z_t(s)$ pour tout pas de temps $t = 1, \dots, \tau$, N le nombre de simulation et $nGrid$ le nombre de points de la grille $\mathcal{G}$. Les cartes des niveaux non-stationnaires obtenues, montrent que pour Tmax, il n'y a pas de différence significative entre les niveaux de retour non-stationnaires à $\tau = 2$ ans et à $\tau = 10$ ans, pour la même raison mentionnée dans chapitre 2(paramètre de forme $\gamma < 0$). Par contre, les niveaux de retour non-stationnaires des températures minimales ont une tendance à la hausse, et comme conséquence, il faudrait s'attendre à des nuits de plus en plus chaudes dans les prochaines années sur la région. De plus, il faut ajouter que pour Tmin et Tmax,

les niveaux de retour marginaux non stationnaires estimés en tenant compte de la structure de dépendance spatiale sont similaires, voire identiques, aux niveaux de retour estimés aux stations. Mais les incertitudes associées ne sont pas du tout les mêmes en dehors des stations. En d'autres termes, notre modèle quantifie mieux les insertions spatio-temporelles et estime de manière fiable les quantiles extrêmes non stationnaires contrairement au krigeage spatial mis en œuvre dans le chapitre 2. Pour des raisons de temps de calculs, nous n'avons considéré dans notre analyse que des horizons de temps relativement courts. Mais nous sommes en train de travailler sur notre programme de calculs afin de l'optimiser, ce qui devrait nous permettre sans doute d'envisager des périodes de temps assez longues et de faire une comparaison assez approfondie des résultats avec ceux obtenus dans la deuxième partie.

Le chapitre 4, présente une contribution majeure à la statistique des extrêmes spatio-temporels. Dans cette méthodologie, la non-stationnarité dans l'espace et dans le temps est traitée par une fonction de scédasticité introduite dans le cadre des "extrêmes hétéroscédastique". Cette fonction de scédasticité relie la distribution de queue du processus spatio-temporel non-stationnaire $\{X_t(s)\}_{s \in S, t \in T}$ à un processus spatial latent $\{Z(s)\}_{s \in S}$. Nous supposons que les distributions marginales de Z partagent le même point d'extrémité droit commun x^* que les distributions marginales de X de sorte que la condition de queue proportionnelle suivante soit vérifiée :

$$\lim_{x \rightarrow x^*} \frac{P(X_t(s_j) > x)}{P(Z(s_j) > x)} = c_\theta \left(\frac{t}{n}, s_j \right) \quad \text{with} \quad \int_0^1 c_\theta(u, s_j) du = 1,$$

où $c_\theta : [0, 1] \times S \rightarrow]0, \infty[$ est une fonction supposée continue et positive appelée fonction de tendance ou de scédasticité. Elle évalue de façon conjointe les tendances des événements extrêmes dans l'espace et dans le temps. De plus les marginales de Z sont supposées appartenir au domaine d'attraction d'une distribution des valeurs extrêmes généralisée G_γ d'indice de queue γ , i.e qu'il existe des fonctions de normalisations $a_n(\cdot) > 0$ et $b_n(\cdot)$ de sorte que les pseudo-observations du processus spatial latent Z se déduisent à partir des observations du processus spatio-temporel non-stationnaire X à partir de la relation suivante :

$$Z_t(s_j) = \left\{ c_\theta \left(\frac{t}{n}, s_j \right) \right\}^{-\gamma} \left[X_t(s_j) - \frac{\left\{ c_\theta \left(\frac{t}{n}, s_j \right) \right\}^\gamma - 1}{\gamma} (a_{Z(s_j)} - \gamma b_{Z(s_j)}) \right],$$

$j = 1, \dots, m ; t = 1, \dots, n$

Les paramètres du processus spatio-temporel non-stationnaire et le processus spatial latent sont reliés entre eux par les expressions suivantes :

$$\begin{cases} b_{t,s_j} \approx b_{Z(s_j)} + a_{Z(s_j)} \frac{c_\theta(t/n, s_j)^\gamma - 1}{\gamma} \\ a_{t,s_j} \approx a_{Z(s_j)} c_\theta(t/n, s_j)^\gamma \end{cases}$$

où b_{t,s_j} et a_{t,s_j} sont les paramètres marginaux de X pour chaque pas de temps t et pour tout $s_j \in S$. L'idée dans cette modélisation consiste à ajuster un processus

1.C. ASYMPTOTIC DISTRIBUTION OF ℓ -PARETO PROCESS

ℓ -Pareto au processus spatial latent Z , et se ramener au processus spatio-temporel non-stationnaire via les transformations inverses. Pour cela on estime d'abord les paramètres marginaux et ceux de la structure dépendance. Du point de vu marginal, le paramètre θ de la fonction de scédasticité est estimé en utilisant la méthode du maximum de vraisemblance sous l'hypothèse que les dates des dépassements de seuil $b_Z(s_j)$ ont pour fonction densité de probabilité c_θ . Par contre dans notre analyse $b_Z(s_j)$ est fixé au quantile d'ordre $1 - \alpha$ de l'ensemble des observations de $X_{\cdot, s_j}$, i.e $b_Z(s_j) = q_{1-\alpha}\{X_{\cdot, s_j}\}$ avec $\alpha \in]0, 1[$. Les paramètres $a_Z(s_j)$ et γ sont estimés en utilisant la log-vraisemblance indépendance tout en tirant profit des relations qui existent entre les paramètres du processus spatio-temporel non-stationnaire et le processus spatial latent en maximisant cette relation :

$$\mathcal{L}(\gamma, a_Z(s_j)) = \sum_{t,j=1}^{n,m} \mathbb{1}_{\{x_t(s_j) \geq b_Z(s_j)\}} \log \left[\frac{1}{a_{t,s}} \left\{ 1 + \gamma \frac{x_t(s_j) - b_{t,s_j}}{a_{t,s_j}} \right\}_+^{-1/\gamma-1} \right]$$

Comme dans le chapitre 3, les paramètres de la structure de dépendance spatiale du processus spatial latent sont estimés en se plaçant dans le cadre du modèle Brown Resnick tout en ajustant cette structure de dépendance par une fonction d'extrémogramme. La méthode du score du gradient est mise à profit pour estimer les paramètres sur des marginales *GPD* standard de Z . Dans ce cadre de modélisation, nous avons étendu la notion du niveau de retour ponctuel *ENE* aux processus stochastiques spatio-temporels. Autrement dit, nous avons défini une surface de retour non-stationnaire, comme étant une fonction paramétrique $f_\tau : S \rightarrow \mathbb{R}$ de sorte que f_τ soit solution de ce problème d'optimisation

$$f_\tau = \operatorname{argmin}_{f \in \mathcal{C}(S)} \left(1 - \sum_{t=1}^{\tau} P(X_t > f) \right)^2$$

Ainsi grâce à la relation entre le processus spatio-temporel non-stationnaire et le processus spatial latent, puis l'approximation du processus spatial latent par un processus ℓ -Pareto $W_{\ell, \gamma, \Lambda}$ qu'on sait simuler, nous sommes en mesure d'approximer la probabilité $P(X_t > f)$, i.e

$$\begin{aligned} \mathbb{P}(X_t(s) > f(s)) &= P(Z(s) > \varphi_t(f)) \\ &= P\left(Z(s) > \varphi_t(f) | \ell(Z) > u\right) P(\ell(Z) > u) \\ &\approx P\left(W_{\ell, \gamma, \Lambda}(s) > \varphi_t(f)\right) P(\ell(Z) > u) \end{aligned}$$

ou $\varphi_t(f) = \varphi_t(f, c_\theta, a_Z, b_Z, \gamma, s)$, $P(W_{\ell, \gamma, \Lambda}(s) > \varphi_t(f))$ est approximée de manière empirique en utilisant les simulations de $W_{\ell, \gamma, \Lambda}$ avec $P(W_{\ell, \gamma, \Lambda}(s) > \varphi_t(f)) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{W^i > \varphi_t(f)}$ et la probabilité $P(\ell(Z) > u)$ est aussi estimée empiriquement en grâce aux pseudo-observations de Z . Etant donnée une fonction paramétrique $f_{\tau, \beta}$, nous déduisons une surface de retour non-stationnaire en optimisant le problème suivant grâce aux simulations :

$$f_{\tau, \beta} = \operatorname{argmin}_{\beta} \left(1 - \sum_{t=1}^{\tau} \hat{P}(\ell(Z) > u) \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{W^i > \varphi_t(f_\beta)} \right)^2$$

Cette surface $f_{\tau,\beta}$ de retour non-stationnaire s'interprète comme étant une surface susceptible d'être dépassée par le processus au moins une fois sur toute la région d'étude pour une période de temps τ donné. Cet outil original est appliqué à des données de températures maximales quotidiennes du Burkina Faso.

Le dernier chapitre 5 est un article publié dans "International Journal of Mathematics and Mathematical Sciences". Cet article combine deux approches statistiques de la théorie des valeurs extrêmes à savoir l'approche des extrêmes hétéroscédastiques et l'approche généralisé des dépassements de seuil spatiaux. La non-stationnarité est traitée comme dans le chapitre 4, mais en changeant certaines hypothèses tandis que la dépendance spatiale est maintenant modélisé par un processus ℓ -Pareto généralisé au lieu de l'ajuster par un processus ℓ -Pareto standard. L'hypothèse de la variation régulière simple faite sur Z est maintenant remplacée par une variation régulière fonctionnelle, i.e

$$nP \left[\left(1 + \gamma \left(\frac{Z - b_n}{a_n} \right) \right)^{1/\gamma} \in \cdot \right] \rightarrow \Lambda(\cdot), n \rightarrow +\infty,$$

où Λ est une mesure non nulle sur $\mathcal{C}_+(S)$ telle que, $\Lambda(tA) = t^{-1}\Lambda(A)$, $A \subset \mathcal{C}_+(S)$. L'hypothèse d'homogénéité sur ℓ est également étendue à toute fonction de risque. Sous des hypothèses minimales sur la fonction de risque ℓ et pour n suffisamment grand, la distribution conditionnelle des ℓ -dépassements pour un certain seuil $u \geq 0$, du processus normalisé peut être approximée par un processus ℓ -Pareto généralisé, c'est-à-dire que pour certaines fonctions $a_n(\cdot) > 0$ et $b_n(\cdot) \in \mathcal{R}$,

$$P \left\{ \left\lfloor \frac{Z - b_n}{\ell(a_n)} \right\rfloor \in A \mid \ell \left(\frac{Z - b_n}{\ell(a_n)} \right) \geq u \right\} \rightarrow P \{W_\ell \in A\}, n \rightarrow \infty,$$

où W_ℓ est un processus stochastique non dégénéré défini sur S et appartenant à la famille des processus ℓ -Pareto généralisé d'indice de queue γ . Le processus ℓ -Pareto généralisé W_ℓ a des marginales *GPD* pour tout $s_0 \in S$ et est défini par

$$W_\ell = \begin{cases} a(Y_\ell^\gamma - 1)/\gamma + b, & \gamma \neq 0 \\ a \log Y_\ell + b, & \gamma = 0 \end{cases}$$

où $a > 0$ est une fonction d'échelle, b la fonction de localisation et Y_ℓ un processus stochastique ℓ -Pareto standard à variation régulière sur $\mathcal{C}_+(S)$ dont la mesure de probabilité est complètement caractérisée par la mesure limite Λ . Les paramètres marginaux et de la structure de dépendance sont alors estimés comme précédemment. Nous avons réadapté les définitions des niveaux de retour ponctuel non-stationnaire *ENE* et *EWT* en introduisant la fonction de scédasticité et la loi des ℓ -dépassements, i.e que pour un niveau x_τ , $P(X_t(s) > x_\tau) \approx c_\theta \left(\frac{t}{n}, s \right) P(Z(s) > x) P(Z(s) > x_\tau \mid Z(s) > x)$, $x < x_\tau$. Dans ce cas, le niveau de retour non-stationnaire x_τ vérifié respectivement les équations suivantes :

$$ENE : 1 = \sum_{t=t_0}^{t_0+n_x\tau} \left\{ c_\theta \left(\frac{t}{n}, s \right) \phi_x(s) \bar{F}_{Z(s)|Z(s)>x}(x_\tau - x) \right\},$$

$$EWT : n_x \tau = 1 + \sum_{i=1}^{\infty} \prod_{t=1}^i \left\{ 1 - c_{\theta}\left(\frac{t}{n}, s\right) \phi_x(s) \bar{F}_{Z(s)|Z(s)>x}(x_{\tau} - x) \right\},$$

La relation entre le processus spatio-temporel et le processus spatial latent a également été mise à profit pour estimer les quantiles extrêmes (QE) non-stationnaires.

$$QE : x_{\tau}(s) = z_{\tau}(s) \hat{c}_{\theta}(t_{\tau}, s)^{\hat{\gamma}} + \frac{\hat{c}_{\theta}(t_{\tau}, s)^{\hat{\gamma}} - 1}{\hat{\gamma}} (\hat{a}_n - \hat{\gamma} \hat{b}_n), \quad s \in S$$

Cette nouvelle méthode est illustrée sur des données de précipitations journalières au Burkina Faso. Ce qui nous a permis de mettre en évidence les récentes tendances des précipitations extrêmes au Burkina Faso. Les résultats obtenus sont très intéressants et particulièrement importants pour appréhender l'influence des changements climatiques sur les événements météorologiques extrêmes au Burkina-Faso et toute la région d'Afrique Sub-Sahélienne. En perspective, nous comptons étendre notre méthodologie à d'autres techniques d'homogénéisation, intégrer en plus de la non-stationnarité de queue, des formes de structures de dépendances spatiales non-stationnaires et l'appliquer à d'autres types de données un peu plus large. Nous envisageons également d'étendre les autres définitions du niveau de retour ponctuel décrites dans le chapitre 2 aux processus stochastiques spatio-temporels. Pour terminer ces perspectives nous regarderons l'impact du changement climatique sur les rendements agricoles en Afrique Sub-sahélienne, cas du Burkina Faso.

Bibliographie

- F. Acero, J. A. García, M. C. Gallego, S. Parey, et D. Dacunha-Castelle. Trends in summer extreme temperatures over the iberian peninsula using nonurban station data. *Journal of Geophysical Research : Atmospheres*, 119(1) :39–53, 2014.
- R. Adler. The geometry of random fields. *The Geometry of Random Fields*, 1981.
- A. Alhassane, S. Salack, M. Ly, I. Lona, S. B. Traoré, et B. Sarr. Évolution des risques agroclimatiques associés aux tendances récentes du régime pluviométrique en afrique de l’ouest soudano-sahélienne. *Science et changements planétaires/Sécheresse*, 24(4) :282–293, 2013.
- A.-N. Amenah, P. Kevei, et L. Viharos. *Estimation of Tail Indices of Heavy-Tailed Distributions with Application Ph. D. thesis*. PhD thesis, University of Szeged, Bolyai Institute, 2021.
- A. Arns, T. Wahl, I. Haigh, J. Jensen, et C. Pattiaratchi. Estimating extreme water level probabilities : A comparison of the direct methods and recommendations for best practise. *Coastal Engineering*, 81 :51–66, 2013.
- P. Asadi, A. C. Davison, S. Engelke, et al. Extremes on river networks. *Annals of Applied Statistics*, 9(4) :2023–2050, 2015.
- S. Aulbach et M. Falk. Local asymptotic normality in δ -neighborhoods of standard generalized pareto processes. *Journal of Statistical Planning and Inference*, 142(6) :1339–1347, 2012.
- A. A. Balkema et L. De Haan. Residual life time at great age. *The Annals of probability*, pages 792–804, 1974.
- F. Ballani et M. Schlather. A construction principle for multivariate extreme value distributions. *Biometrika*, 98(3) :633–645, 09 2011.
- A. Bechler, L. Bel, et M. Vrac. Conditional simulations of the extremal t process : application to fields of extreme precipitation. *Spatial statistics*, 12 :109–127, 2015.
- J. Beirlant, Y. Goegebeur, J. Segers, et J. Teugels. *Statistics of Extremes : Theory and Applications*. Wiley, 2004.
- L. Belzile et al. *mev : Modelling Extreme Values*, 2022. R package version 1.14.
- B. Beranger, S. A. Padoan, et S. A. Sisson. Models for extremal dependence derived from skew-symmetric families. *Scandinavian Journal of Statistics*, 44(1) :21–45, 2017.
- B. Beranger, A. Stephenson, et S. Sisson. High-dimensional inference using the extremal skew-t process. *Extremes*, pages 1–33, 2020.

- E. Bernard, P. Naveau, M. Vrac, et O. Mestre. Clustering of maxima : Spatial dependencies among heavy rainfall in france. *Journal of climate*, 26(20) :7929–7937, 2013. <https://doi.org/10.1175/JCLI-D-12-00836.1>.
- P. Billingsley. *Probability and Measure*. John Wiley and Sons, Inc, New-York, third edition. ISBN 978-8126517718., 1995.
- N. Bingham. Regular variation. *Encyclopedia of Mathematics and its Applications*, 27, 1987.
- J. Blanchet, C. Aly, T. Vischel, G. Panthou, Y. Sané, et M. Diop Kane. Trend in the co-occurrence of extreme daily rainfall in west africa since 1950. *Journal of Geophysical Research : Atmospheres*, 123(3) :1536–1551, 2018.
- B. M. Brown et S. I. Resnick. Extreme values of independent stochastic processes. *Journal of Applied Probability*, pages 732–739, 1977.
- T. Buishand, L. De Haan, C. Zhou, *et al.* On spatial extremes : with application to a rainfall problem. *The Annals of Applied Statistics*, 2(2) :624–642, 2008.
- R. Cabral, A. Ferreira, et P. Friederichs. Space–time trends and dependence of precipitation extremes in north-western germany. *Environmetrics*, 31(3) :e2605, 2020.
- G. Chagnaud, G. Panthou, T. Vischel, et T. Lebel. A synthetic view of rainfall intensification in the west african sahel. *Environmental Research Letters*, 17(4) : 044005, 2022.
- L. Cheng, A. AghaKouchak, E. Gilleland, et R. W. Katz. Non-stationary extreme value analysis in a changing climate. *Climatic change*, 127(2) :353–369, 2014.
- C. Chevalier, O. Martius, et D. Ginsbourger. Modeling non-stationary extreme dependence with stationary max-stable processes and multidimensional scaling. *Journal of Computational and Graphical Statistics*, pages 1–29, 2020.
- S. Coles, J. Heffernan, et J. Tawn. Dependence measures for extreme value analyses. *Extremes*, 2(4) :339–365, 1999.
- S. Coles, J. Bawa, L. Trenner, et P. Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.
- S. G. Coles et J. A. Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society : Series B (Methodological)*, 53(2) :377–392, 1991.
- D. Cooley. Return periods and return levels under climate change. In *Extremes in a changing climate*, pages 97–114. Springer, 2013.
- D. Dacunha-Castelle, T. T. H. Hoang, et S. Parey. Modeling of air temperatures : preprocessing and trends, reduced stationary process, extremes, simulation. *Journal de la Société Française de Statistique*, 156(1) :138–168, 2015.

BIBLIOGRAPHIE

- M. N. Dalhoumi. *Sur l'estimation de probabilités de queues multivariées*. PhD thesis, Université Montpellier, 2017.
- R. A. Davis, T. Mikosch, et Y. Zhao. Measures of serial extremal dependence and their estimation. *Stochastic Processes and their Applications*, 123(7) :2575–2602, 2013.
- A. C. Davison et R. L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society : Series B (Methodological)*, 52(3) :393–425, 1990.
- A. C. Davison, S. A. Padoan, M. Ribatet, et al. Statistical modeling of spatial extremes. *Statistical science*, 27(2) :161–186, 2012.
- A. P. Dawid, M. Musio, et L. Ventura. Minimum scoring rule inference. *Scandinavian Journal of Statistics*, 43(1) :123–138, 2016.
- R. de Fondeville et A. Davison. Functional peaks-over-threshold analysis, 2020. Preprint.
- R. de Fondeville et A. C. Davison. High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3) :575–592, 2018.
- L. de Haan et A. Ferreira. Extreme value theory : An introduction, 2006.
- L. de Haan, A. K. Tank, et C. Neves. On tail trend detection : modeling relative risk. *Extremes*, 18(2) :141–178, 2015.
- L. De Haan et al. A spectral representation for max-stable processes. *The annals of probability*, 12(4) :1194–1204, 1984.
- A. L. Dekkers, J. H. Einmahl, et L. De Haan. A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, pages 1833–1855, 1989.
- A. Deme, S. Galle, et F. Hourdin. Les projections du climat en afrique de l'ouest : évidences et incertitudes. *Les sociétés rurales face aux changements climatiques et environnementaux en Afrique de l'Ouest*, Editions IRD, Paris, France, 2015.
- B. Diakarya, B. Koté, et S. Moussa. Spatial stochastic framework for sampling time parametric max-stable processes. *International Journal of Statistics and Probability*, 1(2) :203, 2012.
- B. Diakarya, N. S. P. Clovis, et D. Moumouni. Asymptotic dependence modeling for spatio-temporal max-stable processes. *European Journal of Pure and Applied Mathematics*, 10(5) :1035–1049, 2017.
- B. Diallo. *Perceptions endogènes, analyses agroclimatiques et stratégies d'adaptation aux variabilités et changements climatiques des populations dans trois zones climatiques du Burkina Faso*. PhD thesis, Département Formation et Recherche, Comité permanent inter-états de lutte ... , 2010.

- C. Dombry et M. Ribatet. Functional regular variations, pareto processes and peaks over threshold. *Statistics and its Interface*, 8(1) :9–17, 2015.
- M. G. Donat, A. L. Lowry, L. V. Alexander, P. A. O’Gorman, et N. Maher. More extreme precipitation in the world’s dry and wet regions. *Nature Climate Change*, 6(5) :508–513, 2016.
- S. Dossou-Gbete, B. Some, et D. Barro. Modlling the dependence of parametric bivariate extreme value copulas. *Asian Netwrk for Scientific Information*, 2(3) : 41–54, 2009. DOI : 10.3923/ajms.2009.41.54.
- W. H. DuMouchel. Estimating the stable index α in order to measure tail thickness : a critique. *the Annals of Statistics*, pages 1019–1031, 1983.
- E. F. Eastoe et J. A. Tawn. Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 58(1) :25–45, 2009.
- E. F. Eastoe et J. A. Tawn. Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika*, 99(1) :43–55, 2012.
- J. H. J. Einmahl, L. de Haan, et C. Zhou. Statistics of heteroscedastic extremes. *J. R. Stat. Soc., Ser. B, Stat. Methodol.*, 78(1) :31–51, 2016.
- P. Embrechts, C. Kluppelberg, et T. Mikosch. *Modelling extremal events for insurance and finance*. Springer, Berlin, 1997.
- P. Embrechts, C. Klüppelberg, et T. Mikosch. *Modelling extremal events : for insurance and finance*, volume 33. Springer Science & Business Media, 2013.
- S. Engelke, A. Malinowski, Z. Kabluchko, et M. Schlather. Estimation of hüsler–reiss distributions and brown–resnick processes. *Journal of the Royal Statistical Society : Series B : Statistical Methodology*, pages 239–265, 2015.
- S. Engelke, R. de Fondeville, et M. Oesting. Extremal behaviour of aggregated data with an application to downscaling. *Biometrika*, 106(1) :127–144, 2019.
- M. Falk, J. Hüsler, et R.-D. Reiss. *Laws of small numbers : extremes and rare events*, 1994.
- C. Faye. Caractéristiques spatio-temporelles de la pluviométrie et de l’écoulement extrême dans un contexte de changements climatiques en afrique de l’ouest : l’exemple du sénégal. *LJEE*, 32 & 33, 2018.
- A. Ferreira, L. de Haan, et C. Zhou. Exceedance probability of the integral of a stochastic process. *Journal of Multivariate Analysis*, 105(1) :241–257, 2012.
- A. Ferreira, L. De Haan, *et al.* The generalized pareto process ; with a view towards application and simulation. *Bernoulli*, 20(4) :1717–1737, 2014.

BIBLIOGRAPHIE

- A. Ferreira, P. Friederichs, L. de Haan, C. Neves, et M. Schlather. Estimating space-time trend and dependence of heavy rainfall. *arXiv preprint arXiv :1707.04434*, 2017.
- R. A. Fisher et L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2) :180–190, 1928.
- J. Galambos. *The asymptotic theory of extreme order statistics*. Wiley, 1978.
- J. Gaume, N. Eckert, G. Chambon, M. Naaim, et L. Bel. Mapping extreme snowfalls in the french alps using max-stable processes. *Water Resources Research*, 49(2) : 1079–1098, 2013.
- A. E. Gelfand et E. M. Schliep. Spatial statistics and gaussian processes : A beautiful marriage. *Spatial Statistics*, 18 :86–104, 2016.
- E. Giné, M. G. Hahn, et P. Vatan. Max-infinitely divisible and max-stable sample continuous processes. *Probability theory and related fields*, 87(2) :139–165, 1990.
- B. Gnedenko. Sur la distribution limite du terme maximum d’une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.
- T. Gneiting et A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477) :359–378, 2007.
- I. Grabemann et R. Weisse. Climate change impact on extreme wave conditions in the north sea : an ensemble study. *Ocean Dynamics*, 58(3-4) :199–212, 2008.
- R. Grotjahn, R. Black, R. Leung, M. F. Wehner, M. Barlow, M. Bosilovich, A. Gershunov, W. J. Gutowski, J. R. Gyakum, R. W. Katz, *et al.* North american extreme temperature events and related large scale meteorological patterns : a review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics*, 46(3) :1151–1184, 2016.
- Z. Hao, V. P. Singh, et F. Hao. Compound extremes in hydroclimatology : a review. *Water*, 10(6) :718, 2018.
- Y. Hu, Z. Liang, V. P. Singh, X. Zhang, J. Wang, B. Li, et H. Wang. Concept of equivalent reliability for estimating the design flood under non-stationary conditions. *Water resources management*, 32(3) :997–1011, 2018.
- H. Hult et F. Lindskog. Extremal behavior of regularly varying stochastic processes. *Stochastic Processes and their applications*, 115(2) :249–274, 2005.
- R. Huser et A. Davison. Space—time modelling of extreme events. *Journal of the Royal Statistical Society : Series B : Statistical Methodology*, pages 439–461, 2014.
- R. Huser et A. C. Davison. Composite likelihood estimation for the brown–resnick process. *Biometrika*, 100(2) :511–518, 2013.

- R. Huser et M. G. Genton. Non-stationary dependence structures for spatial extremes. *Journal of agricultural, biological, and environmental statistics*, 21(3) : 470–491, 2016a.
- R. Huser et M. G. Genton. Non-stationary dependence structures for spatial extremes. *Journal of agricultural, biological, and environmental statistics*, 21(3) : 470–491, 2016b.
- A. Hyvärinen. Some extensions of score matching. *Computational statistics & data analysis*, 51(5) :2499–2512, 2007.
- A. Hyvärinen et P. Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.
- B. Ibrahim, H. Karambiri, J. Polcher, H. Yacouba, et P. Ribstein. Changes in rainfall regime over burkina faso under the climate change conditions simulated by 5 regional climate models. *Climate Dynamics*, 42(5) :1363–1381, 2014.
- IPCC. *Climate Change 2014 : Impacts, Adaptation, and Vulnerability. Part B :Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014.
- Y. IPCC, Lautre. 6e rapport du giec : Accélérer l’action climatique, c’est possible. 2022.
- A. Janßen et P. Wan. k -means clustering of extremes. *Electronic Journal of Statistics*, 14(1) :1211–1233, 2020. DOI : 10.1214/20-EJS1689.
- A. F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348) :158–171, 1955.
- C. Jiang, L. Xiong, L. Yan, J. Dong, et C.-Y. Xu. Multivariate hydrologic design methods under nonstationary conditions and application to engineering practice. *Hydrology and Earth System Sciences*, 23(3) :1683–1704, 2019.
- H. Joe, R. L. Smith, et I. Weissman. Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society : Series B (Methodological)*, 54(1) :171–183, 1992.
- Z. Kabluchko, M. Schlather, L. De Haan, et al. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5) :2042–2065, 2009.
- O. Kallenberg. Large deviations. In *Foundations of Modern Probability*, pages 537–560. Springer, 2002.

BIBLIOGRAPHIE

- S. Khan, G. Kuhn, A. R. Ganguly, D. J. Erickson III, et G. Ostrouchov. Spatio-temporal variability of daily and weekly precipitation extremes in south america. *Water Resources Research*, 43(11), 2007.
- C. Klüppelberg et S. I. Resnick. The pareto copula, aggregation of risks, and the emperor's socks. *Journal of Applied Probability*, 45(1) :67–84, 2008.
- M. Leadbetter. Extremes and local dependence in stationary sequences, z. wahr-scheinlichkeit, 65, 291–306, 1983.
- Q. Lejeune et F. Saeed. Étude de l'impact des changements climatiques futurs sur les ressources en eau au burkina faso. *Report produced under the project "Pro-jet d'Appui Scientifique aux processus de Plans Nationaux d'Adaptation dans les pays francophones les moins avancés d'Afrique subsaharienne"*, Climate Analytics gGmbH, Berlin, 2019.
- T. Lin et L. de Haan. On convergence toward an extreme value distribution in $c [0, 1]$. *The Annals of Probability*, 29(1) :467–483, 2001.
- F. Lindskog, S. I. Resnick, J. Roy, et al. Regularly varying measures on metric spaces : Hidden regular variation and hidden jumps. *Probability Surveys*, 11 : 270–314, 2014.
- H. Madsen, D. Lawrence, M. Lang, M. Martinkova, et T. Kjeldsen. Review of trend analysis and climate change projections of extreme precipitation and floods in europe. *Journal of Hydrology*, 519 :3634–3650, 2014.
- V. Maric. Regular variation and differential equations-introduction. *Regular Variation and Differential Equations*, 1726 :1–+, 2000.
- G. Matheron. *Traité de géostatistique appliquée*. Editions Technip, 1962.
- A. Mefleh, R. Biard, C. Dombry, et Z. Khraibani. Trend detection for heterosce-dastic extremes. *Extremes*, 23(1) :85–115, 2020.
- L. Mentaschi, M. Voudoukas, E. Voukouvalas, L. Sartini, L. Feyen, G. Besio, et L. Alfieri. The transformed-stationary approach : a generic and simplified metho-dology for non-stationary extreme value analysis. *Hydrology and Earth System Sciences*, 20(9) :3527–3547, 2016.
- L. Mouhamed, S. B. Traore, A. Alhassane, et B. Sarr. Evolution of some observed climate extremes in the west african sahel. *Weather and climate extremes*, 1 : 19–25, 2013. <https://doi.org/10.1016/j.wace.2013.07.005>.
- R. B. Nelsen. An introduction to copulas. springer, new york. *MR2197664*, 2006.
- G. Nicolet, N. Eckert, S. Morin, et J. Blanchet. Decreasing spatial dependence in extreme snowfall in the french alps since 1958 under climate change. *Journal of Geophysical Research : Atmospheres*, 121(14) :8297–8310, 2016.

- G. Nicolet, N. Eckert, S. Morin, et J. Blanchet. Assessing climate change impact on the spatial dependence of extreme snow depth maxima in the french alps. *Water Resources Research*, 54(10) :7820–7840, 2018.
- A. K. Nikoloulopoulos, H. Joe, et H. Li. Extreme value properties of multivariate t copulas. *Extremes*, 12(2) :129–148, 2009.
- F. Nkrumah, T. Vischel, G. Panthou, N. A. B. Klutse, D. C. Adukpo, et A. Diedhiou. Recent trends in the daily rainfall regime in southern west africa. *Atmosphere*, 10(12) :741, 2019. <https://doi.org/10.3390/atmos10120741>.
- J. R. Olsen, J. H. Lambert, et Y. Y. Haimes. Risk of extreme events under nonstationary conditions. *Risk Analysis*, 18(4) :497–510, 1998.
- T. Opitz. Extremal t processes : Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, 122 :409–413, 2013.
- F. Palacios-Rodríguez, G. Toulemonde, J. Carreau, et T. Opitz. Generalized pareto processes for simulating space-time extreme events : an application to precipitation reanalyses. *Stochastic Environmental Research and Risk Assessment*, 34(12) :2033–2052, 2020.
- G. Panthou. *Analyse des extrêmes pluviométriques en Afrique de l'Ouest et de leurs évolution au cours des 60 dernières années*. PhD thesis, Université de Grenoble, 2013.
- G. Panthou, T. Vischel, et T. Lebel. Recent trends in the regime of extreme rainfall in the central sahel. *International Journal of Climatology*, 34(15) :3998–4006, 2014.
- G. Panthou, T. Lebel, T. Vischel, G. Quantin, Y. Sane, A. Ba, O. Ndiaye, A. Diongue-Niang, et M. Diopkane. Rainfall intensification in tropical semi-arid regions : The sahelian case. *Environmental Research Letters*, 13(6) :064013, 2018. <http://dx.doi.org/10.1088/1748-9326/aac334>.
- S. Parey, F. Malek, C. Laurent, et D. Dacunha-Castelle. Trends and climate evolution : statistical approach for very high temperatures in france. *Climatic Change*, 81(3-4) :331–352, 2007.
- S. Parey, T. T. H. Hoang, et D. Dacunha-Castelle. Different ways to compute temperature return levels in the climate change context. *Environmetrics*, 21(7-8) :698–718, 2010.
- PAS-PNA. *Communication sur l'adaptation aux changements climatiques du Burkina Faso*. Gouvernement du Burkina Faso, 2021.
- M. D. Penrose. Semi-min-stable processes. *The Annals of Probability*, pages 1450–1463, 1992.
- J. Pickands. Statistical inference using extreme order statistics. *the Annals of Statistics*, 3(1) :119–131, 1975.

BIBLIOGRAPHIE

- J.-L. Puget, R. Blanchet, J. Salençon, et A. Carpentier. Le changement climatique. *Académie des sciences. Synthèse des débats sur le climat*, 20, 2010.
- N. Pujol, L. Neppel, et R. Sabatier. Regional tests for trend detection in maximum precipitation series in the french mediterranean region. *Hydrological Sciences Journal*, 52(5) :956–973, 2007.
- L. K. Read et R. M. Vogel. Reliability, return periods, and risk under nonstationarity. *Water Resources Research*, 51(8) :6381–6398, 2015.
- S. I. Resnick. Extreme values, regular variation and point processes. *Springer*, 1987.
- S. I. Resnick. *Heavy-tail phenomena : probabilistic and statistical modeling*. Springer, New York, 2007.
- M. Ribatet et M. Sedki. Extreme value copulas and max-stable processes. *Journal de la Société Française de Statistique*, 154(1) :138–150, 2013.
- J. Richards et J. L. Wadsworth. Spatial deformation for nonstationary extremal dependence. *Environmetrics*, 32(5) :e2671, 2021.
- J. Richards, J. A. Tawn, et S. Brown. Modelling extremes of spatial aggregates of precipitation using conditional methods. *arXiv preprint arXiv :2102.10906*, 2021.
- H. Rootzén et R. W. Katz. Design life level : quantifying risk in a changing climate. *Water Resources Research*, 49(9) :5964–5972, 2013.
- H. Rootzén, N. Tajvidi, *et al.* Multivariate generalized pareto distributions. *Bernoulli*, 12(5) :917–930, 2006.
- H. Rootzén, J. Segers, et J. L. Wadsworth. Multivariate generalized pareto distributions : Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165 :117–131, 2018a.
- H. Rootzén, J. Segers, et J. L. Wadsworth. Multivariate peaks over thresholds models. *Extremes*, 21(1) :115–145, 2018b. [http ://dx.doi.org/10.1007/s10687-017-0294-4](http://dx.doi.org/10.1007/s10687-017-0294-4).
- S. Salack, B. Sarr, S. K. Sangare, M. Ly, I. S. Sanda, et H. Kunstmann. Crop-climate ensemble scenarios to improve risk assessment and resilience in the semi-arid regions of west africa. *Climate Research*, 65 :107–121, 2015.
- S. Salack, C. Klein, A. Giannini, B. Sarr, O. N. Worou, N. Belko, J. Bliefernicht, et H. Kunstman. Global warming induced hybrid rainy seasons in the sahel. *Environmental Research Letters*, 11(10) :104008, 2016.
- J. D. Salas et J. Obeysekera. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering*, 19(3) :554–568, 2014.

- M. Sall, S. M. Tall, A. Tandian, et A. Samb. Changements climatiques, stratégies d'adaptation et mobilités. evidence à partir de quatre sites au sénégal, 2011.
- G. Salvadori, C. D. Michele, et F. Durante. On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences*, 15(11) :3293–3305, 2011.
- M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5(1) : 33–44, 2002.
- R. L. Smith. Max-stable processes and spatial extremes. *Unpublished manuscript*, 205 :1–32, 1990.
- B. Sultan, R. Lalou, A. Oumarou, M. A. Soumaré, A. M. Sanni, *et al.* *Les sociétés rurales face aux changements climatiques et environnementaux en Afrique de l'Ouest*. IRD éditions, 2017.
- M. B. Sylla, F. Giorgi, J. S. Pal, P. Gibba, I. Kebe, et M. Nikiema. Projected changes in the annual cycle of high-intensity precipitation events over west africa for the late twenty-first century. *Journal of Climate*, 28(16) :6475–6488, 2015.
- J. Tawn, R. Shooter, R. Towe, et R. Lamb. Modelling spatial extreme events with environmental applications. *Spatial Statistics*, 28 :39–58, 2018. <https://doi.org/10.1016/j.spasta.2018.04.007>.
- C. M. Taylor, D. Belušić, F. Guichard, D. J. Parker, T. Vischel, O. Bock, P. P. Harris, S. Janicot, C. Klein, et G. Panthou. Frequency of extreme sahelian storms tripled since 1982 in satellite observations. *Nature*, 544(7651) :475–478, 2017.
- E. Thibaud et T. Opitz. Efficient inference and simulation for elliptical pareto processes. *Biometrika*, 102(4) :855–870, 2015.
- T. L. Thorarinsdottir, T. Gneiting, et N. Gissibl. Using proper divergence functions to evaluate climate models. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1) :522–534, 2013.
- T. Vischel, T. Lebel, G. Panthou, G. Quantin, A. Rossi, M. Martinet, B. Sultan, R. Lalou, M. Sanni, A. Oumarou, *et al.* Le retour d'une période humide au sahel? observations et perspectives. *Les sociétés rurales face aux changements climatiques et environnementaux en Afrique de l'Ouest*, pages 43–60, 2015.
- R. Von Mises. La distribution de la plus grande de n valeurs. in (ed.), selected papers (vol. ii, pp. 271–294). providence, ri. *American Mathematical Society*, 1954.
- J. L. Wadsworth et J. A. Tawn. Efficient inference for spatial extreme value processes associated to log-gaussian random functions. *Biometrika*, 101(1) :1–15, 2014.

BIBLIOGRAPHIE

- C. Wilcox, T. Vischel, G. Panthou, A. Bodian, J. Blanchet, L. Descroix, G. Quantin, C. Cassé, B. Tanimoun, et S. Kone. Trends in hydrological extremes in the senegal and niger rivers. *Journal of hydrology*, 566 :531–545, 2018. <https://doi.org/10.1016/j.jhydrol.2018.07.063>.
- J. Xia, D. She, Y. Zhang, et H. Du. Spatio-temporal trend and statistical distribution of extreme precipitation events in huaihe river basin during 1960–2009. *Journal of Geographical Sciences*, 22(2) :195–208, 2012.
- L. Yan, L. Xiong, S. Guo, C.-Y. Xu, J. Xia, et T. Du. Comparison of four nonstationary hydrologic design methods for changing environment. *Journal of Hydrology*, 551 :132–150, 2017.
- X. Zhang, F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, et T. Nozawa. Detection of human influence on twentieth-century precipitation trends. *Nature*, 448(7152) :461–465, 2007.

Titre : Analyse de tendances dans les extrêmes dans un cadre spatio-temporel par dépassements de seuil. Application aux pluies et températures extrêmes au Burkina Faso.

Mots clés : Modèles POT fonctionnels, Extrêmes spatio-temporels, Détection de tendances, Non-stationnarités, Dépendance spatiale, Statistique Environnemental, Changements climatiques.

Résumé : Dans cette thèse nous avons étudié les tendances des événements climatiques extrêmes dans le contexte actuel de changement climatique. Les récentes évolutions des températures et des précipitations extrêmes sont d'abord évaluées en utilisant un modèle POT paramétrique de détection de tendances. Dans ce modèle paramétrique les paramètres marginaux sont des fonctions polynômes limitées aux ordres un et deux. Nous avons ensuite proposé des modèles originaux spatio-temporels qui constituent notre contribution majeure à la statistique des extrêmes spatio-temporels. Une des approches intègre des tendances paramétriques dans les distributions marginales du processus comme décrit précédemment et modélise la structure de dépendance spatiale par un processus ℓ -Pareto standard. Les autres approches développées combinent la modélisation

des extrêmes hétéroscédastiques avec l'approche des processus ℓ -Pareto généralisé. Dans ces approches, la non-stationnarité dans l'espace et le temps est gouvernée par une fonction de scédasticité ou de tendance. Cette fonction de scédasticité introduite dans le cadre des extrêmes hétéroscédastiques établit une relation entre la distribution de queue du processus non-stationnaire avec un processus spatial latent. Nous avons étendu dans ce cadre la notion du niveau de retour univarié aux processus stochastiques spatio-temporels. Ces méthodes nous permettent de simuler le processus étudié dans l'espace et le temps et d'en déduire les niveaux ou surfaces de retour non-stationnaires. Enfin les différentes méthodes proposées sont illustrées sur des données de températures et de précipitations au Burkina Faso.

Title : Trend detection in extremes : pointwise and spatial approaches by Peaks-Over-Thresholds. Application to extreme temperature and precipitation in Burkina Faso. **Keywords :**

Functional POT models, Space-time statistics, Trend detection in extremes, Non-stationarities, Spatial dependence, Environmental Statistics, Climate Change.

Abstract : In this thesis we studied trends in extreme weather events in the current context of climate change. Recent changes in temperature and precipitation extremes are first evaluated using a parametric trend detection POT model. In this parametric model the marginal parameters are polynomial functions limited to order one and two. Then we proposed original spatiotemporal models that constitute our major contribution to the statistics of spatiotemporal extremes. One approach incorporates parametric trends in the marginal distributions of the process as described above and models the spatial dependence structure by a standard ℓ -Pareto process. The other approaches developed combine the modeling of heteroskedastic

extremes with the generalized ℓ -Pareto process approach. In these approaches, non-stationarity in space and time is governed by a scedasticity or trend function. This scedasticity function introduced in the framework of heteroskedastic extremes establishes a relation between the tail distribution of the non-stationary process with a latent spatial process. In this context, we have extended the notion of the univariate return level to space-time stochastic processes. These methods allow us to simulate the studied process in space and time and to deduce the levels or non-stationary return surfaces. Finally, the different methods proposed are illustrated on temperature and rainfall data in Burkina Faso.